

Article

Combinatorial Models of the Distribution of Prime Numbers

Vito Barbarani ^{1,†}  0000-0002-4953-7162,¹ European Physical Society; vito.barbarani@gmail.com

† Current address: via Cancherini 85, 51039 Quarrata, Italy

Abstract: This work is divided into two parts. In the first one the combinatorics of a new class of randomly generated objects, exhibiting the same properties as the distribution of prime numbers, is solved and the probability distribution of the combinatorial counterpart of the n -th prime number is derived, together with an estimate of the prime-counting function $\pi(x)$. A proposition equivalent to the Prime Number Theorem (*PNT*) is proved to hold, while the equivalent of the Riemann Hypothesis (*RH*) is proved to be false with probability 1 (*w.p. 1*) for this model. Many identities involving Stirling numbers of the second kind and harmonic numbers are found, some of which appear to be new. The second part is dedicated to generalizing the model to investigate the conditions enabling both *PNT* and *RH*. A model representing a general class of random integer sequences is found, for which *RH* holds *w.p. 1*. The prediction of the number of consecutive prime pairs, as a function of the gap d , is derived from this class of models and the results are in agreement with empirical data for large gaps. A heuristic version of the model, directly related to the sequence of primes, is discussed and new integral lower and upper bounds of $\pi(x)$ are found.

Keywords: set partitions; Stirling numbers of the second kind; harmonic numbers; prime number distribution; Riemann hypothesis; Gumbel distribution

1. Introduction. The bingo bag of primes.

This work aims to investigate two main objectives: what random means in the case of the distribution of prime numbers and the general conditions ensuring that what we know as *Prime Number Theorem* and *Riemann Hypothesis* also hold in the case of a generic random sequence of natural numbers. To achieve these goals, we'll make use of models based on combinatorics and probability theory. Of course, this is not the first time that models are used to derive new conjectures and theorems about the prime sequence. In the mid-1930s, Cramér [14]-[15] proposed the first stochastic model of the sequence of primes, as a way to formulate conjectures about their distribution. This model can be briefly summarized as follows.

Denote by $\{X_n, n = 1, 2, 3, \dots\}$ a sequence of independent random variables X_n , taking values 0 and 1 according to

$$\text{Prob}\{X_n = 1\} = \frac{1}{\ln(n)}, n > 2$$

(the probability may be arbitrarily chosen for $n = 1, 2$). Then the random sequence of natural numbers

$$S = \{n : X_n = 1, n > 1\}$$

is a stochastic model of the sequence of primes. This means that if $\Omega = \{\omega\}$ is the set of all possible sequences ω realizations of S , then the prime sequence, say ω_p , is obviously an element of Ω . We can therefore study the properties of the particular sequence ω_p , through the methods of probability theory applied to the sample space Ω , and conjecture with Cramér that if a certain property holds in Ω with probability 1 (*w.p. 1* in that follows), that is for almost all the sequences, the same property holds for

the prime sequence ω_p . Obviously the results of these conjectures lay heavily on the assumption about the probability of the event $\{X_n = 1\}$, which connects the model with the distribution of prime numbers as stated by the *Prime Number Theorem (PNT)*

$$\pi(x) \sim \frac{x}{\ln(x)} \sim Li(x) = \int_2^x \frac{dx}{\ln(x)} \quad (1)$$

$\pi(x)$ being the prime-counting function. We see from (1) that $\frac{1}{\ln(n)}$ can be interpreted as the expected density of primes around n , or the probability that a given integer n should be a prime.

As reported by Granville [25] in his review article dedicated to the work of Cramér in this area, he imagined building the set of random sequences S through a game of chance consisting of a series of repeated draws (independent trials) from an ordered sequence of urns (U_1, U_2, U_3, \dots) , such that the n -th urn contains $\ln(n)$ black balls and 1 white ball, starting from $n = 3$ forward. In this game the event $\{X_n = 1\}$ is then realized by drawing the white ball from the urn number n .

The heuristic method based on Cramér's model has been playing an essential role since the mid-1930s to the present day when formulating conjectures concerning primes. The first serious difference, between the prediction of the model and the actual behavior of the distribution of primes, was found in 1985 [32]. As pointed out in [25], this difference arises because of the assumptions made by heuristic procedures based on the probabilistic model: the hypothesis, when applying the 'Sieve of Eratosthenes', that sieving by the different primes $\leq \sqrt{n}$ are independent events and Gauss's conjecture about the density of primes around n .

The problem was known well before Cramér's work. In their 1923's paper [27] Hardy and Littlewood, while discussing some asymptotic formulas of Sylvester and Brun, containing erroneous constant factors, functions of the number $e^{-\gamma}$ (γ is the Euler-Mascheroni constant), observe that "any formula in the theory of primes, deduced from considerations of probability, is likely to be erroneous just in this way". They explain this error arises because of the different answers to the question about the chance that a large number n should be prime. Following the *PNT*, and Gauss's conjecture, this chance is approximately $1/\ln(n)$, while if we consider the chance n should not be divisible by any prime $\leq \sqrt{n}$, assuming independent events for any prime (Hardy and Littlewood don't mention this condition, but it is implicit), this chance is asymptotically equivalent to

$$\prod_{p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right) \sim 2e^{-\gamma} \frac{1}{\ln(n)}, p \text{ prime.}$$

Hence, they conclude, any inference based on the previous formula is "incorrect to the extent of a factor $2e^{-\gamma} = 1.123\dots$ ". More recently new contradictions have been found, between the predictions of the model and the actual distribution of primes, that no Cramér-type model, assuming the hypothesis of independent random variables, can overcome [35].

Cramér's model finds its connection with the distribution of prime numbers on Gauss's conjecture; hence it assumes the *PNT* as a priori hypothesis. In this work, a new model is proposed, based on a class of combinatorial objects I call *First Occurrence Sequences (FOS)*, reproducing the stochastic structure of the prime sequence as a result of the straightforward random structure of the model, based on a sequence of independent equally probable trials. The *PNT* emerges however, without any a priori *ad hoc* assumption about the probability of events. Perhaps the best way to summarize the aim of this work is that it tries to answer the question Granville cites at the beginning of his paper [25]: "It is evident that the primes are randomly distributed but, unfortunately, we don't know what 'random' means."¹

¹ The quote is of Prof. R. C. Vaughan.

Following Cramér's method, we can describe the pinpoints of our model as a game of chance. Suppose we have an urn with $n = 2$ balls of different colors, say black (B) and white (W), and let's consider the following sampling with replacement game: choose a ball, say B and replace it; repeat drawing until the white ball W occurs for the first time. In general, the game ends when you draw the ball with a color other than the first ball drawn. The sequences of B and W balls we get from this game are *FOS* of different lengths k . Some examples:

$BBW, k = 3;$

$BW, k = 2;$

$WWWB, k = 4;$

$WB, k = 2.$

Assuming each ball has the same probability of being drawn, it is easy to see the average numbers of steps we need to get both colors for the first time, or the average length of *FOS* when $n = 2$, is equal to 3. If we repeat this simple game with different values of n (number of colors), and the rule that the game ends when all colors appear in the sequence for the first time and compare the average length of *FOS* of order n , say $L'_{n/n}$, with the prime number p_n of the same order, we would find the results reported in Table 1 for $n = 2, 3, 4, 5$ and 6.

Table 1. *FOS* average length and primes p_n .

n	$L'_{n/n}$	p_n
2	3	3
3	5.5	5
4	8.333	7
5	11.416	11
6	14.7	13

The correlation between *FOS* average lengths and the sequence of primes is a general property of the model. Indeed assuming each colored ball has the same probability of being drawn, and every sampling step with replacement is independent of each other, we can easily find

$$L'_{n/n} = n \sum_{k=1}^n \frac{1}{k} \sim n \ln n. \quad (2)$$

The above equation shows the close connection existing between the combinatorial objects I have called *FOS* and the primes. In particular the average length of the n order *FOS*, obtained through the sampling with replacement game from an urn with n colored balls, is asymptotic to the expression of the n -th prime obtained as a consequence of the *PNT*. In this sense, the n -th order *FOS* can be defined as the combinatorial counterpart of the n -th prime number.

The following sections are devoted to developing of the theory of *FOS* and their analogies with the distribution of primes. In Section 2 a formal definition of this kind of objects is given, together with the probability spaces we need to define probability relations and probability recursive formulas. In Section 3 the combinatorics of *FOS* is completely solved, demonstrating its close connections with set partitions and Stirling numbers of the second kind. In Section 4, the probability equations derived from the combinatorial analysis lead to new identities for Stirling numbers of the second kind and for harmonic numbers. The combinatorial model of primes based on *FOS* is defined together with a discrete estimate $\hat{\pi}(x)$ of the prime-counting function $\pi(x)$, which is studied and numerically tested. We prove the *PNT* holds for this model while the *Riemann Hypothesis* doesn't. In Section 5, the model is generalized to investigate the conditions enabling both the *PNT* and the *Riemann Hypothesis*. We find a general

model of random sequences for which the *Riemann Hypothesis* holds *w.p. 1*, and apply this model to the problem of counting consecutive prime pairs, as function of the gap between successive primes. The theoretical results obtained from the model agree with the empirical data for large gaps and show the importance of the correlation between primes in the case of small gaps. Finally this model is directly related to the sequence of primes, leading to new integral lower and upper bounds of the prime-counting function $\pi(x)$. The last Section 6 summarizes this work's general significance and hints at its links with other research areas, particularly with research on physical models.

2. First Occurrence Sequences. A new class of combinatorial objects.

In this section, we will be concerned about defining probability spaces and *First Occurring Sequences* as events, and deriving some fundamental relations among the probability of certain events, while the evaluation of these probabilities will be treated in Section 3.

2.1. Definitions and probability spaces.

Each FOS can be seen as a particular result of a sampling with replacement random experiment, consisting of a repeated sequence of independent trials with more than one outcome. The proper probability space for such an experiment can be defined as follows (see [41, Lect. 1, 2] and [29, Chap. 10, 13]). Let \mathcal{A}_n be a collection of n distinct symbols

$$\mathcal{A}_n = \{a_1, a_2, \dots, a_n\}, n \geq 2$$

and suppose our experiment of random draws from \mathcal{A}_n is repeated k times. Then an outcome of the compound experiment is a k -length sequence of elements of \mathcal{A}_n

$$\omega_n(k) = \{(x_1, x_2, \dots, x_k) : x_j \in \mathcal{A}_n, j = 1, 2, \dots, k\}, k \geq 1.$$

The probability space of the experiment is the triplet $(\Omega_n^k, \mathcal{F}_n^k, P_n^k)$, where

$$\Omega_n^k = \{\omega_n(k)\}$$

is the sample space of all possible n^k outcome sequences given by the k -fold Cartesian product of \mathcal{A}_n with itself

$$\Omega_n^k = \mathcal{A}_n \times \mathcal{A}_n \times \dots \times \mathcal{A}_n, k \text{ times};$$

\mathcal{F}_n^k is the σ -algebra of subsets of Ω_n^k generated by finite-dimensional cylinder sets C_n of the form

$$C_n(B_1, B_2, \dots, B_k) = \{\omega_n(k) : x_1 \in B_1, x_2 \in B_2, \dots, x_k \in B_k\} \quad (3)$$

with $B_i \in \mathcal{B}$ $i=1, 2, \dots, k$, \mathcal{B} a σ -algebra on \mathcal{A}_n of subsets of \mathcal{A}_n .

To define the probability measure $P_n^k : \mathcal{F}_n^k \rightarrow [0, 1]$ let's consider a collection of numbers $\{p_i, i = 1, 2, \dots, n\}$ such that

$$p_i \geq 0, i = 1, 2, \dots, n$$

and

$$\sum_{i=1}^n p_i = 1.$$

The collection of numbers p_i defines the probability P of any outcome a_i of the single trial experiment and of any subset $B \in \mathcal{B}$ through the equations

$$P[a_i] = p_i \quad (4)$$

$$P[B] = \sum_{a_i \in B} p_i. \quad (5)$$

Assuming each trial is independent, the probability of an outcome $\omega_n(k)$ of the compound experiment is given by

$$P[\omega_n(k)] = \prod_{i=1}^k P[x_i]. \quad (6)$$

The above equation defines a probability distribution on the sample space Ω_n^k , since it is [41, par. 2.3]

$$P[\Omega_n^k] = \sum_{\omega_n(k)} P[\omega_n(k)] = 1$$

and the probability measure P_n^k on the σ -algebra \mathcal{F}_n^k can be generated by P through the formula

$$P_n^k[E] = \sum_{\omega_n(k) \in E} P[\omega_n(k)], \text{ for every event } E \in \mathcal{F}_n^k. \quad (7)$$

I recall here two properties of P_n^k , which will be used throughout this paper.

Remark 2.1. [σ -additivity] Given $C_i \in \mathcal{F}_n^k$, $i = 1, 2, \dots$, and $C_i \cap C_j = \emptyset$ for $i \neq j$, then

$$P_n^k\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} P_n^k(C_i).$$

Remark 2.2. Given any finite-dimensional cylinder set (3) the probability measure P_n^k has the property [41, Corollary 2.1]

$$P_n^k[C_n(B_1, B_2, \dots, B_k)] = \prod_{i=1}^k P[B_i]. \quad (8)$$

where P is the probability distribution defined by (4)-(6) (obviously $\mathcal{A}_n \in \mathcal{B}$ and if $B_i = \mathcal{A}_n$ for some i , it is $P[\mathcal{A}_n] = 1$).

Definition 2.1. A collection $O_i = \{a_{j_1}, a_{j_2}, \dots, a_{j_i}\} \subset \mathcal{A}_n$ is an ordered choice of i distinct symbols from \mathcal{A}_n , ($1 \leq i \leq n$), if $a_{j_l} \neq a_{j_m}$ for $l \neq m$, $l = 1, 2, \dots, i$, $m = 1, 2, \dots, i$.

Among the events of the σ -algebra \mathcal{F}_n^k we are interested in the following ones.

Definition 2.2. We denote with $\mathcal{S}_{i/n}(k)$, k -length sequences with i/n distinct symbols ($1 \leq i \leq n$, $k \geq i$), the following events

$$\mathcal{S}_{i/n}(k) = \{\omega_n(k) = (x_1, x_2, \dots, x_k) : \omega_n(k) \subset O_i \text{ and } O_i \subset \omega_n(k)\}$$

for some i -distinct ordered choice O_i .

Definition 2.3. We denote with $\mathcal{S}'_{i/n}(k)$, k -length First Occurrence Sequences (FOS) with i/n distinct symbols ($1 \leq i \leq n$, $k \geq i$), the events $\mathcal{S}_{i/n}(k)$ such that $x_j \neq x_k$, $j = 1, 2, \dots, (k-1)$.

Hence if we choose among the elementary outcomes in Ω_n^k , those with only i distinct symbols and denote them with $\omega_{i/n}(k)$, the event $\mathcal{S}_{i/n}(k)$ is the subset $\mathcal{S}_{i/n}(k) = \{\omega_{i/n}(k)\}$. If we denote further with $\omega'_{i/n}(k)$ those outcomes in the above subset, where

the i -th distinct symbol is unique (occurs once only) and occupies the last k -th position in the sequence, then the event FOS is the subset $\mathcal{S}'_{i/n}(k) = \{\omega'_{i/n}(k)\}$. The following remarks follow obviously from the above definitions.

Remark 2.3. For every elementary event $\omega_{i/n}(k)$ and $\omega'_{i/n}(k)$, there is one and only one ordered choice O_i replicating the order of the first occurrence of the i distinct symbols in the elementary sequence.

Remark 2.4. In general it is $\mathcal{S}'_{i/n}(k) \subset \mathcal{S}_{i/n}(k)$, the two events coincide when $i = k = 1$ and $i = k = n$. In these cases:

$$\mathcal{S}'_{1/n}(1) = \mathcal{S}_{1/n}(1) = \{a_i, i = 1, 2, \dots, n\}$$

$$\mathcal{S}'_{n/n}(n) = \mathcal{S}_{n/n}(n) = \{\omega_{n/n}(n)\}$$

where $\{\omega_{n/n}(n)\}$ is the group of permutations of the set \mathcal{A}_n .

Remark 2.5.

$$\mathcal{S}'_{1/n}(k) = \emptyset, \text{ if } k \geq 2.$$

Let's now derive the general probability relation between events $\mathcal{S}'_{i/n}(k)$, $\mathcal{S}_{i/n}(k)$ and elementary outcome sequences.

Definition 2.4. Given a single k -length outcome with i/n distinct symbols $\omega_{i/n}(k)$ ($1 \leq i \leq n$, $k \geq i$), let's denote with $Y(\omega_{i/n}(k))$ the subset of i symbols occurring in $\omega_{i/n}(k)$

$$Y(\omega_{i/n}(k)) = \{a_j \in \mathcal{A}_n : a_j \in \omega_{i/n}(k)\}$$

and with $Y^c(\omega_{i/n}(k))$ the complementary set of the remaining $(n - i)$ symbols

$$Y^c(\omega_{i/n}(k)) = \mathcal{A}_n - Y(\omega_{i/n}(k)).$$

An analogous definition applies to the symbol set $Y(\omega'_{i/n}(k))$ and the complementary symbol set $Y^c(\omega'_{i/n}(k))$, of an elementary outcome $\omega'_{i/n}(k)$. Obviously both Y and Y^c are subsets of a σ -algebra \mathcal{B} on \mathcal{A}_n of subsets of \mathcal{A}_n .

Theorem 2.1. The probability of the events $\mathcal{S}'_{i/n}(k)$ and $\mathcal{S}_{i/n}(k)$ as a function of elementary outcomes of the compound experiment and their symbol sets and complementary symbol sets, are given by

$$P_n^k[\mathcal{S}'_{i/n}(k)] = \sum_{\omega_{i-1/n}(k-1)} P[\omega_{i-1/n}(k-1)]P[Y^c(\omega_{i-1/n}(k-1))] \quad (9)$$

$$P_n^k[\mathcal{S}_{i/n}(k)] = \sum_{j=i}^k \sum_{\omega'_{i/n}(j)} P[\omega'_{i/n}(j)](P[Y(\omega'_{i/n}(j))])^{k-j}. \quad (10)$$

Proof. For every sequence $\omega_{i-1/n}(k-1)$, the cylinder set

$$C_n[\omega_{i-1/n}(k-1), Y^c(\omega_{i-1/n})] \subset \mathcal{S}'_{i/n}(k)$$

contains $(n - i + 1)$ FOS $\omega'_{i/n}(k) \in \mathcal{S}'_{i/n}(k)$, and we get the whole event as the union of the disjoint cylinder sets

$$\mathcal{S}'_{i/n}(k) = \bigcup_{\omega_{i-1/n}(k-1)} C_n[\omega_{i-1/n}(k-1), Y^c(\omega_{i-1/n})].$$

From equation (8) the probability of the cylinder set is given by

$$P_n^{k-1}[C_n(\omega_{i-1/n}(k-1), Y^c(\omega_{i-1/n}(k-1)))] = P[\omega_{i-1/n}(k-1)]P[Y^c(\omega_{i-1/n}(k-1))]$$

hence equation (9) follows from the σ -additivity property of the probability measure.

For every sequence $\omega'_{i/n}(j)$ ($i \leq j \leq k$), the cylinder set

$$C_n[\omega'_{i/n}(j), \underbrace{Y(\omega'_{i/n}(j)), Y(\omega'_{i/n}(j)), \dots, Y(\omega'_{i/n}(j))}_{(k-j) \text{ times}}] \subset \mathcal{S}_{i/n}(k)$$

(when $j = k$ the cylinder coincides with $\omega'_{i/n}(k)$), contains i^{k-j} sequences $\omega_{i/n}(k) \in \mathcal{S}_{i/n}(k)$ and the whole event is the union of the disjoint cylinder sets

$$\mathcal{S}_{i/n}(k) = \bigcup_{j=i}^k \bigcup_{\omega'_{i/n}(j)} C_n[\omega'_{i/n}(j), \underbrace{Y(\omega'_{i/n}(j)), Y(\omega'_{i/n}(j)), \dots, Y(\omega'_{i/n}(j))}_{(k-j) \text{ times}}].$$

From equation (8) the probability of the cylinder set is

$$P_n^k[C_n[\omega'_{i/n}(j), \underbrace{Y(\omega'_{i/n}(j)), Y(\omega'_{i/n}(j)), \dots, Y(\omega'_{i/n}(j))}_{(k-j) \text{ times}}]] = P[\omega'_{i/n}(j)](P[Y(\omega'_{i/n}(j))])^{k-j}.$$

Equation (10) follows from the σ -additivity of the probability measure P_n^k .

□

Corollary 2.1.1. Given $i = n$ (the number of distinct symbols is equal to the total number of symbols in \mathcal{A}_n) and $k \geq n$ then

$$P_n^k[\mathcal{S}_{n/n}(k)] = \sum_{j=n}^k P_n^j[\mathcal{S}'_{n/n}(j)].$$

Proof. When $i = n$ the symbol set of FOS $\omega'_{n/n}(j)$, $j \geq i$, becomes

$$Y(\omega'_{n/n}(j)) = \mathcal{A}_n$$

and remembering $P[\mathcal{A}_n] = 1$, the thesis follows immediately through simple manipulations of (10), after noting that the measure P_n^k of cylinder sets

$$C_n[\mathcal{S}'_{n/n}(j), \underbrace{\mathcal{A}_n, \mathcal{A}_n, \dots, \mathcal{A}_n}_{(k-j) \text{ times}}]$$

is equal to P_n^j measure of $\mathcal{S}'_{n/n}(j)$. □

Dealing with FOS in general and problems like the average length of sequences, will require considering an infinite number of independent trials or repetitions of the random sampling with replacement experiment. The probability space $(\Omega_n^k, \mathcal{F}_n^k, P_n^k)$ we have defined above for finite values of k , can be generalized to consider sampling sequences of infinite length from the set \mathcal{A}_n , so that the set of all possible outcomes is Ω_n^∞ , the Cartesian product of countable many copies of \mathcal{A}_n . The product measure [29, Theorem 6.3, pag.141] allows us to define an extended probability space $(\Omega_n^\infty, \mathcal{F}_n^\infty, P_n^\infty)$ where sequences $\mathcal{S}_{i/n}(k)$, for a fixed k , are subsets of Ω_n^∞ defined as

$$\mathcal{S}_{i/n}(k) \times \mathcal{A}_n \times \mathcal{A}_n \times \dots \in \mathcal{F}_n^\infty$$

and the probability measure and its properties remain unchanged [29, Par. 13.1]

$$P_n^\infty[\mathcal{S}_{i/n}(k) \times \mathcal{A}_n \times \mathcal{A}_n \times \dots] = P_n^k[\mathcal{S}_{i/n}(k)] = \sum_{\omega_{i/n}(k)} P[\omega_{i/n}(k)]$$

where $P[\omega_{i/n}(k)]$ is defined in (6). The same obviously applies to $\mathcal{S}'_{i/n}(k)$ events.

If we consider only i/n FOS as outcomes of our experiment, We may define an even more straightforward probability space [29, Example 2, pag.270], say $(\Omega'_{i/n}, \mathcal{F}'_{i/n}, P'_{i/n})$, where the sample space is the countable set

$$\Omega'_{i/n} = \{\omega'_{i/n}(k), k = i, i+1, i+2, \dots\}$$

and the probability of each elementary sample $\omega'_{i/n}(k)$ is defined as in (6), while for every event $E \in \mathcal{F}'_{i/n}$ it is

$$P'_{i/n}[E] = \sum_{\omega'_{i/n}(k) \in E} P[\omega'_{i/n}(k)] = \sum_{k=i}^{\infty} \sum_{\omega'_{i/n}(k) \in E} P[\omega'_{i/n}(k)].$$

Hence the probability of the event $\mathcal{S}'_{i/n}(k)$ is still given by $P'_{i/n}[\mathcal{S}'_{i/n}(k)] = P_n^k[\mathcal{S}'_{i/n}(k)]$.

Remembering the definition of partition of a sample space [41, p.4], the following remarks follow immediately from the definitions above.

Remark 2.6. Given $k \geq n$, the collection of events $\{\mathcal{S}_{i/n}(k) \mid i = 1, 2, \dots, n\}$ is a partition of Ω_n^k . Hence

$$\mathcal{S}_{i/n}(k) \cap \mathcal{S}_{j/n}(k) = \emptyset \text{ for } i \neq j, k \geq n$$

$$\Omega_n^k = \bigcup_{i=1}^n \mathcal{S}_{i/n}(k).$$

Remark 2.7. The collection of events $\{\mathcal{S}'_{i/n}(k) \mid k = i, i+1, i+2, \dots\}$ is a partition of $\Omega'_{i/n}$. Hence

$$\mathcal{S}'_{i/n}(k) \cap \mathcal{S}'_{i/n}(m) = \emptyset \text{ for } k \neq m, k \geq i, m \geq i$$

$$\Omega'_{i/n} = \bigcup_{k=i}^{\infty} \mathcal{S}'_{i/n}(k).$$

2.2. Probability relations.

In order to simplify our expressions, let's adopt the following symbols for the probability of events $\mathcal{S}_{i/n}(k)$ and $\mathcal{S}'_{i/n}(k)$, $1 \leq i \leq n, k \geq i$

Notation.

$$P_n^k[\mathcal{S}_{i/n}(k)] = \phi_{i/n}(k)$$

$$P_n^k[\mathcal{S}'_{i/n}(k)] = \phi'_{i/n}(k).$$

This notation will be used throughout the paper.

In the following we will derive some relations which are valid under the assumption of the general probability distribution (4)-(6).

Note that from Remark 2.4 and 2.5 it follows

$$\phi_{1/n}(1) = \phi'_{1/n}(1) = 1 \quad (11)$$

$$\phi_{n/n}(n) = \phi'_{n/n}(n) \quad (12)$$

$$\phi'_{1/n}(k) = 0, k \geq 2, \quad (13)$$

while from Remark 2.6 and 2.7

$$\sum_{i=1}^n \phi_{i/n}(k) = 1, k \geq n \quad (14)$$

$$\sum_{k=i}^{\infty} \phi'_{i/n}(k) = 1. \quad (15)$$

Assuming $k \geq n$, from (14) and Corollary 2.1.1, which can be written using the new notation

$$\phi_{n/n}(k) = \sum_{j=n}^k \phi'_{n/n}(j), \quad (16)$$

it follows

$$\sum_{i=1}^{n-1} \phi_{i/n}(k) + \sum_{j=n}^k \phi'_{n/n}(j) = 1. \quad (17)$$

Since of course $\sum_{i=1}^{n-1} \phi_{i/n}(k) \rightarrow 0$ as $k \rightarrow \infty$, it is interesting to note that the above equation is a heuristic proof of $\sum_{k=n}^{\infty} \phi'_{n/n}(k) = 1$.

From (14) with $k = n$, (12) and (15) with $i = n$, we can write the following equality

$$\sum_{i=1}^{n-1} \phi_{i/n}(n) = \sum_{k=n+1}^{\infty} \phi'_{n/n}(k). \quad (18)$$

From here on throughout this paper we'll assume that the the symbols in the set \mathcal{A}_n are equally probable that is, with reference to the probability distribution (4)-(6), we'll make the following assumption.

Assumption 2.1. (Equally probable symbols)

$$P[a_i] = \frac{1}{n}, i = 1, 2, \dots, n$$

Under the above assumption, the probability (6) of any sequence $\omega_n(k) \in \Omega_n^k$ is given by

$$P[\omega_n(k)] = \frac{1}{n^k}$$

while for the probability of FOS and related events it holds

$$\phi_{i/n}(k) = \frac{\#\mathcal{S}_{i/n}(k)}{n^k} \quad (19)$$

$$\phi'_{i/n}(k) = \frac{\#\mathcal{S}'_{i/n}(k)}{n^k}, \quad (20)$$

where the symbol "# \mathcal{S} " denotes the cardinality of the set \mathcal{S} . Hence equation (12) can be completed in this case as

$$\phi_{n/n}(n) = \phi'_{n/n}(n) = \frac{n!}{n^n}. \quad (21)$$

while, since $\#\mathcal{S}_{1/n}(k) = n$, when $k \geq 1$, it is

$$\phi_{1/n}(k) = \frac{1}{n^{k-1}}. \quad (22)$$

The trivial case $i = 1$ is completely solved by equations (11), (13) and (22), so in the following we shall assume $i \geq 2$.

The following lemma follows directly from the definitions.

Lemma 2.1. *Under Assumption 2.1, the probability of the symbol set and its complementary (see Definition 2.4) of any outcome sequence $\omega_n(k)$ is independent of the sequence and is given by*

$$P[Y(\omega_{i/n}(k))] = \frac{i}{n}$$

$$P[Y^c(\omega_{i/n}(k))] = \frac{n-i}{n}$$

Corollary 2.1.2. *Under Assumption 2.1, from Theorem 2.1 it follows ($2 \leq i \leq n, k \geq i$)*

$$\phi'_{i/n}(k) = \phi_{i-1/n}(k-1) \left(\frac{n-i+1}{n} \right) \quad (23)$$

$$\phi_{i/n}(k) = \sum_{j=i}^k \phi'_{i/n}(j) \left(\frac{i}{n} \right)^{k-j} \quad (24)$$

$$\phi'_{i/n}(k) = \sum_{j=i-1}^{k-1} \phi'_{i-1/n}(j) \left(\frac{i-1}{n} \right)^{k-j-1} \left(\frac{n-i+1}{n} \right) \quad (25)$$

Proof. From Lemma 2.1 we get

$$P[Y^c(\omega_{i-1/n}(k-1))] = \frac{n-i+1}{n}$$

$$P[Y(\omega'_{i/n}(j))] = \frac{i}{n}.$$

By substituting these in (9) and (10) respectively, and remembering the definition (7) of the probability measure, equations (23) and (24) follow.

Equation (25) is obtained simply by putting together the previous results. \square

Note that the last equation (25) provides us with a method to calculate the sequence $(\phi'_{i/n}(k), i = 2, 3, \dots, n)$, starting from (11) and (13), recursively. In particular, for $i = 2$ we get

$$\phi'_{2/n}(k) = \frac{n-1}{n^{k-1}}, k \geq 2. \quad (26)$$

3. Combinatorics of FOS

In the previous section, we have obtained the probability of FOS through the general recursive formula (25) and through closed-form solutions only in a few special cases. The aim of this section is the development of a combinatorial theory of FOS, in order to investigate their connections with other topics of combinatorics, such as set partitions and Stirling numbers of the second kind [37, Ch. 9], then to derive the closed-form probability functions in the next section.

The following simple equation relates the cardinality of the sets $\mathcal{S}_{i/n}(k)$ and $\mathcal{S}'_{i/n}(k)$ defined in Definition 2.2 and 2.3

$$\#\mathcal{S}'_{i+1/n}(k+1) = \#\mathcal{S}_{i/n}(k)(n-i). \quad (27)$$

The combinatorics of objects like sequences $\omega_{i/n}(k) \in \mathcal{S}_{i/n}(k)$ is well known and treated within the more general problem of counting functions between two finite sets in enumerative combinatorics [42, 1.9 p. 71], with applications to set partitions, words and random allocation [20, II.3 p. 106]. The problem of finding $\#\mathcal{S}_{i/n}(k)$ is the same as finding the number of words (or sequences) of length k , over an alphabet of cardinality n , containing i letters (or symbols). As reported in [36, equation (5)], the total number n^k of possible sequences can be obtained through the following sum, involving the Stirling numbers of the second kind $S(k, i)$ (see [42, equation (1.96) p.75 for details])²

$$n^k = \sum_{i=1}^k \binom{n}{i} i! S(k, i). \quad (28)$$

The above equation decomposes the total number of sequences, as the sum over i of the number of such sequences that contain exactly i distinct symbols. The same identity, obtained through classical algebraic methods, is reported in [6, equation (29)]. On the other hand, remembering Remark 2.6, we can also write

$$n^k = \sum_{i=1}^n \#\mathcal{S}_{i/n}(k), \quad (29)$$

hence by comparing (28) and (29), it follows

$$\#\mathcal{S}_{i/n}(k) = \binom{n}{i} i! S(k, i). \quad (30)$$

The same solution is found in [20, II.6 p. 113], under the hypothesis of random allocation of symbols, which is the same as Assumption 2.1. From (30) and (27) we can finally derive an explicit expression for the cardinality of FOS too.

In the rest of this section, we prefer to present an alternative procedure that allows us to derive the number of FOS sequences directly, in order both to find new identities involving the Stirling numbers of the second kind and to show a new combinatorial meaning of these numbers, strictly related to the sequence of primes, which is the main focus of this paper.

3.1. Ordered FOS

Remembering Definition 2.1 and Remark 2.3, let's introduce a new kind of combinatorial objects.

Definition 3.1. Let O_i be an ordered subset of i distinct elements from \mathcal{A}_n , we denote with $\mathcal{S}'_{i/n}(k/O_i)$, k -length ordered First Occurrence Sequences (oFOS) with i/n distinct symbols on the ordered choice O_i , ($1 \leq i \leq n$, $k \geq i$), the events

$$\mathcal{S}'_{i/n}(k/O_i) = \{\omega'_{i/n}(k/O_i)\}$$

where $\omega'_{i/n}(k/O_i)$ is an elementary FOS outcome having O_i as i -distinct ordered choice, replicating the order of the first occurrence of its distinct i symbols.

² I keep here the same symbol with two different spellings, S and S , both to denote the set of events (sequences) and the Stirling numbers of the second kind. The meaning is always clear from the context, in any case when S denotes a set, it is always followed by a subscript. The adoption of the same symbol is justified by the close connection existing between the cardinality of the set of sequences and the Stirling numbers of the second kind.

The following example can help to clarify the above definition.

Example 3.1. Suppose $n = 4$, $i = 3$ and $k = 4$, with $\mathcal{A}_4 = \{1, 2, 3, 4\}$, $O_3 = \{3, 1, 4\}$; then the event $\mathcal{S}'_{3/4}(4/O_3)$ is made up of the following elementary outcomes:

$$\mathcal{S}'_{3/4}(4/O_3) = \{\omega'_{3/4}(4/O_3)\} = \{(3, 3, 1, 4), (3, 1, 3, 4), (3, 1, 1, 4)\}.$$

Remark 3.1. The cardinality of oFOS events depends on parameters i and k only, not on the particular choice of symbols in the ordered set O_i .

The following theorem establishes the relation between FOS and oFOS.

Theorem 3.1. The cardinality of the set of FOS $\mathcal{S}'_{i/n}(k)$ is given by ($1 \leq i \leq n$, $k \geq i$):

$$\#\mathcal{S}'_{i/n}(k) = \frac{n!}{(n-i)!}g(k, i) \quad (31)$$

where $g(k, i)$ is the number of k -length oFOS on an ordered subset of i symbols O_i

$$\#\mathcal{S}'_{i/n}(k/O_i) = g(k, i). \quad (32)$$

Proof. Given the ordered choice of symbols $O_i = \{a_{j_1}, a_{j_2}, \dots, a_{j_i}\}$, any finite subset of i indices

$$M_i = \{m_1, m_2, \dots, m_i\} \subset I_k = \{1, 2, \dots, k\}$$

subject to the constraints

$$1 = m_1 < m_2 < m_3 < \dots < m_i = k,$$

defines a finite-dimensional cylinder set on O_i

$$C_{M_i}(B_1, B_2, \dots, B_k/O_i) = \{\omega_n(k) = (x_1, x_2, \dots, x_k) : x_m \in B_m, m = 1, 2, \dots, k\}$$

through the following assignments:

$$B_m = \{a_{j_1}\} \text{ for } 1 = m_1 \leq m < m_2$$

$$B_m = \{a_{j_2}\} \text{ for } m = m_2$$

$$B_m = \{a_{j_1}, a_{j_2}\} \text{ for } m_2 < m < m_3$$

$$B_m = \{a_{j_3}\} \text{ for } m = m_3$$

$$B_m = \{a_{j_1}, a_{j_2}, a_{j_3}\} \text{ for } m_3 < m < m_4$$

... ..

$$B_m = \{a_{j_{i-1}}\} \text{ for } m = m_{i-1}$$

$$B_m = \{a_{j_1}, a_{j_2}, \dots, a_{j_{i-1}}\} \text{ for } m_{i-1} < m < m_i$$

$$B_m = \{a_{j_i}\} \text{ for } m = m_i = k.$$

Therefore the *oFOS* event is equal to

$$\mathcal{S}'_{i/n}(k/O_i) = \bigcup_{M_i} C_{M_i}(B_1, B_2, \dots, B_k/O_i)$$

and the *FOS* event

$$\mathcal{S}'_{i/n}(k) = \bigcup_{O_i} \mathcal{S}'_{i/n}(k/O_i) = \bigcup_{O_i} \bigcup_{M_i} C_{M_i}(B_1, B_2, \dots, B_k/O_i).$$

As far as the cardinality of the sets from the above equation we get

$$\#\mathcal{S}'_{i/n}(k) = \#\{O_i\}g(k, i).$$

Since

$$\#\{O_i\} = \mathcal{C}(n, i)\mathcal{P}(i) = \binom{n}{i}i!$$

with $\mathcal{C}(n, i)$ i -combinations of n elements, and $\mathcal{P}(i)$ number of permutations of i elements, equation (31) follows. \square

3.2. Counting ordered *FOS*

The problem of finding the number of *FOS* is thus reduced to that of finding the numbers $g(k, i)$. When $i = 1, k \geq i$ we get immediately

$$g(1, 1) = 1, g(k, 1) = 0, k > 1; \quad (33)$$

when $i = 2$, it is

$$g(k, 2) = 1, k \geq 2. \quad (34)$$

Indeed for $i = 2$, there is only one *oFOS*, whatever the value of $k \geq 2$. Note that (33) is a particular case of the general property

$$g(i, i) = 1, i \geq 1. \quad (35)$$

Example 3.2. $i = 2, O_2 = \{1, 2\}$

$$g(2, 2) = 1, \text{ the only oFOS is } (1, 2)$$

$$g(3, 2) = 1, \text{ the only oFOS is } (1, 1, 2)$$

... ..

The calculations are more complex when $i \geq 3$.

Example 3.3. $i = 3, O_3 = \{1, 2, 3\}$, through elementary enumerative combinatorics we find

$$g(3, 3) = 1, \text{ the only oFOS is}$$

$$(1, 2, 3)$$

$$g(4, 3) = 3, \text{ the oFOS are}$$

$(1, 1, 2, 3), (1, 2, 1, 3), (1, 2, 2, 3).$

$g(5, 3) = 7$, the oFOS are

$(1, 1, 1, 2, 3), (1, 1, 2, 1, 3), (1, 1, 2, 2, 3), (1, 2, 1, 1, 3),$

$(1, 2, 1, 2, 3), (1, 2, 2, 1, 3), (1, 2, 2, 2, 3).$

...

Example 3.4. $i = 4, O_4 = \{1, 2, 3, 4\}$, through elementary enumerative combinatorics we find

$g(4, 4) = 1$, the only oFOS is

$(1, 2, 3, 4)$

$g(5, 4) = 6$, the oFOS are

$(1, 1, 2, 3, 4), (1, 2, 1, 3, 4), (1, 2, 3, 1, 4),$

$(1, 2, 2, 3, 4), (1, 2, 3, 2, 4), (1, 2, 3, 3, 4).$

$g(6, 4) = 25$, the oFOS are

$(1, 1, 1, 2, 3, 4), (1, 1, 2, 1, 3, 4), (1, 1, 2, 2, 3, 4), (1, 1, 2, 3, 1, 4), (1, 1, 2, 3, 2, 4),$

$(1, 1, 2, 3, 3, 4), (1, 2, 1, 1, 3, 4), (1, 2, 1, 2, 3, 4), (1, 2, 2, 1, 3, 4), (1, 2, 2, 2, 3, 4),$

$(1, 2, 1, 3, 1, 4), (1, 2, 1, 3, 2, 4), (1, 2, 2, 3, 1, 4), (1, 2, 2, 3, 2, 4), (1, 2, 1, 3, 3, 4),$

$(1, 2, 2, 3, 3, 4), (1, 2, 3, 1, 1, 4), (1, 2, 3, 1, 2, 4), (1, 2, 3, 2, 1, 4), (1, 2, 3, 2, 2, 4),$

$(1, 2, 3, 1, 3, 4), (1, 2, 3, 3, 1, 4), (1, 2, 3, 3, 3, 4), (1, 2, 3, 2, 3, 4), (1, 2, 3, 3, 2, 4).$

...

I'll give in the following some results about the calculations of $g(k, i)$ numbers, starting with the next lemma that states a recursive formula.

Lemma 3.1. The number of k -length oFOS on an ordered subset O_i , as function of the cardinality on an ordered subset O_{i-1} , is

$$g(k, i) = \sum_{j=i-1}^{k-1} g(j, i-1)(i-1)^{k-j-1}, \text{ for } 2 \leq i \leq k, \quad (36)$$

starting with $g(k, 1)$, $k \geq 1$, given by (33).

Proof. Equation (36) follows directly from definition 3.1 and remark 3.1, considering that from any j -length oFOS on an ordered set O_{i-1} ($i-1 \leq j \leq k-1$), one can obtain $(i-1)^{k-j-1}$ k -length oFOS on an ordered set O_i having the i -th symbol in the last position k of the sequence.

Note that this lemma holds independently from Assumption 2.1 of equally probable symbols. \square

The result reported below is simply the sum of the first k terms of a geometric series and will be referred to by the next theorem.

Lemma 3.2. Given q real, $|q| < 1$, then

$$\sum_{j=0}^k q^j = \frac{1 - q^{k+1}}{1 - q}, \quad k \geq 0 \text{ integer.}$$

Theorem 3.2. Let $g(k, i)$ be the number of k -length oFOS on an ordered set of i distinct elements. Then the following statements hold:

$$g(k, i) = \sum_{j=1}^{i-2} a_{i,j} [(i-1)^{k-i+1} - j^{k-i+1}], \quad i \geq 3, k \geq i \quad (37)$$

with coefficients $a_{i,j}$ defined recursively as

$$a_{i+1,j} = -\left(\frac{j}{i-j}\right) a_{i,j}, \quad i \geq 3, j = 1, 2, \dots, (i-2) \quad (38)$$

$$a_{i+1,i-1} = (i-1) \sum_{j=1}^{i-2} a_{i,j} \quad (39)$$

starting with

$$a_{3,1} = 1.$$

Proof. Let's proceed by induction. Equation (37) holds for $i = 3$, indeed from Lemma 3.1 remembering (34) we get

$$g(k, 3) = \sum_{j=2}^{k-1} 2^{k-j-1} = 2^{k-3} \sum_{j=0}^{k-3} \frac{1}{2^j}, \quad k \geq 3,$$

hence from Lemma 3.2

$$g(k, 3) = 2^{k-2} - 1.$$

Let's now prove that if (37) is true for i , $i \geq 3$, then it is true for $(i+1)$. Substituting (37) in (36) for $g(k, i+1)$ we get

$$g(k, i+1) = \sum_{s=i}^{k-1} \sum_{j=1}^{i-2} a_{i,j} [(i-1)^{s-i+1} - j^{s-i+1}] i^{k-s-1}$$

from which, after some manipulations, one gets

$$g(k, i + 1) = i^k \sum_{j=1}^{i-2} \sum_{s=i}^{k-1} \left[\frac{a_{i,j}}{(i-1)^i} \left(\frac{i-1}{i} \right)^{s+1} - \frac{a_{i,j}}{j^i} \left(\frac{j}{i} \right)^{s+1} \right].$$

The single sum

$$\begin{aligned} & i^k \sum_{s=i}^{k-1} \left[\frac{a_{i,j}}{(i-1)^i} \left(\frac{i-1}{i} \right)^{s+1} - \frac{a_{i,j}}{j^i} \left(\frac{j}{i} \right)^{s+1} \right] = \\ & = i^k \left[\frac{a_{i,j}}{(i-1)^i} \left(\frac{i-1}{i} \right)^{i+1} \sum_{s=0}^{k-i-1} \left(\frac{i-1}{i} \right)^s - \frac{a_{i,j}}{j^i} \left(\frac{j}{i} \right)^{i+1} \sum_{s=0}^{k-i-1} \left(\frac{j}{i} \right)^s \right] \end{aligned}$$

after the application of Lemma 3.2

$$\begin{aligned} \sum_{s=0}^{k-i-1} \left(\frac{i-1}{i} \right)^s &= \frac{1 - \left(\frac{i-1}{i} \right)^{k-i}}{1 - \left(\frac{i-1}{i} \right)} \\ \sum_{s=0}^{k-i-1} \left(\frac{j}{i} \right)^s &= \frac{1 - \left(\frac{j}{i} \right)^{k-i}}{1 - \left(\frac{j}{i} \right)} \end{aligned}$$

and some manipulations, can be written as

$$\begin{aligned} & i^k \sum_{s=i}^{k-1} \left[\frac{a_{i,j}}{(i-1)^i} \left(\frac{i-1}{i} \right)^{s+1} - \frac{a_{i,j}}{j^i} \left(\frac{j}{i} \right)^{s+1} \right] = \\ & = a_{i,j}(i-1)(i^{k-i} - (i-1)^{k-i}) - a_{i,j} \left(\frac{j}{i-j} \right) (i^{k-i} - j^{k-i}). \end{aligned}$$

Finally we can write

$$g(k, i + 1) = \sum_{j=1}^{i-2} [a_{i,j}(i-1)(i^{k-i} - (i-1)^{k-i}) - a_{i,j} \left(\frac{j}{i-j} \right) (i^{k-i} - j^{k-i})]$$

$$g(k, i + 1) = (i-1)(i^{k-i} - (i-1)^{k-i}) \sum_{j=1}^{i-2} a_{i,j} - \sum_{j=1}^{i-2} a_{i,j} \left(\frac{j}{i-j} \right) (i^{k-i} - j^{k-i})$$

hence

$$g(k, i + 1) = \sum_{j=1}^{i-1} a_{i+1,j} (i^{k-i} - j^{k-i})$$

with $a_{i+1,j}$ given by (38), $j = 1, 2, \dots, i-2$, $a_{i+1,i-1}$ given by (39). \square

The last theorem gives us a method to calculate each coefficient of the row $i + 1$ from the coefficients of the previous row i . The following corollary state an alternative way to calculate the last term of each row, as function of the other coefficients of the same row.

Corollary 3.2.1. *It is*

$$a_{i+1,i-1} = 1 - \sum_{j=1}^{i-2} (i-j)a_{i+1,j}, \quad i \geq 3. \quad (40)$$

Proof. The result follows directly from (35) and (37):

$$\sum_{j=1}^{i-2} a_{i,j}(i-j-1) = 1, \quad i \geq 3,$$

$$(i-1) \sum_{j=1}^{i-2} a_{i,j} = 1 + \sum_{j=1}^{i-2} ja_{i,j}.$$

Hence, remembering (38) and (39), after some manipulations, one gets (40). \square

The first rows of coefficients $a_{i,j}$ are reported in Table 2

Table 2. Coefficients $a_{i,j}$, $3 \leq i \leq 8$.

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
$i = 3$	1					
$i = 4$	$-\frac{1}{2}$	2				
$i = 5$	$\frac{1}{6}$	-2	$\frac{9}{2}$			
$i = 6$	$-\frac{1}{24}$	$\frac{4}{3}$	$-\frac{27}{4}$	$\frac{32}{3}$		
$i = 7$	$\frac{1}{120}$	$-\frac{2}{3}$	$\frac{27}{4}$	$-\frac{64}{3}$	$\frac{625}{24}$	
$i = 8$	$-\frac{1}{720}$	$\frac{4}{15}$	$-\frac{81}{16}$	$\frac{256}{9}$	$-\frac{3125}{48}$	$\frac{324}{5}$

Theorem 3.2 and the subsequent corollary state a method of calculating coefficients $a_{i,j}$ by rows. The next proposition establishes another way to find them, proceeding by columns.

Corollary 3.2.2. Given the coefficient $a_{j+2,j}$, head of column j , the successive coefficients in the same column can be calculated as

$$a_{i,j} = \frac{(-j)^{i-j-2}}{(i-j-1)!} a_{j+2,j}, \quad j \geq 1, \quad i \geq j+3. \quad (41)$$

The column head coefficients can be calculated as a function of the previous ones only, through the following formula

$$a_{j+2,j} = j \sum_{l=1}^{j-1} \frac{(-l)^{j-l-1}}{(j-l)!} a_{l+2,l}, \quad j \geq 2, \quad (42)$$

starting with

$$a_{3,1} = 1.$$

Proof. Equation (41) is derived simply through a recursive application of equation (38), from the row index $j+3$ to the required row index i .

By substituting (41) into (39), we obtain the second equation (42). \square

The next theorem states a general non-recursive formula to get the coefficients $a_{i,j}$ directly as a function of the indices i, j .

Theorem 3.3. *The following equation holds for each coefficient $a_{i,j}$*

$$a_{i,j} = (-1)^{i-j-2} \frac{j^{i-3}}{(i-j-1)!(j-1)!}, \quad i \geq 3, j = 1, 2, \dots, (i-2). \quad (43)$$

Proof. First of all, let's prove by induction, the formula holds in the case of column head coefficients. In this case, using the notation of Corollary 3.2.2, equation (43) becomes

$$a_{l+2,l} = \frac{l^{l-1}}{(l-1)!}, \quad l \geq 1. \quad (44)$$

The above equation is true for $l = 1$, since we know it is $a_{3,1} = 1$. Assuming it is true for $l = 1, 2, \dots, j$, from (42) we get the identity

$$j \sum_{l=1}^{j-1} \frac{(-1)^{j-l-1} l^{j-2}}{(j-l)!(l-1)!} = \frac{j^{j-1}}{(j-1)!}, \quad (45)$$

where the term on the left-hand side of the equation is the value of $a_{j+2,j}$ calculated through (42), the one on the right-hand side is the value of the same coefficient given by (44). Then the same identity holds for the next head of column coefficient $a_{j+3,j+1}$; indeed equation (42) leads to the result

$$a_{j+3,j+1} = (j+1) \sum_{l=1}^j \frac{(-1)^{j-l} l^{j-1}}{(j-l+1)!(l-1)!} = \frac{(j+1)^j}{j!}.$$

The proof of (43) in the case of non-head of column coefficients, follows simply from (41) after assuming (44). \square

The identity (45) deserves some more attention: it expresses the normalization property of the FOS probabilities $\sum_{j=n}^{\infty} \phi'_{n/n}(j) = 1$ (see forward equation (76) and Remark 4.1). After some simple manipulations it can be written as

$$\sum_{l=0}^j (-1)^l \binom{j+1}{l} l^j = (-1)^j (j+1)^j, \quad (46)$$

that appears to be "complementary" of the well-known identity (see [23, equation (1.13)])

$$\sum_{l=0}^j (-1)^l \binom{j}{l} l^j = (-1)^j j!. \quad (47)$$

3.3. Ordered FOS and Stirling numbers of the second kind

The solution we have found about the number of elementary sequences of the event $\mathcal{S}'_{i/n}(k/O_i)$, presented in Definition 3.1 as oFOS of length k with i distinct symbols on the ordered choice O_i , allows us to show the equivalence of $g(k, i)$ numbers and Stirling numbers of the second kind. Indeed from equation (37) of Theorem 3.2 it follows

$$g(k, i) = (i-1)^{k-i+1} \sum_{j=1}^{i-2} a_{i,j} - \sum_{j=1}^{i-2} a_{i,j} j^{k-i+1}, \quad i \geq 3, k \geq i$$

and remembering (39) of the same theorem

$$g(k, i) = a_{i+1,i-1} (i-1)^{k-i} - \sum_{j=1}^{i-2} a_{i,j} j^{k-i+1}. \quad (48)$$

Substituting in the above equation $a_{i,j}$ and $a_{i+1,i-1}$ as given by (43) and (44) of Theorem 3.3 we get

$$g(k, i) = \sum_{j=1}^{i-1} (-1)^{i-j-1} \frac{j^{k-2}}{(i-j-1)!(j-1)!}.$$

Since

$$\frac{1}{(i-j-1)!(j-1)!} = \binom{i-1}{j} \frac{j}{(i-1)!}$$

we finally arrive at

$$g(k, i) = \frac{(-1)^{i-1}}{(i-1)!} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} j^{k-1}, \quad i \geq 3, k \geq i. \quad (49)$$

The right-hand side of the previous equation is the explicit formula for the Stirling number of the second kind $S(k-1, i-1)$ (see [37, equation (9.21)]). The equivalence can be extended to values $i = 1$ and $i = 2$, since from equation (34) for $i = 2$ it follows

$$S(k-1, 1) = 1, k \geq 2$$

and for $i = 1$ from (33)

$$S(k-1, 0) = 0, k > 1,$$

$$S(0, 0) = 1.$$

Note that the last equation solves the "bit tricky" case of the Stirling number $S(0, 0)$ [24, p. 258], by calculating it instead of assuming it "by convention" (this assumption is common to all the treatments of the subject, see for example [42, p. 73]). We have thus proved

Theorem 3.4. *Given $g(k, i)$, the cardinality of the oFOS set $S'_{i/n}(k/O_i)$ defined in Definition 3.1, and the Stirling number of the second kind $S(k-1, i-1)$, the following equation holds true*

$$g(k, i) = S(k-1, i-1), \quad i \geq 1, k \geq i. \quad (50)$$

Remark 3.2. *Due to the equivalence established by the previous theorem, equation (36) of Lemma 3.1 can be rewritten as*

$$S(k, i) = \sum_{j=i}^k S(j-1, i-1) i^{k-j},$$

that can be obtained through the repeated application of the recursive equation of the Stirling numbers of the second kind (see [37, equation (9.1)])

$$S(k+1, n) = nS(k, n) + S(k, n-1). \quad (51)$$

When $k = i$ we know it is (see equation (35)) $g(i, i) = 1$, hence from the general formula (49) we get

$$g(i, i) = \frac{(-1)^{i-1}}{(i-1)!} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} j^{i-1} = 1$$

or

$$\sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} j^{i-1} = (-1)^{i-1} (i-1)!$$

which is the identity (47) mentioned above. This identity is thus closely related to the Stirling numbers and to the general property $g(i, i) = S(i - 1, i - 1) = 1, i \geq 1$ (see [6, equation (4) and the references therein]). Note that by substituting (31), with $g(k + 1, i + 1)$ given by (50), into equation (27) we obtain for $\#S_{i/n}(k)$ the same equation (30).

We know Stirling numbers of the second kind have a combinatorial meaning, directly related to the problem of partitioning a set into a fixed number of subsets: $S(n, m)$ counts the number of ways a set of n elements, can be partitioned into m nonempty disjoint subsets ([37, Ch. 9, p. 113], [42, Ch. 1.9, p. 73], see also [24] for a complete combinatorial treatment of the Stirling numbers). The analysis performed in this section highlighted an alternative combinatorial interpretation of these numbers, together with the close connection of *oFOS* with set partitions, that will be shown to be strictly related to the sequence of primes in the next section.

4. First Occurrence Sequences, set partitions and the sequence of primes

In this section, we will delve into the *FOS* sequences as a model of the distribution of prime numbers. The simple *oFOS* sequences can already act as a first-level model of the prime number distribution, as we will see in the next subsection, while the complete model based on *FOS* will be developed at the end of this section. In order to develop the tools to build this model, we continue in the following to deepen the combinatorial implications of Theorem 3.4, exploring some interesting identities involving Stirling numbers of the second kind and harmonic numbers, derived through the probabilities of *FOS*.

4.1. Ordered *FOS* and the Prime Number Theorem

The equivalence (50) stated by Theorem 3.4 and the combinatorial meaning associated with S and g numbers suggest the following analogy with the distribution of primes. Given any integer set of the type

$$I_{k+1} = \{1, 2, 3, 4, \dots, k, k + 1\}, k \geq 2, (k + 1) \text{ prime,}$$

there exists a sequence of $(n + 1)$ successive primes

$$O_{n+1} = \{p_1 = 2, p_2 = 3, \dots, p_n, p_{n+1} = k + 1\} \subset I_{k+1}$$

such that the prime-counting function $\pi(k) = n$. The set O_{n+1} can be viewed as an ordered choice of $n + 1$ integers and the set I_{k+1} as an *oFOS* over O_{n+1} (remember Definition 3.1). Note that the integers i between two consecutive primes, $p_j < i < p_{j+1}$, are multiples of the primes from p_1 to p_j , hence they can be considered as repetitions of these symbols, thus confirming the *oFOS* schema. In the light of this analogy, some classical results of the theory of random partitions of finite sets can be reinterpreted as a general form of the *PNT*, which holds for all *oFOS*-like sequences.

The *oFOS* model of the distribution of prime numbers can be defined as follows. Let consider the number of primes less than or equal to k , represented by the prime-counting function $\pi(k)$, as a random variable $\pi_o(k)$ with probability mass function defined by

$$P_k[\pi_o(k) = n] = \frac{g(k + 1, n + 1)}{\sum_{j=0}^k g(k + 1, j + 1)}, n = 1, 2, \dots, k, \quad (52)$$

hence, from Theorem 3.4

$$P_k[\pi_o(k) = n] = \frac{S(k, n)}{\sum_{j=0}^k S(k, j)}, n = 1, 2, \dots, k. \quad (53)$$

Equation (53) defines a uniform probability distribution on the class Π_k of partitions of a set of k elements, obtained by assigning the same probability B_k^{-1} to each partition $\sigma \in \Pi_k$,

$$P_k(\sigma) = \frac{1}{B_k},$$

where the total number of partitions B_k is equal to the k -th Bell number defined as [37, 9.4, p. 133]

$$B_k = \sum_{j=1}^k S(k, j). \quad (54)$$

The combinatorics of Π_k is studied in [28] and [39]³ in connection with the asymptotic ($k \rightarrow \infty$) distribution of the probability measure P_k .⁴ From this approach it follows that the random variable $\pi_o(k)$ with uniform probability (53), defined as a model of $\pi(k)$, has mathematical expectation and variance asymptotically equal to (see [40, Ch. 4, p. 114-115])

$$\mathbb{E}[\pi_o(k)] = \frac{k}{r}(1 + o(1))$$

$$\text{Var}[\pi_o(k)] = \frac{k}{r^2}(1 + o(1)),$$

where $r = W(k)$, the unique positive solution of $re^r = k$, $W(x)$ being the Lambert function [13] defined in the domain $x \geq -1/e$. Since for the function $W(k)$ we know it is $r \sim \ln k$ as $k \rightarrow \infty$, the above equations say the average and variance of the random variable $\pi_o(k)$ are asymptotic to

$$E[\pi_o(k)] \sim \frac{k}{\ln k} \quad (55)$$

$$\text{Var}[\pi_o(k)] \sim \frac{k}{\ln^2 k}. \quad (56)$$

Equation (55), and the related variance (56), represents the form assumed by the PNT when considering *oFOS*-like sequences with uniform probability distribution of set partitions. This is not the only admissible distribution. About this problem and the general conditions in order that a uniform distribution is obtained, through random allocation algorithms (or random urn models), see [43] and the references therein. The following result (see [40, Theorem 1.1, p. 115]) completes the application of the methods derived from the theory of random partitions of sets, to the *oFOS* model of the prime number distribution. The probability distribution of the normalized random variable

$$\eta_k = \frac{\pi_o(k) - E[\pi_o(k)]}{(\text{Var}[\pi_o(k)])^{1/2}}$$

converges to that of the standard normal distribution as $k \rightarrow \infty$, that is:

$$\lim_{k \rightarrow \infty} P\{\eta_k < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

³ See also [20, Ch. IX, p. 692] for a brief summary of these results.

⁴ Harper points out the methods used are based on the combinatorial implications of probability theory and ascribes to W. Feller (see [19, Ch. X, p. 256]) and V. Goncharov (see [28, and the references therein]), the first applications in this field.

4.2. FOS probabilities and some combinatorial identities involving Stirling numbers

Let's collect in this subsection some formulas about the probability function $\phi'_{i/n}(k)$, which will be referred to in the rest of this paper. Under Assumption 2.1, from equation (31) of Theorem 3.1 the probability of FOS is given by

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} g(k, i), 1 \leq i \leq n, k \geq i. \quad (57)$$

By substituting $g(k, i)$ with equations (37) and (39) of Theorem 3.2, we get the following expressions for $3 \leq i \leq n, k \geq i$

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} \sum_{j=1}^{i-2} a_{i,j} [(i-1)^{k-i+1} - j^{k-i+1}] \quad (58)$$

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} [(i-1)^{k-i} a_{i+1, i-1} - \sum_{j=1}^{i-2} a_{i,j} j^{k-i+1}]. \quad (59)$$

Finally from (43) of Theorem 3.3 the explicit expression for $\phi'_{i/n}(k)$ follows

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} \sum_{j=1}^{i-1} (-1)^{i-j-1} \frac{j^{k-2}}{(i-j-1)!(j-1)!}. \quad (60)$$

If we set $l = i - j - 1$, the above equation can be written

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} \sum_{l=0}^{i-2} (-1)^l \frac{(i-l-1)^{k-2}}{(i-l-2)!!}. \quad (61)$$

When $i = n$ ($n \geq 2, k \geq n$), equation (60) becomes⁵

$$\phi'_{n/n}(k) = \frac{n!}{n^k} \sum_{j=1}^{n-1} (-1)^{n-j-1} \frac{j^{k-2}}{(n-j-1)!(n-1)!}, \quad (62)$$

which, observing that

$$\frac{n!}{n^k (n-j-1)!(n-1)!} = \binom{n-1}{j} \left(\frac{j}{n}\right)^{k-1},$$

can be written as

$$\phi'_{n/n}(k) = \sum_{j=1}^{n-1} (-1)^{n-j-1} \binom{n-1}{j} \left(\frac{j}{n}\right)^{k-1}. \quad (63)$$

Equation (61) when $i = n$ ($n \geq 2, k \geq n$), becomes

$$\phi'_{n/n}(k) = \frac{n!}{n^k} \sum_{l=0}^{n-2} (-1)^l \frac{(n-l-1)^{k-2}}{(n-l-2)!!}, \quad (64)$$

which, observing that

$$\frac{n!}{n^k (n-l-2)!!} = \binom{n-1}{l} \left(1 - \frac{l+1}{n}\right)^{k-1}$$

can be written as

⁵ It's easy to check this equation, and (64) below, applies to the case $n = 2$ too.

$$\phi'_{n/n}(k) = \sum_{l=0}^{n-2} (-1)^l \binom{n-1}{l} \left(1 - \frac{l+1}{n}\right)^{k-1}. \quad (65)$$

Let's now derive the sequence probabilities expressed through the Stirling numbers of the second kind. Remembering the equivalence (50), equation (57) can be written as

$$\phi'_{i/n}(k) = \frac{n!}{(n-i)!n^k} S(k-1, i-1), 1 \leq i \leq n, k \geq i. \quad (66)$$

From this equation and (27), always under the hypothesis of equally probable symbols, hence (19), (20), it is easy to get

$$\phi_{i/n}(k) = \frac{n!}{(n-i)!n^k} S(k, i), 1 \leq i \leq n, k \geq i. \quad (67)$$

The above equations, combined with the general properties of the probability functions (11)-(18), give rise to a series of combinatorial identities involving the Stirling numbers of the second kind. The following corollary reports the less trivial ones.

Corollary 4.1. *The following identities hold true.*

From (14) it follows

$$\sum_{i=1}^n \frac{S(k, i)}{(n-i)!} = \frac{n^k}{n!}, k \geq n. \quad (68)$$

From (15) it follows

$$\sum_{k=i}^{\infty} \frac{S(k-1, i-1)}{n^k} = \frac{(n-i)!}{n!}, 1 \leq i \leq n. \quad (69)$$

From (16) it follows

$$\frac{S(k, n)}{n^k} = \sum_{j=n}^k \frac{S(j-1, n-1)}{n^j}. \quad (70)$$

From (17) it follows

$$\sum_{i=1}^{n-1} \frac{S(k, i)}{(n-i)!} + n^k \sum_{j=n}^k \frac{S(j-1, n-1)}{n^j} = \frac{n^k}{n!}. \quad (71)$$

From (18) it follows

$$\sum_{k=n+1}^{\infty} \frac{S(k-1, i-1)}{n^k} = \frac{1}{n^n} \sum_{i=1}^{n-1} \frac{S(n, i)}{(n-i)!}. \quad (72)$$

Note that identity (68) is the same as (28) and identities (68) (71) imply (70).

I report in the following two more results about the sum of probabilities $\phi'_{n/n}(k)$, that clarify the probabilistic meaning of identity (46) and the close connection between these probabilities and the Stirling numbers of the second kind.

Corollary 4.2. *The following equations hold true*

$$\sum_{s=n}^k \phi'_{n/n}(s) = 1 - \sum_{j=0}^{n-2} (-1)^j \binom{n}{j+1} \left(1 - \frac{j+1}{n}\right)^k, \quad (73)$$

$$\sum_{s=n}^k \phi'_{n/n}(s) = \frac{n!}{n^k} S(k, n). \quad (74)$$

of equation (73). From (65)

$$\sum_{s=n}^k \phi'_{n/n}(s) = \sum_{j=0}^{n-2} (-1)^j \leq \binom{n-1}{n-j-1} \left(\frac{n}{n-j-1} \right) \sum_{s=n}^k \left(\frac{n-j-1}{n} \right)^s.$$

By applying Lemma 3.2

$$\left(\frac{n}{n-j-1} \right) \sum_{s=n}^k \left(\frac{n-j-1}{n} \right)^s = \frac{\left(1 - \frac{j+1}{n}\right)^{n-1} - \left(1 - \frac{j+1}{n}\right)^k}{\left(\frac{j+1}{n}\right)}$$

hence

$$\sum_{s=n}^k \phi'_{n/n}(s) = \sum_{j=0}^{n-2} (-1)^j \binom{n}{j+1} \left[\left(1 - \frac{j+1}{n}\right)^{n-1} - \left(1 - \frac{j+1}{n}\right)^k \right]. \quad (75)$$

Since we know that $\sum_{s=n}^{\infty} \phi'_{n/n}(s) = 1$ and

$$\left(1 - \frac{j+1}{n}\right)^k \rightarrow 0 \text{ as } k \rightarrow \infty$$

it is

$$\sum_{j=0}^{n-2} (-1)^j \binom{n}{j+1} \left(1 - \frac{j+1}{n}\right)^{n-1} = 1. \quad (76)$$

Finally from (75) and (76) equation (73) follows. \square

of equation (74). Remembering (66) we can write

$$\sum_{j=n}^k \phi'_{n/n}(j) = n! \sum_{j=n}^k \frac{S(j-1, n-1)}{n^j}$$

hence through identity (70) we simply get (74). \square

Remark 4.1. Note that equation (76) above is the same as (45), (46). Indeed after some manipulations it can be written as

$$\frac{(-1)^{n-1}}{n^{n-1}} \sum_{j=0}^{n-1} (-1)^j \binom{n}{j} j^{n-1} = 1.$$

This identity has thus a probabilistic meaning, connected with the normalization property of FOS probabilities.

Remark 4.2. We know [34, Eq. 26.8.42] the Stirling number of the second kind, for fixed value of n , as $k \rightarrow \infty$ is asymptotic to

$$S(k, n) \sim \frac{n^k}{n!}.$$

This asymptotic behavior follows very simply from equation (74), due to the connection established between FOS probabilities and $S(k, n)$.

4.3. Bounds of $g(k, i)$ numbers and Stirling numbers of the second kind

We speak of $g(k, i)$ numbers since the results are directly related to their formulation through the coefficients $a_{i,j}$, as stated by Theorem 3.2. The results are obviously applicable to the Stirling numbers of the second kind.

Definition 4.1. Let's define for $i \geq 3, 1 \leq j \leq (i - 2)$

$$\sigma_{i,j} = \sum_{s=1}^j (-1)^{j-s} |a_{i,s}| = |a_{i,j}| - |a_{i,j-1}| + \dots \pm |a_{i,1}|$$

and

$$\tau_{i,j}(k) = \sum_{s=1}^j (-1)^{j-s} |a_{i,s}| s^{k-i+1} = |a_{i,j}| j^{k-i+1} - |a_{i,j-1}| (j-1)^{k-i+1} + \dots \pm |a_{i,1}|,$$

where the sign of $|a_{i,1}|$ is negative if j is an even number, positive if it's an odd number.

Lemma 4.1. The following statements hold true:

$$\lim_{i \rightarrow \infty} \frac{a_{i+1,i-1}}{a_{i,i-2}} = e; \quad (77)$$

$$\lim_{i \rightarrow \infty} \frac{\sigma_{i,i-3}}{a_{i,i-2}} = 1; \quad (78)$$

$$a_{i,i-2} > \sigma_{i,i-3}, i \geq 5; \quad (79)$$

$$\sigma_{i,i-3} > 0, i \geq 5; \quad (80)$$

$$\sigma_{i,i-2} > a_{i,i-2}; \quad (81)$$

$$|a_{i,i-3}| > \sigma_{i,i-4}, i \geq 5. \quad (82)$$

Proof. Equation (77) follows simply from the values of the coefficients given by (44).

Equation (39) of Theorem 3.2 gives

$$\frac{a_{i+1,i-1}}{(i-1)a_{i,i-2}} = 1 - \frac{\sigma_{i,i-3}}{a_{i,i-2}}$$

from that and (77) equation (78) follows.

Considering Definition 4.1, from (39) we can write

$$a_{i,i-2} - \sigma_{i,i-3} = \frac{a_{i+1,i-1}}{(i-1)}$$

with $a_{i+1,i-1} > 0$, hence (79).

From equation (44) it is easy to prove that

$$0 < \frac{a_{i+1,i-1}}{(i-1)} < a_{i,i-2} \text{ for } i \geq 4$$

hence the previous equation implies (80) too.

Since

$$\sigma_{i,i-2} = a_{i,i-2} - \sigma_{i,i-3} \text{ and}$$

$$\sigma_{i,i-3} = |a_{i,i-3}| - \sigma_{i,i-4}$$

(80) implies both (81) and (82). \square

Lemma 4.2. Given $\sigma_{i,j}$ and $\tau_{i,j}(k)$ of Definition 4.1, it is

$$\tau_{i,j}(k) > \sigma_{i,j}.$$

Proof. Suppose j even, then we can write

$$|a_{i,s}|s^{k-i+1} - |a_{i,s-1}|(s-1)^{k-i+1} > |a_{i,s}| - |a_{i,s-1}|, s=2,4,\dots,j.$$

By adding the pairs of the previous inequality, we obtain the thesis.

In the case when j is odd, we can repeat the above procedure considering the pairs for $s = 3, 5, 7, \dots, j$, then adding up these pairs and the last term $|a_{i,1}|$ taken with a positive sign. \square

Theorem 4.1. Given the numbers $g(k, i)$, $i \geq 3, k \geq i$, the following bounds hold

$$a_{i+1,i-1}(i-1)^{k-i} - a_{i,i-2}(i-2)^{k-i+1} < g(k, i) < a_{i+1,i-1}(i-1)^{k-i} \quad (83)$$

and, for i fixed

$$\lim_{k \rightarrow \infty} \frac{g(k, i)}{a_{i+1,i-1}(i-1)^{k-i}} = 1 \quad (84)$$

Proof. From the Definition 4.1 we can write in general

$$\sigma_{i,j} = |a_{i,j}| - \sigma_{i,j-1}$$

$$\tau_{i,j}(k) = |a_{i,j}|j^{k-i+1} - \tau_{i,j-1}(k),$$

and, remembering (48),

$$g(k, i) = a_{i+1,i-1}(i-1)^{k-i} - \tau_{i,i-2}(k)$$

$$g(k, i) = a_{i+1,i-1}(i-1)^{k-i} - a_{i,i-2}(i-2)^{k-i+1} + \tau_{i,i-3}(k).$$

The thesis (83) follows from Lemma 4.2 and (80), (81) of Lemma 4.1.

The inequalities (83) can be rewritten as

$$1 - (i-2) \frac{a_{i,i-2}}{a_{i+1,i-1}} \frac{(i-1)^i (i-2)^k}{(i-2)^i (i-1)^k} < \frac{g(k, i)}{a_{i+1,i-1}(i-1)^{k-i}} < 1,$$

from that equation (84) follows. \square

Remark 4.3. The closed-form (44) of coefficients allows us to rewrite (83) and (84) as

$$\frac{(i-1)^{k-1}}{(i-1)!} - \frac{(i-2)^{k-1}}{(i-2)!} < g(k, i) < \frac{(i-1)^{k-1}}{(i-1)!}, \quad (85)$$

$$g(k, i) \sim \frac{(i-1)^{k-1}}{(i-1)!}. \quad (86)$$

When speaking of Stirling numbers of the second kind, the previous bounds become

$$\frac{i^k}{i!} - \frac{(i-1)^k}{(i-1)!} < S(k, i) < \frac{i^k}{i!}. \quad (87)$$

Note that (87) gives an alternative proof of the asymptotic behavior of $S(k, i)$ for fixed i (see Remark 4.2).

Remark 4.4. Assuming the left-hand side of the above inequality is positive both for $S(k, i)$ and $S(k-1, i-1)$, that is

$$a - \frac{\ln(i-a)}{\ln(1-\frac{1}{i-a})} < k, a = 1, 2$$

we get the bounds of the ratio

$$\frac{(i-1)^{k-1}}{i^{k-1}} - (i-1) \frac{(i-2)^{k-1}}{i^{k-1}} < \frac{S(k-1, i-1)}{S(k, i)} < \frac{(i-1)^{k-1}}{i^{k-1} - (i-1)^k}$$

that for $i \gg 1$ becomes

$$e^{-\frac{k}{i}} (1 - ie^{-\frac{k}{i}}) < \frac{S(k-1, i-1)}{S(k, i)} < e^{-\frac{k}{i}} \frac{1}{(1 - ie^{-\frac{k}{i}})}, \quad (88)$$

under condition

$$i \ln i < k.$$

4.4. FOS average length and harmonic numbers

As anticipated in the Introduction, calculating the average length of FOS sequences, that is of elementary events $\omega'_{i/n}(k)$, $k \geq i$, of the countable space $\Omega'_{i/n}$, is a simple task leading to the result (2) in the case $i = n$. This calculation is based on the following well-known results, in particular on Lemma 4.4 expressing the average number of trials one has to wait before obtaining the first success, in a sequence of independent Bernoulli trials with p probability of success.

Lemma 4.3. Given q real, $|q| < 1$, then

$$\sum_{k=1}^{\infty} kq^k = \frac{q}{(1-q)^2}.$$

Lemma 4.4. Given p real, $0 < p \leq 1$, then

$$\sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}.$$

Therefore if we denote in general with $L'_{i/n}$ the quantity we are looking for, we can establish the following recursive relation

$$L'_{i/n} = L'_{i-1/n} + \frac{n}{n-i+1}, i = 1, 2, \dots, n \quad (89)$$

starting with

$$L'_{0/n} = 0.$$

It is easy to see the above equation leads to

$$L'_{i/n} = n(H_n - H_{n-i}), i = 1, 2, \dots, n, \quad (90)$$

where H_n is the n -th harmonic number

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$$

and

$$H_0 = 0.$$

If we consider the set of discrete random variables $q_{i/n}$ with probability mass function

$$\phi'_{i/n}(k) = \text{Prob}\{q_{i/n} = k\}, 1 \leq i \leq n, k \geq i,$$

then the quantity $L'_{i/n}$ can be also obtained as the mathematical expectation of the corresponding random variable $q_{i/n}$, hence through

$$L'_{i/n} = E[q_{i/n}] = \sum_{k=i}^{\infty} k \phi'_{i/n}(k). \quad (91)$$

From (91) and (90), remembering (66), we get the following relation between harmonic numbers and Stirling numbers of the second kind.

Corollary 4.3.

$$H_n - H_{n-i} = \frac{(n-1)!}{(n-i)!} \sum_{k=i}^{\infty} \frac{k}{n^k} S(k-1, i-1), 1 \leq i \leq n. \quad (92)$$

Other combinatorial identities involving harmonic numbers and finite term series, can be derived from the direct calculation of (91), as established by the following proposition.⁶

Theorem 4.2. Assuming n and i positive integers, $n \geq 3$ and $3 \leq i \leq n$, the following identity holds true

$$\begin{aligned} H_n - H_{n-i} &= \\ &= \frac{n!}{(n-i)!n^i} \left[\frac{a_{i+1,i-1}}{(n-i+1)^2} [ni - (i-1)^2] - \sum_{j=1}^{i-2} a_{i,j} \left(\frac{j}{n-j}\right) \left(\frac{ni+j-ij}{n-j}\right) \right] \end{aligned} \quad (93)$$

where $a_{i,j}$ are the coefficients defined by Theorem 3.2.

Proof. By substituting in (91) the term $\phi'_{i/n}(k)$ as given by (58), we get

$$L'_{i/n} = \sum_{k=i}^{\infty} \frac{n!}{(n-i)!n^k} k \sum_{j=1}^{i-2} a_{i,j} [(i-1)^{k-i+1} - j^{k-i+1}]$$

from which it follows

$$L'_{i/n} = \frac{n!}{(n-i)!n^{i-1}} \sum_{j=1}^{i-2} a_{i,j} \left[\sum_{k=i}^{\infty} k \left(\frac{i-1}{n}\right)^{k-i+1} - \sum_{k=i}^{\infty} k \left(\frac{j}{n}\right)^{k-i+1} \right]. \quad (94)$$

We can write

$$\sum_{k=i}^{\infty} k \left(\frac{i-1}{n}\right)^{k-i+1} = \sum_{k=1}^{\infty} k \left(\frac{i-1}{n}\right)^k + \frac{(i-1)^2}{n} \sum_{k=0}^{\infty} \left(\frac{i-1}{n}\right)^k$$

⁶ These identities appear to be new, like the method of deriving them. Of course, it's not simple to establish the novelty of an identity involving harmonic numbers because of the historical interest of this subject and the great development of theory and applications. For an extensive collection of these identities, see [9], [10], [11] and the references therein.

and remembering Lemma 4.3 and the sum of geometric series

$$\sum_{k=i}^{\infty} k \left(\frac{i-1}{n} \right)^{k-i+1} = \frac{(i-1)}{(n-i+1)^2} [ni - (i-1)^2].$$

In a similar way the second term becomes

$$\sum_{k=i}^{\infty} k \left(\frac{j}{n} \right)^{k-i+1} = \left(\frac{j}{n-j} \right) \left(\frac{j+ni-ij}{n-j} \right).$$

Inserting the previous formulas into (94) we get

$$L'_{i/n} = \frac{n!}{(n-i)!n^{i-1}} \sum_{j=1}^{i-2} a_{i,j} \left[\frac{(i-1)}{(n-i+1)^2} [ni - (i-1)^2] - \left(\frac{j}{n-j} \right) \left(\frac{j+ni-ij}{n-j} \right) \right]$$

from which and equations (39), (90), (93) follows. \square

The following corollary is a simple application of the formulas of Theorem 3.3 to that of the previous theorem.

Corollary 4.2.1. For n and i positive integers, $n \geq 3$ and $3 \leq i \leq n$, the following identity holds true

$$\begin{aligned} H_n - H_{n-i} &= \\ &= \frac{n!}{(n-i)!n^i} \left[\frac{(i-1)^{i-2}}{(i-2)!} \frac{[ni - (i-1)^2]}{(n-i+1)^2} - \sum_{j=1}^{i-2} (-1)^{i-j-2} \frac{j^{i-2}}{(i-j-1)!(j-1)!} \frac{(ni+j-ij)}{(n-j)^2} \right]. \end{aligned}$$

When $i = n$ we get for the n -th harmonic number

$$H_n = \frac{n!}{n^n} \left[\frac{(n-1)^{n-2}}{(n-2)!} (2n-1) - \sum_{j=1}^{n-2} (-1)^{n-j-2} \frac{j^{n-2}}{(n-j-1)!(j-1)!} \frac{(n^2+j-nj)}{(n-j)^2} \right].$$

4.5. FOS as a model of the distribution of primes

This subsection aims to analyse analogies and limits of FOS with respect to the true sequence of prime numbers. There is a correspondence between prime numbers and FOS, as stated by the following remark.

Remark 4.5. Given any prime p_n , $n \geq 2$, there exists an event FOS

$$\mathcal{S}'_{n/n}(p_n) \in \Omega'_{n/n} = \bigcup_{k=n}^{\infty} \mathcal{S}'_{n/n}(k)$$

(see Remark 2.7), and the whole sequence of prime numbers corresponds to the element

$$(\mathcal{S}'_{n/n}(p_n), n = 2, 3, \dots) \in \mathcal{U}$$

of the product set

$$\mathcal{U} = \prod_{n=2}^{\infty} \Omega'_{n/n},$$

we can define as FOS universe.

The model based on *FOS* takes then into consideration the set $\Omega'_{n/n}$ as the combinatorial counterpart of the n -th prime p_n , considered as a discrete random variable assuming integer values between n and ∞ . The statistics of this random variable are then used to define a "counting variable", which is the *FOS* model equivalent of the prime-counting function. In order to avoid any confusion when dealing with our model, we keep the symbol p_n to denote the "true" value of the n -th prime, and adopt the symbol q_n for the random variable modeling it. The basic assumption of the model is the following.

Definition 4.2. *The discrete random variable q_n has probability mass function*

$$\phi'_{n/n}(k) = \text{Prob}\{q_n = k\}, k = n, n+1, n+2, \dots,$$

with $\phi'_{n/n}(k)$, defined in equation (20).

The results of the model are referred to the sequence of random variables $\{q_n\}$ and confronted with the analytical or numerical results we know about the true prime sequence $\{p_n\}$. These results, some of which are reported in two sample tables, justify *a posteriori* the model's adoption and the not completely rigorous approach we follow in deriving it.

From (66) and (90) respectively we get the probability

$$\phi'_{n/n}(k) = \frac{n!}{n^k} S(k-1, n-1), k \geq n$$

and the mean value of q_n

$$E[q_n] = L'_{n/n} = nH_n \sim n \ln n. \quad (95)$$

Hence, given $k \geq n$, k integer, the probability distribution function of q_n

$$\text{Prob}\{q_n \leq k\} = \Phi_n(k) \quad (96)$$

with

$$\Phi_n(k) = \sum_{j=n}^k \phi'_{n/n}(j) = n! \sum_{j=n}^k \frac{S(j-1, n-1)}{n^j}, \quad (97)$$

is simply expressed through the number of ways a set of k elements can be partitioned into n disjoint nonempty subsets

$$\Phi_n(k) = \frac{(n-1)!}{n^{k-1}} S(k, n), \quad (98)$$

where the last equation follows from (74).

Let's define the discrete random variable ζ_k whose probability distribution is induced by q_n through the following

Definition 4.3.

$$\text{Prob}\{\zeta_k \geq n\} = \text{Prob}\{q_n \leq k\} = \Phi_n(k) \quad (99)$$

and obviously

$$\text{Prob}\{\zeta_k < n\} = 1 - \Phi_n(k). \quad (100)$$

In our combinatorial model, ζ_k is associated with the value $\pi(k)$ of the prime-counting function, considered as a random variable assuming integer values between 1 and k .

Theorem 4.3. The discrete random variable ξ_k with probability distribution (99) assumes values between 1 and k , has probability mass function

$$\text{Prob}\{\xi_k = n\} = \Phi_n(k) - \Phi_{n+1}(k), \quad n = 1, 2, \dots, k, \quad (101)$$

and mathematical expectation

$$E[\xi_k] = \sum_{n=1}^k \Phi_n(k). \quad (102)$$

Proof. From (98), due to the properties of $S(k, n)$ numbers, we get for $n = 1$

$$\text{Prob}\{\xi_k \geq 1\} = \Phi_1(k) = 1$$

$$\text{Prob}\{\xi_k < 1\} = 1 - \Phi_1(k) = 0;$$

for $n = k$

$$\text{Prob}\{\xi_k \geq k\} = \Phi_k(k) = \frac{(k-1)!}{k^{k-1}}$$

$$\text{Prob}\{\xi_k < k\} = 1 - \Phi_k(k) = 1 - \frac{(k-1)!}{k^{k-1}};$$

for $n > k$

$$\text{Prob}\{\xi_k \geq n\} = \Phi_n(k) = 0$$

$$\text{Prob}\{\xi_k < n\} = 1 - \Phi_n(k) = 1.$$

This proves the first statement.

Given n and m positive integers with $n < m$, from (100) we get

$$\text{Prob}\{n \leq \xi_k < m\} = \Phi_n(k) - \Phi_m(k). \quad (103)$$

From this relation, assuming $m = n + 1$ the probability mass function (101) follows. The mathematical expectation of ξ_k is given by

$$E[\xi_k] = \sum_{n=1}^k n(\Phi_n(k) - \Phi_{n+1}(k)) = \sum_{n=1}^k \Phi_n(k) - k\Phi_{k+1}(k),$$

that is the same as equation (102) since $\Phi_{k+1}(k) = 0$. \square

Remark 4.6. The mean value of the discrete random variable ξ_k is equal to

$$E[\xi_k] = \sum_{n=1}^k \text{Prob}\{q_n \leq k\} = \sum_{n=1}^k \sum_{j=n}^k \phi'_{n/n}(j).$$

The previous theorem suggests to assume $E[\xi_k]$ as an estimate $\hat{\pi}(k)$ derived from the combinatorial model, of the prime-counting function $\pi(k)$, that is, remembering the equality (98)

$$\hat{\pi}(k) = E[\xi_k] = \sum_{n=1}^k \frac{(n-1)!}{n^{k-1}} S(k, n). \quad (104)$$

An alternative expression for $\hat{\pi}(k)$ can be derived from Remark 4.6 and equation (73)

$$\hat{\pi}(k) = k - \sum_{j=0}^{k-2} (-1)^j \sum_{n=j+2}^k \binom{n}{j+1} \left(1 - \frac{j+1}{n}\right)^k.$$

To check the quality of the function $\hat{\pi}(k)$ as a numerical estimate of $\pi(k)$, equation (104) looks more suitable since it can be written in a recursive form. Indeed if we put

$$\hat{\pi}(k) = \sum_{n=1}^k w(k, n)$$

with

$$w(k, n) = \frac{(n-1)!}{n^{k-1}} S(k, n),$$

from the recursive equation of the Stirling numbers of the second kind (51), we get the following recurrence for the terms $w(k, n)$

$$w(k+1, n) = w(k, n) + \left(\frac{n-1}{n}\right)^k w(k, n-1), k \geq 1, 1 \leq n \leq (k+1)$$

initialized as

$$w(k, 1) = 1, k \geq 1$$

$$w(1, n) = 0, n > 1.$$

Table 3 reports a series of values of the recursive estimate $\hat{\pi}(k)$, compared with the prime-counting function $\pi(k)$ and the logarithmic integral function $li(k)$. The last column of the table reports the values obtained through continuous approximation formulas I'll expose hereafter.

Table 3. $\hat{\pi}(k)$ estimate through recursive and continuous formulas

k	$\pi(k)$	$li(k)$	$\hat{\pi}(k) = \sum_{n=1}^k w(k, n)$	$\hat{\pi}(k) = \int_1^k e^{-ane^{-\frac{k}{n}}} dn$
100	25	30.1261	26.9462	23.1623
200	46	50.1921	47.0309	41.3716
300	62	68.3336	65.50659	58.2265
400	78	85.4178	83.06021	74.2987
500	95	101.7939	99.98131	89.8312
600	109	117.6465	116.4275	104.9572
700	125	133.0889	132.4971	119.7596
800	139	148.1967	148.2565	134.2952
900	154	163.0236	163.7537	148.6046
1000	168	177.6097	179.0246	162.7185
2000	303	314.8092	323.4725	296.6907
3000	430	442.7592	458.9438	422.8337
4000	550	565.3645	589.1406	544.3509
5000	669	684.2808	715.6488	662.6328
6000	783	800.4141		778.4514
7000	900	914.3308		892.2944
8000	1007	1026.416		1004.4960
9000	1117	1136.949		1115.2989
10000	1229	1246.137		1224.8866

In order to analyze the asymptotic behavior of $\hat{\pi}(k)$, we look for a continuous approximation of the probability function $\Phi_n(k)$. After promoting k and n to be continuous

variables, which is a justifiable assumption for $k \gg 1$ and considering the relation (98) between Φ and S , equation (97) is written as

$$\Phi_n(k) = \int_n^k \left(\frac{n-1}{n}\right)^{u-1} \Phi_{n-1}(u-1) du.$$

Taking the derivative of the above equation, we get the differential equation for Φ

$$\frac{\partial \Phi_n(k)}{\partial k} = \left(\frac{n-1}{n}\right)^{k-1} \Phi_{n-1}(k-1) \quad (105)$$

that can be written as

$$\frac{1}{\Phi_n(k)} \frac{\partial \Phi_n(k)}{\partial k} = \frac{S(k-1, n-1)}{S(k, n)}.$$

We set

$$\frac{S(k-1, n-1)}{S(k, n)} = \frac{\lambda_n(k)}{n} e^{-\frac{k}{n}} \quad (106)$$

where $\lambda_n(k)$ is a parameter function depending on n and k . Then the general solution of the differential equation is

$$\Phi_n(k) = C e^{\frac{1}{n} \int \lambda_n(k) e^{-\frac{k}{n}} dk} \quad (107)$$

and the value of the constant C is determined by the asymptotic condition $\Phi_n(k \rightarrow \infty) = 1$.

Remark 4.7. It is interesting to note the method, based on probabilistic arguments, that led to equation (107), provides an alternative approach to the problem of exploring the asymptotics ($k \rightarrow \infty$) of Stirling numbers of the second kind. Since this topic is outside the scope of this work, I cite only the following example. If we choose

$$\lambda_n(k) = \left(n + 1 - \frac{k}{n}\right)$$

then it is

$$\Phi_n(k) = e^{\left(\frac{k}{n} - n\right) e^{-\frac{k}{n}}}$$

and from (98) we get the well known asymptotic approximation of $S(k, n)$ (see [31, and the references therein])

$$S(k, n) \approx \frac{n^k}{n!} e^{\left(\frac{k}{n} - n\right) e^{-\frac{k}{n}}},$$

which is valid in the region $n < \frac{k}{\ln k}$, coinciding asymptotically with the region of validity of (88), $n \ln n < k$.

As a continuous approximation of the probability function $\Phi_n(k)$ we are looking for, let's choose a parameter function depending on n only through the following simple relation

$$\lambda_n(k) = \lambda_n = an, \quad a > 0 \text{ constant},$$

where the constant a varies between 1 and e as suggested by the boundary conditions of the Stirling number ratio. Indeed for $n = k$ the ratio (106) is

$$\frac{S(k-1, k-1)}{S(k, k)} = 1,$$

that implies $\lambda_n = en$, while for $k \gg n$ we know it is

$$\frac{S(k-1, n-1)}{S(k, n)} \sim e^{-\frac{k}{n}},$$

hence $\lambda_n = n$.

Therefore (107) becomes

$$\Phi_n(k) = e^{-ane^{-\frac{k}{n}}} \quad (108)$$

and from (102) and (104) we get the continuous approximation of $\hat{\pi}(k)$

$$\hat{\pi}(k) = \int_1^k \Phi_n(k) dn = \int_1^k e^{-ane^{-\frac{k}{n}}} dn. \quad (109)$$

The last column of Table 3 reports some examples from this approximation formula with $a = (1+e)/2$. Table 4 reports the values calculated through (109) with $a = e$, compared with the values of $\pi(k)$ up to $k = 10^{27}$.⁷

Table 4. $\hat{\pi}(k)$ estimate through continuous approximation

k	$\pi(k)$	$\hat{\pi}(k) = \int_1^k e^{-ane^{-\frac{k}{n}}} dn$	$\hat{\pi}(k)/\pi(k)$
10^5	9,592	9,428.02	0.98291
10^6	78,498	78,480.93	0.99978
10^7	664,579	671,099.45	1.00981
10^8	5,761,455	5,855,689.76	1.01636
10^9	50,847,534	51,900,660.41	1.02071
10^{10}	455,052,511	465,792,892.49	1.02360
10^{11}	4,118,054,813	4,223,145,802.17	1.02552
10^{12}	37,607,912,018	38,614,679,105.06	1.02677
10^{13}	346,065,536,839	355,603,668,431.86	1.02756
10^{14}	3,204,941,750,802	3,294,779,143,238.6	1.02803
10^{15}	29,844,570,422,669	30,688,289,307,555	1.02827
10^{16}	279,238,341,033,925	287,153,196,808,146	1.02834
10^{17}	2,623,557,157,654,233	2.69779945552531E + 15	1.02830
10^{18}	24,739,954,287,740,860	2.54367476772712E + 16	1.02816
10^{19}	234,057,667,276,344,607	2.40603623244741E + 17	1.02797
10^{20}	2,220,819,602,560,918,840	2.28238863108907E + 18	1.02772
10^{21}	21,127,269,486,018,731,928	2.17071405049150E + 19	1.02745
10^{22}	201,467,286,689,315,906,290	2.06936330707283E + 20	1.02715
10^{23}	1,925,320,391,606,803,968,923	1.97697565027884E + 21	1.02683
10^{24}	18,435,599,767,349,200,867,866	1.89241852576407E + 22	1.02650
10^{25}	176,846,309,399,143,769,411,680	1.81474179606716E + 23	1.02617
10^{26}	1,699,246,750,872,437,141,327,603	1.74314253336239E + 24	1.02583
10^{27}	16,352,460,426,841,680,446,427,399	1.676937646378480E + 25	1.02550

The following proposition, which is the FOS equivalent of the PNT, can be stated for the function $\hat{\pi}(k)$ defined by (109).

Theorem 4.4. *Given $\epsilon > 0$ however small, it holds*

⁷ Data with $\pi(k)$ values are taken from N. J. A. Sloane, 'The On-Line Encyclopedia of Integer Sequences. Sequence A006880', <http://oeis.org> Wikipedia contributors, 'Prime-counting function', Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Prime-counting_function&oldid=987382574

$$1 - \epsilon < \frac{\hat{\pi}(k)}{\frac{k}{\ln k}} < 1 + \epsilon,$$

definitively for $k > k_0$.

Proof. $\Phi_n(k)$, $n > 0$, $k > 0$, is a decreasing function with n . Indeed

$$\frac{\partial \Phi_n(k)}{\partial n} = -a\Phi_n(k)e^{-\frac{k}{n}}\left(1 + \frac{k}{n}\right) < 0.$$

Given n_0 and n_1 such that $1 < n_0 < n_1 < k$, we can write

$$\hat{\pi}(k) = \int_1^{n_0} \Phi_n(k)dn + \int_{n_0}^{n_1} \Phi_n(k)dn + \int_{n_1}^k \Phi_n(k)dn$$

and, due to the decreasing of $\Phi_n(k)$ with n ,

$$\begin{aligned} \Phi_{n_0}(k)(n_0 - 1) + \Phi_{n_1}(k)(n_1 - n_0) + \Phi_k(k)(k - n_1) &\leq \\ &\leq \hat{\pi}(k) \leq \end{aligned}$$

$$\Phi_1(k)(n_0 - 1) + \Phi_{n_0}(k)(n_1 - n_0) + \Phi_{n_1}(k)(k - n_1).$$

Given ϵ , $0 < \epsilon < 1$, let $\frac{1}{\gamma} = 1 + \frac{\epsilon}{2}$, hence $\frac{2}{3} < \gamma < 1$, and

$$n_0 = n_0(k) = \frac{k}{\ln k}$$

$$n_1 = n_1(k) = \frac{k}{\gamma \ln k}$$

then the previous inequalities become (note that $n_1(k) < k$ for $\frac{1}{\gamma} < \ln k$)

$$\begin{aligned} \left(1 - \frac{\ln k}{k}\right)e^{-\frac{a}{\ln k}} + \left(\frac{1}{\gamma} - 1\right)e^{-a\left(\frac{k^{1-\gamma}}{\gamma \ln k}\right)} + \left(\ln k - \frac{1}{\gamma}\right)e^{-\frac{a}{\gamma}k} &\leq \\ &\leq \frac{\hat{\pi}(k)}{\frac{k}{\ln k}} \leq \\ &\leq \left(1 + \frac{\ln k}{k}\right)e^{-ae^{-k}} + \left(\frac{1}{\gamma} - 1\right)e^{-\frac{a}{\ln k}} + \left(\ln k - \frac{1}{\gamma}\right)e^{-a\left(\frac{k^{1-\gamma}}{\gamma \ln k}\right)}. \end{aligned}$$

The limit for $k \rightarrow \infty$ of the left-hand and the right-hand side of the previous inequality, say $L(k)$ and $R(k)$ respectively, gives

$$\lim_{k \rightarrow \infty} L(k) = 1,$$

that is

$$1 - \epsilon < L(k) < 1 + \epsilon, \text{ for } k > k_L,$$

and

$$\lim_{k \rightarrow \infty} R(k) = \frac{1}{\gamma},$$

that is

$$\frac{1}{\gamma} - \frac{\epsilon}{2} < R(k) < \frac{1}{\gamma} + \frac{\epsilon}{2}, \text{ for } k > k_R.$$

Let $k_0 = \text{Max}\{k_L, k_R\}$, then for $k > k_0$ we get the thesis. \square

What does the model say about the Riemann Hypothesis [12]? We know the most famous *millennium problem*⁸ is equivalent to establish the bound of the prime-counting function $\pi(x)$ [30]

$$|\pi(x) - \text{li}(x)| < M\sqrt{x} \ln x$$

with $M > 0$ constant, definitively for x sufficiently large.⁹ We can prove this bound, I'll refer to as the *RH rule* throughout this work, occurs w.p. 0, or better that the *RH rule* doesn't hold w.p. 1 for the random variable ζ_k , defined above as the stochastic counterpart of $\pi(k)$ in our model.

Theorem 4.5. Let ζ_k then random variable with probability distribution (99), then it is

$$\text{Prob}\{|\zeta(k) - \text{li}(k)| < M\sqrt{k} \ln k\} \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Proof. Let

$$n(k) = \text{li}(k) - M\sqrt{k} \ln k$$

and

$$m(k) = \text{li}(k) + M\sqrt{k} \ln k.$$

Since

$$\text{li}(k) = \frac{k}{\ln k} + \mathcal{O}\left(\frac{k}{\ln^2 k}\right)$$

it is easy to see that both $\Phi_{n(k)}(k)$ and $\Phi_{m(k)}(k)$

$$\Phi_{n(k)}(k) \sim \Phi_{m(k)}(k) \sim e^{-e^{\frac{1}{\ln k}}} \rightarrow 1 \text{ as } k \rightarrow \infty.$$

From (103) then it follows

$$\lim_{k \rightarrow \infty} \text{Prob}\{n(k) \leq \zeta_k < m(k)\} = \lim_{k \rightarrow \infty} (\Phi_{n(k)}(k) - \Phi_{m(k)}(k)) = 0.$$

\square

Theorems 4.4 and 4.5 say what we know as *PNT* when dealing with primes, holds for the counting variable ζ_k associated with the set of sequences generated through the random sampling with replacement game I proposed in the Introduction, while the Riemann Hypothesis requires compliance with much stricter requirements such that it has zero probability to occur for the same counting variable.

5. Model generalization and the connection with the Riemann Hypothesis

This section deals with the generalization of the *FOS* model so that we can discriminate between two classes of models, compliant and not compliant with the *RH rule*. This allows us to use the different stochastic properties, depending on different models, of the random variables q_n and ζ_k to deduce some properties of the whole sequence of primes. In particular, we apply the model to infer the counting of successive prime pairs and then compare the result with actual sequences of prime pairs.

⁸ E. Bombieri, "Problems of the Millennium: The Riemann Hypothesis", Clay Math. Institute. (2000) online at http://www.claymath.org/sites/default/files/official_problem_description.pdf

⁹ $\text{li}(x)$ is the logarithmic integral function defined for $x > 1$ as $\text{li}(x) = \int_0^x \frac{dt}{\ln t}$, where the integral is to be intended as Cauchy principal value.

5.1. Model generalization

In this section we will consider the random variables q_n of Definition 4.2 as a continuous random variable, assuming values in $[n, \infty)$ according to the following probability distribution function

$$\text{Prob}\{q_n \leq x\} = \begin{cases} 0 & , \quad x < n \\ \Phi_n(x) & , \quad x \geq n \end{cases} \quad (110)$$

with $\Phi_n(x)$ given by the solution of the differential equation

$$\frac{1}{\Phi_n(x)} \frac{\partial \Phi_n(x)}{\partial x} = \frac{\lambda_n}{f(n)} e^{-\frac{x}{f(n)}}$$

obtained by considering a more general expression for the ratio (106), where n is substituted by $f(n)$, a positive function of n .

Notation. The probability density function of q_n , that is the function $\phi'_{n/n}(x)$ of the FOS model, will be denoted as follows:

$$\phi_n(x) = \phi'_{n/n}(x), \quad x \geq n.$$

Therefore the probability functions of q_n become

$$\Phi_n(x) = C e^{-\lambda_n e^{-\frac{x}{f(n)}}} + \text{const.}$$

$$\phi_n(x) = \frac{\partial \Phi_n(x)}{\partial x} = C \frac{\lambda_n}{f(n)} e^{-\frac{x}{f(n)}} e^{-\lambda_n e^{-\frac{x}{f(n)}}}.$$

The constant C is determined by the normalization condition

$$\int_n^\infty \phi_n(x) dx = C [\Phi_n(x)]_n^\infty = 1$$

hence

$$C = \frac{1}{(1 - e^{-\lambda_n e^{-n/f(n)}})}.$$

The complete expressions of the probability density and distribution function are

$$\phi_n(x) = \frac{1}{(1 - e^{-\lambda_n e^{-n/f(n)}})} \frac{\lambda_n}{f(n)} e^{-\frac{x}{f(n)}} e^{-\lambda_n e^{-\frac{x}{f(n)}}} \quad (111)$$

$$\Phi_n(x) = \frac{e^{-\lambda_n e^{-\frac{x}{f(n)}}} - e^{-\lambda_n e^{-n/f(n)}}}{1 - e^{-\lambda_n e^{-n/f(n)}}}. \quad (112)$$

Note that $\phi_n(x)$ has a single maximum point at

$$x = x_M = f(n) \ln(\lambda_n). \quad (113)$$

The mathematical expectation of q_n (this calculation is reported in [4] for the case $f(n) = n$) is

$$E[q_n] = \int_n^\infty x \phi_n(x) dx = n + \frac{f(n)}{1 - e^{-\lambda_n e^{-n/f(n)}}} \left[\ln(\lambda_n) + E_1(\lambda_n e^{-\frac{n}{f(n)}}) + \gamma - \frac{n}{f(n)} \right] \quad (114)$$

where $\gamma = 0.57721\dots$ is the Euler-Mascheroni constant and $E_1(x)$ is the exponential integral function with positive real argument $x > 0$, sometimes referred to as upper

(or complementary) incomplete Gamma function $\Gamma(0, x)$, defined by (see [1, func. 5.1.1, 6.5.3])

$$E_1(x) = \Gamma(0, x) = \int_x^\infty \frac{e^{-t}}{t} dt.$$

We require the mathematical model to satisfy a similar condition to (95) about the mean of q_n , that leads to the following formula

$$E[q_n] = f(n) [\ln(\lambda_n) + \gamma] \sim n \ln n \quad (115)$$

obtained from (114), assuming $n \gg 1$ sufficiently large, so we can neglect the terms $e^{-\lambda_n e^{-n/f(n)}}$ and $E_1(\lambda_n e^{-\frac{n}{f(n)}})$. Note that the function $E_1(x)$ decreases rapidly as $x \rightarrow \infty$, in particular [2, (1.4)]

$$\frac{e^{-x}}{2} \ln\left(1 + \frac{2}{x}\right) \leq E_1(x) \leq e^{-x} \ln\left(1 + \frac{1}{x}\right).$$

Remark 5.1. The FOS combinatorial model studied in Subsection 4.5 corresponds to the function $f(n) = n$. In this case (115) is satisfied for

$$\lambda_n = an, a > 0, \text{ constant}$$

(in particular when $a = e^{-\gamma}$ it is $E[q_n] = n \ln n$).

We know from Theorem 4.5 that the RH rule does not hold w.p. 1 for this class of models, which I'll refer to as n-Model.

Another class of models of particular interest to our study, as we'll see in the following, is the one determined by the choice $f(n) = \ln n$. In this case condition (115) can be satisfied using different functions having opposing implications as far as the compliance of the model with the RH rule. We focus in particular on a solution ensuring the RH rule holds w.p. 1.

Remark 5.2. The model with $f(n) = \ln n$ will be referred to as $\log(n)$ -Model.

The simplest $\log(n)$ -Model is the one defined by the choice

$$E[q_n] = n \ln n \quad (116)$$

hence by the parameter value

$$\lambda_n = ae^n, a = e^{-\gamma}. \quad (117)$$

Through the methods of the previous section, it is possible to prove the same statements as Theorems 4.4 and 4.5, that is this $\log(n)$ -Model is compliant with the PNT and not with the RH rule.

Let us consider instead the $\log(n)$ -Model defined by the assumption

$$E[q_n] = ali(n) \quad (118)$$

where $ali: \mathbb{R} \rightarrow (1, +\infty)$ is the inverse function of the logarithmic integral function $li(x)$.¹⁰ In this case we get

$$\lambda_n = ae^{\frac{ali(n)}{\ln n}}, a = e^{-\gamma}, \quad (119)$$

and condition (115) is satisfied since we know it is [17, Theorem 3.3]

¹⁰ I use for this function the same notation as in [17].

$$\frac{ali(n)}{n \ln n} = 1 + O\left(\frac{\ln \ln n}{\ln n}\right). \quad (120)$$

Equations (111) and (112), under the approximation for $n \gg 1$, become:

$$\phi_n(x) = \frac{a}{\ln n} e^{\frac{-x+r_n}{\ln n}} e^{-ae^{\frac{-x+r_n}{\ln n}}} \quad (121)$$

$$\Phi_n(x) = e^{-ae^{\frac{-x+r_n}{\ln n}}}, \quad (122)$$

with $r_n = ali(n)$.

The double exponential distribution function (122) is often called Gumbel distribution [26],[5].

Remark 5.3. In the case of the model with $f(n) = \ln n$, the ratio (106) is equal to

$$\frac{a}{\ln n} e^{\frac{-k+r_n}{\ln n}};$$

it is $o(e^{-\frac{k}{n}})$ as $k \rightarrow \infty$ and deviates asymptotically from the Stirling number ratio which is $\sim e^{-\frac{k}{n}}$.

Briefly speaking, the $\log(n)$ -Model defined above corresponds to a subset of the FOS universe \mathcal{U} which is different from the one of the n -Model, containing sequences conforming to both the PNT and the RH rule, as stated by the following theorems.

The continuous approximation function $\hat{\pi}(k)$ becomes

$$\hat{\pi}(k) = \int_1^k \Phi_n(k) dn = \int_1^k e^{-ae^{\frac{-k+r_n}{\ln n}}} dn. \quad (123)$$

Also for this function we can enunciate a proposition, which is the equivalent of the PNT.

Theorem 5.1. Given $\epsilon > 0$ however small, there exists k_0 such that for $k > k_0$ it holds

$$1 - \epsilon < \frac{\hat{\pi}(k)}{\frac{k}{\ln k}} < 1 + \epsilon,$$

where the function $\hat{\pi}(k)$ is defined by (123).

The proof is completely analogous to that of Theorem 4.4.

Remark 5.4. Associated with the $\log(n)$ -Model we can define the counting random variable ζ_k as in Definition 4.3, after assuming the probability distribution (110). It is easy to prove the analog of Theorem 4.3 for this variable, in particular equation (101) and (103). We can therefore apply the probability method developed for the n -Model also in the case of the $\log(n)$ -Model.

The RH rule holds w.p. 1 for the $\log(n)$ -Model with $E[q_n]$ defined by (118).

Lemma 5.1. Let $f(x) = ali(x) : \mathbb{R} \rightarrow (1, +\infty)$ be the inverse function of the logarithmic integral function $li(x)$ and

$$f^{(n)}(x) = \frac{d^n f(x)}{dx^n}$$

its n -th derivative. Then we have

$$f^{(1)}(x) = \ln(ali(x))$$

and for $n \geq 2$

$$f^{(n)}(x) = \frac{1}{f^{n-1}(x)} \sum_{k=1}^{n-1} a_{k,n} (f^{(1)}(x))^k$$

with

$$a_{1,n} = 1$$

$$a_{k,n} = ka_{k,n-1} - (n-2)a_{k-1,n-1}, \quad k = 2, 3, \dots, n-2$$

$$a_{n-1,n} = -(n-2)a_{n-2,n-1}.$$

Proof. Note that the function $f(x) = ali(x)$ is real analytic. The first derivative follows directly from the definition, since $li(ali(x)) = x$. We proceed by induction. The thesis is true for $n = 2$, indeed

$$f^{(2)}(x) = \frac{1}{f(x)} a_{1,2} f^{(1)}(x) = \frac{f^{(1)}(x)}{f(x)}.$$

Let's assume it is true for $(n-1)$

$$f^{(n-1)}(x) = \frac{1}{f^{n-2}(x)} \sum_{k=1}^{n-2} a_{k,n-1} (f^{(1)}(x))^k$$

then we have for the n -th derivative

$$f^{(n)}(x) = \frac{1}{f^{n-1}(x)} \sum_{k=1}^{n-2} a_{k,n-1} k (f^{(1)}(x))^k - \frac{(n-2)}{f^{n-1}(x)} \sum_{k=1}^{n-2} a_{k,n-1} (f^{(1)}(x))^{k+1} \text{text},$$

from which, grouping the terms by powers of $f^{(1)}(x)$, the thesis follows. \square

Theorem 5.2. Let ξ_k be the random counting variable of Definition 4.3 with probability distribution (110) associated with the $\log(n)$ -Model (118), then it is

$$\text{Prob}\{|\xi(k) - li(k)| < M\sqrt{k} \ln k\} \rightarrow 1 \text{ as } k \rightarrow \infty, \quad M > 0, \text{ constant.}$$

Proof. Let

$$n(k) = li(k) - M\sqrt{k} \ln k$$

and

$$m(k) = li(k) + M\sqrt{k} \ln k,$$

then from (103) we know it is

$$\text{Prob}\{n(k) \leq \xi_k < m(k)\} = (\Phi_{n(k)}(k) - \Phi_{m(k)}(k))$$

with probability distribution functions given by (122),

$$\Phi_{n(k)}(k) = e^{-ae^{\frac{-k+ali(n(k))}{\ln n(k)}}},$$

$$\Phi_{m(k)}(k) = e^{-ae^{\frac{-k+ali(m(k))}{\ln m(k)}}}.$$

Since we know $ali(x)$ is real analytic, consider its Taylor series $ali(x_0 + h)$ at

$$x_0 = li(k)$$

$$h = -M\sqrt{k} \ln k,$$

from Lemma 5.1

$$ali(n(k)) = k - M\sqrt{k} \ln^2 k + \frac{1}{2} M^2 \ln^3 k - \frac{1}{6} \frac{(1 - \ln k)}{\sqrt{k}} M^3 \ln^4 k + \dots,$$

which can be written

$$ali(n(k)) = k - M\sqrt{k} \ln^2 k + O(\ln^3 k). \quad (124)$$

From the asymptotic expansion of $li(x)$ we get

$$li(x) = \frac{k}{\ln k} \left(1 + O\left(\frac{1}{\ln k}\right) \right)$$

which leads to the following expression for $\ln n(k)$

$$\ln n(k) = \ln [li(k) - M\sqrt{k} \ln k] = \ln k - \ln \ln k + O\left(\frac{1}{\ln k}\right). \quad (125)$$

From (124) and (125) we finally get

$$\frac{-k + ali(n(k))}{\ln n(k)} = \frac{-M\sqrt{k} \ln k + O(\ln^2 k)}{1 - \frac{\ln \ln k}{\ln k} + O\left(\frac{1}{\ln^2 k}\right)}$$

$$\Phi_{n(k)}(k) \sim e^{-ae^{-M\sqrt{k} \ln k}}$$

and the limits for $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} \frac{-k + ali(n(k))}{\ln n(k)} = -\infty$$

$$\lim_{k \rightarrow \infty} \Phi_{n(k)}(k) = 1.$$

Through analogous calculations for $m(k)$ we find

$$\frac{-k + ali(m(k))}{\ln m(k)} = \frac{M\sqrt{k} \ln k + O(\ln^2 k)}{1 - \frac{\ln \ln k}{\ln k} + O\left(\frac{1}{\ln^2 k}\right)}$$

$$\Phi_{m(k)}(k) \sim e^{-ae^{M\sqrt{k} \ln k}}$$

and the limits for $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} \frac{-k + ali(m(k))}{\ln m(k)} = \infty$$

$$\lim_{k \rightarrow \infty} \Phi_{m(k)}(k) = 0.$$

From (103) then it follows

$$\lim_{k \rightarrow \infty} Prob\{n(k) \leq \zeta_k < m(k)\} = \lim_{k \rightarrow \infty} (\Phi_{n(k)}(k) - \Phi_{m(k)}(k)) = 1.$$

□

We can get an estimate of the speed of convergence to 1 of the probability of the *RH rule* as follows. If we denote with $\mathcal{U}' \subset \mathcal{U}$ the subset of random sequences of positive integers distributed according to the *log(n)-Model* with mean $ali(n)$, under the assumption that the sequence of prime numbers $\{p_n, n = 1, 2, 3, \dots\}$ belongs to the set \mathcal{U}' , we can use the properties of the Gumbel distribution and Chebyshev's inequality to get the probability of the *RH rule*, more generally the probability of any difference $|p_n - ali(n)|$, as function of n . In other words we can consider \mathcal{U}' as a stochastic process $\{q_n, n = 1, 2, 3, \dots\}$ of random variables differently distributed and the sequence of primes as particular *realization* of the process.

Different bounds of the difference

$$|p_n - ali(n)| \leq \delta_n$$

have been established for the sequence of prime numbers, depending on the conditions assumed. From [33, (1.3)] we know that it holds unconditionally for

$$\delta_n = O(ne^{-c\sqrt{\ln n}}), c > 0,$$

while under the *Riemann Hypothesis* we get [17, Theorem 6.2]

$$\delta_n = \frac{1}{\pi} \sqrt{n} (\ln n)^{5/2}, n \geq 11.$$

From the properties of the Gumbel distribution we know the variance of q_n is ¹¹

$$E[(q_n - ali(n))^2] = \frac{\pi^2}{6} (\ln n)^2$$

hence from Chebyshev's inequality (see [19, Chapter IX, p. 233]) it follows that the probability of the unconditioned bound converges to 1 with a quadratic law

$$\text{Prob}\{|p_n - ali(n)| \leq Cne^{-c\sqrt{\ln n}}\} \geq 1 - \frac{\pi^2 (\ln n)^2 e^{2c\sqrt{\ln n}}}{6C^2 n^2}, C, c > 0 \text{ const.},$$

while in the case the *Riemann Hypothesis* holds

$$\text{Prob}\{|p_n - ali(n)| \leq \frac{1}{\pi} \sqrt{n} (\ln n)^{5/2}\} \geq 1 - \frac{\pi^4}{6} \frac{1}{n(\ln n)^3}.$$

5.2. Counting successive prime pairs through the *log(n)-Model*

The general problem of prime pairs can be stated as follows: *given d and N positive integers, d even, what's the number $H_N(d)$ of pairs of primes (not necessarily consecutive), $p, p' \leq N$ such that $p' = p + d$?* Connected with this there is the following one, we can call as the problem of counting successive prime pairs: *given d and N positive integers, d even, what's the number $h_N(d)$ of pairs of primes $p_i, p_{i+1} \leq N$, separated by the gap d , that is such that $p_{i+1} = p_i + d$?*

Mathematics has no certain results about these two problems which are different from the so called "bounded gaps problem", solved recently by mathematician Yitang Zhang [45], who proved that as $i \rightarrow \infty$ there are infinitely many prime pairs $p_{i+1} - p_i = d$, with $d < 7.0 \times 10^7$. Successively this bound has been lowered to 246 through the extension of the method of Zhang and other contributors. Nevertheless, the twin primes conjecture which requires the bound to be reduced to $d = 2$, remains unproven. Important results towards this goal can be found in [21], [22], showing in particular that

¹¹ Note that the probability distribution (112) is not exactly a Gumbel distribution, because of the cut at $x = n$ of definition (110). Assuming n sufficiently large, so that (121) and (122) hold, we can consider the probability distribution of the model as a Gumbel distribution.

$\liminf_{i \rightarrow \infty} \frac{d}{\ln p_i} = 0$, that is there always be pairs of primes $p_{i+1} = p_i + d$ with d less than any fraction of $\ln p_i$, the average spacing between successive primes.

In 1923 Hardy and Littlewood proposed the following formula as a solution of the general problem, stressing that it was impossible for them "to offer anything approaching a rigorous proof" (see [27, Conjecture B]):

$$H_N(d) \sim 2c_2 J(d) \frac{N}{\ln^2 N} \quad (126)$$

where

$$c_2 = \prod_{p>2} \left(1 - \frac{1}{(p-1)^2}\right) \quad (127)$$

$$J(d) = \prod_{p|d} \frac{p-1}{p-2}. \quad (128)$$

The product is over all odd primes in (127), over odd primes dividing d in (128). The constant $C_2 = 2c_2 = 1.32032363 \dots$ is called the twin prime constant, while the term $J(d)$ is responsible for the irregularities and the oscillating pattern of the counting functions. A formula similar to (126) for the counting of successive prime pairs separated by a gap d , was suggested by Wolf [44, and the references therein] after enumerating, by means of a computer program, all gaps between consecutive primes up to $N = 2^{44} \approx 1.76 \times 10^{13}$. Starting from the observed pattern, which showed an exponential decay of the number of pairs depending on d , at the end of a heuristic procedure including approximations due to the *PNT*, he deduced the complete formula¹²

$$h_N(d) \sim 2c_2 J(d) \frac{\pi^2(N)}{N} e^{-d\pi(N)/N}. \quad (129)$$

The functional dependence on N of the formula above, can be transformed into a dependence on the sequence number n of the greatest prime p_n interested in the count. Indeed we can set $N = p_n \sim n \ln n$ and $\pi(N) = n$, hence it is easy to get

$$\frac{\pi^2(N)}{N} e^{-d\pi(N)/N} \sim F_n(d) = \frac{n}{\ln n} e^{-\frac{d}{\ln n}} \quad (130)$$

and the counting formula becomes

$$h_n(d) \sim 2c_2 J(d) \frac{n}{\ln n} e^{-\frac{d}{\ln n}} \quad (131)$$

where I changed the capital letter N to lower case n so that to remember the different meaning of the functional dependence.

The original Hardy and Littlewood conjecture was empirically tested by several authors and used to calculate the number of pairs $h_N(d)$ in the case of small gaps d between successive primes [7]. I show in the following that the conjecture of Wolf, which is based on a large amount of empirical data, can be derived from the combinatorial model we defined as *log(n)-Model* in Remark 5.3.

Remark 5.5. Note that the conjecture (131) can be written as

$$h_n(d) \propto \int_1^n \frac{\partial F_t(d)}{\partial t} dt$$

where $F_n(d)$ is the function defined in (130) and

$$\frac{\partial F_n(d)}{\partial n} = \frac{1}{\ln n} e^{-\frac{d}{\ln n}} (1 + o(1)). \quad (132)$$

¹² The procedure is reported in: M. Wolf (2018), "Some heuristics on the gaps between consecutive primes". Available at <https://arxiv.org/abs/1102.0481>

Definition 5.1. Given the sequence of primes $\{p_n, n = 1, 2, 3, \dots\}$, let's define the gap g_n between successive primes and the number of successive prime pairs $h_n(d)$ separated by a gap d depending on the sequence number n as

$$g_n = p_{n+1} - p_n$$

$$h_n(d) \# \{p_i : g_i = d, i \leq n\}.$$

The gap g_n can be modeled as the difference between the random variables

$$g_n = q_{n+1} - q_n$$

with probability mass function

$$\sigma_n(d) = \int_n^\infty \phi_{n,n+1}(x, x+d) dx \quad (133)$$

where $\phi_{n,n+1}(x, y)$ is the joint probability density function of the couple (q_n, q_{n+1}) . We can obtain the estimate of the function $h_n(d)$, as resulting from the combinatorial model adopted, through the equation

$$h_n(d) = \sum_{i=1}^n \sigma_i(d) \approx \int_1^n \sigma_t(d) dt. \quad (134)$$

We do the following assumption.

Assumption 5.1. (Large gaps independence law) For $d \gg 2$ sufficiently large, the two events $q_n = x$ and $q_{n+1} = x + d$ are independent.

Under the previous assumption the joint probability $\phi_{n,n+1}(x, x+d)$, can be written as the product of the two probability density functions

$$\phi_{n,n+1}(x, x+d) = \phi_n(x)\phi_{n+1}(x+d), d \gg 1 \quad (135)$$

and its value depends on the type of model we choose.

Let's calculate the joint probability (135) for the $\log(n)$ -Model defined in Remark 5.2. We adopt the following approximations

$$\ln(n+1) \approx \ln n$$

and

$$r_{n+1} \approx r_n + \ln n, n > 1,$$

valid for both ¹³

$$r_n = n \ln n$$

and

$$r_n = \text{ali}(n).$$

Then (135) becomes

$$\phi_n(x)\phi_{n+1}(x+d) = \frac{ea^2}{\ln^2 n} e^{-\frac{d}{\ln n}} e^{-2\left(\frac{x-r_n}{\ln n}\right)} e^{-ae^{-\left(\frac{x-r_n}{\ln n}\right)}(1+e^{1-\frac{d}{\ln n}})},$$

¹³ This means the results we are going to obtain depend on the PNT only, not on the Riemann Hypothesis.

and (133), after the variable change

$$\frac{x - r_n}{\ln n} = t,$$

is written

$$\sigma_n(d) = \frac{ea^2}{\ln n} e^{-\frac{d}{\ln n}} \int_{\frac{n-r_n}{\ln n}}^{\infty} e^{-2t} e^{-ae^{-t}(1+e^{1-\frac{d}{\ln n}})} dt.$$

Let α be

$$\alpha = a(1 + e^{1-\frac{d}{\ln n}}),$$

then the integral

$$\int_{\frac{n-r_n}{\ln n}}^{\infty} e^{-2t} e^{-e^{-t}\alpha} dt = \frac{1}{\alpha^2} \left(1 - e^{-\alpha e^{-\frac{n}{\ln n} + n}} - \alpha e^{-\frac{n}{\ln n} + n} e^{-\alpha e^{-\frac{n}{\ln n} + n}} \right) \approx \frac{1}{\alpha^2}, \text{ for } n \gg 1.$$

Finally, the probability mass function (133) resulting from the $\log(n)$ -Model under Assumption 5.1 and $n \gg 1$ is (note it is independent of parameter a)

$$\sigma_n(d) \approx \frac{e}{(1 + e^{1-\frac{d}{\ln n}})^2} \frac{e^{-\frac{d}{\ln n}}}{\ln n}. \quad (136)$$

Comparing the previous equation with (132), we see the $\log(n)$ -Model confirms Wolf's conjecture, in particular the exponential decay of the pair number depending on d , at least for large values of the gap d , when Assumption 5.1 can be considered valid. Note that in the case of the n -Model we would get the same expression as in (136), but with the function n in place of $\ln n$. The fact that only the RH rule compliant model agrees with the counting function of consecutive prime pairs, derived from empirical data, is a notable achievement of the theory, since it is a strong indication that the counting problem is deeply connected with the Riemann Hypothesis.

Figures 1 and 2 show the results obtained from the $\log(n)$ -Model with respect to the actual numbers of pairs in two cases: $49 \times 10^6 < n \leq 50 \times 10^6$ and $2^{47} \leq p_n \leq 2^{48}$.

5.3. A heuristic model of the distribution of prime numbers

The generalized model we presented in Subsection 5.1 can be directly related to the sequence of primes by changing condition (115) into the following

$$E[q_n] = f(n) [\ln(\lambda_n) + \gamma] = p_n \quad (137)$$

which leads to the value

$$\lambda_n = ae^{\frac{p_n}{f(n)}}, \text{ with } a = e^{-\gamma}$$

while the point of maximum of the probability density function (113) becomes

$$x_M = p_n - \gamma f(n).$$

It is interesting to note we would obtain the same expressions, with $a = 1$, if we substitute the condition (137) with the following one, on the point of maximum of the function $\phi_n(x)$

$$x_M = f(n) \ln(\lambda_n) = p_n. \quad (138)$$

In this case the mean value of q_n changes to

$$E[q_n] = p_n + \gamma f(n).$$

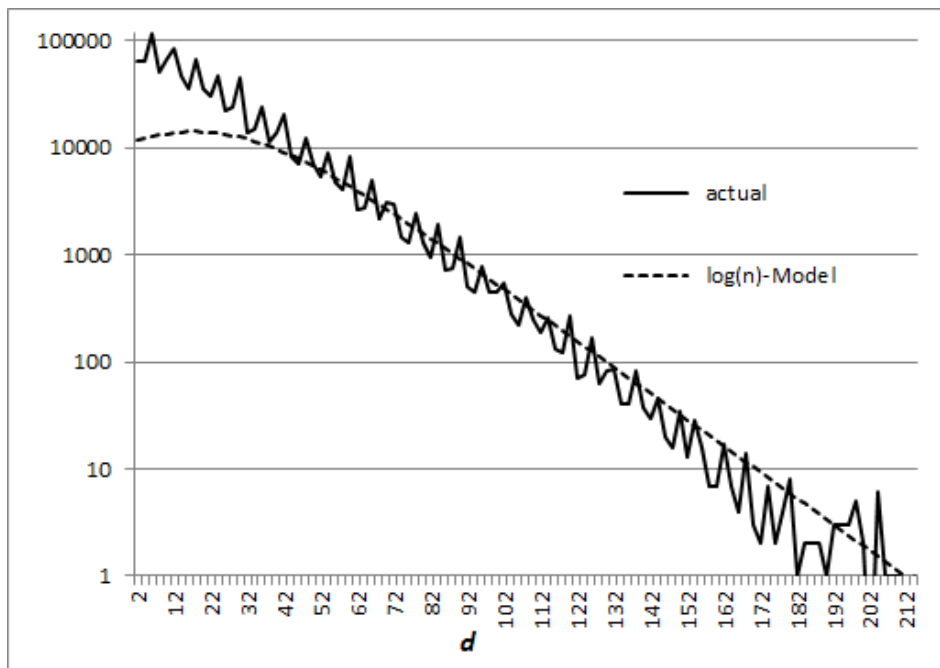


Figure 1. Numbers of successive prime pairs as function of the gap $d = p_{n+1} - p_n$ in $49,000,000 \leq n \leq 50,000,000$ (logarithmic scale)

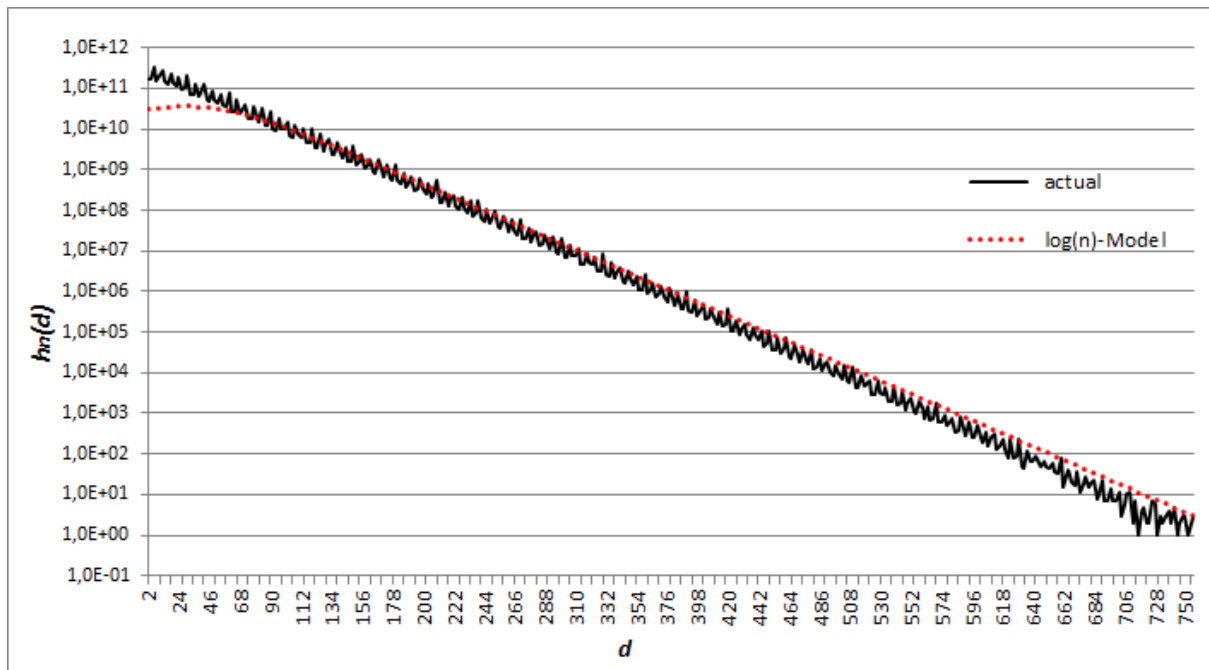


Figure 2. Numbers of successive prime pairs as function of the gap $d = p_{n+1} - p_n$ in $4,461,632,979,716 \leq n \leq 8,731,188,863,469$ (logarithmic scale)

This model has the property to express the prime counting function $\pi(k)$ through the values of the sequence of primes p_n from $n = 1$ to $n = k$. Let consider indeed the case when $f(n) = \sigma$ is constant. In this case the average of the random counting variable (102) is

$$E[\zeta_k] = \hat{\pi}(k, \sigma) = \sum_{n=1}^k e^{-ae^{-\frac{-k+p_n}{\sigma}}} \quad (139)$$

and it approximates the value of $\pi(k)$ with any precision depending on σ .

Remark 5.6. It holds

$$\sum_{n=1}^{\pi(k)} e^{-ae^{-\frac{-k+p_n}{\sigma}}} \leq \hat{\pi}(k, \sigma) \leq \sum_{n=1}^{\pi(k)} e^{-ae^{-\frac{-k+p_n}{\sigma}}} + 1 \quad (140)$$

for

$$\sigma < \frac{2}{\ln \ln(k - \pi(k)) + \gamma}$$

and

$$\lim_{\sigma \rightarrow 0^+} \hat{\pi}(k, \sigma) = \lim_{\sigma \rightarrow 0^+} \sum_{n=1}^{\pi(k)} e^{-ae^{-\frac{-k+p_n}{\sigma}}} = \pi(k). \quad (141)$$

Indeed

$$\hat{\pi}(k, \sigma) = \sum_{n=1}^l e^{-ae^{-\frac{-k+p_n}{\sigma}}} + \sum_{n=l+1}^k e^{-ae^{-\frac{-k+p_n}{\sigma}}}$$

since

$$\frac{l+1}{\sigma} > e^{-ae^{-\frac{-k+p_n}{\sigma}}}, \quad n = l+2, \dots, k$$

we can write

$$\sum_{n=1}^l e^{-ae^{-\frac{-k+p_n}{\sigma}}} \leq \hat{\pi}(k, \sigma) \leq \sum_{n=1}^l e^{-ae^{-\frac{-k+p_n}{\sigma}}} + e^{-ae^{-\frac{-k+p_{l+1}}{\sigma}}} (k-l).$$

Assuming $l = \pi(k)$ and $e^{-ae^{-\frac{-k+p_{l+1}}{\sigma}}} (k-l) \leq 1$, (140) easily follows. Note that for each term of the sum (139) it is

$$\lim_{\sigma \rightarrow 0^+} e^{-ae^{-\frac{-k+p_n}{\sigma}}} = \begin{cases} 1 & , \quad p_n < k \\ 0 & , \quad p_n > k \end{cases}$$

hence (141).

From equation (137) we get in the case of the $\log(n)$ -Model, $f(n) = \ln n$:

$$\phi_n(x) = \frac{a}{\ln n} e^{-\frac{x-p_n}{\ln n}} e^{-ae^{-\frac{x-p_n}{\ln n}}}, \quad (142)$$

$$\Phi_n(x) = e^{-ae^{-\frac{x-p_n}{\ln n}}}; \quad (143)$$

in the case of the n -Model, $f(n) = n$:

$$\phi_n(x) = \frac{a}{n} e^{-\frac{x-p_n}{n}} e^{-ae^{-\frac{x-p_n}{n}}}, \quad (144)$$

$$\Phi_n(x) = e^{-ae^{-\frac{x-p_n}{n}}}. \quad (145)$$

The probability functions above are valid for $x \geq n$ (remember (110)), $n > 1$ sufficiently large, and $a = e^{-\gamma}$.

Definition 5.2. Equations (142), (143) and (144), (145) define the heuristic log(n)-Model and the heuristic n-Model respectively.

Examples of different probability distribution functions $\phi_n(x)$ resulting for these heuristic models are given in Figure 3, for $n = 20, 30, 40$.

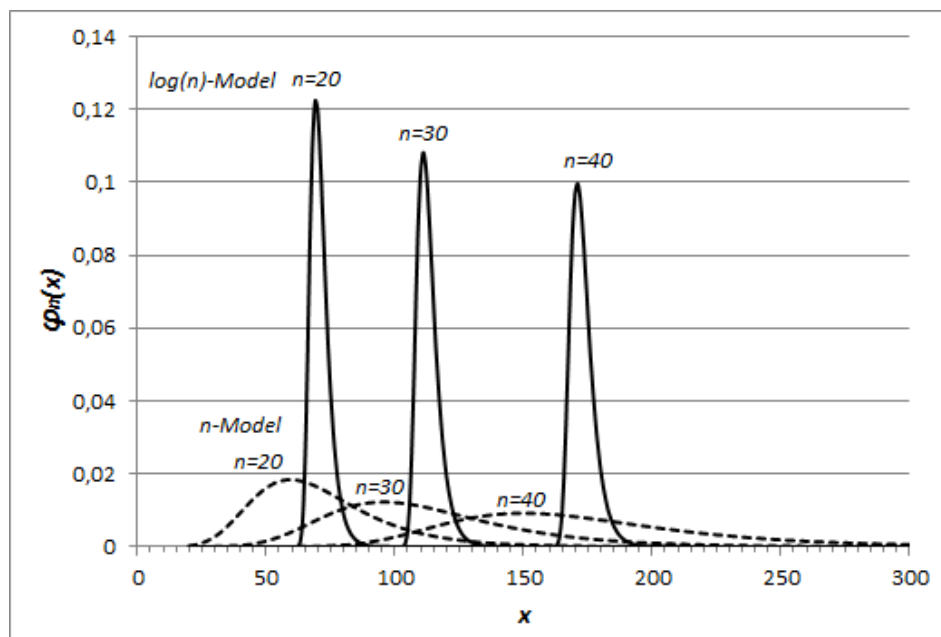


Figure 3. Heuristic model probability distribution functions $\phi_n(x)$ for $n = 20, 30, 40$

From Figure 3, we see the dispersion of the random variable q_n around its mean value p_n is very small in the case of the *heuristic log(n)-Model*, so that we can assume the estimate $\hat{\pi}(k)$ is very close to the prime-counting function itself $\pi(k)$ for this model, as stated by Remark 5.6 in the particular case $f(n) = \text{constant}$.

Table 5 shows some examples of estimate of $\pi(k)$, calculated through the *heuristic log(n)-Model* with $a = e^{-\gamma}$, for different orders of magnitude.

Table 5. $\pi(k)$ estimate through heuristic log(n)-Model

k	$\pi(k)$	$\hat{\pi}(k)$
10	4	4.33
100	25	25.50
1,000	168	167.65
10,000	1,229	1,229.34
100,000	9,592	9,592.09
1,000,000	78,498	78,498.23

The following considerations we propose in the case of the *heuristic log(n)-Model* because of its relevance, are obviously valid in the case $f(n) = \sigma$, constant.

The following theorem gives a quantitative evaluation of the probability of error we do, when considering the random variable ζ_k (in practice its mean value for small variances) as an estimate of $\pi(k)$.

Theorem 5.3. Given the random variable ξ_k of Definition 4.3, $\epsilon > 0$, let

$$\xi_{p_n - \lceil \epsilon \rceil} \geq n = \pi(p_n)$$

be the first kind error,

$$\xi_{p_n + \lceil \epsilon \rceil} < n = \pi(p_n)$$

the second kind error of ξ_k at $k = p_n$ and

$$W_{\xi}(n, \epsilon) = \text{Prob}\{\xi_{p_n - \lceil \epsilon \rceil} \geq n\} + \text{Prob}\{\xi_{p_n + \lceil \epsilon \rceil} < n\}$$

the sum of the two error probabilities. Then for the heuristic $\log(n)$ -Model defined by equations (142), (143) with $a = e^{-\gamma}$ it is

$$W_{\xi}(n, \epsilon) \leq \frac{\pi^2 \ln^2 n}{6 \epsilon^2}. \quad (146)$$

Proof. From the theory of Gumbel distribution we know the variance of q_n is

$$E[(q_n - p_n)^2] = \frac{\pi^2}{6} \ln^2 n$$

hence from Chebyshev's inequality

$$\text{Prob}\{|q_n - p_n| \geq \epsilon\} \leq \frac{\pi^2 \ln^2 n}{6 \epsilon^2}.$$

It is

$$\text{Prob}\{|q_n - p_n| \geq \epsilon\} = \Phi_n(p_n - \epsilon) + (1 - \Phi_n(p_n + \epsilon))$$

and

$$\Phi_n(p_n - \lceil \epsilon \rceil) + (1 - \Phi_n(p_n + \lceil \epsilon \rceil)) < \Phi_n(p_n - \epsilon) + (1 - \Phi_n(p_n + \epsilon)),$$

therefore we can write also

$$\Phi_n(p_n - \lceil \epsilon \rceil) + (1 - \Phi_n(p_n + \lceil \epsilon \rceil)) \leq \frac{\pi^2 \ln^2 n}{6 \epsilon^2}.$$

From (99), (100) it follows

$$\Phi_n(p_n - \lceil \epsilon \rceil) = \text{Prob}\{\xi_{p_n - \lceil \epsilon \rceil} \geq n\}$$

$$(1 - \Phi_n(p_n + \lceil \epsilon \rceil)) = 1 - \text{Prob}\{\xi_{p_n + \lceil \epsilon \rceil} \geq n\} = \text{Prob}\{\xi_{p_n + \lceil \epsilon \rceil} < n\}$$

$$\text{Prob}\{\xi_{p_n - \lceil \epsilon \rceil} \geq n\} + \text{Prob}\{\xi_{p_n + \lceil \epsilon \rceil} < n\} \leq \frac{\pi^2 \ln^2 n}{6 \epsilon^2}.$$

□

Remark 5.7. Through the heuristic model we can derive new lower and upper bounds of the prime-counting function $\pi(k)$:

$$\int_1^k e^{-ae^{-\frac{-k+\beta(n)}{\ln n}}} dn \leq \pi(k) \leq \int_1^k e^{-ae^{-\frac{-k+\alpha(n)}{\ln n}}} dn, \quad (147)$$

where

$$\alpha(n) = n(\ln n + \ln \ln n - 1), \quad (148)$$

$$\beta(n) = n(\ln n + \ln \ln n). \quad (149)$$

We know ([18], [38]) the following lower and upper bounds of the n -th prime p_n hold

$$\alpha(n) < p_n < \beta(n)$$

for $n \geq 2$ the left-hand side and $n \geq 6$ the right-hand side, hence for $\Phi_n(k)$ given by (143) it follows

$$e^{-ae^{-\frac{-k+\beta(n)}{\ln n}}} \leq \Phi_n(k) \leq e^{-ae^{-\frac{-k+\alpha(n)}{\ln n}}}, \quad n \geq 6, \quad (150)$$

and assuming $k \gg 6$ sufficiently large so that the condition $n \geq 6$ may be neglected, we can write (147) for $\hat{\pi}(k)$ given by

$$\hat{\pi}(k) = \sum_{n=1}^k \Phi_n(k).$$

From Remark 5.6 and Theorem 5.3 we know we can consider as valid for the heuristic model the approximation $\hat{\pi}(k) \approx \pi(k)$ from which we get (147).

6. Concluding remarks

The first aim of this work was to deepen the problem of randomness in the distribution of prime numbers, through such simple combinatorial objects as *First Occurrence Sequences*, showing new analogies between the classical set partition problem and the distribution of primes themselves. *First Occurrence Sequences* define a general class of objects for which the *Prime Number Theorem* holds, like for the prime sequence, but they fail to represent more stringent constraints, required by the *Riemann Hypothesis*, like the equivalent condition established by Helge von Koch I called *RH rule*. In order to investigate this second step, the simple model must be generalized (or the class of combinatorial objects must be restricted) to discriminate between *RH rule* compliant and not compliant models. The analysis based on probability methods shows the *Riemann Hypothesis* holds *w.p. 1* (together with the *Prime Number Theorem*, of course) for the class of random sequences represented by the *log(n)-Model* with mean equal to $ali(n)$, the inverse function of the logarithmic integral function. Therefore, we can conclude that the property represented by the *Riemann Hypothesis* through the *RH rule* is largely independent of the primes and belongs to a large class of random sequences. The question of whether this class also contains the sequence of prime numbers cannot be decided by the model except statistically. Something similar happens, within the context of analytical theory, when speaking of zeta function $\zeta(s)$, Dirichlet L-functions $L(s, \chi)$ [3, Chap. 12] (s is the complex variable $s = \sigma + it$) and the *Generalized Riemann Hypothesis* "that is the hypothesis that not only $\zeta(s)$ but all the functions $L(s, \chi)$ have their zeros in the critical strip on the line $\sigma = \frac{1}{2}$ " [16, p.124].

Attempts to prove the Riemann hypothesis through the methods of physics go back as far as the so-called Hilbert-Polya conjecture, which associates the imaginary part of non-trivial zeros of $\zeta(s)$ to the eigenvalues of some self-adjoint (Hermitian) operator that might be considered like the Hamiltonian of a physical system.¹⁴ Each type of combinatorial model we presented in the previous sections, based on probability equations like (111) and more generally (25), can be transformed into a single particle Hamiltonian of

¹⁴ The story of Hilbert-Polya conjecture is documented on Odlyzko's personal website. See: <http://www.dtc.umn.edu/~odlyzko/>

an equivalent quantum system that emerges as a solution to an underlying combinatorial problem. This topic is outside the scope of this work and will not be treated here, I just want to mention that the analogy with the physical problem allows us to overcome some critical points of the combinatorial model and eliminate any arbitrariness in the choice of the mean of the random variable q_n (the combinatorial counterpart of the n -th prime). The solution $ali(n)$ emerges as a general asymptotic solution for such models due to concepts like interaction potential and energy levels, that play an important role also in describing integer sequences like Fibonacci numbers and prime numbers, and the application of the Hellmann-Feynman theorem to the whole system (see [8, for a mathematical foundation of the theorem]). These methods result in computational benefits improving the estimates obtained from the combinatorial model and may suggest new conjectures about the distribution of primes. In [4] the quantum model derived from what we have called n -Model is developed and applied to explore the region beyond the Skewes number, in connection with the numerical results known in the literature derived through analytical methods.

These two topics (the parallelism between the treatment from the point of view of random sequences and of Riemann zeta and Dirichlet L-functions; the transposition of the problem into a quantum physics framework), seem worth to be addressed in future work.

Acknowledgments: I am grateful to prof. Marek Wolf (Cardinal Stefan Wyszyński University in Warsaw, Department of Physics) for his data about the number of successive prime pairs up to $p_n \leq 2^{48}$.

References

1. Abramowitz, M.; Stegun, I. A. editors. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*; Nat. Bureau of Standards: Washington, 1964.
2. Alzer, H. On Some Inequalities for the Incomplete Gamma Function. *Mathematics of Computation* **1997**, *66* 218, 771–778.
3. Apostol, T.M. *Introduction to analytic number theory. Undergraduate Texts in Mathematics*; Springer-Verlag: New York-Heidelberg, 1976.
4. Barbarani, V. A Quantum Model of the Distribution of Prime Numbers and the Riemann Hypothesis. *Int J Theor Phys* **2020**, *59*, 2425–2470.
5. Beirlant, J.; Goegebeur, Y.; Segers, J.J.J.; Teugels, J. *Statistics of Extremes: Theory and Applications*; Wiley: Chichester, 2004.
6. Boyadzhiev, K. N. Close Encounters with the Stirling Numbers of the Second Kind. *Mathematics Magazine* **2012**, *85*, 252–266.
7. Brent, R. P. The Distribution of small Gaps Between Successive Primes. *Mathematics of Computation* **1974**, *28* 12, 315–324.
8. Carfi, D. The Pointwise Hellmann-Feynman Theorem. *AAPP. Physical, Mathematical and Natural Sciences* **2010**, *88* 1.
9. Choi, J.; Srivastava, H. M. Some summation formulas involving harmonic numbers and generalized harmonic numbers. *Mathematical and Computer Modelling* **2011**, *54*, 2220–2234.
10. Chu, W.; De Donno, L. Hypergeometric series and harmonic number identities. *Advances in Applied Mathematics* **2005**, *34*, 123–137.
11. Chu, W.; De Donno, L. Identità Binomiali e Numeri Armonici. *Bollettino dell'Unione Matematica Italiana, Serie 8* **2007**, *10-B* 1, 213–235.
12. Conrey, J. B. The Riemann Hypothesis. *Notices of the AMS* **2003**, *50* 3, 341–353.
13. Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; Jeffrey, D. J.; Knuth, D. E. On the Lambert Function. *Advances in Computational Mathematics* **1996**, *5* 1, 329.
14. Cramér, H. Prime numbers and probability. *Skand. Math. Kongr.* **1935**, *8*, 107–115.
15. Cramér, H. On the order of magnitude of the difference between consecutive prime numbers. *Acta Arith.* **1936**, *2*, 23–46.
16. Davenport, H. *Multiplicative Number Theory. Graduate Texts in Mathematics*; Springer-Verlag: New York, 2000.
17. De Reyna, J. A.; Toulisse, J. The n -th prime asymptotically. *Journal de theorie de nombres de Bordeaux* **2013**, *25* 3, 521–555.
18. Dusart, P. The k -th prime is greater than $k(\ln k + \ln \ln k - 1)$ for $k \geq 2$. *Mathematics of Computation* **1999**, *68* 225, 411–415.
19. Feller, W. *An Introduction to Probability Theory and its Applications. Vol. I*, 3rd ed.; John Wiley and Sons: New York, 1968.
20. Flajolet, P.; Sedgewick, R. *Analytic Combinatorics*; Cambridge University Press: Cambridge, 2009.
21. Goldston, D. A.; Pintz, J.; Yıldırım, C. Y. Primes in tuples I. *Ann. of Math.* **2009**, *170*, 819–862.
22. Goldston, D. A.; Pintz, J.; Yıldırım, C. Y. Primes in tuples II. *Acta Math.* **2010**, *204*, 1–47.
23. Gould, H. W. *Combinatorial Identities: A Standardized Set of Tables Listing 500 Binomial Coefficient Summations*, Revised ed.; Morgantown Printing and Binding Company: Morgantown, West Virginia, 1972.
24. Graham, R. L.; Knuth, D. E.; Patashnik, O. *Concrete Mathematics*; Addison-Wesley: New York, 1994.
25. Granville, A. Harald Cramér and the Distribution of Prime Numbers. *Scand. Actuarial J.* **1995**, *1*, 12–28.
26. Gumbel, E. J. Les valeurs extrêmes des distributions statistiques. *Annales de l'I. H. P.* **1935**, *5* 2, 115–158.

27. Hardy, G. H.; Littlewood, J. E. Some problems of "Partitio numerorum", III: On the expression of a number as a sum of primes. *Acta Math.* **1923**, *44*, 1–70.
28. Harper, L. H. Stirling Behaviour is Asymptotically Normal. *The Annals of Mathematical Statistics* **1967**, *38* 2, 410–414.
29. Kingman, J. F. C.; Taylor, S. J. *Introduction to Measure and Probability*; Cambridge University Press: Cambridge, 1966.
30. Koch, von H. Sur la distribution des nombres premiers. *Acta Mathematica* **1901**, *24* 1, 159–182.
31. Louchard, G. Asymptotics of the Stirling numbers of the second kind revisited. *Applicable Analysis and Discrete Mathematics* **2013**, *7* 2, 193–218.
32. Maier, H. Primes in short intervals. *Michigan Math. J.* **1985**, *32*, 221–225.
33. Massias, J-P.; Robin, G. Bornes effectives pour certaines fonctions concernant les nombres premiers. *Journal de Theorie des Nombres de Bordeaux* **1996**, *8*, 215–242.
34. NIST. *Digital Library of Mathematical Functions*; Olver, F. W. J., Olde Daalhuis, A. B., Lozier, D. W., Schneider, B. I., Boisvert, R. F., Clark, C. W., Miller, B. R., Saunders, B. V., Cohl, H. S., McClain, M. A., Eds.; <http://dlmf.nist.gov/>. Release 1.0.28 of 2020-09-15.
35. Pintz, J. Cramér vs. Cramér. On Cramér's probabilistic model for primes. *Functiones et Approximatio Commentarii Mathematici* **2007**, *37*, DOI:10.7169/facm/1229619660.
36. Pitman, J. Some Probabilistic Aspects of Set Partitions. *The American Mathematical Monthly.* **1997**, *104* 3, 201–209.
37. Quaintance, J.; Gould, H. W. *Combinatorial Identities for Stirling Numbers: The Unpublished Notes of H. W. Gould*; World Scientific Publishing: Singapore, 2016.
38. Rosser, J. B. Explicit bounds for some functions of prime numbers. *American Journal of Mathematics* **1941**, *63* 1, 211–232.
39. Sachkov, V. N. Random Partitions of Sets. *SIAM Theory of Probability and its applications* **1974**, *19* 1, 184–190.
40. Sachkov, V. N. *Probabilistic Methods in Combinatorial Analysis*; Cambridge University Press: Cambridge, 1997.
41. Sinai, Y. G. *Probability Theory. An Introductory Course*; Springer-Verlag: Berlin Heidelberg, 1992.
42. Stanley, R. P. *Enumerative Combinatorics. Vol. 1*, 2nd ed.; Cambridge University Press: Cambridge, 2012.
43. Stam, A. J. Generation of a Random Partition of a Finite Set by a Urn Model. *Journal of Combinatorial Theory Series A* **1983**, *35*, 231–240.
44. Wolf, M. Application of statistical mechanics in number theory. *Physica A* **1999**, *274*, 149–157.
45. Zhang, Y. Bounded gaps between primes. *Annals of Mathematics* **2014**, *179* 3, 1121–1174.