Journal de Théorie des Nombres de Bordeaux

anciennement Séminaire de Théorie des Nombres de Bordeaux

Tanguy RIVOAL

On the distribution of Hawkins' random "primes"

Tome 20, n° 3 (2008), p. 799-809.

<http://jtnb.cedram.org/item?id=JTNB_2008__20_3_799_0>

Journal de Théorie des Nombres
de Bordeaux **20** (2008), 799-809

On the distribution of Hawkins' random "primes"

par Tanguy RIVOAL

Dedicated to Henri Cohen on the occasion of his 60th birthday

Résumé. Hawkins a défini une version probabiliste du crible d'Ératosthène et étudié la suite des nombres "premiers" aléatoires $(p_k)_{k\geq 1}$ ainsi créés. Au moyen de diverses techniques probabilistes, de nombreux auteurs ont ensuite obtenu des résultats très fins sur ces "premiers", souvent en accord avec des théorèmes ou conjectures classiques sur les nombres premiers usuels. Dans ce papier, on prouve que le nombre d'entiers $k \leq n$ tel que $p_{k+\alpha} - p_k = \alpha$ est presque sûrement équivalent à $n/\log(n)^\alpha$, pour tout entier $\alpha \geq 1$ fixé. C'est un cas particulier d'un travail récent de Bui and Keating (exprimé autrement) mais notre méthode est différente et fournit un terme d'erreur. On montre également que le nombre d'entiers $k \leq n$ tel que $p_k \in a\mathbb{N}+b$ est presque sûrement équivalent à n/a, pour tous entiers $a \geq 1$ et $0 \leq b \leq a-1$ fixés, ce qui peut être vu comme un analogue du théorème de Dirichlet.

Abstract. Hawkins introduced a probabilistic version of Erathosthenes' sieve and studied the associated sequence of random "primes" $(p_k)_{k\geq 1}$. Using various probabilistic techniques, many authors have obtained sharp results concerning these random "primes", which are often in agreement with certain classical theorems or conjectures for prime numbers. In this paper, we prove that the number of integers $k \leq n$ such that $p_{k+\alpha} - p_k = \alpha$ is almost surely equivalent to $n/\log(n)^\alpha$, for a given fixed integer $\alpha \geq 1$. This is a particular case of a recent result of Bui and Keating (differently formulated) but our method is different and enables us to provide an error term. We also prove that the number of integers $k \leq n$ such that $p_k \in a\mathbb{N}+b$ is almost surely equivalent to n/a, for given fixed integers $a \geq 1$ and $0 \leq b \leq a-1$, which is an analogue of Dirichlet's theorem.

1. Introduction

The simplest method for determining a not too large list of prime numbers is Erathosthenes' sieve. Legendre found an analytical formula for this sieve which can theoretically be used to compute any desired value of

Manuscrit reçu le 5 novembre 2007.

$\pi(x) := \#\{1 \leq k \leq x : k \text{ is prime}\}$. Furthermore, a variation of Legendre's formula can be used to prove that $\pi(x) \leq (e^{-\gamma} + o(1))x/\log\log(x)$ as $x \to +\infty$ (see [13, p. 57]), which is a non-trivial bound but far from the prime number theorem $\pi(x) \sim x/\log(x)$. Here, γ is Euler's constant which appears because of Mertens' theorem

$$\prod_{p\leq x}\Big(1-\frac{1}{p}\Big)^{-1} \sim e^{\gamma}\,\log(x) \quad (x \to +\infty),$$

where the product is over all prime numbers $p \leq x$. Modern sieve methods have partially fixed the flaws in Erathosthenes' sieve and enabled us to obtain results for $\pi(x)$ which are closer to the truth, as well as many other important results in analytical number theory.

Hawkins [5, 6] wondered what would be the behavior of the following random version of Erathosthenes' sieve. Let A_1 be the set of integers ≥ 2, set $p_1 = 2$ and delete independently the elements of $A_1 \setminus \{p_1\}$ with probability $1/p_1$. Denote A_2 the set of the remaining integers and by p_2 the smallest element of A_2 which is $> p_1$ and delete independently the elements of $A_2 \setminus \{p_1, p_2\}$ with probability $1/p_2$ and so on. This generates an increasing sequence $(p_n)_{n\geq 1}$ of random integers which mimics the usual prime numbers [1]). A natural problem is to estimate the asymptotic behavior of these random primes p_n and of the Mertens-like product

$$m_n := \prod_{1\leq k\leq n}\Big(1-\frac{1}{p_k}\Big)^{-1}.$$

There exist two methods for formalizing Hawkins' random sieve. The first method (which is combinatorial) was developped by Hawkins [5, 6] then Wunderlich [14, 15] and the second one was developped by Neudecker and Williams [12]. The latter noticed that $(p_n, m_n; \mathscr{F}_n)_{n\geq 1}$ is a markovian process for the natural filtration $\mathscr{F}_n = \sigma(p_k, k = 1, \ldots, n)$ defined by $p_1 = 2$, $m_1 = 2$ and

$$\mathbf{P}\big(p_{n+1} - p_n = j \big| \mathscr{F}_n\big) = \frac{1}{m_n}\Big(1-\frac{1}{m_n}\Big)^{j-1}$$

for all integers $j \geq 1$ and $n \geq 1$. The behaviour of p_n and m_n has been extensively studied and is now much better known than the behaviour of the prime numbers, for which many of the results quoted below and in the final section are still conjectures:

– Hawkins [6] proved a "prime number theorem" in L^1 and in probability: with $\Pi(n) = \sum_{p_k\leq n} 1$, we have $\mathbf{E}(\Pi(n)) \sim n/\log(n)$ and for all functions

[1]Those will not be used anywhere in the sequel and there is no problem in denoting Hawkins' "primes" by p_n. Of course, Hawkins' (or random) "primes" have no reason to be prime numbers but most of the time we will drop the quotation marks since there will not be any ambiguity.

ψ such that $\psi(n) = o(\log(n))$, $\mathbf{P}\big(|\Pi(n) - \mathbf{E}(\Pi(n))| \geq n\psi(n)/\log(n)^2\big) \ll 1/\psi(n)^2$.

– Wunderlich [14, 15] obtained almost sure (a.s.) analogues of Mertens' theorem and of the prime number theorem: almost surely, $m_n \sim \log(n)$, $\Pi(n) \sim n/\log(n)$ and $p_n \sim n\log(n)$ as $n \to +\infty$. He also proved a random analogue of the twin prime conjecture:

$$\#\{1 \leq k \leq n \,: p_{k+1} - p_k = 1\} \sim n/\log(n) \ \text{ a.s.},$$

– More recently, Bui and Keating [1] addressed the question of the number of Hawkins' primes p_n such that for example $p_{n+k} - p_n$ is bounded (for a given fixed $k \geq 1$). Lorch [10] also deviced a generalised random sieve to deal with such questions.

As a corollary of their results, which we do not quote, Bui and Keating proved in particular that, for any integers $\beta \geq 1$ and $d \geq 1$, we have that

$$\#\{j \leq x : j, j+d, j+2d, \dots, j+(\beta-1)d \ \text{ are Hawkins' primes}\} \sim \frac{x}{\log(x)^\beta} \ \text{ a.s.} \tag{1.1}$$

as $x \to +\infty$. Of course, for the prime numbers, the exact analogue of (1.1) is meaningless most of the time. However, Dickson [3] gave necessary conditions on positive integers $d_1 < d_2 < \cdots < d_{\beta-1}$ such that $j, j+d_1, \dots, j+d_{\beta-1}$ are simultaneously primes for infinitely many j and he conjectured that these conditions are sufficient. Furthermore, Hardy and Littlewood [4] conjectured that, in this case, there exists a constant $C(\underline{d}) > 0$ such that

$$\#\{j \leq x : j, j+d_1, j+d_2, \dots, j+d_{\beta-1} \ \text{ are primes}\} \sim C(\underline{d})\,\frac{x}{\log(x)^\beta}$$

as $x \to +\infty$. Both conjectures are still open.

In the first part of this paper, we give a proof of (a different formulation of) the case where $d = 1$ in (1.1), with an explicit error term which was not given in [1].

Theorem 1. *For any fixed integer $\alpha \geq 1$, we have*

$$\#\{1 \leq k \leq n : p_{k+\alpha} = p_k + \alpha\} = \frac{n}{\log(n)^\alpha} + \mathcal{O}\Big(\frac{n\log\log(n)}{\log(n)^{\alpha+1}}\Big) \quad \text{a.s.} \tag{1.2}$$

as $n \to +\infty$, where the implicit constant depends on α.

The result is also trivially true for $\alpha = 0$. When, $\alpha = 1$, we get one of Wunderlich's results mentioned before, plus an error term. Note that

$$\#\{1 \leq k \leq n : p_{k+\alpha} = p_k + \alpha\} = \#\{2 \leq j \leq p_n : j, j+1, \dots, j+\alpha \ \text{ are Hawkins' primes}\}. \tag{1.3}$$

By (1.2), the left hand side of (1.3) is equivalent to $n/\log(n)^\alpha$ a.s. while the right hand side (1.3) is equivalent to $p_n/\log(p_n)^{\alpha+1}$ a.s. by (1.1) with $\beta = \alpha + 1$ and $d = 1$: the almost sure random prime number theorem proved by Wunderlich provides the bridge between both estimates. Bui and Keating essentially use Wunderlich's combinatorial approach while we use here the markovian approach mentioned above.

In a second part, we consider the distribution of Hawkins' primes in the subsequences of the form $ak+b$ (a, b fixed integers) and prove the following result.

Theorem 2. *For any fixed integers $a \geq 1, b$ with $0 \leq b \leq a-1$, as $n \to +\infty$, we have*

$$\#\{1 \leq k \leq n : p_k \in a\mathbb{N} + b\} = \frac{1}{a}\, n + \mathcal{O}\Big(\frac{n}{\log(n)}\Big) \quad \text{a.s.} \tag{1.4}$$

as $n \to +\infty$, where the implicit constant depends on a.

This can be viewed as the analogue of Dirichlet-de la Vallée Poussin's theorem for the prime numbers in arithmetic progressions, where a similar estimate holds with $1/\varphi(a)$ (φ is Euler's totient) instead of $1/a$ and with the further assumption that a and b are coprime. Of course, it is not surprising that Hawkins' sieve cannot detect arithmetical facts such as coprimality or Dickson's conditions.

The proofs of both theorems will use the following result, which gives the speed of convergence in a generalisation of the strong law of large numbers. We will find that the right hand side of (1.5) is easier to control than the left hand side, the latter corresponding to the quantity we want to estimate.

Proposition 1. *Consider a process $(X_n; \mathscr{F}_n)_{n\geq 1}$ and an increasing sequence $(b_n)_{n\geq 1}$ of real numbers such that $\sum_{n=1}^\infty b_n^{-2}\operatorname{Var}(X_n) < +\infty$ and $b_n \to +\infty$. Then, as $k \to +\infty$,*

$$\sum_{n\leq k} X_n = \sum_{n\leq k} \mathbf{E}\big(X_n|\mathscr{F}_{n-1}\big) + o(b_k) \quad \text{a.s..} \tag{1.5}$$

A proof can be found in Loeve's book [9, p. 387, E].

Acknowledgement. I warmly thank the referee for his/her very careful reading of the paper.

2. Proof of Theorem 1

As already implicit in the Introduction, $\mathbf{P}$ and $\mathbf{E}$ denote respectively the probability and (conditional) expectation on the probability space on which the Markov process $(p_n, m_n; \mathscr{F}_n)_{n\geq 1}$ and all the other processes considered below are defined.

We need two lemmas to prove Theorem 1.

Lemma 1. *For all $\alpha \geq 1$ and $n \geq 1$, we have*

$$\mathbf{P}\big(p_{n+\alpha} - p_n = \alpha \big| \mathscr{F}_n\big) = \frac{1}{m_n^\alpha} \prod_{j=1}^{\alpha-1} \Big(1 - \frac{1}{p_n + j}\Big)^{\alpha-j}, \tag{2.1}$$

where, by convention, the empty product is 1 *for* $\alpha = 1$.

Proof. For simplicity, we note $\mathbf{I}_{n+1} = \mathbf{1}_{\{1\}}(p_{n+1} - p_n)$ where $\mathbf{1}_A(x)$ is the indicator function of a given set A.

We prove (2.1) by induction on $\alpha \geq 1$. When $\alpha = 1$, the identity (2.1) reduces to the equality $\mathbf{E}(\mathbf{I}_{n+1}|\mathscr{F}_n) = m_n^{-1}$, which is true for all $n \geq 1$ by definition. Let us suppose that (2.1) holds for $\alpha - 1$ and for all $n \geq 1$. Note that since $(p_n)_{n\geq 1}$ is strictly increasing, we have

$$\{p_{n+\alpha} - p_n = \alpha\} = \bigcap_{j=1}^{\alpha} \{p_{n+j} - p_{n+j-1} = 1\}.$$

Therefore, the basic properties of conditional expectations justify that the following chain of equalities holds for any integer $n \geq 1$:

$$\begin{aligned}
&\mathbf{P}\big(p_{n+\alpha} - p_n = \alpha \big| \mathscr{F}_n\big)\\
&= \mathbf{E}\Big(\mathbf{P}\big(p_{n+\alpha} - p_n = \alpha \big| \mathscr{F}_{n+1}\big) \Big| \mathscr{F}_n\Big)\\
&= \mathbf{E}\Big(\mathbf{E}\big(\mathbf{I}_{n+1}\mathbf{I}_{n+2}\cdots\mathbf{I}_{n+\alpha} \big| \mathscr{F}_{n+1}\big) \Big| \mathscr{F}_n\Big)\\
&= \mathbf{E}\Big(\mathbf{I}_{n+1}\, \mathbf{E}\big(\mathbf{I}_{n+2}\cdots\mathbf{I}_{n+\alpha} \big| \mathscr{F}_{n+1}\big) \Big| \mathscr{F}_n\Big)\\
&= \mathbf{E}\Big(\mathbf{I}_{n+1}\, \mathbf{P}\big(p_{n+1+\alpha-1} - p_{n+1} = \alpha - 1 \big| \mathscr{F}_{n+1}\big) \Big| \mathscr{F}_n\Big). \qquad (2.2)
\end{aligned}$$

We can apply the induction hypothesis to the probability $\mathbf{P}(\cdots|\mathscr{F}_{n+1})$ occuring inside (2.2) and we get

$$\begin{aligned}
&\mathbf{P}\big(p_{n+\alpha} - p_n = \alpha \big| \mathscr{F}_n\big)\\
&= \mathbf{E}\Big(\frac{\mathbf{I}_{n+1}}{m_{n+1}^{\alpha-1}} \prod_{j=1}^{\alpha-2} \Big(1 - \frac{1}{p_{n+1} + j}\Big)^{\alpha-j-1} \Big| \mathscr{F}_n\Big)\\
&= \mathbf{E}\Big(\frac{\mathbf{I}_{n+1}}{m_n^{\alpha-1}} \Big(1 - \frac{1}{p_{n+1}}\Big)^{\alpha-1} \prod_{j=1}^{\alpha-2} \Big(1 - \frac{1}{p_{n+1} + j}\Big)^{\alpha-j-1} \Big| \mathscr{F}_n\Big)\\
&= \mathbf{E}\Big(\frac{\mathbf{I}_{n+1}}{m_n^{\alpha-1}} \Big(1 - \frac{1}{p_n + 1}\Big)^{\alpha-1} \prod_{j=1}^{\alpha-2} \Big(1 - \frac{1}{p_n + j + 1}\Big)^{\alpha-j-1} \Big| \mathscr{F}_n\Big)\\
&= \frac{\mathbf{E}(\mathbf{I}_{n+1}|\mathscr{F}_n)}{m_n^{\alpha-1}} \prod_{j=1}^{\alpha-1} \Big(1 - \frac{1}{p_n + j}\Big)^{\alpha-j} = \frac{1}{m_n^\alpha} \prod_{j=1}^{\alpha-1} \Big(1 - \frac{1}{p_n + j}\Big)^{\alpha-j},
\end{aligned}$$

where we used the definition of m_{n+1} in the second equality. This finishes the induction. $\square$

Lemma 2. *For all $\alpha \geq 1$, we have*

$$\sum_{n=1}^{k} \frac{1}{m_n^\alpha} = \frac{k}{\log(k)^\alpha} + \mathcal{O}\Big(\frac{k \log\log(k)}{\log(k)^{\alpha+1}}\Big) \quad \text{a.s.} \tag{2.3}$$

as $k \to +\infty$. The implicit constant depends on α.

Proof. All the implicit constants in the $\mathcal{O}$ and $\ll$ symbols below depend (at most) on α.

Heyde proved in [8] that $m_n - \log(n) \sim \log\log(n)$ a.s. as $n \to +\infty$. Hence, we have

$$\frac{1}{m_n^\alpha} - \frac{1}{\log(n)^\alpha} = \frac{\log(n)^\alpha - m_n^\alpha}{m_n^\alpha \log(n)^\alpha} \ll \frac{\log\log(n)}{\log(n)^{\alpha+1}} \quad \text{a.s.}$$

and thus

$$\sum_{n=3}^{k} \frac{1}{m_n^\alpha} = \sum_{n=3}^{k} \frac{1}{\log(n)^\alpha} + \mathcal{O}\Big(\sum_{n=3}^{k} \frac{\log\log(n)}{\log(n)^{\alpha+1}}\Big) \quad \text{a.s..}$$

Since

$$\sum_{n=3}^{k} \frac{1}{\log(n)^\alpha} = \frac{k}{\log(k)^\alpha} + \mathcal{O}\Big(\frac{k}{\log(k)^{\alpha+1}}\Big)$$

and

$$\sum_{n=3}^{k} \frac{\log\log(n)}{\log(n)^{\alpha+1}} \ll \log\log(k) \sum_{n=3}^{k} \frac{1}{\log(n)^{\alpha+1}} \ll \frac{k \log\log(k)}{\log(k)^{\alpha+1}},$$

the result follows. $\square$

We are now ready to prove Theorem 1. We have to estimate the asymptotic behavior of

$$\Pi_\alpha(k) := \#\{1 \leq n \leq k : p_{n+\alpha} = p_n + \alpha\} = \sum_{n=1}^{k} X_{n+1},$$

where $X_{n+1} = \mathbf{1}_{\{\alpha\}}(p_{n+\alpha} - p_n)$. Using Lemma 1, we have

$$\mathbf{E}(X_{n+1} | \mathscr{F}_n) = \mathbf{P}(p_{n+\alpha} - p_n = \alpha | \mathscr{F}_n) = x_n m_n^{-\alpha},$$

where

$$x_n = \prod_{j=1}^{\alpha-1} \Big(1 - \frac{1}{p_n + j}\Big)^{\alpha-j}.$$

Furthermore, since $X_{n+1} = 0$ or 1, we have

$$\mathrm{Var}(X_{n+1}) = \mathbf{E}(X_{n+1}^2) - \mathbf{E}(X_{n+1})^2 \leq 1.$$

Despite what is suggested by the notation, the random variable X_{n+1} is not $\mathscr{F}_{n+1}$-measurable if $\alpha \geq 2$ but only $\mathscr{F}_{n+\alpha}$-measurable. Therefore, if $\alpha \geq 2$, we cannot apply Proposition 1 directly to estimate $\sum_{n=1}^{k} X_{n+1}$ in terms of $\sum_{n=1}^{k} \mathbf{E}(X_{n+1}|\mathscr{F}_n)$, as we would like to do. To solve this problem, let us define recursively the random variables $X_{n+1}^{(j)}$, $j = 0, \ldots, \alpha$, $n \geq 1$, by $X_{n+1}^{(0)} = X_{n+1}$ and

$$X_{n+1}^{(j)} = \mathbf{E}\big(X_{n+1}^{(j-1)}|\mathscr{F}_{n+\alpha-j}\big).$$

For any given j, n, the random variable $X_{n+1}^{(j)}$ is $\mathscr{F}_{n+\alpha-j}$-measurable and we also have $X_{n+1}^{(j)} = \mathbf{E}\big(X_{n+1}|\mathscr{F}_{n+\alpha-j}\big)$. In particular, $X_{n+1}^{(\alpha)} = \mathbf{E}\big(X_{n+1}|\mathscr{F}_n\big)$. Furthermore, the inequality $\mathbf{E}(Z|\mathscr{G})^2 \leq \mathbf{E}(Z^2|\mathscr{G})$, which is a special case of Jensen's inequality, implies that

$$\operatorname{Var}\big(X_{n+1}^{(\alpha)}\big) \leq \operatorname{Var}\big(X_{n+1}^{(\alpha-1)}\big) \leq \cdots \leq \operatorname{Var}\big(X_{n+1}\big) \leq 1. \tag{2.4}$$

Using (2.4), for a given $j \in \{0, \ldots, \alpha - 1\}$, we can apply Proposition 1 to the process $(X_{n+1}^{(j)}; \mathscr{F}_{n+\alpha-j})_{n\geq 1}$ with, for example, $b_n = n^{1/2+\varepsilon}$ for any fixed $\varepsilon > 0$ to be specified later. We obtain that

$$\begin{aligned}\sum_{n=1}^{k} X_{n+1}^{(j)} &= \sum_{n=1}^{k} \mathbf{E}\big(X_{n+1}^{(j)}|\mathscr{F}_{n+\alpha-j-1}\big) + o(b_k)\\ &= \sum_{n=1}^{k} X_{n+1}^{(j+1)} + o(b_k) \quad \text{a.s.,}\end{aligned}$$

where the constant in the o depends on α and j. Hence,

$$\sum_{j=0}^{\alpha-1}\sum_{n=1}^{k}\big(X_{n+1}^{(j)} - X_{n+1}^{(j+1)}\big) = \sum_{j=0}^{\alpha-1} o(b_k) = o(b_k) \quad \text{a.s..}$$

On the left hand side, we have a telescoping sum (on j) and after simplifications, we get

$$\sum_{n=1}^{k} X_{n+1} = \sum_{n=1}^{k} \mathbf{E}\big(X_{n+1}|\mathscr{F}_n\big) + o(b_k) \quad \text{a.s..}$$

The last equality can be rewritten as

$$\Pi_\alpha(k) = \sum_{n=1}^{k} \frac{x_n}{m_n^\alpha} + \mathcal{O}(k^{1/2+\varepsilon}) \quad \text{a.s.}$$

as $k \to +\infty$. Since $p_n \to +\infty$, we have $x_n = 1 + \mathcal{O}(p_n^{-1})$, where the implicit constant only depends on α, and thus

$$\Pi_\alpha(k) = \sum_{n=1}^{k} \frac{1}{m_n^\alpha} + \mathcal{O}\Big(\sum_{n=1}^{k} \frac{1}{m_n^\alpha p_n}\Big) + \mathcal{O}(k^{1/2+\varepsilon}) \quad \text{a.s..}$$

To finish the proof of the theorem, we note that the series of term $1/(m_n^\alpha p_n)$ is almost surely convergent by the results of Wunderlich quoted in the introduction (remember that $\alpha \geq 1$). Therefore, using Lemma 2 we get that, almost surely,

$$\begin{aligned}\Pi_\alpha(k) &= \frac{k}{\log(k)^\alpha} + \mathcal{O}\Big(\frac{k\log\log(k)}{\log(k)^{\alpha+1}}\Big) + \mathcal{O}(1) + \mathcal{O}(k^{1/2+\varepsilon}) \\ &= \frac{k}{\log(k)^\alpha} + \mathcal{O}\Big(\frac{k\log\log(k)}{\log(k)^{\alpha+1}}\Big),\end{aligned}$$

provided that $\varepsilon < 1/2$. This finishes the proof of Theorem 1.

3. Proof of Theorem 2

Given a real number x, the smallest integer $\geq x$ will be denoted as usual by $\lceil x \rceil$ and we set $R_n = 1 - m_n^{-1}$. Consider the random variable Y_n which takes the value 1 if $p_n \in a\mathbb{N} + b$ and 0 otherwise. The process $(Y_n)_{n\geq 1}$ is $\mathscr{F}_n$-adapted. We want to estimate $\Pi_D(k) := \sum_{n=1}^k Y_n$ as $k \to +\infty$ and for this we follow the same approach as previously: we seek an increasing sequence b_k, as small as possible, such that

$$\sum_{n=1}^{k} Y_{n+1} = \sum_{n=1}^{k} \mathbf{E}(Y_{n+1}|\mathscr{F}_n) + o(b_k).$$

We have

$$\begin{aligned}\mathbf{E}\big(Y_{n+1}|\mathscr{F}_n\big) &= \mathbf{P}(p_{n+1} \in a\mathbb{N} + b|\mathscr{F}_n) \\ &= \mathbf{P}(p_{n+1} - p_n \in a\mathbb{N} - p_n + b|\mathscr{F}_n) \\ &= \frac{1}{m_n} \sum_{k \geq \lceil (p_n - b+1)/a\rceil} R_n^{ak-p_n+b-1} \\ &= \frac{1}{m_n(1-R_n^a)}\, R_n^{a\lceil (p_n-b+1)/a\rceil - (p_n-b+1)},\end{aligned}$$

where we summed the geometric series to get the last equality.

Since $Y_n = 0$ or 1, we have $\mathrm{Var}(Y_n) \leq 1$. Therefore, we can apply Proposition 1 to the process $(Y_n; \mathscr{F}_n)_{n\geq 1}$ with $(b_k)_k$ any increasing sequence of real numbers such that $\sum_k 1/b_k^2 < +\infty$, for example $b_k = k^{1/2+\varepsilon}$ for any

$\varepsilon > 0$ to be further specified later. Then, $\Pi_D(k) = S_k + o(k^{1/2+\varepsilon})$, where

$$S_k = \sum_{n=1}^{k} \frac{1}{m_n(1-R_n^a)}\, R_n^{a\lceil (p_n-b+1)/a\rceil - (p_n-b+1)}.$$

It is not clear why it is simpler to estimate S_k rather than $\Pi_D(k)$ but this turns out to be the case, as we will now show.

Since

$$0 \le a\lceil (p_n-b+1)/a\rceil - (p_n-b+1) \le a,$$

we have

$$\sum_{n=1}^{k} \frac{R_n^a}{m_n(1-R_n^a)} \le S_k \le \sum_{n=1}^{k} \frac{1}{m_n(1-R_n^a)}.$$

Furthermore, since $0 < R_n < 1$, we trivially have that

$$\frac{1}{a} \le \frac{1-R_n}{1-R_n^a} \le \frac{1}{aR_n^{a-1}}.$$

Hence, together with the fact that $1 - R_n = m_n^{-1}$, we obtain that

$$\frac{1}{a}\sum_{n=1}^{k} R_n^a \le S_k \le \frac{1}{a}\sum_{n=1}^{k} \frac{1}{R_n^{a-1}}. \tag{3.1}$$

For any real number d, we have $R_n^d = 1 + \mathcal{O}(m_n^{-1})$ as $n \to +\infty$, where the implicit constant depends on d. Thus, we deduce from (3.1) that

$$S_k = \frac{1}{a}\sum_{n=1}^{k}\Big(1 + \mathcal{O}\Big(\frac{1}{m_n}\Big)\Big) = \frac{k}{a} + \mathcal{O}\Big(\sum_{n=1}^{k}\frac{1}{m_n}\Big) = \frac{k}{a} + \mathcal{O}\Big(\frac{k}{\log(k)}\Big) \quad \text{a.s.},$$

where we used (a weak form of) Lemma 2 with $\alpha = 1$ to estimate $\sum_{n=1}^{k} m_n^{-1}$. Hence, we have

$$\Pi_D(k) = \frac{k}{a} + \mathcal{O}\Big(\frac{k}{\log(k)}\Big) + \mathcal{O}(k^{1/2+\varepsilon}) \quad \text{a.s.}$$

and the result follows provided we choose $\varepsilon < 1/2$.

4. Further readings and some problems

We conclude by mentioning other results in the literature, which were not necessarily used in this paper but which might interest the reader. We set $\mathrm{li}(x) = \int_2^\infty \frac{\mathrm{d}t}{\log(t)}$, which is equivalent to $x/\log(x)$ when $x \to +\infty$.

– Neudecker and Williams [12] proved the following analogue of the Riemann hypothesis: there exists a finite non zero random variable L such that $L = \lim_n p_n \exp(-m_n)$ a.s. and $L\,\mathrm{li}(p_n/L) = n + \mathcal{O}(n^{1/2+\varepsilon})$ a.s.. Remember that one of the many formulations of the Riemann's hypothesis is that $\mathrm{li}(p_n) = n + \mathcal{O}(n^{1/2+\varepsilon})$.

– Heyde [7] improved the previous result and showed that

$$\limsup_{n\to+\infty} \frac{|L\,\mathrm{li}(p_n/L)-n|}{\sqrt{2n\log\log(n)}} \leq 3 \ \text{ a.s.},$$

In [8], he also proved that $p_n - n\log(n) \sim n\log\log(n)$ and $m_n - \log(n) \sim \log\log(n)$ a.s. as $n \to +\infty$.

– Neudecker [11] showed that

$$\limsup_{n\to+\infty} \frac{p_{n+1}-p_n}{\log(p_n)^2} = 1 \quad \text{a.s.,} \tag{4.1}$$

which is an analogue of Cramér's conjecture [2] (i.e., that (4.1) holds if the p_n are replaced by the prime numbers). Cramér made his conjecture in the setting of his famous probabilistic model for the primes, which has nothing to do with Hawkins' model.

It would be interesting to continue the study of Hawkins' sieve. For example, is it possible to obtain non trivial bounds for the repartition of random primes which are values of a given polynomial of degree $d \geq 2$? What about random primes which are of the form $a^n + b$, for fixed $a \geq 2, b \in \mathbb{Z}$? What can be said about the integers which can be written as the sum of two random primes?

Bibliography

[1] H. M. Bui, J. P. Keating, *On twin primes associated with the Hawkins random sieve.* J. Number Theory **119**.2 (2006), 284–296.

[2] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers.* Acta Arith. **2** (1936), 23–46.

[3] L. Dickson, *A new extension of Dirichlet's theorem on prime numbers.* Messenger of Math. **33** (1904), 155–161.

[4] G. H. Hardy, J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes.* Acta Math. **44** (1923), 1–70.

[5] D. Hawkins, *The random sieve.* Math. Mag. **31** (1957/1958), 1–3.

[6] D. Hawkins, *Random sieves. II.* J. Number Theory **6** (1974), 192–200.

[7] C. C. Heyde, *A loglog improvement to the Riemann hypothesis for the Hawkins random sieve.* Ann. Probab. **6** (1978), no. 5, 870–875.

[8] C. C. Heyde, *On asymptotic behavior for the Hawkins random sieve.* Proc. AMS **56** (1976), 277–280.

[9] M. Loève, *Probability theory*, Third edition. D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto, Ont.-London, 1963.

[10] J. D. Lorch, *A generalized Hawkins sieve and prime k-tuplets.* Rocky Mountain J. Math. **37** (2007), no. 2, 533–550.

[11] W. Neudecker, *On twin "primes" and gaps between successive "primes" for the Hawkins random sieve.* Math. Proc. Cambridge Philos. Soc. **77** (1975), 365–367.

[12] W. Neudecker, D. Williams, *The 'Riemann hypothesis' for the Hawkins random sieve.* Compositio Math. **29** (1974), 197–200.

[13] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Deuxième édition. Cours Spécialisés, Société Mathématique de France, Paris, 1995.

[14] M. C. Wunderlich, *A probabilistic setting for prime number theory.* Acta Arith. **26** (1974), 59–81.

[15] M. C. Wunderlich, *The prime number theorem for random sequences*. J. Number Theory **8** (1976), no. 4, 369–371.

Tanguy Rivoal
Institut Fourier,
CNRS UMR 5582, Université Grenoble 1,
100 rue des Maths, BP 74,
38402 Saint-Martin d'Hères cedex, France.
E-mail: rivoal@ujf-grenoble.fr