The Prime Numbers Hidden Symmetric Structure and its Relation to the Twin Prime Infinitude and an Improved Prime Number Theorem.

Imre Mikoss

Universidad Simón Bolívar, Dep. de Formación Integral y Ciencias Básicas, Valle de Sartenejas, Baruta, Caracas, Venezuela. email: mikoss@hotmail.com

Due to the sieving process represented by a Secondary Sieving Map; during the generation of the prime numbers, geometric structures with definite symmetries are formed which become evident through their geometrical representations. The study of these structures allows the development of a constructive prime generating formula. This defines a mean prime density yielding a second order recursive and discrete prime producing formula and a second order differential equation whose solutions produce an improved Prime Number Theorem. Applying these results to twin prime pairs is possible to generate a "Twin Prime Number Theorem" and important conclusions about the infinitude of the twin primes.

Introduction

Most of the knowledge about the sequence of numbers named primes is a set of unproved theorems and conjectures[1]. The reason of this fact remains as elusive as the very proofs. The approach of this work proposes a new and heuristic way of treating these problems. As it is well known, sieving algorithms are the only efficient way to produce primes. This fact should be taken as an indication that sieving is the natural way of producing primes. A myth has been generated about the sequence of primes, and many attempts have been undertaken to find some properties which should be intrinsic to the sequence itself, despite its generating procedure[2]. Most of them (perhaps all of them) are based on the famous Euler formula which relates the sequence of primes with the Zeta function[3]. This formula is nothing else than an analytic representation of the sieving process[4]. Through the construction of this formula a limit is taken, which eliminates the very heart of the process. Due to this limit, the relevance of the erased part has remained hidden from the scientific community through centuries. In the present work the structure of this hidden part is made evident through the iteration of a Secondary Sieving Map. Through the structure created by the construction of the primes, the reason why the twin primes should be infinite becomes clear. This last is known as the Twin Prime Conjecture[1], which is one of the unproved icons in the modern number theory. Using a mean prime density derived from the geometric construction, a discrete second order equation is obtained as well as a continuous version of it, which is a second order non-linear differential equation. This is a first approximation for a prime differential equation and is used to demonstrate constructively and to improve the best known approximation to the primes, known as the Prime Number Theorem.

Sieving as a Recursive Map

Usually recursive maps act on subsets of the real numbers, although some of them are defined on geometrical objects[5]. A recursive map which acts on infinite and discrete sequences of numbers is proposed here. This map called Secondary Sieving Map (SSM) is denoted with the Greek letter β. Given:

$$\eta=\{\eta_\alpha,\eta_\beta,\eta_\chi,\eta_\delta,....\infty\}$$

An infinite and discrete sequence of natural numbers which satisfies:

$$\eta_\alpha<\eta_\beta<\eta_\chi<\eta_\delta<....<\infty$$

Then β act on this sequence η in the following way:

$$\beta(\eta)=\eta-\eta^*=\eta-\eta_\beta\cdot\eta$$

$$\eta^*=\{\eta_\beta\cdot\eta_\alpha,\eta_\beta\cdot\eta_\beta,\eta_\beta\cdot\eta_\gamma,\eta_\beta\cdot\eta_\delta...\infty\}$$

The minus sign means element extraction. The temptation to factorize η should be avoided because the resulting equation is not the original; η_β is a number whereas η is a set. In order to make the last definition operational, η and η^* should fulfill $\eta\supseteq\eta^*$. The second element of the original sequence η_β, is called the "pivot number". The outcomes of applying β is named generation. The natural numbers set (without the cero) ℕ* = n complies with all the features required to be an argument of β. Applying β on n recursively is similar, but not equal, to the Eratosthenes sieve, because the last one is not applied on an infinite sequence and it was not conceived as an iterative mapping. The SSM applied to n, could be written in Mathematica ® as "Nest[Complement[#,#[[2]]*#]&,Range[m1],m2]", where m1 is the size of the natural numbers subset on which it will be applied, and m2 is the desired iterations number (it is impossible to act on a infinite set with a computer). Acting once whit β on n produces the first generation, the set of odd numbers:

$$\beta(\text{n})=\text{n}^1_{(2)}=\{1,3,5,7,...\}=2.\text{n}+1$$

Using this last notation, n^0 = n (generation cero). Observe that the first pivot number (the number 2 which appears as a sub index in parentheses) used to generate de odd numbers, is not present in them. However the number one does appear. This is a persisting characteristic in β's successive applications on n: the number one always survives and the pivot number used to generate the last iteration obviously disappears from it, because it was multiplied by one and extracted. In the second generation, the main features of these sequences start to emerge:

$$\beta^2(n)=\beta(\beta(n))=\beta\left(n^1_{(2)}\right)=n^2_{(2,3)}=$$
$$\{1,5,7,11,13,17,19,23,25,29,...\}=$$
$$\{1,7,13,19,25,...\}\cup\{5,11,17,23,29,...\}=$$
$$(6\cdot n+1)\cup(6\cdot n+5)=6\cdot n+\begin{pmatrix}1\\5\end{pmatrix}$$

Where **U** means as usual the union of two sets and the last equation is a convenient way to write the existence of two overlapped linear behaviors. Observe that the set of pivot numbers (which record is kept in n's sub index) starts to form the set of prime numbers. Note also that the sequences $\text{n}_{(2,\ 3)}^2$ and $\text{n}_{(2)}^1$ are qualitatively different: a splitting has occurred in the first one. Instead of one linear function of n, there are two overlapped and simultaneous. In order to understand this segregation, the symmetric features of the SSM should be examined through a heuristic geometrical representation of it, shown in the next section. As a final sentence for this section, it should be noted that in the last decades some insight has been gained about a fractal structure of the primes[6], and it is well known in dynamical systems that fractal structures are produced by iterative mappings[5]. The SSM acting on n could be the basis for the fractal structure of the primes.

Mirror Symmetry and Periodicity in Generations, as a Geometrical Image of a Multiple Linear Representation

The SSM can be represented geometrically using an infinite chain of curves (jumps) which connects the sequence of numbers denoted with η^* (see equations 1 and 2). The non-touched numbers corresponds to the

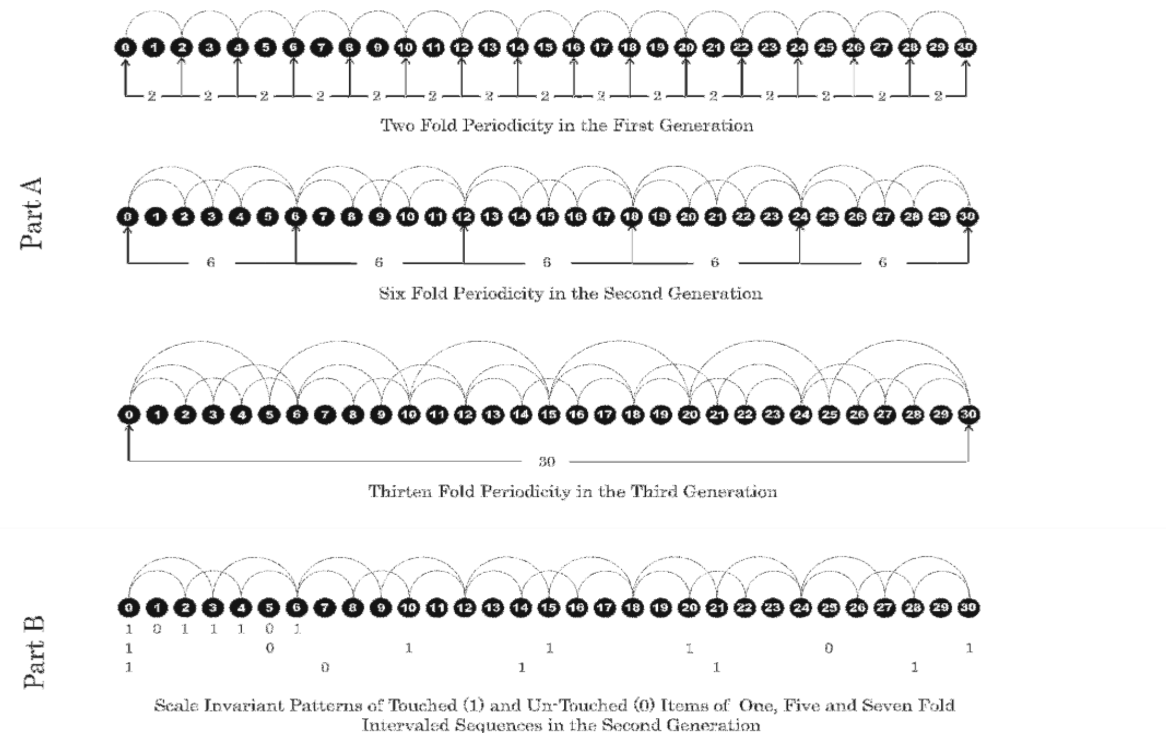

Figure 1 | Generations and Discrete Scale Invariance Graphical Representations. In Part A the first three generations geometric structure is shown through an analogy between extracted and touched numbers. The periodicity in each generation and their mirror symmetries are evident in the structures of the touching curves. In Part B the DSI of the Second Generation is shown.

first generation. Further iterative applications of β can be made on the same graph simply overlapping the geometrical representations of the corresponding η^*. For example in Figure 1, the representations of three generations are seen in separate lines. The touched numbers under β's action, are just touched once whereas in the graphical representation multiple touching is allowed. In this way the geometrical features are better observed. From these drawings can be inferred that due to the successive applications of the SSM, periodic structures form spontaneously. The period of a particular generation is given by the multiplication of all the previous pivot numbers (the formal demonstration will be published elsewhere). Actually these structures are periodic both in the generations but also in the superposition of the η^* produced and extracted during each generation. In fact this last superposition is the most notorious in the geometrical representations from Figure 1, Part A. As it can be observed in the same figure, the splitting mentioned in the past section is caused by the incommensurability of the first pivot number (2) and the second (3) which sets the first untouched numbers (1 and 5) to lie symmetrically around the number three. Then, due to the six-fold periodicity, the two linear behaviours represented in equation (3) are produced. Note that mirror symmetry around the numbers 3, 9, 15… as well as around the numbers 6, 12, 18… starts to emerge. This symmetry will become more evident in the third generation and afterwards:

$$\beta^3(n) = n^3_{(2,3,5)} = \{1, 7, 11, 13, 17, 19, 23, 29, 31\ldots\} =$$

$$\begin{bmatrix} 1 & 31 & 61 & 91 & 121 & 151 & 181 & \ldots \\ 7 & 37 & 67 & 97 & 127 & 157 & 187 & \ldots \\ 11 & 41 & 71 & 101 & 131 & 161 & 191 & \ldots \\ 13 & 43 & 73 & 103 & 133 & 163 & 193 & \ldots \\ 17 & 47 & 77 & 107 & 137 & 167 & 197 & \ldots \\ 19 & 49 & 79 & 109 & 139 & 169 & 199 & \ldots \\ 23 & 53 & 83 & 113 & 143 & 173 & 203 & \ldots \\ 29 & 59 & 89 & 119 & 149 & 179 & 209 & \ldots \end{bmatrix} = 30 \cdot n + \begin{pmatrix} 1 \\ 7 \\ 11 \\ 13 \\ 17 \\ 19 \\ 23 \\ 29 \end{pmatrix}$$

This last equation is again a condensed and convenient way to represent the third generation. In Figure 1 the corresponding representation shows a new period (30 = 2.3.5) and a new mirror symmetry around the multiples of 30 and their halves. In order to advance in a description of the subject, some definitions are needed. The numbers in the one-column matrix (1,5 for the second generation and 1,7,11,13,17,19,23,29 for the third generation) are called seeds. Each seed form a branch through the proper periodicity of its generation. A branch represents a linear mapping of n with the correspondent period between their elements and its seed as starting phase. In the big matrix of the third generation, their 8 infinite branches are written starting with their corresponding 8 seeds. The mirror symmetry mentioned earlier can be easily identified in the seeds structure. For example, the third generation could be written as:

$$n^3_{(2,3,5)} = \left(30 \cdot n + \begin{pmatrix} 1 \\ 7 \\ 13 \\ 19 \end{pmatrix} \right) \cup \left(30 \cdot n + \begin{pmatrix} 11 \\ 17 \\ 23 \\ 29 \end{pmatrix} \right)$$

This seeds separation is quite interesting because each has a six-fold periodicity which is obviously a previous generation remnant. Each four member sequence is defined as a stem. The mirroring between these two stems is quite easily expressed mathematically: the sum of the first and the second stem inverted, results in the generation's period, 1 + 29 = 7 + 23 = 13 + 17 = 19 + 11 = 30. Stems of higher generations become more complicated. For example the stems of the fourth generation are not anymore so ordered as of the third ones. This has to do with an important generation's feature; they operate on two characteristics lengths: one length is the period and the other is the prime confidence interval (PCI). The PCI is defined through the most important characteristic of the generations: their ability to "forecast" new primes. It was already mentioned that the set of pivot numbers tend to form the set of primes. But each pivot number through its generation will certainly forecast all the primes between itself and its square, this interval is called PCI. The PCI is depleted from the pivot number multiples and the other numbers contained within have their squares, cubes, etc certainly in the outside of this interval. In this sense, the generations are successively better approximations to the sequence of primes. The pristine natural numbers n, are the generation cero. They already predict the number 3 as the second prime. In the following table, the beginning from the first three generations, are seen with the pivot numbers and their squares in *italic*, and the predicted primes in **bold**:

1,*2*,**3**,*4*
1,*3*,**5**,**7**,*9*
1, *5*, **7,11,13,17,19,23**,*25*

After the third generation each branch has its own PCI, and then the number 29 is also predicted. The PCI grows with the square of the pivot numbers and the period grows as a factorial (the product of all previous pivot numbers, this prime factorial is usually called primorial in the literature). The PCI is bigger that the period (super-periodic region) at the first generations and the period outgrow the PCI after the fourth generation (sub-periodic region). In the super-periodic region the stems are composed of seeds which are primes, in the sub-periodic region this is not the case anymore, the seeds can be primes or not. This has important consequences on the structure of higher generations due to what is called internal sieving. A last sentence (disclaimer) for this section: probably many of the features mentioned here are already known in modular algebra, but it has been preferred not to link these results with any established mathematical theory in order to avoid confusions.

Those links, if needed, surely will be constructed in future works.

Relation with the Euler Identity

After the exposition of the main ideas developed until now, a question arises: does this approach shed new light on the prime numbers? The answer is yes. The symmetric structures produced by the SSM actually are also present in the Euler equation:

$$\zeta(s)=1+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\frac{1}{5^s}+\frac{1}{6^s}+....=$$
$$\left(1-2^{-s}\right)^{-1}\left(1-3^{-s}\right)^{-1}\left(1-5^{-s}\right)^{-1}\left(1-7^{-s}\right)^{-1}....$$

The function in the left is the Riemann Zeta Function. During the deduction of this equality a limit is taken in which the third step is shown here:

$$\left(1-\frac{1}{2^s}\right)\left(1-\frac{1}{3^s}\right)\left(1-\frac{1}{5^s}\right)\zeta(s)=$$
$$1+\frac{1}{7^s}+\frac{1}{11^s}+\frac{1}{13^s}+\frac{1}{17^s}+\frac{1}{19^s}+\frac{1}{23^s}+\frac{1}{29^s}+\frac{1}{31^s}+\ldots$$

The limit consists in repeating infinitely the procedure which produced the last equality[4]. When the first inverse prime after the number one, in the right side, get extremely big, infinity in fact; the Euler identity is proven. But note that in the exemplified step the sequence of denominators are exactly the same as the sequence of numbers corresponding to the third generation. This is not surprising because as it has been mentioned previously; the SSM as well as the Euler identity are nothing else than prime sieve representations. But in Euler's identity deduction, the structure of these sequences has been neglected due to the limit and following an analytical prime representation goal. The kernel of the problem is the successively broken symmetry of the prime sequence, which is notorious in the end result but gives no hint a priori of its origin, if the partially broken symmetries (the generations in the present work nomenclature) are neglected. Then it is important to know exactly how each generation is produced from the previous through the SSM. Due to the generation's primorial periodicity, it is possible to restrict the element extraction to the zone corresponding to the next primorial during the construction (application of SSM) of the next generation. For instance in the construction of the fourth generation from the third generation:

$$\begin{bmatrix}
1 & 31 & 61 & \langle 91\rangle & 121 & 151 & 181 & \ldots \\
\langle 7\rangle & 37 & 67 & 97 & 127 & 157 & 187 & \ldots \\
11 & 41 & 71 & 101 & 131 & \langle 161\rangle & 191 & \ldots \\
13 & 43 & 73 & 103 & \langle 133\rangle & 163 & 193 & \ldots \\
17 & 47 & \langle 77\rangle & 107 & 137 & 167 & 197 & \ldots \\
19 & \langle 49\rangle & 79 & 109 & 139 & 169 & 199 & \ldots \\
23 & 53 & 83 & 113 & 143 & 173 & \langle 203\rangle & \ldots \\
29 & 59 & 89 & \langle 119\rangle & 149 & 179 & 209 & \ldots
\end{bmatrix}$$

$$n^{4}_{(2,3,5,7)} = 210.n + \begin{pmatrix} 1 \\ 11 \\ 13 \\ 17 \\ 19 \\ 23 \\ 29 \\ 31 \\ 37 \\ 41 \\ 43 \\ 47 \\ 53 \\ 59 \\ 61 \\ 67 \\ 71 \\ 73 \\ 79 \\ 83 \\ 89 \\ 97 \\ 101 \\ 103 \\ 107 \\ 109 \\ 113 \\ 121 \\ 127 \\ 131 \\ 137 \\ 139 \\ 143 \\ 149 \\ 151 \\ 157 \\ 163 \\ 167 \\ 169 \\ 173 \\ 179 \\ 181 \\ 187 \\ 191 \\ 193 \\ 197 \\ 199 \\ 209 \end{pmatrix}$$

The enclosed numbers correspond to the sequence η^* of the fourth generation. Note that the branches have a cut-off just before the first period (2.3.5.7 = 210) because further elements are un-interesting; they follow a repetitive pattern due to the periodicity. Note also that the new pivot number is 7, and there are 7 columns in the matrix. But one element will be extracted in each row for the next generation, leaving 6. This gives the hint that the seed number in each generation is the product of each pivot number minus one: 1, 2, 8, 48, etc. (the formal demonstration will be published elsewhere). From the previous matrix, once the enclosed elements have been extracted and the two dimensional matrix reduced to one dimension, one obtains the large multi-linear representation in the previous page. All what was written before suggest that is possible to recast the SSM in a form which involves no infinite sequence. This would mean an equation which is closer than ever to a prime generating formula. In fact this goal can be achieved and it is done in the next section.

Discrete Scale Invariance and Internal Sieving Leading to a Finite and Constructive Formula for Prime Generation

The SSM is based on element extraction. This is done through η^* in the defining equations of β. But η^* and η are both infinite sequences. In preceding sections strong arguments have been given indicating that all features of the SSM can be reduced to the first period, due precisely to the intrinsic periodicity of this mapping's results. Can a new mapping based in the symmetries from the SSM constructed with finite sets or sequences as argument? The affirmative answer to this question was partially given in the last section when the fourth generation was obtained from the third in just the new generation interval (210). But how should be reduced the element extraction to one period? The solution to this question lies in a new symmetry of the SSM. This symmetry could be expressed as follows: The patterns produced by the segregation (with primorial periodicity and mirror symmetries as already mentioned) between extracted and non-extracted elements of n, after β's successive applications, are the same seen from the perspective of a 1-periodic infinite sequence (the original n) as well as from the perspective of any r-periodic infinite sequence if r is relative prime with the set of all the pivot numbers used during the mappings of β. In particular all the "future" or still not used pivot numbers (which in fact are all the rest of the prime numbers) make sequences which show the same patterns or symmetric structures. In Figure 1, Part B, two examples of this symmetry are shown in the last row. This last definition have all the features of a "Discrete Scale Invariance" (DSI)[7] but one: the known DSI has a preferred scale and the rest of the scales are powers of the fundamental one. In the present case there is no preferred scale and no invariant discrete scale is a power of any other because all are relative primes. The proof of the DSI for the sequences studied in this work will be published elsewhere. This powerful symmetry has dramatic consequences on what is now defined as "internal sieving". The SSM contains a normal sieve which is performed with, and on, infinite sequences. The DSI symmetry allows to find the extracted elements restricted to the first period. As the already extracted elements, by all the previous mappings of β, leaved the same patterns on all the still unused pivot numbers; the pattern of the set of numbers which will be extracted by the next pivot number is known. It is simply the pattern embedded in the sequence of seeds multiplied by the pivot number. For example; in the previous construction of the fourth generation, the extracted numbers for the next generation are: (7, 49, 77, 91, 119, 133, 161, 203) = 7.(1, 7, 11, 13, 17, 19, 23, 29). The finite sequence in the last parenthesis is the set of the seeds from the third generation. Then the Internal Sieving (IS) is defined as the sieving restricted to the first period. The set of the elements which will be extracted is constructed with the product between the seeds from the previous generation and the next pivot number. With these "intra-period" rules the Ω mapping will be defined, acting on finite sequences and producing primes in the form of "used" pivot numbers. Omega acts initially on two numbers {{S}, T} where T is the initial period T=1, and S is the initial seed S=1 (the cero generation). Because 2 is the start and the end of the first branch (2 is the first pivot number), Omega adds to the only seed "1" the period "1" just once, forming the sequence: (S, S+T) = (1, 2). Then the output from the first mapping of omega is (the first generation):

$$\Omega\left(\left\{\{1\},1\right\}\right)=\left\{\{1,2\}-\{2\},1.2\right\}=\left\{\{1\},2\right\}$$

Repeating the procedure on the first generation:

$$\Omega\left(\left\{\{1\},2\right\}\right)=\left\{\{1,3,5\}-\{3\},2.3\right\}=\left\{\{1,5\},6\right\}$$

Here again to the unique seed 1, the period 2 is successively added until the last number (5) is still smaller than the period (6 = 2.3); then 3 times the set of the seeds is extracted. The only element from this last set which is contained in the unique branch is 3. Then the number 3 is extracted leaving the already known bilinear superposition. Observe that it is redundant to maintain a record of the period (the primorial right part) because it can be obtained summing the two extremes of the seeds. However it is kept for the sake of clarity. After the first generation, the action of Ω becomes systematic:

$$\Omega\left(\left\{\{1,5\},6\right\}\right)=\left\{\{1,7,13,19,25,5,11,17,23,29\}-\{5,25\},30\right\}=$$
$$\left\{\{1,7,11,13,17,19,23,29\},30\right\}$$

And in general:

$$\Omega(\sigma)=\left[\sigma+0,\sigma+\sigma_{-1}+1,\sigma+2(\sigma_{-1}+1),\ldots,\sigma+(\sigma_2-1)(\sigma_{-1}+1)\right]-\sigma_2\sigma$$

Where σ_2 and σ_{-1}, are σ's second and last elements, and σ is used as column vector forming the matrix between the brackets. Here again the minus sign after the closing bracket means element extraction. A vectorial version for Ω is given as follows:

$$\Omega(\sigma)=\overline{\left(\sigma\wedge\nu_{\sigma_2-1}\right)\bullet\left(1,\sigma_{-1}+1\right)}-\sigma_2\cdot\sigma$$

Here σ is the vector of seeds, ∧ is the matrix external generalized product (in which the last operation between individual elements is a two dimensional vector instead the usual product). The vector ν is composed from the natural numbers starting at cero until the second seed $\sigma 2$ minus one. The number σ_{-1} is the last seed and the bold point is a scalar product. The line over the entire operation means that the resulting matrix is projected in one dimension ("vectorized") and ordered. The minus sign is a set operation and means element extraction. For example operating with Omega on (1, 5):

$$\Omega\left((1,5)\right)=\overline{\left((1,5)\wedge(0,1,2,3,4)\right)\bullet(1,6)}-5\cdot(1,5)=$$
$$\overline{\begin{pmatrix}(1,0),(1,1),(1,2),(1,3),(1,4)\\(5,0),(5,1),(5,2),(5,3),(5,4)\end{pmatrix}\bullet(1,6)}-(5,25)=$$
$$\overline{\begin{pmatrix}1,7,13,19,25\\5,11,17,23,29\end{pmatrix}}-(5,25)=(1,7,11,13,17,19,23,29)$$

These operations can be resumed in one line of Mathematica ® code:

```
Nest[Complement[Flatten[Outer[List,#,Range[#[[2]]-1].{1,#[[-1]]+1}],#[[2]]*#]&,{1,5},3]
```

This command has already the iterative order Nest embedded as the last operation. Observe that in this example 3 iterations are produced. Note also that the starting generation is the second one; {1, 5} instead the cero generation {1}. This is due to the Ω's singular behaviour when it is started at the cero generation. After the second generation the branch production becomes systematic. A final reflection on Ω; reducing the action of the SSM to the first period of each generation, made evident that the primes are the remnant of a decimating machine. This machine when

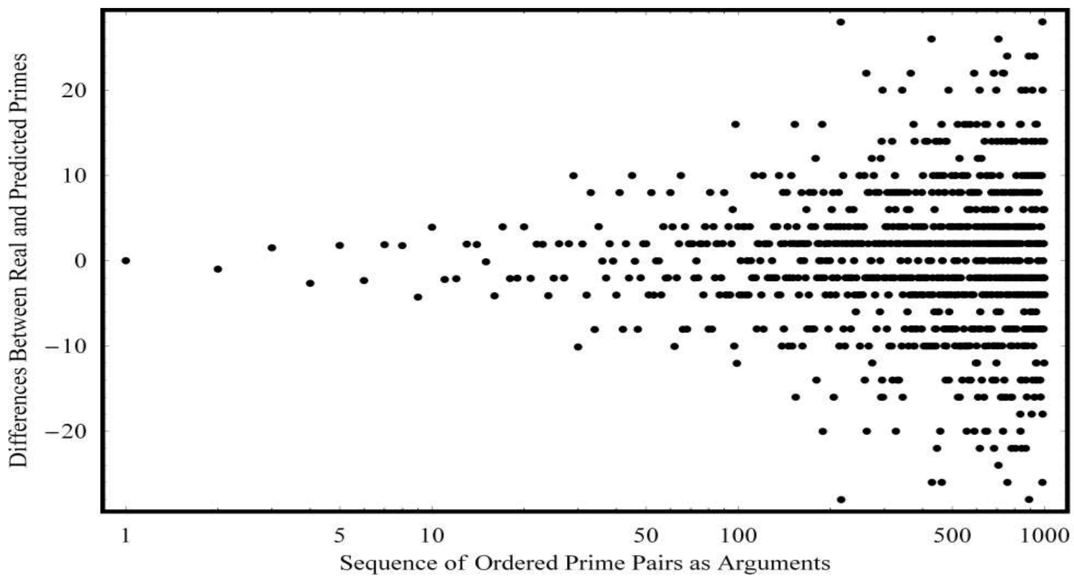

Figure 2 | Second Order Discrete Prime Producing Equation Goodness. The closeness of the results is quantized in stripes below and over the real prime values. The difference values look like entire discrete quantities, but in fact they are fractional values very close to the integers.

applied to the natural numbers eliminates all of them but the number one. The difference between the primes and the no-primes is the "door" through they are leaving the set of the numbers modified by Ω; those which leaves through the last door (activated by the multiplication by 1, which is always in the set of seeds) are primes, those leaving through the rest of the internal sieving, are not. This prime generating decimation machine is self-regulated: If the density of the possible primes is high (low) in any region this will mean that in the next generation the decimation will be high (low), lowering (elevating) in this way the density. This property is notorious at the very first generations where the density of possible primes is the highest ever, leading to a fierce decimation and to a rapid stabilization of the density of primes.

A "Mean Field Theory" for Primes leads to an Improved Prime Number Theorem.

Using some properties from Omega, it is easy to obtain the mean behavior of the prime numbers. Each time Ω is applied; the period is increased by the pivot number p used at that iteration and the number of seeds by (p-1). In the PCI all the seeds are primes and one can make the supposition that their density is the same there and in the zone outside the PCI. Then applying Omega changes the mean density of primes from Δ_{n-1} to Δ_n through the following factors:

$$\Delta_n = \Delta_{n-1}\frac{(p_n - 1)}{p_n} = \Delta_{n-1}\left(1 - \frac{1}{p_n}\right)$$

But Δ_n and Δ_{n-1} can be expressed as the inverse of the difference of two consecutive prime numbers:

$$\Delta_{n-1} = \frac{1}{(p_n - p_{n-1})};\Delta_n = \frac{1}{(p_{n+1} - p_n)}$$

As in a lattice of atoms, between two sites one can count the distance as corresponding to one atom, one prime in this case. Substituting these equations in the previous one and solving for p_{n+1}, one obtains a second order recursive equation for the next prime:

$$p_{n+1} = \frac{p_n(2p_n - p_{n-1} - 1)}{(p_n - 1)}$$

The accuracy of the former equation is variable, but in a fraction of the cases it gives the next prime almost exactly. In Figure 2 the differences between the true primes and the numbers produced through the former formula with the two previous true primes as arguments, are shown for the first 1000 primes. This is a logarithmic plot in the abscissa and shows that the “errors” in the prime prediction are approximately symmetrically distributed around the cero. In a fraction of the trials, quasi-exact (the difference to the true prime is a fraction less than one) solutions are obtained. Besides the first few cases, all the “predicted” primes are almost entire numbers and separated from the real ones by an even almost integer difference (producing the quantized stripes from the graph). The conversion of the former discrete equation in a continuous one with derivatives; has the goal of reproducing the results of the Prime Number Theorem and even to improve them. If the function γ(n) is able to generate sequentially the prime numbers with natural numbers as arguments, then $\gamma(n) = p_n$ where p_n is the nth prime. This means that the minimal measurable difference between two arguments has to be 1. Then the best approximation for a derivative is:

$$\gamma'(\mathrm{n}) = \frac{\gamma(n+1) - \gamma(n)}{1} = p_{n+1} - p_n$$

Substituting these results in the discrete equation of prime densities, the following equation is obtained:

$$\gamma'(\mathrm{n\text{-}1}) = \gamma'(\mathrm{n})\left(1 - \frac{1}{\gamma(\mathrm{n})}\right)$$

Calculating the second derivative of γ in a similar way:

$$\gamma''(\mathrm{n}) = \gamma'(\mathrm{n}) - \gamma'(\mathrm{n\text{-}1})$$

Substituting this last result in the former equation, the following simple, non-linear differential equation is obtained (these are approximations to γ and are called γ_o) :

$$\gamma_o''(n) = \frac{\gamma_o'(n)}{\gamma_o(n)}$$

The general solution to this equation can be found in terms of exponential integrals:

$$\gamma_o(\mathrm{n}) = e^{\left[-A + Ei^{(-1)}\left(e^A (n+B)\right)\right]}$$

Where A and B, are the integration constants and Ei $^{(-1)}$ is the inverse function of the exponential integral given by:

$$Ei(x) = \int_{-x}^{\infty} \frac{e^{-t}}{t} dt$$

If the constants are set to cero; A = B = 0 then the simplest version is obtained:

$$\gamma_o(\mathrm{n}) = e^{\left[Ei^{(-1)}(n)\right]}$$

The function π (n) is defined as the number of primes which exist in the interval (0, n)[8]. If p_n is the nth prime, then there are n primes in the interval (0, p_n). But $p_n = \gamma(n)$, applying π to both sides:

$$\pi(\mathrm{p_n}) = \pi(\gamma(n)) = n$$

Then π and γ, are the inverse of each other. Knowing the simplest approximate solution for γ, one can obtain the correspondent π inverting it:

$$n = Ei\left[\ln(\gamma_o)\right] = -\int_{-\ln(\gamma_o)}^{\infty} \frac{e^t}{t} dt$$

Substituting t = - ln[z] the following equation is obtained:

$$n = \int_0^{\gamma_o} \frac{1}{\ln(z)} dz = Li(\gamma_o)$$

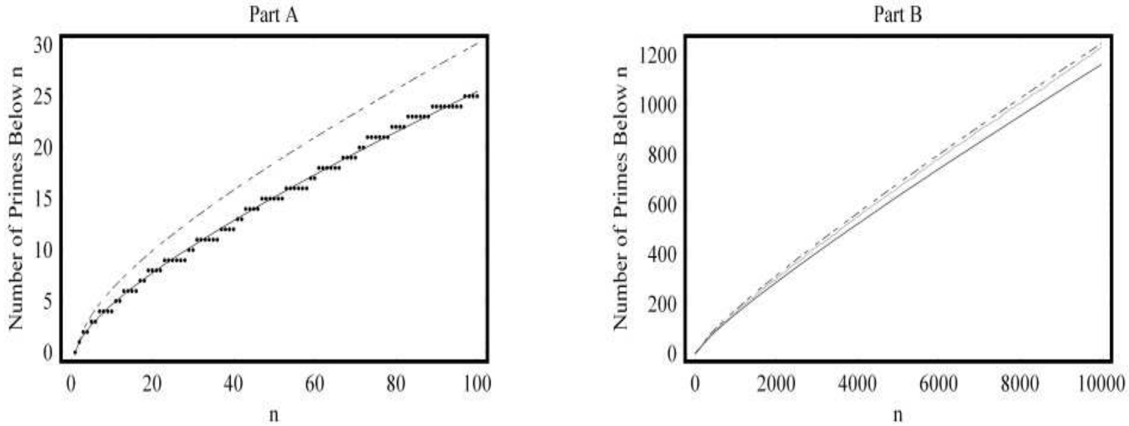

Figure 3 | Comparison of the Old and New Approximations to π(n). The values of π(n) are shown as isolated points, the new approximation as a continuous line and Li(x) as a doted line. In the Part B; Li(x) remains close to π(n), whereas the new approximation diverges slowly.

But n is the number of primes less or equal than $\gamma = p_n$, then n is $\pi(p_n)$, which constitutes the first constructive demonstration of the Prime Number Theorem (the right side defines the Logarithm Integral function). Even more, it is the first construction of an approximate differential equation for the prime numbers. This is the simplest version of the solution. If the constants A and B are left free, then an improved Prime Number Theorem is obtained:

$$\pi_\Omega(x) = \frac{Ei\left[\ln(x)+A\right]}{e^A} - B = \frac{1}{e^A}\int_{-\ln(x)-A}^{\infty} \frac{e^{-t}}{t} dt - B$$

This equation has a curious behavior depending on the values of A and B. With appropriate values given to A and B (A = .52; B = -0.71, obtained through the best fit for the 100 first primes) the curve on the graph of Figure 3, Part A, is remarkably closer to π than the Logarithm Integral. But this closeness is limited; if one looks deeper the function crosses π (n) and then remains far away from it. This means that certain values of A, produces almost exact results for bounded regions of the argument. This confirms the fact that the previous differential equation is an approximation to the real differential equation of γ(n). For a 100 times bigger domain window the approximation remains below of the real π (n) after certain point, as it shown in Figure 3, Part B. The approximate differential equation for γ can be transformed on an approximate differential equation for π just applying the chain rule to the inverse function:

$$\frac{d^2\pi_o(y)}{dy^2} + \frac{\left(\frac{d\pi_o(y)}{dy}\right)^2}{y} = 0$$

This last equation has the same solution as the previous.

On the Infinitude of the Twin Primes

As it was mentioned, the successive generations are finer and finer approximations to the prime numbers. In this sense the first generation states that all the primes are odd numbers. It also states that all the rest of the primes after the number two should be sieved out from an infinite sequence of twin primes. The first generation is the cradle of all twin primes. But in the first generation, version Omega, the seed 1 doesn’t express the presence of twin primes. Just in the second generation with the seeds 1 and 5 the twin primes are explicitly present in between the seeds (in a sort of "periodic boundaries" the branches corresponding to 5 first and then to 1 have all their elements separated by two units). Then, all twin primes bigger than the pair (3, 5), are contained in the sequence of pairs (6.n + 5, 6.n + 7). Only in the third generation the existence of twin primes is evident in the seed structure: 1, 7, **11, 13**, *17*, *19*, 23, 29. The first twin pair is shown in bold and the second in italic. As in the previous generation the limiting branches corresponding to the seeds 1 and 29 also constitutes a pair of twin prime producing branches. If the internal sieving is neglected, these number of potential twin primes would grow at a similar rate as the branches, with (p-1)#. The twin prime structure from the third generation ensures that all the forthcoming twin primes will belong to one of the three sets: (30.n + 11, 30.n + 13), (30.n + 17, 30.n + 19) or (30.n + 29, 30.n + 31). During the synthesis of a new generation each branch looses one member due to the internal sieving. Then a simple arithmetic proves that the number of twin primes is infinity; starting at the third generation, the number of possible twin primes will grow in each generation with a factor p-2, where p is the generation's pivot number. This is because each branch grows like the pivot number minus one (due to the internal sieving). But each member of a twin prime pair belongs to different branches, then the internal sieving eliminates 2 twins instead of one (repetition of the internal sieving in the same twin pair is impossible because this would mean that two integers are far apart in lees than one unit). This last sentence, expressed in equations, means that the mean density of twin primes will change as a function of the generations in the following way:

$$\Delta_\tau^{n+1} = \Delta_\tau^{n}\left(\frac{p_n - 2}{p_n}\right)$$

Where Δ_τ is the mean density of twin primes and the superscript indicates its generation, being p_n the generation's pivot number. This last equation shows that the twin primes never fade out: By construction of either β or Ω the number of possible twin primes in each generation is never cero. If the pivot numbers limits to infinity, the mean density of twin primes in each generation tends to be the same non cero quantity. As well as in the case of the plain prime numbers, this discrete equation can be converted in a differential equation:

$$\frac{\tau(n)''}{\tau(n)'} = \frac{2}{\gamma_o(n)}$$

Where τ(n) represents a hypothetical function which gives the nth prime number belonging to a twin

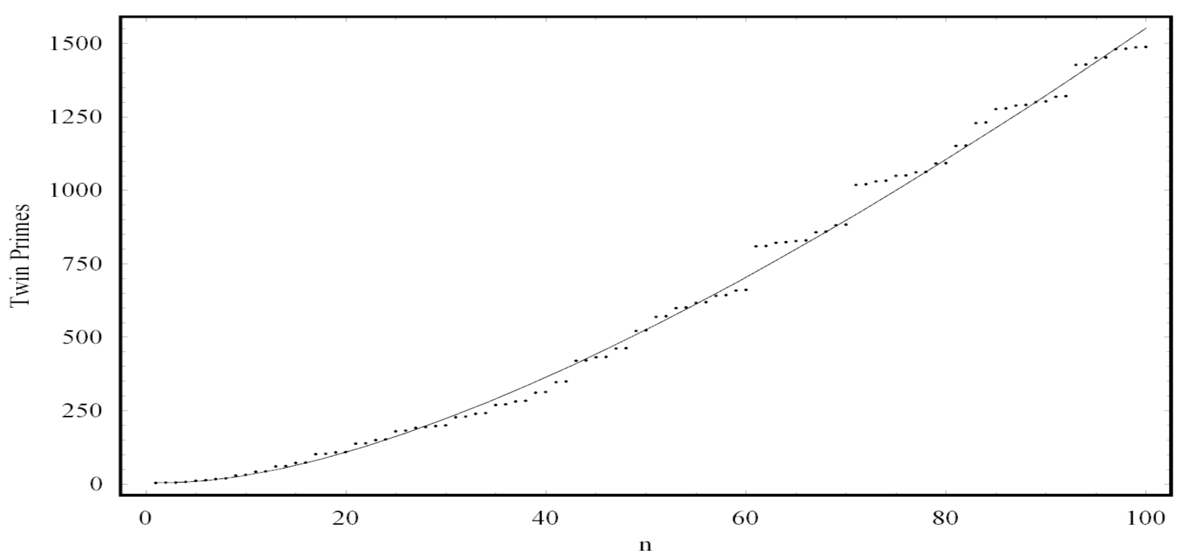

Figure 4| Goodness of the τ(n) Approximation to the nth Twin Prime. The real positions of the Twin Primes are shown with discrete points and the approximation with a continuous line.

prime and $\gamma_o(n)$ is already known. Using the simplest solution for: $\gamma_o(n) = Ei^{(-1)}(n)$; the following solution for τ is obtained:

$$\tau(n) = A + B.e^{Ei^{(-1)}(n)}.\left(Ei^{(-1)}(n) - 1\right)$$

Where A and B are the integration constants. Setting A = 4.5 and B = 0.61 the curve shown in Figure 4 is produced. In the same plot, the true ordered values of the prime numbers belonging to twin primes are shown as points. If the general solution for $\gamma_o(n)$ is used, then 4 integrations constants are present:

$$A + \frac{B}{2}\left(C + n + \frac{e^{C + Ei^{(-1)}\left(-e^{-D}(n+C)\right)}\left(1 + Ei^{(-1)}\left(-e^{-D}(n+C)\right)\right)}{Ei^{(-1)}\left(-e^{-D}(n+C)\right)^2}\right)$$

Finally it should be noted that it is impossible to invert the solution for $\tau(n)$ (looking for an equivalent π function of the twin primes) due to the multiplicity of the function $Ei^{(-1)}(n)$ in it.

Conclusions

In none of the classical papers about the prime numbers, it is stated from where comes the insight to hypothesize and then to prove that the function Li(x) is a good approximation for $\pi(x)$. There were no reasons beyond the experimental fact and its plotting in a logarithmic scale[8]. A direct consequence of this was the extreme difficulty to prove anything about this fact and about the prime numbers in general. A lack of a working model for prime numbers has built the general sense that they were beyond the reach of the human intellectual powers[2]. Without such a model the mathematicians relied on feeble hints hidden in the apparently not understandable structure of the prime numbers. This made the few available proofs into huge efforts very difficult to elaborate and practically just on the level of some specialists. In this work a simple and beautiful model of the prime numbers is constructed starting from the geometrical symmetries formed during the sieving process. These symmetries are related with the relative position of the prime numbers and not with the numbers itself or with their restricted divisibility. The study of this restricted divisibility, as a fundamental fact about the Primes, was another misleading path. The restricted divisibility is more a consequence than a cause in the set of features which characterize the prime numbers. Following the right path of this model to its logical consequences, it was possible to elaborate a theory for the mean quantities of the primes. This theory leads to a second order discrete and approximate equation and to a second order continuous differential equation for the primes. This is the first time that such equations are presented. The simplest solution of this equation produces the Li(x) approximation as well as a general solution with two constants associated with the second order differential equation. The behaviors of the solutions under the adjustment from these constants demonstrate that the theory elaborated in this work is incomplete, and that even deeper symmetries are still lurking in the structure of the prime numbers. Perhaps the biggest advance of this work is the elaboration of a recursive and finite formula for the production of prime numbers. In this formula the infinitude of the twin primes is explicitly shown through its structure. This leads, in combination with the formerly mentioned differential equation for the primes; to an approximation of a differential equation for the twin primes which is analytically solvable. All these successes show that the model for the prime numbers presented in this work is the correct one. Furthermore, some new results, which were reported as breakthroughs, are simply natural consequences of the present model. For example the existence of arithmetic progressions of primes is obligatory due to the structure of the branches[9]. This model is straightforwardly connected with the Riemann Hypothesis through the Euler identity. However, as was shown in this work, the presence of the unique argument “s” in the Riemann function seems to be irrelevant to the structure of the primes, at last in relation with their mean properties. In this reasoning, the zeros related to this argument, are also irrelevant. In fact the symmetric structures formed in β and Ω are the same which form the Euler identity with s = 1 or with any other value. It is the author's belief that this work has just scratched the surface of the Primes real structure. Following the lead set by it, it will be possible to finally find an exact solution for the prime problem. Some of the deeper symmetries are already envisioned by the author and will be published soon elsewhere. Finally a possible bonus: the already discovered symmetries suggest that unveiling the complete prime number structure would be the same as obtaining the unique natural counting system, not binary, not decimal, but prime.

Acknowledgements

The author wishes to thank Dr. Eduardo Greaves for the revision of this document and also the creators of the wonderful tool Mathematica ®. Without this amazing piece of software it would be impossible to do this work.

1 Guy, R.K. *Unsolved Problems in Number Theory*. 2nd ed. (Springer-Verlag, New York, 1994).

2 Watkins, M. R. Number Theory and Physics Archive. http://www.secamlocal.ex.ac.uk/~mwatkins/zeta/physics.htm .

3 Riemann, B. Über die Anzahl der Primazahlen unter einer Gegeben Gröβe, *Gesammelte Matematische Werke* **VII**, 2nd ed. (Teubner, Leipzig, 1982).

4 Derbyshire, J. *Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematic.*,1st ed. (Joseph Henry Press,Washington D.C., 2003) p. 102.

5 Peitgen H., Jürgens H., Saupe D. *Chaos and Fractals: New Frontiers of Science*. 1st ed. (Springer-Verlag, New York, 1992) Chapter 1.

6 Wolf M., Multifractality of Prime Numbers. *Physica A* **160,** 24-42 (1989).

7 Sornette D., Discrete scale invariance and complex dimensions. *Phys. Rep.* **297,** (1998) p. 239.

8 Lagarias J. C., Odlyzko A. M., Computing pi(x): An Analytic Method. *J. Algorithms,* **8** (1987) p. 173-191.

9 Grenn B., Tao T., The Primes Contain Arbitrarily Long Arithmetic Progressions. http://arxiv.org/PS_cache/math/pdf/0404/0404188.pdf