PERIODICITY IN THE INTERVALS BETWEEN PRIMES

SCOTT FUNKHOUSER, CARLA COTWRIGHT, DIPENDRA SENGUPTA, AND BRYAN WILLIAMS

Abstract. We consider the $n(n-1)/2$ uniquely defined positive intervals among the first $n \leq 10^6$ prime numbers as a probe of the global nature of the sequence of primes. A statistically strong periodicity is identified in the counting function giving the total number of intervals of a certain size. The nature of the periodic signature implies that the sequences of intervals spanning fixed numbers of gaps repeat quasi-cyclically. From the distribution of intervals we extract also the characteristic period of the repetition, which increases with n in a step-wise manner between consecutive primorial numbers and coincides with the most commonly occuring interval. The relationship between the most common interval and the primorial numbers is noteworthy independently of the periodic behaviors.

1. Introduction

The sequence $\mathbf{p} = 2, 3, \ldots$ of prime numbers exhibits a perplexing hybrid of locally unpredictable but globally regular behaviors [10]. The Prime Number Theorem (PNT) represents the most significant global property of the sequence. The PNT may be stated concisely in terms of the prime counting function $\pi(y)$, defined to give the number of primes no greater than y, as

$$\lim_{y\to\infty} \frac{\log(y)\pi(y)}{y} = 1. \quad (1)$$

Date: July 2, 2015.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2 JUL 2015	**technical**	**01-06-2011 to 01-09-2011**

4. TITLE AND SUBTITLE
Periodicity in the intervals between primes

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Carla Cotwright; Scott Funkhouser; Dipendra Sengupta; Bryan Williams

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
SPAWAR Systems Center Atlantic,Multidisciplinary Research Center,Building 3147, Joint Base Charleston,North Charleston,SC,29406

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We consider the n(n-1)/2 uniquely defined positive intervals among the first n<=10^6 prime numbers as a probe of the global nature of the sequence of primes. A statistically strong periodicity is identified in the counting function giving the total number of intervals of a certain size. The nature of the periodic signature implies that the sequences of intervals spanning fixed numbers of gaps repeat quasi-cyclically. From the distribution of intervals we extract also the characteristic period of the repetition, which increases with n in a step-wise manner between consecutive primorial numbers and coincides with the most commonly occuring interval. The relationship between the most common interval and the primorial numbers is noteworthy independently of the periodic behaviors.

15. SUBJECT TERMS
number theory, prime numbers

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**22**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

It follows from Eq.(1) and $n = \pi(p_n)$ that

$$p_n \sim n \log\left(n \log(n)\right), \tag{2}$$

where p_n is the n-th largest prime and '$\sim$' denotes equality in the limit as n diverges. Throughout the remainder of this work n is implicitly very large.

The global behavior identified in the PNT may be interpreted as an emergent, aggregate property of the sequence $g_1, g_2, \ldots$ of prime gaps, where

$$g_n = p_{n+1} - p_n \tag{3}$$

is the n-th prime gap. From Eqs.(1) and (3) we have

$$g_n \sim \frac{p_n}{n} \sim \log(p_n), \tag{4}$$

which implies that the average gap in the vicinity of a given p_n is approximately $\log(p_n)$. The local distribution of gaps with respect to the average is irregular and similar in nature to a probabilistically determined distribution [3], [6], [5]. An effective measure of the global distribution of gaps is the counting function

$$G_{\mathbf{p}}(n, g) \equiv \#\left\{p_m < p_n : g_m = g\right\} \tag{5}$$

giving the number of gaps of a given size g among the first n primes. $G_{\mathbf{p}} = G_{\mathbf{p}}(n, g)$ naturally exhibits a period-6 oscillation with respect to g for a given n. A representative illustration of the periodicity is shown in Fig.(1), which contains a plot of $G_{\mathbf{p}}(10^6, g)$ for all even g from 2 to 44. Similar periodicity is observed in association with the differences between consecutive gaps [4].

It is historically noteworthy and epistemologically significant that the PNT evolved from empirically motivated conjectures. In particular, no later than 1798 Legendre surmised from tabulated data a logarithmic property that is asymptotically equivalent to Eq.(1). Modern computational capabilities have facilitated dramatically such empirical studies

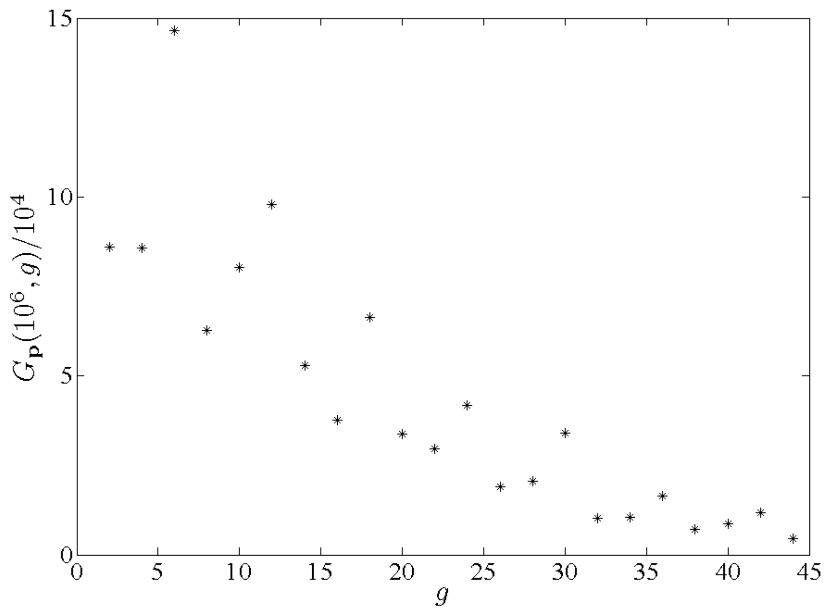

FIGURE 1.

of the properties of the primes. See, for instance, references [1], [2], [4],[7], [8] and [9]. Although it is not generally recognized as a legitimate field within the mathematical community, the empirical study of the nature of numbers has been perhaps as important for the advancement of number theory as experimental physics has been for the advancement of physical theory.

The purpose of the work is to demonstrate empirically a new global property of the primes whose nature is both self-similar and periodic. Rather than consider only the prime gaps, we consider here the entire inventory of positive intervals among the first n primes. In the following Section we present a simple generalized formalism for analyzing the distribution of intervals in an arbitrary increasing sequence of real numbers. In Section 3 the formalism is applied to the intervals among the first n primes, for n as large as 10^6. The distribution exhibits a strong, quantifiable periodicity which is identified as the signature of cyclical repetition in the sequence of intervals spanning a given number of prime gaps. Moreover we find that the characteristic period of the repetition is nearly always the most common interval size, both of which increase with n in a step-wise manner and are always primorial numbers (*i.e.* $2, 6, 30, 210, 2310, \dots$). Section 4 contains a summary of the primary conclusions followed by several related conjectures.

2. Positive intervals within an increasing real sequence

Let $\mathbf{x} = x_1, x_2, \dots$ be an increasing sequence of real numbers which may be either finite or infinitely long. Throughout the following every bold-faced lower-case Roman letter represents a particular type of increasing real sequence whose most general form is $\mathbf{x}$. Among the first n numbers in $\mathbf{x}$ there are n^2 uniquely defined intervals which may be organized naturally in the form of a matrix. For any given $\mathbf{x}$ and n we therefore define $\Delta_{\mathbf{x},n}$ to be an $n \times n$ matrix with elements $x_k - x_j$, where k and j designate respectively the column and row. The elements along the primary diagonal $j = k$ vanish trivially, leaving only $n(n-1)$ uniquely defined non-vanishing intervals. Furthermore because $\Delta_{\mathbf{x},n}$ is antisymmetric a complete description of the intervals requires only the $n(n-1)/2$ positive elements populating the upper-diagonal region of $\Delta_{\mathbf{x},n}$. Let the positive intervals be defined separately as

$$d_{j,k} \equiv x_k - x_j \ ; \quad \forall j < k. \tag{6}$$

In the interest of formality let $\mathcal{D}_{\mathbf{x},n}$ be the set of all $n(n-1)/2$ uniquely defined $d_{j,k}$ among the first n numbers in a given $\mathbf{x}$.

Throughout the following l is reserved for the number $k-j$ of gaps spanned by some $d_{j,k} \in \mathcal{D}_{\mathbf{x},n}$ and is therefore never greater than $n-1$. For a fixed l all intervals $d_{j,\,j+l}$ lie on the same diagonal in $\Delta_{\mathbf{x},n}$, to which the term 'l-th diagonal' refers henceforth. The term 'gap' is used here in reference to an interval between consecutive numbers in any given $\mathbf{x}$, hence

$$g_j \equiv d_{j,j+1}. \tag{7}$$

It is important to note that the elements of a given $\mathcal{D}_{\mathbf{x},n}$ are uniquely defined but are not necessarily uniquely valued. In order to characterize the distribution of positive intervals of various sizes among the first n numbers in a given $\mathbf{x}$ we define the counting function

$$D_{\mathbf{x}}(n,\delta) \equiv \#\left\{d_{j,k} \in \mathcal{D}_{\mathbf{x},n} : d_{j,k} = \delta\right\}. \tag{8}$$

It is also useful to define

$$D^{(l)}_{\mathbf{x}}(n,\delta) \equiv \#\left\{d_{j,j+l} \in \mathcal{D}_{\mathbf{x},n} : d_{j,j+l} = \delta\right\} \tag{9}$$

to give the total number of intervals of size δ which span exactly l gaps, hence

$$D_{\mathbf{x}}(n,\delta) = \sum_{l=1}^{n-1} D^{(l)}_{\mathbf{x}}(n,\delta). \tag{10}$$

In order to represent the most numerous interval size among the first n numbers n $\mathbf{x}$ let $\lambda_{\mathbf{x}} = \lambda_{\mathbf{x}}(n)$ be defined such that

$$D_{\mathbf{x}}(n,\lambda_{\mathbf{x}}) \geq D_{\mathbf{x}}(n,\delta) \tag{11}$$

for all δ. In any case where $D_{\mathbf{x}}(n,\delta)$ has the same maximal value for more than one different δ then let $\lambda_{\mathbf{x}}(n)$ be defined as $\lambda_{\mathbf{x}}(n-1)$.

The following generalized examples are presented in order to establish the basic properties of $D_{\mathbf{x}} = D_{\mathbf{x}}(n,\delta)$ relevant to the present investigation. In so doing it is useful to specify an ordered set $\mathfrak{g} = \{\gamma_1, \gamma_2, \ldots \gamma_\nu\}$ containing the gaps that may be found in a particular sequence. The variable ν is reserved for the number of elements in $\mathfrak{g}$. It is also convenient to define

$$\Gamma \equiv \sum_{m=1}^{\nu} \gamma_m \tag{12}$$

and

$$\gamma \equiv \frac{\Gamma}{\nu}, \tag{13}$$

such that γ is the mean value among all elements in a given $\mathfrak{g}$. If $\nu > 1$ then the manner in which the gaps are selected from $\mathfrak{g}$ determines the nature of the sequence and is specified accordingly. Note that the elements of $\mathfrak{g}$ are not necessarily listed in ascending order, nor are they necessarily all uniquely valued. If $n \gg \nu$ and if the elements of $\mathfrak{g}$ occur in the sequence with approximately uniform frequency then

$$d_{1,n} \approx (n-1)\gamma. \tag{14}$$

2.1. **Uniformly spaced sequence.** Consider first the rudimentary case of a sequence $\mathbf{u} = u_1, u_2, \ldots$ of uniformly spaced numbers. By definition the associated set of available gaps is simply $\mathfrak{g} = \{\gamma\}$ for some $\gamma = \gamma_1$. We therefore have

$$d_{j,k} = (k-j)\gamma. \tag{15}$$

All elements spanning l gaps are consequently identical. The total number of intervals of a given size in $\mathcal{D}_{\mathbf{u},n}$ is simply the length $n-l$ of the diagonal in $\Delta_{\mathbf{u},n}$ on which they are found, hence

$$D_{\mathbf{u}}(n,\delta) = \begin{cases} n - \delta/\gamma & \text{for } \delta = \gamma, 2\gamma, \ldots, (n-1)\gamma \\ 0 & \text{otherwise} \end{cases} . \tag{16}$$

Note that the non-vanishing values of $D_{\mathbf{u}}$ vary linearly with respect to δ.

2.2. **Sequence with randomly ordered gaps.** Next let us consider a sequence $\mathbf{r} = r_1, r_2, \ldots$ in which each gap is selected in an effectively random manner from some $\mathfrak{g}$ containing $\nu > 1$ elements, each of which is equally probable. An interval $d_{j,k}$ spanning more than ν consecutive gaps in $\mathbf{r}$ is, on average, given by

$$d_{j,k} \approx (k-j)\gamma \tag{17}$$

analogously to Eq.(15). If $n \gg \nu$ then the indices j associated with intervals $d_{j,k}$ equal to a particular $\delta > \Gamma$ are found with uniform probability over the range $[1, n - \delta/\gamma]$, where δ/γ is the average number of gaps spanned by an interval equal to δ. Moreover, each allowed value of $d_{j,k}$ is equally likely to be found in association with a given j. It follows that the average density $D_{\mathbf{r}}/(n - \delta/\gamma)$ is approximately constant among all δ for a given $\mathbf{r}$ and $n \gg \nu$. We therefore have

$$D_{\mathbf{r}}(n, \delta) \mathrel{\widetilde{\propto}} n - \delta/\gamma, \tag{18}$$

where $\widetilde{\propto}$ signifies global proportionality. Equation (18) implies that the non-vanishing values of $D_{\mathbf{r}}$ for $\delta > \Gamma$ are correlated to a single line, approximating the precise linearity in Eq.(16).

As an example let $\mathbf{r}'$ be a particular sequence of the form $\mathbf{r}$ with $\mathfrak{g} = \{2, 4, 6\}$. Fig.(2) is a plot of all non-vanishing $D_{\mathbf{r}'}(10^3, \delta)$. The largest interval is 3998, which is expected from Eq.(14). The plotted points are strongly clustered about the line $500 - \delta/8$, with the associated correlation coefficient being approximately 0.997. Although the details of the linear trend are immaterial to the present investigation it is worthwhile to mention that the constant of proportionality in Eq.(18) is approximately ω/γ in any case where $\mathfrak{g}$ is of the form

$$\mathfrak{g}_\omega \equiv \{\gamma_1, (\gamma_1 + \omega), (\gamma_1 + 2\omega), \ldots, (\gamma_1 + (\nu - 1)\omega)\}. \tag{19}$$

In the present example we have $\omega = 2$ and $\gamma = 4$, hence $D_{\mathbf{r}'}$ is broadly correlated to the line $n/2 - \delta/8$.

2.3. **Sequence with cyclically repeating gaps.** Finally let us consider a third generalized sequence $\mathbf{c} = c_1, c_2, \ldots$ in which the gaps repeat cyclically such that

$$g_j = \gamma_\kappa \quad ; \quad \kappa = 1 + \mathrm{mod}(j - 1, \nu) \tag{20}$$

for a given $\mathfrak{g}$ with $\nu > 1$ and for all $j = 1, \ldots, n - 1$. We therefore have

$$g_j = g_{j+\nu}, \tag{21}$$

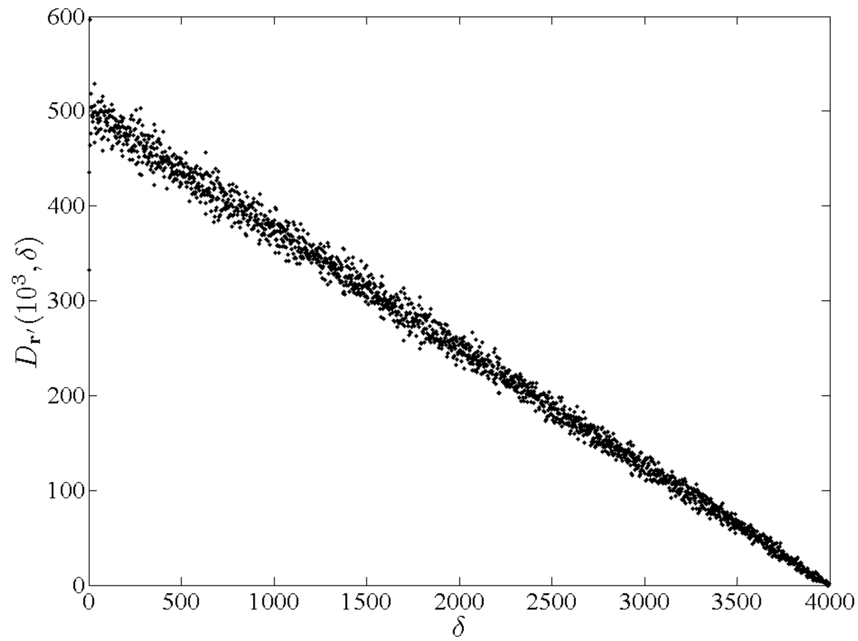

FIGURE 2. Distribution of intervals in a sequence $\mathbf{r}'$ with gaps randomly selected from $\mathfrak{g} = \{2, 4, 6\}$.

which implies the more general periodicity

$$d_{j,k} = d_{j+\nu,k+\nu}. \tag{22}$$

Because every interval spanning ν gaps is equal to Γ we also have

$$d_{j,k+\nu} = d_{j,k} + \Gamma. \tag{23}$$

Every element on the $b\nu$-th diagonal in $\Delta_{\mathbf{c},n}$ is equal to $b\Gamma$, for any positive integer b no greater than $(n-1)/\nu$, and thus

$$D_{\mathbf{c}}(n, b\Gamma) = n - b\nu. \tag{24}$$

Furthermore for $\mathrm{mod}(l, \nu) \neq 0$ and $n-l > \nu$ Eq.(22) implies that the sequence $d_{1,1+l}, d_{2,2+l}, \ldots, d_{n-l,n}$ of all intervals spanning l gaps, *i.e.* the l-th diagonal in $\Delta_{\mathbf{c},n}$, repeats after every ν terms.

Because every interval spanning ν gaps in $\mathbf{c}$ is equal to Γ, the most common interval size is also Γ for sufficiently large n. We may prove that property by considering an upper bound on $D_{\mathbf{c}}$ for δ not equal to an integer multiple of Γ. For convenience let $\sigma_{\mathbf{c}}^{(l)}(n,\delta)$ be the total number of intervals equal to a given δ in a single cycle of ν consecutive rows on the l-th diagonal. If $\mathrm{mod}(\delta,\Gamma)\neq 0$ then $\sigma_{\mathbf{c}}^{(l)}(n,\delta)$ could be no greater than $\nu-1$. The total number of uniquely defined $d_{j,j+l}\neq\delta$ on the l-th diagonal could be therefore no fewer than the number of full cycles on the diagonal, hence

$$D_{\mathbf{c}}^{(l)}(n,\delta) \leq n-l-\left(\frac{n-l-\mathrm{mod}(n-l,\nu)}{\nu}\right). \tag{25}$$

Since no two intervals of the same size could occur on the same row in any $\Delta_{\mathbf{x},n}$, the sum of $\sigma_{\mathbf{c}}^{(l)}(n,\delta)$ over all l could be no greater than $\nu-1$. Consequently the upper bound on $D_{\mathbf{c}}$ occurs when $\sigma_{\mathbf{c}}^{(l)}(n,\delta)=\nu-1$ and when all intervals equal to δ span just one gap. From Eq.(25) we therefore have

$$D_{\mathbf{c}}(n,\delta) \leq (n-1)\left(\frac{\nu-1}{\nu}\right); \quad \forall \quad \mathrm{mod}(\delta,\Gamma)\neq 0. \tag{26}$$

It follows from Eqs. (24) and (26) that

$$\lambda_{\mathbf{c}}(n)=\Gamma; \quad \forall \quad n>\nu^2-\nu+1. \tag{27}$$

Note that Eq.(27) represents the most general and thus the strongest possible requirement. In a case where ν is not significantly greater than the number of unique values among the elements of $\mathfrak{g}$, the requirement for $\lambda_{\mathbf{c}}(n)=\Gamma$ is typically just $n>\nu$.

The period-Γ repetition of intervals in $\mathbf{c}$ also effects periodicity in the δ-dependence of $D_{\mathbf{c}}$. With $k=j+l$ Eq.(23) implies

$$\sigma_{\mathbf{c}}^{(l)}(n,\delta)=\sigma_{\mathbf{c}}^{(l+\nu)}(n,\delta+\Gamma) \tag{28}$$

and thus also

$$\frac{D_{\mathbf{c}}^{(l+\nu)}(n,\delta+\Gamma)}{n-l-\nu} \approx \frac{D_{\mathbf{c}}^{(l)}(n,\delta)}{n-l}, \tag{29}$$

provided that there are at least ν intervals spanning $l+\nu$ gaps among the first n numbers in $\mathbf{c}$, *i.e.* for $n-l-\nu\geq\nu$. With both sides multiplied by $n-l-\nu$, and in summation over all allowed l, Eq.(29) becomes

$$\sum_{l=1}^{n-2\nu} D_{\mathbf{c}}^{(l+\nu)}(n,\delta+\Gamma) \;\approx\; D_{\mathbf{c}}(n,\delta) - \nu \sum_{l=1}^{n-2\nu} \left(\frac{D_{\mathbf{c}}^{(l)}(n,\delta)}{n-l} \right). \tag{30}$$

If $n-2\nu$ is at least as large as δ/γ then the difference between the left side of Eq.(30) and $D_{\mathbf{c}}(n,\delta+\Gamma)$ is negligible on average. For all δ no greater than $(n-2\nu)\gamma \approx c_n - 2\Gamma$ we therefore have

$$D_{\mathbf{c}}(n,\delta+\Gamma) \;\approx\; D_{\mathbf{c}}(n,\delta) - \nu \sum_{l=1}^{n-2\nu} \left(\frac{D_{\mathbf{c}}^{(l)}(n,\delta)}{n-l} \right). \tag{31}$$

Note that Eq.(31) is valid for all but a comparatively small range of δ whenever $n \gg \nu$. Furthermore in the first-order approximation for $l \ll n$, which corresponds to $\delta \ll c_n$, Eq.(31) becomes

$$D_{\mathbf{c}}(n,\delta+\Gamma) \approx \left(1 - \frac{\nu}{n}\right) D_{\mathbf{c}}(n,\delta). \tag{32}$$

Equation (31) implies that $D_{\mathbf{c}}$ varies with δ in a cyclical manner with period Γ for all δ spanning on average no more than $n-2\nu$ gaps. For $\delta \ll c_n$ the correlation reduces to the simple proportionality in Eq.(32), such that the basic form of $D_{\mathbf{c}}$ over the range $(0,\Gamma]$ is repeated in progressively smaller scale over subsequent periods of δ.

As an example let $\mathbf{c}'$ be a particular sequence of the form $\mathbf{c}$ with $\mathfrak{g} = \{2,4,6\}$, hence

$$\Delta_{\mathbf{c}',n} = \begin{matrix} 0 & 2 & 6 & 12 & 14 & 18 & 24 \\ -2 & 0 & 4 & 10 & 12 & 16 & 22 \\ -6 & -4 & 0 & 6 & 8 & 12 & 18 \\ -12 & -10 & -6 & 0 & 2 & 6 & 12 \\ -14 & -12 & -8 & -2 & 0 & 4 & 10 \\ -18 & -16 & -12 & -6 & -4 & 0 & 6 \\ & & & \cdots & & & \end{matrix} \quad \ldots .$$

Figure (3) is a plot of non-vanishing values of $D_{\mathbf{c}'}(10^2,\delta)$. In accordance with Eq.(27) the most common interval among the first 100 numbers

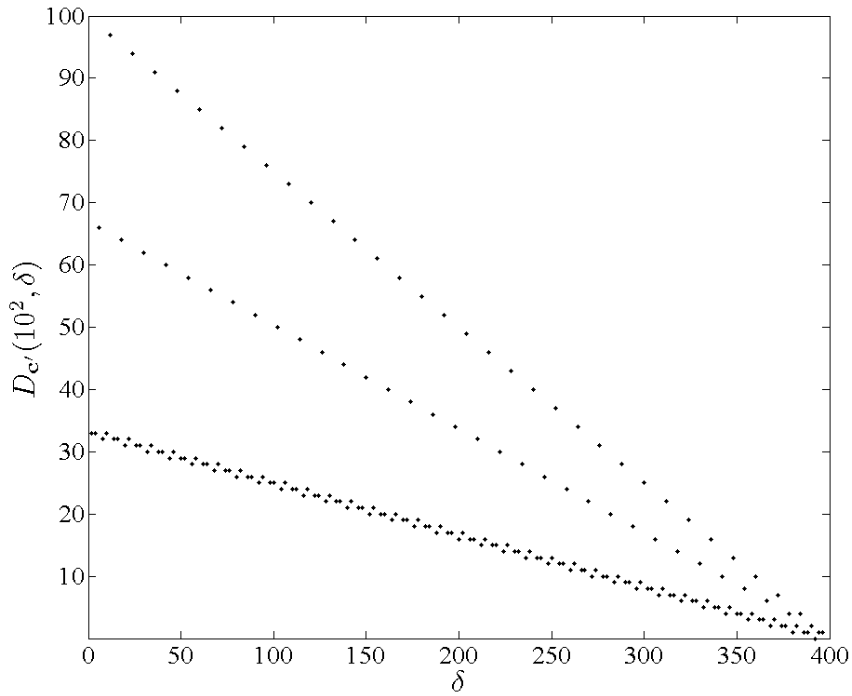

FIGURE 3. Distribution of intervals in a sequence $\mathbf{c}'$ with cyclically repeating gaps $2, 4, 6, 2, 4, 6, \ldots$.

in $\mathbf{c}'$ is $\Gamma = 12$, with $D_{\mathbf{c}'}(10^2, 12) = 97$. The distinctive fanned pattern in Fig.(3) is a signature of the periodicity in Eq.(32). The periodic structure for $l \ll n$ may be seen clearly in Fig.(4), which is a zoomed-in view of Fig.(3) for $\delta \leq 36$. Note how the non-vanishing values for $\delta \in (0, 12]$ are repeated in proportion over the range $(12, 24]$, *etc.*

3. Positive intervals within the sequence of primes

Let us now employ the basic measures developed in the preceding Section to analyze the distribution of positive intervals among the first n primes. Because of the solitary odd prime gap $g_1 = 1$ there are $n - 1$ odd $d_{1,k} \in \mathcal{D}_{\mathbf{p},n}$, each of which is uniquely valued. The anomalous odd intervals are not important in the present investigation and are ignored

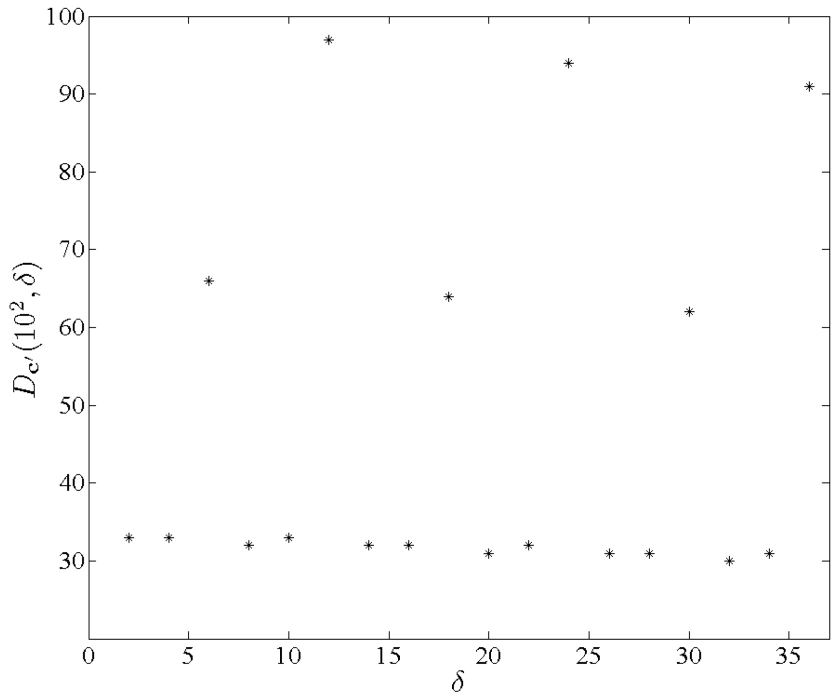

Figure 4.

throughout the following. Although it is not mathematically meaningful to construct $\mathbf{p}$ in terms of some $\mathfrak{g}$, for notational convenience let $\mathfrak{g} = \{g_1, \ldots, g_{n-1}\}$ be specified implicitly in association with the first n primes such that γ represents the average prime gap.

Figure (5) is a plot of $D_{\mathbf{p}}(10^4, \delta)$ *vs.* δ for all even δ from 2 to 104726, which is the largest even interval among the first 10^4 primes. The points are strongly concentrated along several distinct lines which converge on the abscissa near $\delta = 104726$, forming a fanned distribution in the manner of Fig.(3). We should emphasize that every $D_{\mathbf{x}}(n, \delta)$ is a singly-valued function of δ for a given n, and no two points in Fig.(5) share the same δ. The fanned striations are an aggregate visual effect that could be produced only from finely-grained periodicity in the δ-dependence of $D_{\mathbf{p}}$. Figure (6) is a zoomed-in view of Fig.(5) showing

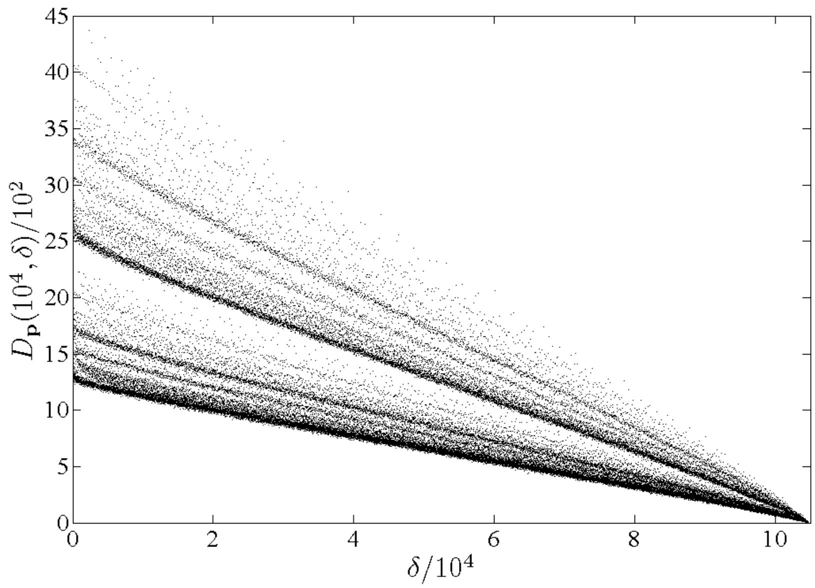

FIGURE 5. Distribution of intervals among the first 10^4 primes.

non-vanishing $D_{\mathbf{p}}(10^4, \delta)$ for $\delta \leq 460$. The quasi-periodic structure observed in Fig.(6) is representative of the type of structure observed throughout the full range of δ. We have verified that $D_{\mathbf{p}}$ has the same basic form, with similarly well-articulated periodicity, for arbitrary n as large as 10^6, which is the limit imposed by computational feasibility.

Recall that the periodicity in $D_{\mathbf{c}}$ is a direct consequence of the cyclical repetition of gaps, which effects a broader cyclical repetition in the intervals spanning more than one gap. The prime gaps, however, are not ordered in the manner of repeating sequence. It is nonetheless possible for intervals spanning many prime gaps to repeat quasi-cyclically, mimicking the global properties of a sequence of the form $\mathbf{c}$. Specifically, suppose that there exists some $\mu = \mu(n)$ such that the prime intervals are globally subject to the periodic correlation

$$d_{j,k} \approx d_{j+\mu,\, k+\mu}, \tag{33}$$

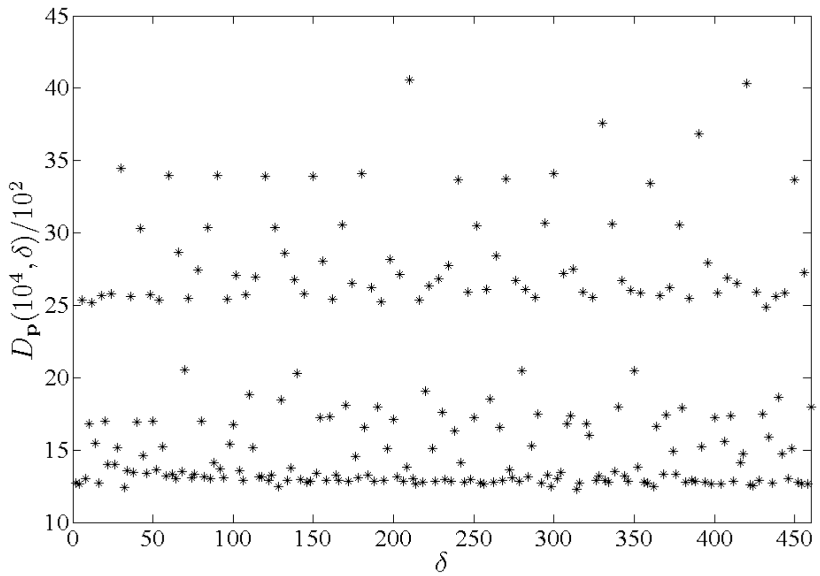

FIGURE 6.

in approximation of the precise periodicity in Eq.(22). The n-dependence of μ has been allowed in anticipation of the outcome of the following analysis. Let $T \approx \mu\gamma$ represent the corresponding period. In order for Eq.(22) to be meaningful $T = T(n)$ must correspond to a common interval size among the first n primes. If the putative quasi-cyclical nature of the prime intervals is strongly pronounced then we may expect $T \approx \lambda_{\mathbf{p}}$. In any case Eq.(33) would imply that intervals of a given size δ spanning l prime gaps occur with approximately the same cyclical regularity as do intervals of size $\delta + T$ spanning $l + \mu$ prime gaps. We would accordingly have

$$\frac{D_{\mathbf{p}}^{(l+\mu)}(n, \delta + T)}{n - l - \mu} \approx \frac{D_{\mathbf{p}}^{(l)}(n, \delta)}{n - l} \tag{34}$$

for $l \leq n - 2\mu$, analogously to Eq.(34). Summation over l produces

$$D_{\mathbf{p}}(n, \delta + T) \approx D_{\mathbf{p}}(n, \delta) - \mu \sum_{l=1}^{n-2\mu} \left(\frac{D_{\mathbf{p}}^{(l)}(n, \delta)}{n - l} \right) \tag{35}$$

for all δ spanning on average no more than $n-2\mu$ gaps. Equations (33) through (35) could be consistent with the PNT and the other proven properties of the primes while naturally effecting the periodicity observed qualitatively in $D_{\mathbf{p}}$.

Ideally one should test for the hypothesized quasi-cyclical behaviors by quantifying the statistical strength of the global correlation implied in Eq.(33). The PNT guarantees, however, that $d_{j,k}$ increases broadly with increasing j, for any fixed $k-j$. Because the putative oscillations associated with Eq.(33) could only be small in comparison to the global trend associated with the PNT, a strong correlation of the form Eq.(33) is observed for nearly every meaningful value of μ. A reliable direct test of Eq.(33) is therefore infeasible. In contrast, the oscillations in $D_{\mathbf{p}}$ associated with Eq.(35) may be large in comparison to the broad variation of $D_{\mathbf{p}}$ with respect to δ. A reliable and effective indirect test of Eq.(33) may be therefore formulated by quantifying the strength of the correlation implied in Eq.(35).

In order to maintain consistency with Section 2, the following measure is defined generally in terms of an arbitrary $D_{\mathbf{x}}$. We substitute for T the independent variable t and define $\rho \approx t/\gamma$ to be the average number of gaps spanned by an interval of size t among the first n numbers in $\mathbf{x}$. For a given t let $\delta_1, \ldots, \delta_M$ represent all of the $M = M(t)$ unique values of δ in the range $(0, p_n - 2t]$ for which $D_{\mathbf{x}}(n, \delta)$ is non-vanishing. With

$$u_i \equiv D_{\mathbf{x}}(n, \delta_i + t) \tag{36}$$

and

$$v_i \equiv D_{\mathbf{x}}(n, \delta_i) - \rho \sum_{l=1}^{n-1} \frac{D_{\mathbf{x}}^{(l)}(n, \delta_i)}{n-l} \tag{37}$$

the standard coefficient measuring the strength of the correlation between the left and right sides of Eq.(35) is

$$\chi_{\mathbf{x}}(n,t) \equiv \frac{\sum_{i=1}^{M}(u_i-\overline{u})(v_i-\overline{v})}{\sqrt{\sum_{i=1}^{M}(u_i-\overline{u})^2\sum_{i=1}^{M}(v_i-\overline{v})^2}}, \tag{38}$$

where $\overline{u}$ and $\overline{v}$ are the respective means of u_i and v_i over all δ_i. Finally, in order to represent the t for which $\chi_{\mathbf{x}} = \chi_{\mathbf{x}}(n,t)$ is maximized for a given n, let $\tau_{\mathbf{x}} = \tau_{\mathbf{x}}(n)$ be defined such that

$$\chi_{\mathbf{x}}(n,\tau_{\mathbf{x}}) \geq \chi_{\mathbf{x}}(n,t) \tag{39}$$

for all t. If there are more than one t for which $\chi_{\mathbf{x}}(n,t)$ has the same maximal value then let $\tau_{\mathbf{x}}(n)$ be the smallest among them.

If the prime intervals repeat quasi-cyclically in the manner of Eq.(33), thereby effecting periodicity in $D_{\mathbf{p}}$ of the form Eq.(35), then $\chi_{\mathbf{p}}(n,t)$ should exhibit pronounced local maxima near unity for all t equal to an integer multiple of T, with T typically producing the global maximum. We therefore expect $T \approx \tau_{\mathbf{p}}$. If the periodicity is sufficiently strong and if n is sufficiently large then we also expect $\tau_{\mathbf{p}} \approx \lambda_{\mathbf{p}}$.

Figure (7) is a plot of $\chi_{\mathbf{p}}(10^3,t)$ for even t from 6 to 2370. The calculations indicate a strong periodic correlation for all t equal to an integer multiple of 6, and an appreciably less strong correlation for all other t. Among multiples of 6, clearly pronounced local maxima are observed at all multiples of $\tau_{\mathbf{p}}(10^3) = 210$, which produces the global maximum $\chi_{\mathbf{p}}(10^3,210) \simeq 0.9914$. We therefore conclude that the non-vanishing values of $D_{\mathbf{p}}(10^3,\delta)$ and $D_{\mathbf{p}}^{(l)}(10^3,\delta)$ are mutually correlated in the manner of Eq.(35) with period $T(10^3) = 210$. Furthermore the most common interval size among the first 10^3 primes is $\lambda_{\mathbf{p}}(10^3) = 210$, indicating an additional degree of consistency with the cyclical behaviors of $\mathbf{c}$.

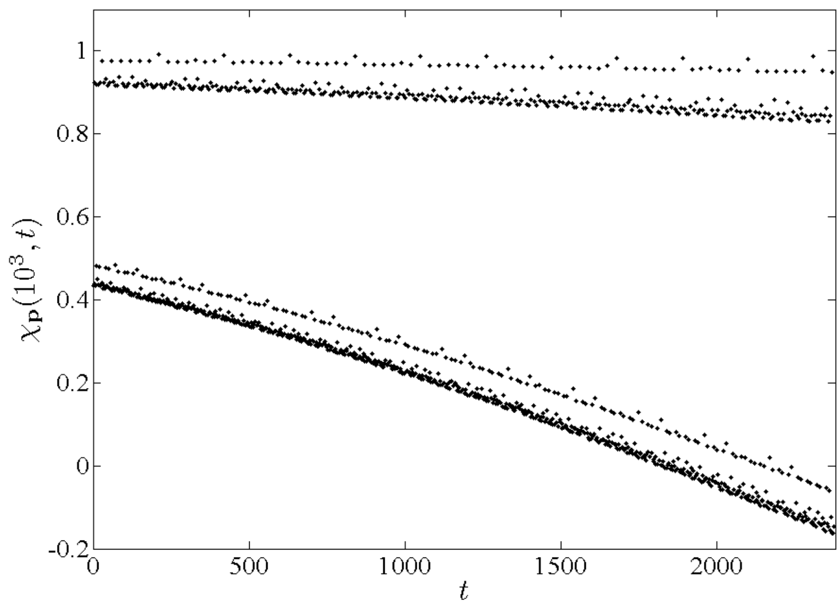

FIGURE 7. Coefficient measuring period-t correlation of the form Eq.(35) in $D_{\mathbf{p}}(10^3, \delta)$.

We report similarly conclusive results for arbitrary n as large as 10^6. Figure (8) contains a plot of $\tau_{\mathbf{p}}$ for logarithmically spaced n from 10^2 to 10^6. In each case the global maximum $\chi_{\mathbf{p}}(n, \tau_{\mathbf{p}})$ is greater than 0.97, and strongly pronounced local maxima are found at all integer multiples of $\tau_{\mathbf{p}}$. Included in Fig.(8) is a plot of $\lambda_{\mathbf{p}}$ for all n over the same range. Both $\chi_{\mathbf{p}}$ and $\lambda_{\mathbf{p}}$ increase with n in a remarkably well articulated step-wise manner, remaining constant for a broad range of n between each sharp transition. Over the entire range of n considered here $\chi_{\mathbf{p}}$ and $\lambda_{\mathbf{p}}$ coincide all but precisely, differing only near the transition points. Moreover all of the observed values of $\chi_{\mathbf{p}}$ and $\lambda_{\mathbf{p}}$ among the first 10^6 primes are of the form

$$y\# \equiv \prod_{a=1}^{\pi(y)} p_a, \tag{40}$$

which is the primorial of some real $y \geq 2$. The 'steps' in Fig.(8) correspond to $p_3\# = 30$, $p_4\# = 210$, $p_5\# = 2310$, $p_6\# = 30030$ and

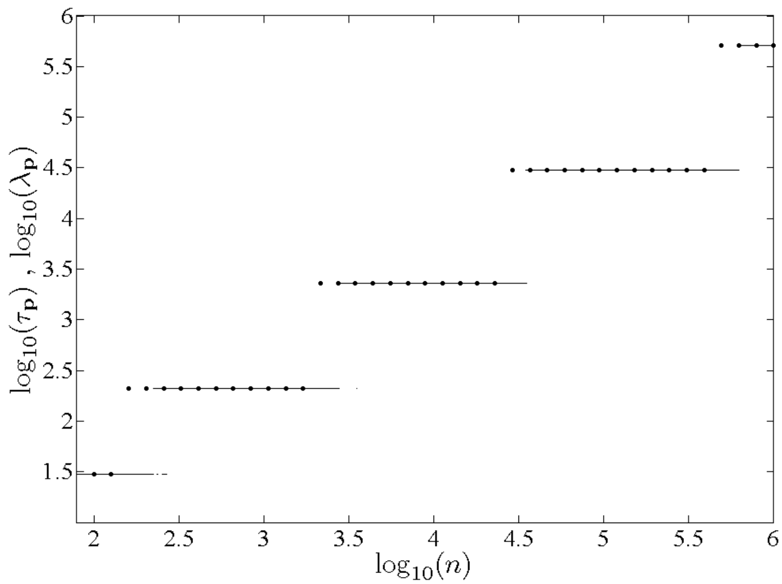

FIGURE 8. $\log_{10}$ plots of the correlation period $\tau_{\mathbf{p}}$, in heavy points, and the most common intervals size $\lambda_{\mathbf{p}}$, in the broken line.

$p_7\#=510510$. Although $\chi_{\mathbf{p}}$ is not statistically meaningful for n smaller than approximately 100, it is worth noting that $\lambda_{\mathbf{p}}$ exhibits the same step-wise behavior for $n<100$ among the primorial numbers $p_1\#=2$, $p_2\# = 6$ and $p_3\#$. We also observe that each step-wise increase in $\chi_{\mathbf{p}} \approx \lambda_{\mathbf{p}}$ from a given $p_i\#$ to $p_{i+1}\#$ occurs where n is near $p_{i+1}\#$.

The calculations presented here imply unequivocally that the non-vanishing values of $D_{\mathbf{p}}$ and $D_{\mathbf{p}}^{(l)}$ are correlated mutually according to Eq.(35), with $T\approx\tau_{\mathbf{p}}$. Given that Eq.(35) is a specific and direct consequence of Eq.(33), it is reliable to conclude that the prime intervals repeat quasi-cyclically in the manner of Eq.(33). The broad agreement between $\tau_{\mathbf{p}}$ and $\lambda_{\mathbf{p}}$ reinforces the reliability of that conclusion.

Lastly it is worth mentioning that the periodic properties identified in this Section are associated with the particular order of the sequence of prime gaps and are not determined simply by the net distribution

$G_{\mathbf{p}}$. In order to demonstrate that point let the sequence $\widetilde{\mathbf{p}} = \widetilde{\mathbf{p}}(n) = \widetilde{p}_1, \widetilde{p}_2, \ldots, \widetilde{p}_n$ be constructed by randomly rearranging the sequence of prime gaps $g_2, \ldots, g_{n-1}$, with $\widetilde{p}_1$ and $\widetilde{p}_2$ always equal to 2 and 3 respectively. As a natural consequence of the construction of $\widetilde{\mathbf{p}}$ we also have $\widetilde{p}_n = p_n$. On average a comparatively small number of other $\widetilde{p}_j$ happen to be prime as well. Figure (9) is a plot of $D_{\widetilde{\mathbf{p}}}(10^4, \delta)$ for a particular $\widetilde{\mathbf{p}}$. The present example is representative of the form of $D_{\widetilde{\mathbf{p}}}$ for arbitrary n and typical $\widetilde{\mathbf{p}}$. In sharp contrast to $D_{\mathbf{p}}$, $D_{\widetilde{\mathbf{p}}}$ varies broadly with δ in a simple linear manner and is similar in form to a distribution associated with a sequence like $\mathbf{r}$ whose gaps are randomly selected. In fact the linear trend of $D_{\widetilde{\mathbf{p}}}(10^4, \delta)$ crosses the ordinate near 1900, which is consistent with the intercept $n\omega/\gamma$ associated with $D_{\mathbf{r}}$ in a case where $\mathfrak{g}$ is of the form Eq.(19) with $\omega = 2$ and $\overline{\gamma} = 10.47$, where 10.47 is approximately the average gap among the first 10^4 primes. For the present purposes, however, it is sufficient to note that $D_{\widetilde{\mathbf{p}}}$ generally lacks the distinctive periodic structure observed in $D_{\mathbf{p}}$.

4. Conclusions

By examining the counting function $D_{\mathbf{p}} = D_{\mathbf{p}}(n, \delta)$ we have identified several new global properties of the positive intervals among the first n primes, for n at least as large as 10^6. The variation of $D_{\mathbf{p}}$ with respect to δ exhibits a strongly pronounced periodic correlation of the form Eq.(35), which is the signature of quasi-cyclical repetition in the sequences of intervals spanning fixed numbers of prime gaps *ala* Eq.(33). Over the observed range of n the characteristic period $T \approx \tau_{\mathbf{p}}$ of the repetition, as inferred from the maximal correlation $\chi_{\mathbf{p}}(n, \tau_{\mathbf{p}})$, almost universally coincides with the most common interval $\lambda_{\mathbf{p}}$. Such a congruence is also a signature of quasi-cyclical repetition of the form Eq.(33). Furthermore for all n considered here both $\tau_{\mathbf{p}}$ and $\lambda_{\mathbf{p}}$ are always primorial numbers and are typically the largest primorial number no greater than n. The regular manner in which T increases with n

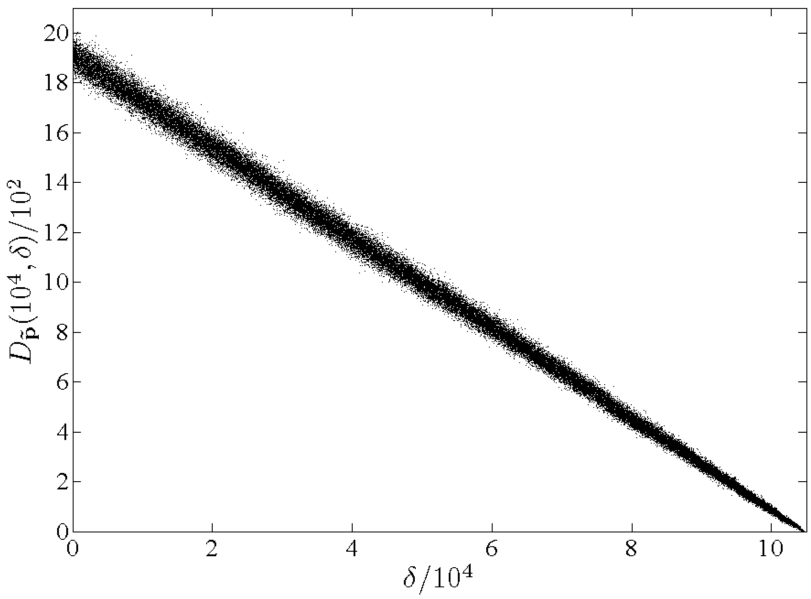

FIGURE 9. Distribution of intervals in a randomized analogue of the first 10^4 primes.

indicates a certain 'Russian dolls'-style self-similarity in the periodicity of the prime intervals, with the characteristics over a certain range of n replicated in scale over consecutively larger ranges.

Although the behaviors reported here have been validated only for n as large as 10^6, it is reasonable to expect that they would be observed for arbitrary n. We therefore offer the following conjectures.

Conjecture 4.1. *For all $n \gg 1$ the positive intervals among the first n primes repeat quasi-cyclically such that*

$$d_{j,\,j+l} \approx d_{j+\mu,\,j+l+\mu}, \tag{41}$$

which implies

$$D_{\mathbf{p}}(n,\delta+T) \approx D_{\mathbf{p}}(n,\delta) - \mu \sum_{l=1}^{n-2\mu} \left(\frac{D_{\mathbf{p}}^{(l)}(n,\delta)}{n-l} \right), \tag{42}$$

where $T = T(n)$ is the characteristic period and $\mu = \mu(n)$ is the positive integer nearest to the average number of gaps spanned by an interval of size T among the first n primes.

Conjecture 4.2. *For all $n \gg 1$ we have*

$$T(n) \approx \tau_{\mathbf{p}}(n) \approx \lambda_{\mathbf{p}}(n). \tag{43}$$

Conjecture 4.3. *The most common interval size $\lambda_{\mathbf{p}}(n)$ among the first $n \gg 1$ primes is always a primorial number. $\lambda_{\mathbf{p}}(n)$ increases with n in a step-wise manner from a given $p_i\#$ to $p_{i+1}\#$, with each transition occurring where n is near $p_{i+1}\#$. We therefore have*

$$\lambda_{\mathbf{p}} \sim n. \tag{44}$$

Conjecture 4.3 is stated without reference to $\tau_{\mathbf{p}}$ so that it may be considered independently of the other Conjectures and independently of any periodic behaviors discussed here.

References

1. S. Ares and M. Castro, *Hidden structure in the randomness of the prime number sequence?*, Physica A **360** (2006), 285–296.
2. A. Bershadskii, *Hidden periodicity and chaos in the sequence of prime numbers*, Advances in Mathematical Physics (2011), 519178.
3. P. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1976), 4 – 9.
4. H. E. Stanley P. Kumar, P. Ch. Ivanov, *Information entropy and correlations in prime numbers*, arXiv:cond-mat/0303110v4 (2003).
5. K. Soundararajan, *Small gaps between prime numbers: The work of goldston-pintz-yildirim*, Bull. Amer. Math. Soc. (new series) **44** (2007), no. 17, 1 – 18.
6. D. Wells, *Prime numbers: the most mysterious figures in math*, John Wiley & Sons, 2005.

7. M. Wolf, *Multifractality of prime numbers*, Physica A **160** (1989), 24–42.
8. ———, *Unexpected regularities in the distribution of prime numbers*, Proc. of the Joint EPS-APS Int. Conf. Physics Computing **8** (1996), 361–367.
9. ———, Physica A **274** (1999), 149.
10. D. Zagier, *The first 50 million prime numbers*, Math. Intelligencer **0** (1997), 7 – 19.

SPAWAR Systems Center Atlantic

Norfolk State University

Elizabeth City State University

Hampton University