



Critical probabilistic characteristics of the Cramér model for primes and arithmetical properties

Michel J. G. Weber

(Dedicated to the memory of J. Kubilius)

Received: 21 February 2023 / Accepted: 13 December 2023 / Published online: 8 January 2024
 © The Indian National Science Academy 2023

Abstract This work is a probabilistic study of the ‘primes’ of the Cramér model, which consists with sums $S_n = \sum_{i=3}^n \xi_i$, $n \geq 3$, where ξ_i are independent random variables such that $\mathbb{P}\{\xi_i = 1\} = 1 - \mathbb{P}\{\xi_i = 1\} = 1/\log i$, $i \geq 3$. We prove that there exists a set of integers $\mathcal{S}$ of density 1 such that

$$\liminf_{\mathcal{S} \ni n \rightarrow \infty} (\log n) \mathbb{P}\{S_n \text{ prime}\} \geq \frac{1}{\sqrt{2\pi e}}, \quad (0.0.1)$$

and that for $b > \frac{1}{2}$, the formula

$$\mathbb{P}\{S_n \text{ prime}\} = \frac{(1 + o(1))}{\sqrt{2\pi B_n}} \int_{m_n - \sqrt{2bB_n \log n}}^{m_n + \sqrt{2bB_n \log n}} e^{-\frac{(t-m_n)^2}{2B_n}} d\pi(t), \quad (0.0.2)$$

in which $m_n = \mathbb{E} S_n$, $B_n = \text{Var } S_n$, holds true for all $n \in \mathcal{S}$, $n \rightarrow \infty$. Further we prove that for any $0 < \eta < 1$, and all n large enough and $\zeta_0 \leq \zeta \leq \exp\left\{\frac{c \log n}{\log \log n}\right\}$, letting $S'_n = \sum_{j=8}^n \xi_j$,

$$\mathbb{P}\{S'_n \zeta\text{-quasiprime}\} \geq (1 - \eta) \frac{e^{-\gamma}}{\log \zeta},$$

according to Pintz’s terminology, where $c > 0$ and γ is Euler’s constant. We also test which infinite sequences of primes are ultimately avoided by the ‘primes’ of the Cramér model, with probability 1. Moreover we show that the Cramér model has incidences on the Prime Number Theorem, since it predicts that the error term is sensitive to subsequences. We obtain sharp results on the length and the number of occurrences of intervals I such as for some $z > 0$,

$$\sup_{n \in I} \frac{|S_n - m_n|}{\sqrt{B_n}} \leq z, \quad (0.0.3)$$

which are tied with the spectrum of the Sturm-Liouville equation.

Keywords Cramér’s model · Riemann hypothesis · Gap between primes · Primes · Divisors · Quasi-prime · Subsequences · Probabilistic models

Mathematics Subject Classification Primary 11A25 · 11N05, Secondary 11B83

Communicated by Rajeeva Karandikar.

M. J. G. Weber (✉)
 IRMA, UMR 7501, 10 Rue du Général-Zimmer, 67084 Strasbourg, Cedex, France
 E-mail: michel.weber@math.unistra.fr

1 Introduction

This article is a probabilistic study of the ‘primes’ of the Cramér model, most of the results obtained have an easy arithmetical interpretation. Let $\mathcal{P} = \{p_i, i \geq 1\}$ denote the sequence of consecutive prime numbers. Cramér’s probabilistic model basically consists with a sequence of independent random variables ξ_i , defined for $i \geq 3$ by

$$\mathbb{P}\{\xi_i = 1\} = \frac{1}{\log i}, \quad \mathbb{P}\{\xi_i = 0\} = 1 - \frac{1}{\log i}, \quad (1.1)$$

and associated partial sums $S_n = \sum_{i=3}^n \xi_i$.

The well-known Cramér twin prime conjecture, which seems largely accepted, does not however assert that there are any primes in the ‘primes’ of the model. This natural question does not appear in the literature and has motivated us. This work is the first, to our knowledge, in which the probability of primality or quasi-primality of the numbers - call them for a while (P_n) - in a probabilistic model for the primes, is thoroughly investigated.

The ‘primes’ (P_n) in Cramér’s model are the instants of jump of the random walk $S = \{S_n, n \geq 3\}$. Prior to doing it, we study (Section 3) the probability $\mathbb{P}\{S_n \text{ prime}\}$ and $\mathbb{P}\{S_n \zeta\text{-quasiprime}\}$ and for, use a result of Selberg [28] on increments of the prime counting function $\pi(\cdot)$. We show that the Cramér ‘primes’ are contained in the set of ‘primes’ of the Bernoulli model, and establish (Section 4) a criterion to test which infinite sequences of primes are with probability 1, ultimately avoided by the ‘primes’ of the model. We also describe new results of different type. We prove that Cramér’s easier prediction on the remainder in the ‘prime number theorem’ cannot be separated from the fact that this one should also be sensitive to the subsequence on which it is observed. Such a property was never emphasized nowhere, and seems hardly conceivable. Using Sakhanenko’s invariance principle [27] and small local oscillation results of the Ornstein-Uhlenbeck process [30], we obtain very sharp estimates on the number of occurrences and the length of intervals I for which

$$\sup_{n \in I} \frac{|S_n - \mathbb{E} S_n|}{\sqrt{\text{Var}(S_n)}} \leq z, \quad (1.2)$$

where z is linked to the spectrum of the Sturm-Liouville equation.

Other more recent probabilistic models are refinements of Cramér’s model in the sense that finer conjectures such as Hardy-Littlewood’s conjecture generalizing the twin-prime conjecture, are *almost surely* true in these models, with *no* consequence for the prime numbers, however. There indeed seems to be an impassable chasm between results which are valid “almost surely” in the model and corresponding results for the sequence of prime numbers. To the best of our knowledge, no probabilistic model for the prime numbers never permitted to establish new properties for these ones.

One of the interesting aspects of the present study, and not the least, is that it can be at least partly adapted to other models for the prime numbers.

Some preliminary facts, at first it is easy to check that for this model, the standard limit theorems from probability theory are fulfilled: the strong law of large numbers (SLLN), the central limit theorem (CLT), the law of the iterated logarithm (LIL), the local limit theorem (LLT), and also an invariance principle (IP) hold (Proposition 2.3). This is true in a wider setting. These points are briefly detailed and completed in Appendix A, where in addition, a sharp estimate of the characteristic function of S_n is proved, as well as new results of the value-distribution of the divisors of S_n . The Appendix further contains a complete and concise description of the value-distribution of the divisors of Bernoulli sums B_n obtained by the author in earlier works. Let $\mathcal{C} = \{j \geq 3 : \xi_j = 1\}$. Note that obviously $S_x = \#\{v \in \mathcal{C} : v \leq x\}$ for all reals $x \geq 3$. In particular, the LIL implies that

$$\#\{v \in \mathcal{C} : v \leq x\} = \int_2^x \frac{dt}{\log t} + \mathcal{O}(\sqrt{x \log \log x}), \quad (1.3)$$

with probability one.

Thus by (1.3),

$$\#\{v \in \mathcal{C} : v \leq x\} = \int_2^x \frac{dt}{\log t} + \mathcal{O}_\epsilon(x^{\frac{1}{2}+\epsilon}), \quad (1.4)$$

with probability one. The same result for the prime sequence $\mathcal{P}$ is equivalent to the Riemann Hypothesis (RH).



The LIL is a consequence of Kolmogorov's LIL, and yields the more precise result

$$\limsup_{x \rightarrow \infty} \frac{\#\{v \in \mathcal{C} : v \leq x\} - \int_2^x \frac{dt}{\log t}}{\sqrt{2 \left(\frac{x}{\log x}\right) \log \log x}} = 1, \quad (1.5)$$

with probability one.

We question the analogy made with (1.4) and prove that this model possesses finer tied properties, enlighting the above analogy somehow differently. We notably prove that if x runs along any increasing subsequence of integers $\mathcal{N}$,

$$\#\{v \in \mathcal{C} : v \leq x\} = \int_2^x \frac{dt}{\log t} + \mathcal{O}(\sqrt{x} \varphi_{\mathcal{N}}(x)), \quad (1.6)$$

with probability one. And we may have that $\varphi_{\mathcal{N}}(x) = o(\sqrt{\log \log x})$, in fact $\varphi_{\mathcal{N}}(x)$ can be as slow as desired, along a suitable subsequence $\mathcal{N}$.

Thus (1.6) implies that 'the prime number theorem' in Cramér's model is sensitive to the subsequence on which x is running, which seems not corroborated with any existing result concerning the counting function $\pi(x)$.

In the limiting case $\varphi_{\mathcal{N}}(x) \equiv \text{Const}$, we study the number of occurrences and the length of intervals I for which

$$\sup_{n \in I} \frac{|S_n - \mathbb{E} S_n|}{\sqrt{\text{Var}(S_n)}} \leq z, \quad (1.7)$$

z being some positive real.

Such a property, namely the maximal duration of small amplitudes of S_n around $\mathbb{E} S_n$, is quite sensitive to the value taken by z , and turns up to be tied with the spectrum of the Sturm-Liouville equation. We obtain in Theorems 2.2, 2.4 quite sharp results.

The proofs combine the IP with small local oscillation results of the Ornstein-Uhlenbeck process.

Write $\mathcal{C} = \{P_j, j \geq 1\}$, where P_j are the instants of jumps of the random walk $\{S_n, n \geq 1\}$, which are recursively defined as follows,

$$P_1 = \inf\{n \geq 3 : S_n = 1\}, \quad P_{v+1} = \inf\{n > P_v : S_n = 1\} \quad v \geq 1. \quad (1.8)$$

The main characteristic of Cramér's model is that heuristically $\mathcal{C}$ should imitate well the sequence $\mathcal{P}$. He proved that with probability one, one has

$$\limsup_{v \rightarrow \infty} \frac{P_{v+1} - P_v}{\log^2 P_v} = 1. \quad (1.9)$$

On the basis of this result, he wrote in [3] p.28, "Obviously we may take this as a suggestion that, for the particular sequence of ordinary prime numbers p_n , some similar relation may hold." He conjectured (Cramér's conjecture) that for some positive constant c ,

$$\limsup_{v \rightarrow \infty} \frac{p_{v+1} - p_v}{\log^2 p_v} = c. \quad (1.10)$$

The almost sure limit result (1.9) has no arithmetical content, as it is purely probabilistic. Further the sequence of differences $\{P_{v+1} - P_v, v \geq 1\}$ is a sequence of independent random variables, which is a very strong property. Impressive numerical evidences (up to 10^{18}) of (1.10) are given [21], see also [9], but depending on the scale of the observed phenomenon, 10^{18} might be a very little number (at least in the fast-growing hierarchy of numbers), and paraphrasing Odlyzko's note [22], that conjecture, if true, can be just barely true. On the other hand, if the conjecture were true, no real singularity should appear, in other words the observed phenomenon is being from the beginning locally 'similar'. If so, one may wonder what could be a reason. We refer the reader to [2], [3], [9], [23], [28] notably, for results conforing or contrary to the Cramér conjecture, which nowadays still appears as a mathematical 'spell'.

As already mentioned, Cramér's model does not assert that there are any primes in the sequence $\mathcal{C}$, and in variants of this model either. An important question, apparently overlooked in the related literature, thus



concerns the possible primality of S_n , namely the study of the probability $\mathbb{P}\{S_n \text{ prime}\}$, prior to the one of probability $\mathbb{P}\{P_v \text{ prime}\}$. The LIL (for instance) shows that such a property is tightly related to the distribution of primes in small intervals, making thereby vain the hope of obtaining definitive results, even on assuming RH. Let $m_n = \mathbb{E} S_n = \sum_{j=3}^n \frac{1}{\log j}$, $B_n = \text{Var} S_n = \sum_{j=3}^n \frac{1}{\log j} (1 - \frac{1}{\log j})$. Let $b > 1$. The intervals

$$I_n = [m_n - \sqrt{2bB_n \log \log n}, m_n + \sqrt{2bB_n \log \log n}],$$

are no longer overlapping as soon as n runs along very moderated growing subsequences. Along such a subsequence S_n can be prime, n large, only if I_n contains a prime number. These intervals are of type

$$[x - c x^{\frac{1}{2}} (\log \log x)^{\frac{1}{2}}, x],$$

for some $c > 0$. It is at present quite out of reach, even on assuming the validity of the RH, to decide for which x , such an interval contains a prime number. One can also makes the similar observation from the sharpened version of the local limit theorem given in Proposition 3.1.

Thus we are in a case where we have a model predicting largest size of gaps between primes, whereas, even on RH, we could not know whether the ‘primes’ of the model are prime. Recall some results on primes in small intervals. Assuming the RH, the best result known to us states as follows,

$$\pi(x) - \pi(x - y) = \int_{x-y}^x \frac{dt}{\log t} + \mathcal{O}\left(x^{\frac{1}{2}} \log \left\{ \frac{y}{x^{\frac{1}{2}} \log x} \right\}\right) \quad (1.11)$$

for y in the range $2x^{\frac{1}{2}} \log x \leq y \leq x$. Thus for $M \geq 2$ fixed,

$$\pi(x) - \pi(x - Mx^{\frac{1}{2}} \log x) \sim x^{\frac{1}{2}} \{M + \mathcal{O}(\log M)\}. \quad (1.12)$$

See Heath-Brown [11], see also the recent paper [10] and the references therein. Without assuming RH, Heath-Brown proved in [11] that if $\epsilon(x) > 0$, $\epsilon(x) \rightarrow 0$ as $x \rightarrow \infty$, then

$$\pi(x) - \pi(x - y) = \frac{y}{\log x} \left(1 + \mathcal{O}(\epsilon^4(x)) + \mathcal{O}\left(\frac{\log \log x}{\log x}\right)^4\right) \quad (1.13)$$

for y in the range $x^{\frac{7}{12} - \epsilon(x)} \leq y \leq \frac{x}{(\log x)^4}$. This slightly improves Huxley’s earlier result in [13] corresponding to $\epsilon(x) = 0$. Huxley’s result shows that the PNT extends to intervals of the type $[x, x + x^\theta]$, $x^{\frac{7}{12}} \leq \theta \leq \frac{x}{(\log x)^4}$, namely that,

$$\#\{[x, x + x^\theta] \cap \mathcal{P}\} \sim \frac{x^\theta}{\log x}. \quad (1.14)$$

We however obtain in Theorem 2.5, without assuming RH, a sharp estimate of $\mathbb{P}\{S_n \text{ prime}\}$, for almost all n , namely for all n , $n \rightarrow \infty$ through a set $\mathcal{S}$ of natural density 1. The lower bound,

$$\liminf_{\mathcal{S} \ni n \rightarrow \infty} (\log n) \mathbb{P}\{S_n \text{ prime}\} > 0, \quad (1.15)$$

is also proved.

Further the property for S_n to be z -quasiprime is investigated. We obtain in Theorem 2.6 a sharp estimate of the probability that S_k be ζ -quasiprime, k large and for the range of values $\zeta_0 \leq \zeta \leq \exp\{c \log(k)/\log \log(k)\}$, $c > 0$.



2 Main results

It is well-known that the LIL (at least for centered square integrable i.i.d. sums) has slower amplitude than the one given by the classical normalizing factor $\sqrt{2n \log \log n}$, when n is restricted to subsequences. For instance, if n runs along the subsequence $\mathcal{N} = \{2^k, k \geq 1\}$, then the LIL restricted to $\mathcal{N}$ holds with normalizing factor $\sqrt{2n \log \log \log n}$. See [31] for a characterization of the LIL for subsequences. The same phenomenon holds in fact - with no additional requirement - for the Cramér model.

Theorem 2.1 *Let $\mathcal{N}$ be any increasing sequence of integers. Then,*

$$\limsup_{\mathcal{N} \ni j \rightarrow \infty} \frac{|S_j - m_j|}{\sqrt{B_j} \varphi_{\mathcal{N}}(j)} = 1,$$

almost surely, where function $\varphi_{\mathcal{N}}(n)$ is defined in (7.1).

Roughly speaking, given $M > 1$, $I_k =]M^k, M^{k+1}]$, $\varphi_{\mathcal{N}}(n)$ is defined as being equal to $\sqrt{2 \log(p+2)}$ if $n \in \mathcal{N} \cap I_{\kappa_p}$, I_{κ_p} being the p -th interval intersecting $\mathcal{N}$.

In the next Theorems we obtain very sharp results on the length, and also the frequencies of the intervals I for which (1.7) holds, namely

$$\sup_{n \in I} \frac{|S_n - m_n|}{\sqrt{B_n}} \leq z,$$

z a positive real (corresponding to $\varphi_{\mathcal{N}} \equiv \text{Const}$).

Theorem 2.2 *Let $f : [1, \infty) \rightarrow \mathbb{R}^+$ be a non-decreasing function such that $f(t) \uparrow \infty$ with t and $f(t) = o_{\rho}(t^{\rho})$. There exists a Brownian motion W such that*

$$\liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f(e^k)} \frac{|S_j - m_j|}{\sqrt{B_j}} = \liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f(e^k)} \frac{|W(B_j)|}{\sqrt{B_j}},$$

with probability 1. Let $f_c(t) = \log^c t$, $c > 0$. Then

$$\liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f_c(e^k)} \frac{|S_j - m_j|}{\sqrt{B_j}} \leq z,$$

with probability 1, if and only if $c \leq 1/\lambda(z)$. And $\lambda(z)$ is the smallest eigenvalue in the Sturm-Liouville equation

$$\psi''(x) - x\psi'(x) = -\lambda\psi(x), \quad \psi(-z) = \psi(z) = 0. \quad (2.1)$$

This is a positive strictly decreasing continuous function of z on $]0, \infty[$. Further,

$$\lambda(z) \sim \frac{\pi^2}{4z^2}, \quad \text{as } z \rightarrow 0. \quad (2.2)$$

Towards this aim, we prove that Cramér's model satisfies an invariance principle:

Proposition 2.3 (IP) *Let $1/\alpha < \beta < 1/2$. There exists a Brownian motion W such that if*

$$\Upsilon = \sup_n \frac{1}{B_n^{\beta}} \sup_{j \leq n} |S_j - m_j - W(B_j)|$$

then, $\mathbb{E} \Upsilon^{\alpha'} < \infty$, $0 \leq \alpha' < \alpha$.



Thanks to the IP above, the question studied in (1.7) can be transferred into a similar one concerning Brownian motion. This is done in section 6, where Theorem 2.2 is proved.

We also obtain a sharp estimate on the number of occurrences of the sets

$$B_k(f, z) = \left\{ \sup_{j \in J_{e^k}} \frac{|S_j - m_j|}{\sqrt{B_j}} \leq z \right\}, \quad k = 1, 2, \dots \quad (2.3)$$

where $J_N = \{j : N \leq B_j < Nf(N)\}$. Let

$$A_k(f, z) = \left\{ \sup_{e^k \leq t \leq e^k f(e^k)} \frac{|W(t)|}{\sqrt{t}} \leq z \right\},$$

and let also $v_n(f, z) = \sum_{k=1}^n \mathbb{P}\{A_k(f, z)\}$.

Theorem 2.4 *Let $0 < z' < z < z''$. Let also $0 < c \leq 1/\lambda(z')$. Then for $a > 3/2$,*

$$\begin{aligned} \mathbb{P}\left\{\sum_{k=1}^n \chi_{B_k(f_c, z)} \leq v_n(f_c, z'') + \mathcal{O}_a\left(v_n^{1/2}(f_c, z'') \log^a v_n(f_c, z'')\right), \quad n \text{ ultimately}\right\} &= 1, \\ \mathbb{P}\left\{v_n(f_c, z') \leq \sum_{k=1}^n \chi_{B_k(f_c, z)} + \mathcal{O}_a\left(v_n^{1/2}(f_c, z') \log^a v_n(f_c, z')\right), \quad n \text{ ultimately}\right\} &= 1. \end{aligned}$$

Further for all n ,

$$K_1(z) \sum_{k=1}^n k^{-c\lambda(z)} \leq v_n(f_c, z) \leq K_2(z) \sum_{k=1}^n k^{-c\lambda(z)}.$$

Estimates of the sums $\sum_{k=1}^n k^{-c\lambda(z)}$ are given in (6.6). The question arises whether the refinements obtained (Theorems 2.1, 2.2, 2.4) may also have an interpretation on the function $\pi(x)$.

The Cramér model is used to ‘predict’ several, sometimes quite elaborated results on the distribution of primes. The example given in (1.6) is

very striking, as the subsequence-LIL is a well-known companion result of the standard LIL, and cannot be dissociated from it. Thus if one uses the Cramér model to make such a prediction concerning the PNT (see after (1.3) and (1.5)) from the standard LIL, probably its most simple prediction, one should also consider the prediction which arises with (1.6), and argue whether this is another deficiency of the model or not.

The same sort of considerations is in order concerning the frequency of large gaps between ‘primes’.

See (A.15) and after in Appendix 7.

Concerning the probability that S_n be prime or quasi-prime, and the primality of P_n , we prove the following results.

2.1 Primality of S_n

Theorem 2.5 (i) *For any constant $b > 1/2$,*

$$\mathbb{P}\{S_n \text{ prime}\} = \frac{1}{\sqrt{2\pi B_n}} \int_{m_n - \sqrt{2bB_n \log n}}^{m_n + \sqrt{2bB_n \log n}} e^{-\frac{(t-m_n)^2}{2B_n}} d\pi(t) + \mathcal{O}\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right), \quad (2.4)$$

as $n \rightarrow \infty$.

(ii) *There exists a set of integers $\mathcal{S}$ of density 1, such that*

$$\mathbb{P}\{S_n \text{ prime}\} = \frac{(1 + o(1))}{\sqrt{2\pi B_n}} \int_{m_n - \sqrt{2bB_n \log n}}^{m_n + \sqrt{2bB_n \log n}} e^{-\frac{(t-m_n)^2}{2B_n}} d\pi(t), \quad (2.5)$$

as $n \rightarrow \infty$, $n \in \mathcal{S}$. Further,

$$\liminf_{\mathcal{S} \ni n \rightarrow \infty} (\log n) \mathbb{P}\{S_n \text{ prime}\} \geq \frac{1}{\sqrt{2\pi e}}. \quad (2.6)$$

The proof uses a result of Selberg [28].



2.2 Quasi-primality of S_n

Let $\Pi_z = \prod_{p \leq z} p$. According to Pintz [23], an integer m is z -quasiprime, if $(m, \Pi_z) = 1$. Let $S'_n = \sum_{j=8}^n \xi_j$, $n \geq 8$. Note that the introduction of S'_n in place of S_n is not affecting Cramér's conjecture, see Remark (A.7). In the next Theorem we study for all n large enough, the probability that S'_n be z -quasiprime.

Theorem 2.6 *We have for any $0 < \eta < 1$, and all n large enough and $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log n}{\log \log n} \right\}$,*

$$\mathbb{P}\{S'_n \zeta - \text{quasiprime}\} \geq (1 - \eta) \frac{e^{-\gamma}}{\log \zeta}.$$

where γ is Euler's constant and c is a positive constant.

The approaches used to prove the above Theorems not apply to the study of the primality of P_v .

2.3 Primality of P_n

We show that when the 'primes' P_v are observed along moderately growing subsequences, then with probability 1, they ultimately avoid any given infinite set of primes satisfying a reasonable tail's condition. We also test which infinite sequences of primes are ultimately avoided by the 'primes' P_v , with probability 1. More precisely we answer the following question:

Question 2.7 *Given an increasing sequence of naturals $\mathcal{K}$ and increasing sequence of primes $\mathfrak{P}$, under which conditions is $\mathfrak{P}$ avoided by all P_v , v large enough, $v \in \mathcal{K}$, with probability 1?*

Theorem 2.8 *Let $\mathcal{K}$ be an increasing sequence of naturals such that the series $\sum_{k \in \mathcal{K}} k^{-\beta}$ converges for some $\beta \in]0, \frac{1}{2}[$. Let $\mathfrak{P}$ be an increasing sequence of primes such that for some $b > 1$,*

$$\sup_{k \in \mathcal{K}} \frac{\#\{\mathfrak{P} \cap [k, bk]\}}{k^{\frac{1}{2}-\beta}} < \infty.$$

Let also $\{\Delta_k, k \geq 1\}$ be the instants of jump of the Bernoulli sequence $\{B_k, k \geq 1\}$ defined in (5.2). Then

$$\mathbb{P}\{\Delta_k \notin \mathfrak{P}, \quad k \in \mathcal{K} \text{ ultimately}\} = 1.$$

Further,

$$\mathbb{P}\{P_v \notin \mathfrak{P}, \quad v \in \mathcal{K} \text{ ultimately}\} = 1.$$

Moreover (case $\beta = 1/2$), let $\mathfrak{P}$ be such that $\sum_{p \in \mathfrak{P}, p > y} p^{-1/2} = \mathcal{O}(y^{-1/2})$, and $\mathcal{K}$ be such that $\sum_{k \in \mathcal{K}} k^{-1/2} < \infty$. Then

$$\mathbb{P}\{P_v \notin \mathfrak{P}, \quad v \in \mathcal{K} \text{ ultimately}\} = 1.$$

The paper is organized as follows. The study of the quasi-primality of S_k is made in section 4, the one of the primality of S_n occupies the whole section 3, and the one of the primality of P_n is made in section 5. These sections are a forming the main body of the paper. Theorem 2.1 is proved in section 7. In section 6, we prove the IP, as well as Theorem 2.2, after some preliminary background, and Theorem 2.4. The standard limit theorems for the Cramér model, statements and proofs, and various results (sharp estimate of the characteristic function of S_n , divisors of Bernoulli sums) used in the course of the proofs are moved to the Appendix A. Some remarks concerning Cramér's proof are also included.



3 Primality of S_n : proof of Theorem 2.5

We need a sharper form of the local limit theorem for S_n than the one given in Lemma A.1.

Proposition 3.1 *We have the following estimate*

$$\left| \mathbb{P}\{S_n = \kappa\} - \frac{e^{-\frac{(\kappa - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} \right| \leq C \frac{(\log n)^{3/2}}{n},$$

for all $\kappa \in \mathbb{Z}$ such that

$$|\kappa - m_n| \leq C \frac{n^{3/4}}{\log n}.$$

The remainder term is of order $\mathcal{O}(\frac{(\log n)^{3/2}}{n})$, which is much better than $o((\frac{\log n}{n})^{1/2})$ in Lemma A.1. This is a consequence of Corollary 1.11 in [8]. For the reader convenience we recall it. Introduce first the necessary notation. Let v_0 and $D > 0$ be real numbers. We denote by $\mathcal{L}(v_0, D)$ the lattice defined by the sequence $v_k = v_0 + Dk$, $k \in \mathbb{Z}$. We associate to any random variable X taking values in $\mathcal{L}(v_0, D)$ with probability one, the following characteristic,

$$\vartheta_X = \sum_{k \in \mathbb{Z}} \mathbb{P}\{X = v_k\} \wedge \mathbb{P}\{X = v_{k+1}\}, \quad (3.1)$$

where $a \wedge b = \min(a, b)$. Note that $\vartheta_X < 1$.

Lemma 3.2 ([8], Cor. 1.11) *Let $X_1, \dots, X_n$ be independent random variables taking almost surely values in a common lattice $\mathcal{L}(v_0, D) = \{v_k, k \in \mathbb{Z}\}$, where $v_k = v_0 + Dk$, $k \in \mathbb{Z}$, v_0 and $D > 0$ are real numbers. We assume that*

$$\vartheta_{X_j} > 0, \quad j = 1, \dots, n. \quad (3.2)$$

Let $S_n = X_1 + \dots + X_n$. Let $\psi : \mathbb{R} \rightarrow \mathbb{R}^+$ be even, convex and such that $\frac{\psi(x)}{x^2}$ and $\frac{x^3}{\psi(x)}$ are non-decreasing on $\mathbb{R}^+$. We assume that

$$\mathbb{E} \psi(X_j) < \infty, \quad j = 1, \dots, n. \quad (3.3)$$

Put

$$L_n = \frac{\sum_{j=1}^n \mathbb{E} \psi(X_j)}{\psi(\sqrt{\text{Var}(S_n)})}.$$

Let $0 < \vartheta_j \leq \vartheta_{X_j}$ and denote $\Theta_n = \sum_{j=1}^n \vartheta_j$. Further assume that $\frac{\log \Theta_n}{\Theta_n} \leq 1/14$. Then, for all $\kappa \in \mathcal{L}(v_0 n, D)$ such that

$$\frac{(\kappa - \mathbb{E} S_n)^2}{\text{Var}(S_n)} \leq \sqrt{\frac{\Theta_n}{14 \log \Theta_n}},$$

we have

$$\left| \mathbb{P}\{S_n = \kappa\} - \frac{De^{-\frac{(\kappa - \mathbb{E} S_n)^2}{2\text{Var}(S_n)}}}{\sqrt{2\pi \text{Var}(S_n)}} \right| \leq C_3 \left\{ D \left(\frac{\log \Theta_n}{\text{Var}(S_n) \Theta_n} \right)^{1/2} + \frac{L_n + \Theta_n^{-1}}{\sqrt{\Theta_n}} \right\}.$$

And C is an absolute constant.



Proof of Proposition 3.1 In our case $D = 1$. Further for $j = 3, \dots, n$, $\mathbb{P}\{\xi_j = k\} \wedge \mathbb{P}\{\xi_j = k + 1\} = \frac{1}{\log j + 2}$, if $k = 0$, and equals 0 for $k \in \mathbb{Z}_*$. Thus $\theta_{\xi_j} = \frac{1}{\log j}$. We choose $\theta_j = \theta_{X_j}$, $\psi(x) = |x|^3$. Then $\Theta_n = \mathbb{E} S_n \sim \frac{n}{\log n}$, $\text{Var}(S_n) = B_n \sim \frac{n}{\log n}$, $L_n \sim (\frac{\log n}{n})^{1/2}$.

Thus $(m_n = \sum_{k=1}^n \frac{1}{\log k}, B_n = \sum_{k=1}^n (1 - \frac{1}{\log k})(\frac{1}{\log k}))$

$$\left| \mathbb{P}\{S_n = \kappa\} - \frac{e^{-\frac{(\kappa - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} \right| \leq C \frac{(\log n)^{3/2}}{n},$$

for all $\kappa \in \mathbb{Z}$ such that $|\kappa - m_n| \leq C \frac{n^{3/4}}{\log n}$. $\square$

Proof of Theorem 2.5 (i) By Lemma 7.1 p. 240 in [23], for $0 \leq x \leq B_n$

$$\begin{aligned} \mathbb{P}\{|S_n - m_n| \geq x\} &= \mathbb{P}\{S_n - m_n \geq x\} + \mathbb{P}\{-(S_n - m_n) \geq x\} \\ &\leq 2 \exp \left\{ -\frac{x^2}{2B_n} \left(1 - \frac{x}{2B_n}\right) \right\}, \end{aligned}$$

noticing that $\{-\xi_j\}_j$ also satisfies the conditions of Kolmogorov's Theorem. Let $b > b' > 1/2$. Then for all sufficiently large n , since $\log B_n \sim \log n$,

$$\mathbb{P}\{|S_n - m_n| \geq \sqrt{2bB_n \log n}\} \leq 2n^{-b'}. \quad (3.4)$$

We have

$$\begin{aligned} &\left| \mathbb{P}\{S_n \in \mathcal{P}\} - \mathbb{P}\{S_n \in \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}]\} \right| \\ &\leq \mathbb{P}\{|S_n - m_n| \geq \sqrt{2bB_n \log n}\} \\ &\leq n^{-b'}. \end{aligned}$$

Further,

$$\begin{aligned} &\left| \mathbb{P}\{S_n \in \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}]\} \right. \\ &\quad \left. - \sum_{\kappa \in \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}]} \frac{e^{-\frac{(\kappa - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} \right| \\ &\leq \sum_{\kappa \in \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}]} \left| \mathbb{P}\{S_n = \kappa\} - \frac{e^{-\frac{(\kappa - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} \right| \\ &\leq C \# \{ \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}] \} \cdot \frac{(\log n)^{3/2}}{n} \\ &\leq C \sqrt{b} \frac{(\log n)^{3/2}}{\sqrt{n}}. \end{aligned}$$

Therefore

$$\left| \mathbb{P}\{S_n \in \mathcal{P}\} - \sum_{\kappa \in \mathcal{P} \cap [m_n - \sqrt{2bB_n \log n}, m_n + \sqrt{2bB_n \log n}]} \frac{e^{-\frac{(\kappa - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} \right| \leq C \sqrt{b} \frac{(\log n)^{3/2}}{\sqrt{n}}. \quad (3.5)$$

By expressing the inner sum as a Riemann-Stieltjes integral [1, p. 77], we get

$$\mathbb{P}\{S_n \in \mathcal{P}\} = \int_{m_n - \sqrt{2bB_n \log n}}^{m_n + \sqrt{2bB_n \log n}} \frac{e^{-\frac{(t - m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} d\pi(t) + \mathcal{O}\left(\frac{(\log n)^{3/2}}{\sqrt{n}}\right). \quad (3.6)$$



(ii) We note that

$$\int_{m_n - \sqrt{2bB_n}}^{m_n + \sqrt{2bB_n}} \frac{e^{-\frac{(t-m_n)^2}{2B_n}}}{\sqrt{2\pi B_n}} d\pi(t) \geq L \frac{\pi(m_n + \sqrt{2bB_n}) - \pi(m_n - \sqrt{2bB_n})}{\sqrt{B_n}}, \quad (3.7)$$

with $L = \frac{e^{-b}}{\sqrt{2\pi}}$.

We use a well-known result of Selberg [28, Th. 1]. Let $\Phi(x)$ be positive and increasing and such that $\frac{\Phi(x)}{x}$ decreasing for $x > 0$. Further assume that

$$(a) \quad \lim_{x \rightarrow \infty} \frac{\Phi(x)}{x} = 0 \quad (b) \quad \liminf_{x \rightarrow \infty} \frac{\log \Phi(x)}{\log x} > \frac{19}{77}. \quad (3.8)$$

Then there exists a (Borel measurable) set $\mathcal{S}$ of positive reals of density one such that

$$\lim_{\mathcal{S} \ni x \rightarrow \infty} \frac{\pi(x + \Phi(x)) - \pi(x)}{(\Phi(x)/\log x)} = 1. \quad (3.9)$$

Let $\Phi(x) = \sqrt{2bx}$. Then the requirements in (3.8) are fulfilled, and so (3.9) holds true. Now let C be some possibly large but fixed positive number, as well as some positive real $\delta < 1/2$.

By (3.9), the set of $x > 0$, call it $\mathcal{S}_\delta$, such that

$$\pi(x + \Phi(x)) - \pi(x) \geq (1 - \delta) \frac{\Phi(x)}{\log x}. \quad (3.10)$$

has density 1. Note that if $\delta' < \delta''$ then $\mathcal{S}_{\delta'} \subseteq \mathcal{S}_{\delta''}$.

Pick $x \in \mathcal{S}_\delta$ and let $\Delta(x) = \pi(x + \Phi(x)) - \pi(x)$. Note that if $|y - x| \leq C$, $|\Phi(y) - \Phi(x)| = o(1)$ for x large. Thus for every $y \in [x - C, x + C]$, $|\Delta(y) - \Delta(x)| \leq C'$, and so

$$\Delta(y) \geq (1 - \delta) \frac{\Phi(x)}{\log x} - C'.$$

the constant C' depending on C only. As $|\frac{\Phi(x)}{\log x} - \frac{\Phi(y)}{\log y}| \leq \frac{C}{2\sqrt{x} \log x}$, we have

$$\Delta(y) \geq (1 - \delta) \frac{\Phi(y)}{\log y} - C' - \frac{C}{2\sqrt{x} \log x}.$$

Thus every $y \in [x - C, x + C]$ also satisfies

$$\Delta(y) \geq (1 - 2\delta) \frac{\Phi(y)}{\log y}, \quad (3.11)$$

if x is large enough.

Let $v = v(x)$ be the unique integer such that $m_{v-1} < x \leq m_v$. As $m_v - m_{v-1} = o(1)$, $v \rightarrow \infty$, it follows that $m_v \in [x - C, x + C]$ provided that x is large enough, in which case we have by (3.11),

$$\frac{\pi(m_v + \Phi(m_v)) - \pi(m_v)}{\Phi(m_v)} \geq \frac{1 - 2\delta}{\log m_v}. \quad (3.12)$$

Let $X \geq 1$ be a large positive integer and ϵ a small positive real. The number $N(X)$ of intervals $[\mu - 1, \mu]$, $\mu \leq X$ such that $\mathcal{S}_\delta \cap [\mu - 1, \mu] \neq \emptyset$ verifies $N(X)/X \sim 1$, $X \rightarrow \infty$, since $\mathcal{S}_\delta$ has density 1.

Given such an $\mu \leq X$, pick $x \in \mathcal{S}_\delta \cap [\mu - 1, \mu]$. We know (recalling that $m_v - m_{v-1} = o(1)$, $v \rightarrow \infty$) that some m_v , $v = v(x)$ belongs to $[\mu - 1 - \epsilon, \mu + \epsilon]$, and that (3.12) is satisfied. The union of these intervals $[\mu - 1 - \epsilon, \mu + \epsilon]$ is contained in $[1 - \epsilon, X + \epsilon]$. It follows that the number of v such that (3.12) is satisfied, forms a set of density 1.

We now use an induction argument in order to replace 2δ in (3.12) by a quantity $\epsilon(v)$ which tends to 0 as v tends to infinity along some other set of density 1, which we shall build explicitly. Let $\mathcal{T}_n$ be the set of v 's of density 1, corresponding to $\delta = \frac{1}{n}$, $n \geq 3$. Let X_3 be large enough so that $\#\{\mathcal{T}_3 \cap [1, X]\} \geq X(1 - 1/3)$ for all



$X \geq X_3$. Next let $X_4 > X_3$ be sufficiently large so that $\#\{\mathcal{T}_4 \cap [X_3, X]\} \geq X(1 - 1/4)$ for all $X \geq X_4$. Like this we manufacture an increasing sequence X_j , verifying for all $j \geq 3$,

$$\#\{\mathcal{T}_j \cap [X_{j-1}, X]\} \geq X(1 - 1/j), \quad \text{for all } X \geq X_j.$$

The resulting set

$$\mathcal{T} = \bigcup_{j=3}^{\infty} \mathcal{T}_j \cap [X_{j-1}, X_j]$$

has density 1 and further we have the inclusions

$$\mathcal{T} \cap [X_{l-1}, \infty) = \bigcup_{j=l}^{\infty} \mathcal{T}_j \cap [X_{j-1}, X_j] \subset \mathcal{T}_l \cap \bigcup_{j=l}^{\infty} [X_{j-1}, \infty) = \mathcal{T}_l \cap [X_{l-1}, \infty), \quad l \geq 4,$$

as the sets $\mathcal{T}_j$ are decreasing with j by definition.

We finally have by (3.7),

$$\frac{\pi(m_v + \Phi(m_v)) - \pi(m_v)}{\Phi(m_v)} \geq \frac{1 - \epsilon(v)}{\log m_v}, \quad (3.13)$$

along $\mathcal{T}$, for some sequence of reals $\epsilon(v) \downarrow 0$ as $v \rightarrow \infty$.

Therefore

$$\begin{aligned} \int_{m_v - \sqrt{2bB_v}}^{m_v + \sqrt{2bB_v}} \frac{e^{-\frac{(t-m_v)^2}{2B_v}}}{\sqrt{2\pi B_v}} d\pi(t) &\geq L \frac{\pi(m_v + \sqrt{2bB_v}) - \pi(m_v - \sqrt{2bB_v})}{\sqrt{B_v}} \\ &\geq L \frac{1 - \epsilon(v)}{\log m_v}, \end{aligned} \quad (3.14)$$

for all $v \in \mathcal{T}$, recalling that $L = \frac{e^{-b}}{\sqrt{2\pi}}$.

Also

$$\liminf_{\mathcal{T} \ni v \rightarrow \infty} (\log v) \mathbb{P}\{S_v \text{ prime}\} \geq \frac{1}{\sqrt{2\pi e}}. \quad (3.15)$$

It further follows from (3.6) that

$$\mathbb{P}\{S_v \text{ prime}\} = (1 + o(1)) \int_{m_v - \sqrt{2bB_v \log v}}^{m_v + \sqrt{2bB_v \log v}} \frac{e^{-\frac{(t-m_v)^2}{2B_v}}}{\sqrt{2\pi B_v}} d\pi(t), \quad (3.16)$$

all $v \in \mathcal{T}$. This achieves the proof of Theorem 2.5. $\square$

Some remarks: estimate (3.9) extends the PNT to smalls intervals $[x, x + \Phi(x)]$ for almost all x . Selberg (developing Cramér's first results [2]) proved with Theorem 4 in [28] a much stronger result since an error term is provided. Assuming the RH, he proved that for any fixed $\vartheta > 0$, (1.14) is true for almost all x . This is an easy consequence of the very sharp Theorem 1 in [28]. The approach used, as well as the alternate approach in Richards [25], seem not allow one to treat the question whether there exists a version of the PNT with an error term valid for almost all integers. This question is in relation with the one on the sensitivity of the error term to subsequences (cf. Introduction, Theorems 2.1, 2.2).



4 Quasi-primality of S_n : proof of Theorem 2.6

Theorem 2.6 is a direct consequence of a more general result, which we shall prove now. Let $2 < \lambda_1 < \lambda_2 < \dots$ be an increasing sequence of reals. Let $\{\zeta_j, j \geq 1\}$ be a sequence of independent binomial random variables defined by $\mathbb{P}\{\zeta_j = 1\} = \frac{1}{\lambda_j} = 1 - \mathbb{P}\{\zeta_j = 0\}$.

Theorem 4.1 Let $T_k = \sum_{j=1}^k \zeta_j$, $k \geq 1$. Assume that $\mu_k = \sum_{j=1}^k \lambda_j^{-1} \uparrow \infty$ with k . For any $0 < \delta < 1$, we have for any $0 < \eta < 1$, and all k large enough and $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{\log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\}$,

$$\mathbb{P}\{T_k \text{ is } \zeta\text{-quasiprime}\} \geq (1 - \eta) \frac{e^{-\gamma}}{\log \zeta},$$

where γ is Euler's constant and c is a positive constant.

The case $\lambda_j = \log(j + 2)$, $j \geq 8$ corresponds to the Cramér model, and we have in particular the following

Corollary 4.2 We have for any $0 < \eta < 1$, and all n large enough and $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log n}{\log \log n} \right\}$,

$$\mathbb{P}\{S'_n \zeta - \text{quasiprime}\} \geq (1 - \eta) \frac{e^{-\gamma}}{\log \zeta}.$$

The proof is based on a randomization argument, and uses the Lemma below.

Lemma 4.3 [19, Theorem 2.3] Let $X_1, \dots, X_k$ be independent random variables, with $0 \leq X_j \leq 1$ for each j . Let $Y_k = \sum_{j=1}^k X_j$ and $\mu = \mathbb{E} Y_k$. For any $\epsilon > 0$,

$$\begin{aligned} (a) \quad & \mathbb{P}\{Y_k \geq (1 + \epsilon)\mu\} \leq e^{-\frac{\epsilon^2 \mu}{2(1+\epsilon/3)}}. \\ (b) \quad & \mathbb{P}\{Y_k \leq (1 - \epsilon)\mu\} \leq e^{-\frac{\epsilon^2 \mu}{2}}. \end{aligned}$$

Proof of Theorem 4.1 Let $\{\epsilon_j, j \geq 1\}$ be a sequence of independent Bernoulli random variables. Let $\{\check{\zeta}_j, j \geq 1\}$ be another sequence of independent random variables, which is independent from the sequence $\{\epsilon_j, j \geq 1\}$, and such that $\zeta_j \stackrel{\text{a.s.}}{=} \check{\zeta}_j \epsilon_j$ for all j . Let $\check{\mathbb{E}}$ (resp. $\check{\mathbb{P}}$) denote the conditional expectation (resp. conditional probability) with respect to the σ -field generated by the sequence $\check{\zeta}_j, j \geq 1$. Write $T_k = \sum_{j=1}^k \check{\zeta}_j \epsilon_j$. We have

$$\mathbb{P}\{P^-(T_k) > \zeta\} = \check{\mathbb{E}} \mathbb{P}\{P^-(\sum_{j=1}^k \check{\zeta}_j \epsilon_j) > \zeta\} = \check{\mathbb{E}} \mathbb{P}\{P^-(B_{\sum_{j=1}^k \check{\zeta}_j}) > \zeta\}. \quad (4.1)$$

According to Theorem A.9, there exist a positive real c and positive constants C_0, ζ_0 such that for k large enough we have,

$$\left| \mathbb{P}\{P^-(B_k) > \zeta\} - \frac{e^{-\gamma}}{\log \zeta} \right| \leq \frac{C_0}{\log^2 \zeta} \quad (\zeta_0 \leq \zeta \leq k^{c/\log \log k}). \quad (4.2)$$

Let $0 < \delta < 1$ and set

$$A_k = \left\{ \exp \left\{ \frac{c \log(\sum_{j=1}^k \check{\zeta}_j)}{\log \log(\sum_{j=1}^k \check{\zeta}_j)} \right\} \geq \zeta \right\}, \quad C_k = \left\{ \sum_{j=1}^k \check{\zeta}_j > 2\delta\mu_k \right\}.$$

Assume that

$$\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\}.$$

On C_k ,

$$\exp \left\{ \frac{c \log(\sum_{j=1}^k \check{\zeta}_j)}{\log \log(\sum_{j=1}^k \check{\zeta}_j)} \right\} > \exp \left\{ \frac{c \log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\} \geq \zeta,$$



and so $C_k \subset A_k$. Therefore on C_k ,

$$\left| \mathbb{P}\{P^-(B_{\sum_{j=1}^k \check{\zeta}_j}) > \zeta\} - \frac{e^{-\gamma}}{\log \zeta} \right| \leq \frac{C_0}{\log^2 \zeta}.$$

We have

$$\begin{aligned} \mathbb{P}\{P^-(T_k) > \zeta\} &= \mathbb{E} \mathbb{P}\{P^-(B_{\sum_{j=1}^k \check{\zeta}_j}) > \zeta\} \geq \mathbb{E} \chi\{C_k\} \mathbb{P}\{P^-(B_{\sum_{j=1}^k \check{\zeta}_j}) > \zeta\} \\ &\geq \left(\frac{e^{-\gamma}}{\log \zeta} - \frac{C_0}{\log^2 \zeta} \right) \mathbb{P}\{C_k\}. \end{aligned}$$

Consequently, for $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\}$,

$$\mathbb{P}\{P^-(T_k) > \zeta\} \geq \left(\frac{e^{-\gamma}}{\log \zeta} - \frac{C_0}{\log^2 \zeta} \right) \mathbb{P}\{C_k\}.$$

We apply Lemma 4.3 with $X_j = \check{\zeta}_j$. The random variables $\check{\zeta}_j$ are independent and verify $\mathbb{P}\{\check{\zeta}_j = 1\} = 1 - \mathbb{P}\{\check{\zeta}_j = 0\} = 2\lambda_j^{-1}$. Further let $\check{\mu}_k = \mathbb{E} \sum_{j=1}^k \check{\zeta}_j = 2 \sum_{j=1}^k \lambda_j^{-1}$. By assumption $\sum_{j=1}^k \lambda_j^{-1} \uparrow \infty$ with k .

Let $0 < \rho < 1$. By Lemma 4.3,

$$\mathbb{P}\left\{ \sum_{j=1}^k \check{\zeta}_j \leq \delta \check{\mu}_k \right\} \leq e^{-\frac{(1-\delta)^2 \check{\mu}_k}{2}} \leq \rho, \quad (4.3)$$

for all $k \geq k_\rho$ say. Thus

$$\mathbb{P}\{C_k\} = \mathbb{P}\left\{ \sum_{j=1}^k \check{\zeta}_j > \delta \check{\mu}_k \right\} \geq 1 - \rho. \quad (4.4)$$

We therefore arrive at

$$\mathbb{P}\{T_k \text{ is } \zeta - \text{quasiprime}\} \geq (1 - \rho) \left(\frac{e^{-\gamma}}{\log \zeta} - \frac{C_0}{\log^2 \zeta} \right)$$

for any $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\}$. As ρ can be as small as we wish, we can state that for any given $0 < \delta < 1$, we have for any $0 < \eta < 1$, and all k large enough and any $\zeta_0 \leq \zeta \leq \exp \left\{ \frac{c \log(2\delta\mu_k)}{\log \log(2\delta\mu_k)} \right\}$,

$$\mathbb{P}\{T_k \text{ is } \zeta - \text{quasiprime}\} \geq (1 - \eta) \frac{e^{-\gamma}}{\log \zeta}.$$

□

Theorem 4.4 Let $(n_k)_{k \geq 1}$ be an increasing sequence of integers such that,

$$\sum_{k \geq 1} \frac{\log \log n_k}{\log n_k} < \infty.$$

Then,

$$\mathbb{P}\{T_{n_k} \text{ not prime, kultimately}\} = 1.$$

In particular,

$$\mathbb{P}\{S'_{n_k} \text{ not prime, kultimately}\} = 1.$$



Proof We also have

$$\mathbb{P}\{T_k \text{ prime}\} = \check{\mathbb{E}} \mathbb{P}\left\{\sum_{j=1}^k \check{\zeta}_j \epsilon_j \text{ prime}\right\} = \check{\mathbb{E}} \mathbb{P}\{B_{\sum_{j=1}^k \check{\zeta}_j} \text{ prime}\} \quad (4.5)$$

By Corollary A.10 in Appendix 7, there exists an absolute constant C_1 such that for all n large enough,

$$\mathbb{P}\{B_n \text{ prime}\} \leq C_1 \frac{\log \log n}{c \log n},$$

(c is the same constant as in Theorem A.9). This along with (4.3), imply

$$\begin{aligned} \mathbb{P}\{T_k \text{ prime}\} &= \check{\mathbb{E}} \mathbb{P}\{B_{\sum_{j=1}^k \check{\zeta}_j} \text{ prime}\} \\ &\leq \check{\mathbb{P}}\left\{\sum_{j=1}^k \check{\zeta}_j \leq \delta \check{\mu}_k\right\} + C_1 \check{\mathbb{E}} \chi\left\{\sum_{j=1}^k \check{\zeta}_j > \delta \check{\mu}_k\right\} \frac{\log \log \sum_{j=1}^k \check{\zeta}_j}{c \log \sum_{j=1}^k \check{\zeta}_j} \\ &\leq e^{-\frac{(1-\delta)^2 \check{\mu}_k}{2}} + C_1 \frac{\log \log \delta \check{\mu}_k}{c \log \delta \check{\mu}_k} \\ &\leq C(c, \delta) \frac{\log \log \delta \check{\mu}_k}{c \log \delta \check{\mu}_k}, \end{aligned} \quad (4.6)$$

for all $k \geq \kappa(c, \delta)$. Theorem 4.4 now follows from Borel-Cantelli lemma. $\square$

5 Primality of P_n

5.1 The inclusion $\mathcal{C} \subset \mathcal{B}$

We use the fact (Introduction) that on a possibly larger probability space, $\xi_j = \check{\xi}_j \epsilon_j$, $j \geq 8$, almost surely, where $\{\epsilon_j, j \geq 8\}$ is a sequence of independent Bernoulli random variables and $\{\check{\xi}_j, j \geq 8\}$, a sequence of independent binomial random variables which is independent from the sequence $\{\epsilon_j, j \geq 8\}$. This is well defined since $2/\log j < 1$, if $j \geq 8$. The indices such that $\check{\xi}_j \epsilon_j = 1$ are obviously contained in the set of indices such that $\epsilon_j = 1$. So that if $\mathcal{C}_1 = \{j \geq 8: \xi_j = 1\}$, $\mathcal{B} = \{j \geq 1: \epsilon_j = 1\}$ and $\mathcal{B}_1 = \{j \geq 8: \epsilon_j = 1\}$, the inclusion

$$\mathcal{C}_1 \subset \mathcal{B}_1, \quad (5.1)$$

is satisfied with probability 1. Whence also,

Proposition 5.1 *The inclusion $\mathcal{C} \subset \mathcal{B}$ holds true with positive probability.*

5.2 Instants of jump in the Bernoulli model

The instants of jump of the sequence $\{B_k, k \geq 1\}$ are defined as follows: Put $\delta_0 = 0$, $\Delta_0 = 0$ and for any integers $\ell \geq 1$, $k \geq 1$,

$$\delta_\ell = \inf\{n \geq 1: \epsilon_{n+\delta_1+\dots+\delta_{\ell-1}} = 1\}, \quad \Delta_k = \delta_1 + \dots + \delta_k. \quad (5.2)$$

We have $\mathcal{B} := \{j \geq 1: \epsilon_j = 1\} = \{\Delta_k, k \geq 1\}$. Proposition 5.1 suggests to first study the probability that Δ_k be prime. We first establish some necessary properties of δ_ℓ and Δ_k .

Theorem 5.2 (i) *The random variables δ_k are i.i.d., exponentially distributed,*

$$\mathbb{P}\{\delta_\ell = m\} = 2^{-m} \text{ for all } \ell \geq 1 \text{ and } m \geq 1.$$

(ii) *Then $\mathbb{E} \delta_1 = 2$, $\text{Var}(\delta_1) = 2$. Further,*

$$\mathbb{E} e^{2i\pi \delta_1 t} = \sum_{m=1}^{\infty} \frac{e^{2i\pi m t}}{2^m}, \quad \mathbb{E} e^{2i\pi \Delta_k t} = \sum_{v=1}^{\infty} \frac{e^{2i\pi v t}}{2^v} C_{v-1}^{k-1}. \quad (5.3)$$



(iii) For all $k \geq 1, m \geq 1, \mathbb{P}\{\Delta_k = m\} = \frac{C_{m-1}^{k-1}}{2^m} = \frac{1}{2} \mathbb{P}\{B_{m-1} = k-1\}$.

(iv) With probability one,

$$\limsup_{n \rightarrow \infty} \frac{\delta_k}{\log k} = 1, \quad \limsup_{n \rightarrow \infty} \frac{\Delta_k - 2k}{2\sqrt{k} \log \log k} = 1.$$

(v) The central limit theorem and the local limit theorem holds, namely $\frac{\Delta_k - 2k}{\sqrt{2k}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$, as $k \rightarrow \infty$ and

$$\sup_n \left| \sqrt{k} \mathbb{P}\{\Delta_k = n\} - \frac{1}{2\sqrt{\pi}} e^{-\frac{(n-2k)^2}{4k}} \right| = \mathcal{O}(k^{-1/2}).$$

Proof (i) As $\{\delta_1 = \mu\} = \{\beta_1 = \dots = \beta_{\mu-1} = 0, \beta_\mu = 1\} \in \sigma(\beta_1, \dots, \beta_\mu)$, we have $\mathbb{P}\{\delta_1 = \mu\} = \mathbb{P}\{\beta_1 = \dots = \beta_{\mu-1} = 0, \beta_\mu = 1\} = 2^{-\mu}$. By recursion on $\ell \geq 1$, one establishes

$$\begin{aligned} \mathbb{P}\{\delta_{\ell+1} = m\} &= \sum_{m_1=1}^{\infty} \dots \sum_{m_\ell=1}^{\infty} \mathbb{P}\{\delta_1 = m_1, \dots, \delta_\ell = m_\ell, \\ &\quad \beta_{m_1+\dots+m_\ell+1} = \dots = \beta_{m_1+\dots+m_\ell+m-1} = 0, \beta_{m_1+\dots+m_\ell+m} = 1\} \\ &= \sum_{m_1=1}^{\infty} \dots \sum_{m_\ell=1}^{\infty} \prod_{i=1}^{\ell} \mathbb{P}\{\delta_i = m_i\} \mathbb{P}\{\beta_{m_1+\dots+m_\ell+1} = \dots = \beta_{m_1+\dots+m_\ell+m-1} = 0, \\ &\quad \beta_{m_1+\dots+m_\ell+m} = 1\} = \sum_{m_1=1}^{\infty} \dots \sum_{m_\ell=1}^{\infty} 2^{-m_1-\dots-m_\ell} 2^{-m} = 2^{-m}, \end{aligned}$$

and also find that $\mathbb{P}\{\delta_1 = a_1, \dots, \delta_{\ell+1} = a_{\ell+1}\} = \prod_{i=1}^{\ell+1} \mathbb{P}\{\delta_i = a_i\}$, for all positive integers $a_i, 1 \leq i \leq \ell+1$. Further $\{\delta_1 = m_1, \delta_2 = m_2, \dots, \delta_{\ell+1} = m_{\ell+1}\} \in \sigma(\beta_1, \dots, \beta_{m_1+\dots+m_{\ell+1}})$.

(ii) The characteristic function of δ_1 being $\mathbb{E} e^{2i\pi\delta_1 t} = \sum_{m=1}^{\infty} \frac{e^{2i\pi m t}}{2^m}$, we have

$$\mathbb{E} e^{2i\pi\Delta_k t} = \left(\sum_{m=1}^{\infty} \frac{e^{2i\pi m t}}{2^m} \right)^k = \sum_{m_1=1}^{\infty} \dots \sum_{m_k=1}^{\infty} \frac{e^{2i\pi(m_1+\dots+m_k)t}}{2^{m_1+\dots+m_k}} = \sum_{v=k}^{\infty} \frac{e^{2i\pi v t}}{2^v} C_{v-1}^{k-1}. \quad (5.4)$$

Let $S(u) = \sum_{a=0}^{\infty} e^{-au}$, then $S(u) = \frac{1}{1-e^{-u}}$, $S'(u) = -\frac{e^{-u}}{(1-e^{-u})^2} = -\sum_{a=1}^{\infty} a e^{-au}$, $S''(u) = \frac{e^{-u}(1+e^{-u})}{(1-e^{-u})^3} = -\sum_{a=1}^{\infty} a^2 e^{-au}$. Thus first and second moments of δ_1 can be computed and one finds that $\mathbb{E} \delta_1 = 2$, $\text{Var}(\delta_1) = 2$.

(iii) Follows from

$$\begin{aligned} \mathbb{P}\{\Delta_k = m\} &= \sum_{\substack{m_i \geq 1, 1 \leq i \leq k \\ m_1+\dots+m_k=m}} \mathbb{P}\{\delta_1 = m_1, \dots, \delta_k = m_k\} \\ &= \sum_{\substack{m_i \geq 1, 1 \leq i \leq k \\ m_1+\dots+m_k=m}} \frac{1}{2^{m_1+\dots+m_k}} = \frac{q_k(m)}{2^m} = \frac{C_{m-1}^{k-1}}{2^m} = \frac{1}{2} \mathbb{P}\{B_{m-1} = k-1\}. \end{aligned} \quad (5.5)$$

(iv) Immediate.

(v) The central limit theorem is obvious since the δ_k 's are i.i.d. and square integrable. By Theorem 6 p.197 in [23], as $\mathbb{E} |\delta_1|^3 < \infty$, we have

$$\sup_n \left| \sqrt{k} \mathbb{P}\{\Delta_k = n\} - \frac{1}{2\sqrt{\pi}} e^{-\frac{(n-2k)^2}{4k}} \right| = \mathcal{O}(k^{-1/2}). \quad (5.6)$$

□



We see that the divisors of the jump's instants Δ_k admit a simple formulation. In particular, we have from (iii),

$$\mathbb{P}\{\Delta_k \text{ prime}\} = \frac{1}{2} \sum_{\substack{v \geq k \\ v \text{ prime}}} \mathbb{P}\{B_{v-1} = k-1\}. \quad (5.7)$$

The formula $\sum_{v=0}^{\infty} C_{v+z}^z x^v = \frac{1}{(1-x)^{z+1}}$, valid for $|x| < 1$, further implies

$$\frac{1}{2} \sum_{v \geq k} \mathbb{P}\{B_{v-1} = k-1\} = 1. \quad (5.8)$$

5.3 Proof of Theorem 2.8

(i) By the local limit theorem for Bernoulli sums

$$\sup_z \left| \mathbb{P}\{B_n = z\} - \sqrt{\frac{2}{\pi n}} e^{-\frac{(2z-n)^2}{2n}} \right| = o\left(\frac{1}{n^{3/2}}\right). \quad (5.9)$$

Besides by Theorem 5.2-(iii), $\mathbb{P}\{\Delta_k = m\} = \frac{1}{2} \mathbb{P}\{B_{m-1} = k-1\}$. Thus

$$\begin{aligned} \mathbb{P}\{\Delta_k \in \mathfrak{P}\} &= \sum_{\substack{v \geq k \\ v \in \mathfrak{P}}} \mathbb{P}\{\Delta_k = v\} = \frac{1}{2} \sum_{\substack{v \geq k \\ v \in \mathfrak{P}}} \mathbb{P}\{B_{v-1} = k-1\} \\ &= \sum_{\substack{v \geq k \\ v \in \mathfrak{P}}} \frac{1}{\sqrt{2\pi(v-1)}} e^{-\frac{(2k-v-1)^2}{2(v-1)}} + o\left(\sum_{\substack{v \geq k \\ v \in \mathfrak{P}}} \frac{1}{v^{3/2}}\right) \\ &= \sum_{\substack{v \geq k \\ v \in \mathfrak{P}}} \frac{1}{\sqrt{2\pi(v-1)}} e^{-\frac{(2k-v-1)^2}{2(v-1)}} + o\left(\frac{1}{\sqrt{k}}\right). \end{aligned}$$

Obviously $\sum_{v \geq 3k} \frac{1}{\sqrt{v}} e^{-\frac{(2k-v-1)^2}{2(v-1)}} \leq e^{-C_1 k}$. Now by assumption, for $k \in \mathcal{K}$,

$$\sum_{\substack{k \leq v \leq bk \\ v \in \mathfrak{P}}} \frac{1}{\sqrt{v}} e^{-\frac{(2k-v-1)^2}{2(v-1)}} \leq C \frac{\#\{\mathfrak{P} \cap [k, bk]\}}{\sqrt{k}} \leq C k^{-\beta}.$$

It easily follows that

$$\mathbb{P}\{\Delta_k \in \mathfrak{P}\} \leq C_b k^{-\beta}, \quad k \in \mathcal{K}.$$

The series $\sum_{k \in \mathcal{K}} \mathbb{P}\{\Delta_k \in \mathfrak{P}\} < \infty$ converges. It follows from Borel-Cantelli lemma that

$$\mathbb{P}\{\Delta_k \notin \mathfrak{P}, \quad k \in \mathcal{K} \text{ ultimately}\} = 1.$$

To prove the same assertion concerning the sequence $\mathcal{P}$, it suffices to argue as before Proposition 5.1, by considering the sequence $\{\check{\xi}_j \in j, j \in \mathcal{K}\}$.



6 Instants of small amplitude in the Cramér model: proofs

Before giving the proofs of Theorem 2.2, Proposition 2.3 and Theorem 2.4, it is necessary for the understanding of the matter to recall some notation and results from [30]. Let $f : [1, \infty) \rightarrow \mathbb{R}^+$ be here and throughout a non-decreasing function such that $f(t) \uparrow \infty$ with t and $f(t) = o_\rho(t^\rho)$. The intervals I considered (cf. (1.7)) are of type $[e^k, e^k f(e^k)]$, $k = 1, 2, \dots$

Put

$$A_k(f, z) = \left\{ \sup_{e^k \leq t \leq e^k f(e^k)} \frac{|W(t)|}{\sqrt{t}} < z \right\}, \quad k = 1, 2, \dots \quad (6.1)$$

Let $U(t) = W(e^t)e^{-t/2}$, $t \in \mathbb{R}$ be the Ornstein-Uhlenbeck process. It will be more convenient to work with U instead of W . Observe that

$$A_k(f, z) = \left\{ \sup_{k \leq s \leq k + \log f(e^k)} |U(s)| \leq z \right\}.$$

And so as U is stationary

$$\mathbb{P}\{A_k(f, z)\} = \mathbb{P}\left\{ \sup_{0 \leq s \leq \log f(e^k)} |U(s)| \leq z \right\}.$$

We say that $f \in \mathcal{U}_z$ whenever $\mathbb{P}\{\limsup_{k \rightarrow \infty} A_k(f, z)\} = 0$, and that $f \in \mathcal{V}_z$ if $\mathbb{P}\{\limsup_{k \rightarrow \infty} A_k(f, z)\} = 1$. By the 0-1 law (since U is strongly mixing), the latter probabilities can only be 0 or 1.

Notice that if $f \in \mathcal{U}_z$, then with probability one

$$J(f) := \liminf_{k \rightarrow \infty} \sup_{k \leq s \leq k + \log f(e^k)} |U(s)| \geq z,$$

whereas $J(f) \leq z$, almost surely if $f \in \mathcal{V}_z$. Introduce also for $n = 1, 2, \dots$ the counting function

$$N_n(f, z) = \sum_{k=1}^n \chi_{A_k(f, z)},$$

with corresponding mean

$$v_n(f, z) := \mathbb{E} N_n(f, z).$$

The classes $\mathcal{U}_z$ and $\mathcal{V}_z$ have been characterized in [30, Th. 1.1], where the following simple convergence criterion is proved.

Theorem 6.1 *Let $\Sigma(f) = \sum_k f(e^k)^{-\lambda(z)}$, where $\lambda(z) > 0$ is defined in Theorem 2.2. Then*

$$f \in \mathcal{U}_z \quad (\text{resp. } \in \mathcal{V}_z) \iff \Sigma(f) < \infty \quad (\text{resp. } = \infty).$$

Further if $\Sigma(f) = \infty$, for any $a > 3/2$,

$$N_n(f, z) = v_n(f, z) + \mathcal{O}(v_n^{1/2}(f, z) \log^a v_n(f, z)),$$

with probability one. And there are positive constants $K_1(z)$, $K_2(z)$ depending on z only, such that for all n

$$K_1(z) \leq \frac{v_n(f, z)}{\sum_{k=1}^n f(e^k)^{-\lambda(z)}} \leq K_2(z).$$

In [30], applications to the Kubilius model are given. The class of functions $f_c(t) = \log^c t$, $c > 0$, is of special interest in view of such applications. For these functions, Theorem 6.1 implies

Corollary 6.2 *If $c > 1/\lambda(z)$, then $f_c \in \mathcal{U}_z$ whereas $f_c \in \mathcal{V}_z$ if $0 < c \leq 1/\lambda(z)$. Further, for any $0 < c \leq 1/\lambda(z)$ and $a > 3/2$,*

$$N_n(f_c, z) \stackrel{a.s.}{=} v_n(f_c, z) + \mathcal{O}(v_n^{1/2}(f_c, z) \log^a v_n(f_c, z)).$$

And for all n , $K_1(z) \leq \frac{v_n(f_c, z)}{\sum_{k=1}^n k^{-c\lambda(z)}} \leq K_2(z)$.



Accordingly, if

$$I(f) := \liminf_{k \rightarrow \infty} \sup_{e^k \leq t \leq e^{k+1}} \frac{|W(t)|}{\sqrt{t}}, \quad (6.2)$$

then $\mathbb{P}\{I(f_c) \leq z\} = 1$ if and only if $0 < c \leq 1/\lambda(z)$. This is clear in view of (6.1). Noticing that $I(f) \leq I(g)$ whenever $f(N) \leq g(N)$ for all N large, we therefore also deduce

Corollary 6.3 *We have $\mathbb{P}\{I(f_c) \leq z\} = 1$ if and only if $0 < c \leq 1/\lambda(z)$. And $\mathbb{P}\{I(f) = \infty\} = 1$ if $f(t) \gg_c f_c(t)$ for all c .*

We need a suitable invariance principle for sums of independent random variables, which will be also used in the next section. This one is due to Sakhanenko (see [27], Theorem 1).

Let $\{\xi_j, j \geq 1\}$ be independent centered random variables with absolute second moments. Let $t_k = \sum_{j=1}^k \mathbb{E} \xi_j^2$, $S_k = \sum_{j=1}^k \xi_j$ and let $\{r_k, k \geq 1\}$ be some non-decreasing sequence of positive reals. Let $\alpha \geq 2$, $y > 0$. Put successively,

$$\begin{aligned} \Delta_n &= \sup_{k \leq n} |S_k - W(t_k)|, \\ \Delta &= \sup_{n \geq 1} \frac{\Delta_n}{r_n}, \\ \bar{\xi} &= \sup_{j \geq 1} \frac{|\xi_j|}{r_j}, \\ L_\alpha(y) &= \sum_{j \geq 1} \mathbb{E} \min \left\{ \frac{|\xi_j|^\alpha}{y^\alpha r_j^\alpha}, \frac{|\xi_j|^2}{y^2 r_j^2} \right\}. \end{aligned} \quad (6.3)$$

Lemma 6.4 *There exists an absolute constant C such that for any fixed α , there exists a Brownian motion W such that for all $x > 0$,*

$$\mathbb{P}\{\Delta \geq C\alpha x\} \leq L_\alpha(x).$$

In our case, we choose $X_i = \xi_i - \mathbb{E} \xi_i$. Let $1/\alpha < \beta < 1/2$. Take $r_j = (\sum_{i=1}^j \mathbb{E} |X_i|^2)^\beta = B_j^\beta$. Since $B_j \asymp \frac{j}{\log j}$, $j \rightarrow \infty$, it follows

$$\sum_{j \geq 1} \frac{\mathbb{E} |X_j|^\alpha}{r_j^\alpha} = \sum_{j \geq 1} \frac{\mathbb{E} |X_j|^\alpha}{B_j^{\alpha\beta}} \leq C \sum_{j \geq 2} \frac{1}{(j/\log j)^{\alpha\beta} (\log j)} < \infty,$$

as $\alpha\beta > 1$. Thus

$$L_\alpha(y) \leq y^{-\alpha} \sum_{j \geq 1} \mathbb{E} \frac{|X_j|^\alpha}{r_j^\alpha} \leq C_\alpha y^{-\alpha}.$$

By Lemma 6.4, there exists a Brownian motion W such that for all $x > 0$,

$$\mathbb{P}\left\{ \sup_n \frac{1}{r_n} \sup_{j \leq n} |S_j - m_j - W(B_j)| \geq C\alpha x \right\} \leq C_\alpha x^{-\alpha}.$$

By a simple use of Tchebycheff's inequality, letting

$$\Upsilon = \sup_n \frac{1}{r_n} \sup_{j \leq n} |S_j - m_j - W(B_j)|,$$

we deduce that

$$\mathbb{E} \Upsilon^{\alpha'} < \infty, \quad (\alpha' < \alpha).$$



We shall now use Theorem 1.5 in [30]. We first note that $\mathbb{E} X_j^2 = \frac{1}{\log j} (1 - \frac{1}{\log j})$, and for $\alpha > 2$, as $|X_j|$ is bounded by $1 + \frac{1}{\log j} \leq C$, say, we have $\mathbb{E} |X_j|^\alpha \leq C^{\alpha-2} \mathbb{E} X_j^2$. Further $\mathbb{E} |X_j|^\alpha \geq (\mathbb{E} |X_j|^2)^{\alpha/2} \geq (\frac{C}{\log j})^{\alpha/2}$. Thus

$$v := \sup_{j \geq 1} \frac{\mathbb{E} |X_j|^\alpha}{\mathbb{E} |X_j|^2} < \infty. \quad (6.4)$$

Therefore assumption (1.4) of Th. 1.5 in [30] is fulfilled. We deduce that there exists a Brownian motion W such that

$$\liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f(e^k)} \frac{|S_j|}{\sqrt{B_j}} = \liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f(e^k)} \frac{|W(B_j)|}{S_j},$$

with probability 1. By Corollary 6.3,

$$\liminf_{k \rightarrow \infty} \sup_{e^k \leq B_j \leq e^k f(e^k)} \frac{|S_j|}{\sqrt{B_j}} \leq z,$$

with probability 1, if and only if $c \leq 1/\lambda(z)$.

We pass to the proof of Theorem 2.4. Let N be for the moment unspecified. Then,

$$\left| \sup_{j \in J_N} \frac{|S_j - m_j|}{\sqrt{B_j}} - \sup_{j \in J_N} \frac{|W(B_j)|}{\sqrt{B_j}} \right| \leq \left(\sup_{j \in J_N} \frac{1}{B_j^{1-2\beta}} \right) \Upsilon \rightarrow 0, \quad (6.5)$$

as N tends to infinity, with probability one.

Recall that

$$A_k(f, z) = \left\{ \sup_{e^k \leq t \leq e^k f(e^k)} \frac{|W(t)|}{\sqrt{t}} \leq z \right\},$$

by (6.1) and that $v_n(f_c, z) = \sum_{k=1}^n \mathbb{P}\{A_k(f_c, z)\}$. Further by Corollary 6.2, for all n ,

$$K_1(z) \sum_{k=1}^n k^{-c\lambda(z)} \leq v_n(f_c, z) \leq K_2(z) \sum_{k=1}^n k^{-c\lambda(z)},$$

and

$$\sum_{k=1}^n k^{-c\lambda(z)} = \begin{cases} \frac{n^{1-c\lambda(z)}}{1-c\lambda(z)} + \zeta(c\lambda(z)) + \mathcal{O}(n^{-c\lambda(z)}) & \text{if } 0 < c\lambda(z) < 1, \\ \log n + \gamma + \mathcal{O}(\frac{1}{n}) & \text{if } c\lambda(z) = 1, \end{cases} \quad (6.6)$$

where γ is Euler's constant, recalling that $\zeta(s) = \lim_{x \rightarrow \infty} (\sum_{n \leq x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s})$, $0 < s < 1$.

Choose $N = e^k$, $k = 1, 2, \dots$ and write now more simply $B_{e^k}(f, z) = B_k(f, z)$. Let also $0 < z' < z < z''$.

By (6.5), on a measurable set of probability as close to one as we please, call it Ω^* , we have for k large enough, $k \geq k_0$ say,

$$A_k(f, z') \subset B_k(f, z) \subset A_k(f, z''). \quad (6.7)$$

As $\lambda(z)$ is a positive strictly decreasing continuous function of z on $]0, \infty[$, we have $0 < \lambda(z'') < \lambda(z) < \lambda(z')$.

Let $f = f_c$ with $0 < c \leq 1/\lambda(z)$, and note that

$c \leq 1/\lambda(z'')$. Let $a > 3/2$. By using Corollary 6.2, we get that on Ω^* , for all n large enough,

$$\sum_{k=k_0}^n \chi_{B_k(f_c, z)} \leq \sum_{k=k_0}^n \chi_{A_k(f_c, z'')} = v_n(f_c, z'') + \mathcal{O}_a(v_n^{1/2}(f_c, z'') \log^a v_n(f_c, z'')). \quad (6.8)$$

We deduce that

$$\mathbb{P} \left\{ \sum_{k=1}^n \chi_{B_k(f_c, z)} \leq v_n(f_c, z'') + \mathcal{O}_a(v_n^{1/2}(f_c, z'') \log^a v_n(f_c, z'')) \right\} = 1. \quad (6.9)$$



Similarly, on Ω^* for all n large enough,

$$\sum_{k=k_0}^n \chi_{A_k}(f_c, z') \leq \sum_{k=k_0}^n \chi_{B_k}(f_c, z). \quad (6.10)$$

Assume that $0 < c \leq 1/\lambda(z')$. By using Theorem 6.1, we get

$$\mathbb{P}\left\{\sum_{k=1}^n \chi_{B_k}(f_c, z) \geq v_n(f_c, z') - \mathcal{O}_a\left(v_n^{1/2}(f_c, z') \log^a v_n(f_c, z')\right), \text{ ultimately}\right\} = 1. \quad (6.11)$$

Note that (6.9) is also valid if $0 < c \leq 1/\lambda(z')$.

Hence we have proved Theorem 2.4.

7 Subsequence LIL results for the Cramér model: proofs

Let $\mathcal{N} = \{n_k, k \geq 1\}$ be any increasing sequence of integers and $M > 1$; the value of M will be irrelevant. Let $I_0 =]0, M]$ and for each integer $k \geq 1$, let $I_k =]M^k, M^{k+1}]$. The subsequence of intervals I_k such that $I_k \cap \mathcal{N} \neq \emptyset$, determines an increasing sequence of indices, which we denote by $\kappa = \{\kappa_p, p \geq 1\}$. For any $n \in \mathcal{N}$ we put

$$\varphi_{\mathcal{N}}(n) = \sqrt{2 \log(p+2)} \quad \text{if } n \in \mathcal{N} \cap I_{\kappa_p}. \quad (7.1)$$

Recall that $r_j = B_j^\beta$ and that

$$\Upsilon = \sup_n \frac{1}{r_n} \sup_{j \leq n} |S_j - m_j - W(B_j)|.$$

Now let $j_p^* = \max\{j : B_j \in \mathcal{N} \cap]2^{p-1}, 2^p]\}$. As

We have

$$\begin{aligned} & \left| \sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{|S_j - m_j|}{\sqrt{B_j} \varphi(j)} - \sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{|W(B_j)|}{\sqrt{B_j} \varphi(j)} \right| \\ & \leq \sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{|S_j - m_j - W(B_j)|}{\sqrt{B_j} \varphi(j)} = \sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{|S_j - m_j - W(B_j)|}{B_j^{1/2-\beta} B_j^\beta \varphi(j)} \\ & \leq \left(\sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{1}{B_j^{1/2-\beta} \varphi(j)} \right) \sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{|S_j - m_j - W(B_j)|}{r_j} \\ & \leq \left(\sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{1}{B_j^{1/2-\beta} \varphi(j)} \right) \frac{2^\beta}{r_{j_p^*}} \sup_{j \leq j_p^*} |S_j - m_j - W(B_j)| \\ & \leq 2^\beta \left(\sup_{\substack{2^{p-1} < B_j \leq 2^p \\ j \in \mathcal{N}}} \frac{1}{B_j^{1/2-\beta} \varphi(j)} \right) \cdot \Upsilon \rightarrow 0, \end{aligned}$$

as $p \rightarrow \infty$ almost surely, since $\beta < 1/2$.

Therefore

$$\limsup_{\mathcal{N} \ni j \rightarrow \infty} \frac{|S_j - m_j|}{\sqrt{B_j} \varphi(j)} = \limsup_{\mathcal{N} \ni j \rightarrow \infty} \frac{|W(B_j)|}{\sqrt{B_j} \varphi(j)},$$

almost surely.



Now that

$$\limsup_{\mathcal{N} \ni j \rightarrow \infty} \frac{|W(B_j)|}{\sqrt{B_j} \varphi(j)} = 1,$$

almost surely, follows from the proof of Theorem 3.3 in [31]. This is rather easy to observe from estimates (3.22), (3.23), (3.24) in [31].

Bibliographic Notes. We mainly referred during this work, to Cramér [2], [3], Granville's exposition in [9], Pintz's systematic analysis of Cramér's model in [23], Ellison's seminar paper [6], which notably contains a detailed proof of Hoheisel's seminal result [12], Ingham [16], Selberg [28], Richards [25] where an interesting alternative approach to derive some of Selberg's results, under a weaker form, is purposed.

Acknowledgements The author is much grateful to Antanas Laurinćikas for his warm appreciation and his interest in this work.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

Appendix A. Some characteristics of the Cramér model

A.1 Classical limit theorems

Let us first briefly look at standard limiting results such as CLT, LIL and ASLLT which are fulfilled by the Cramér model.

Lemma A.1 *The SLLN, CLT, LIL and LLT hold true. More precisely,*

$$\begin{aligned} (\text{SLLN}) \quad & \lim_{n \rightarrow \infty} \frac{S_n}{m_n} = 1, \quad \text{almost surely,} \\ (\text{CLT}) \quad & \frac{S_n - m_n}{\sqrt{B_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad n \rightarrow \infty, \\ (\text{LIL}) \quad & \limsup_{n \rightarrow \infty} \frac{S_n - m_n}{\sqrt{2B_n \log \log B_n}} = 1, \quad \text{almost surely,} \\ (\text{LLT}) \quad & \lim_{n \rightarrow \infty} \sup_k \left| \sqrt{B_n} \mathbb{P}\{S_n = k\} - \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-m_n)^2}{2B_n}} \right| = 0. \end{aligned}$$

This is an immediate consequence of the following Lemma.

Lemma A.2 *Let $\{X_j, j \geq 1\}$ be independent binomial random variables with $\mathbb{P}\{X_j = 0\} = 1 - \mathbb{P}\{X_j = 1\} = p_j$, for all j and let $S_n = \sum_{j=1}^n X_j$, $n \geq 1$. Further let $\sigma_j^2 = \text{Var}(X_j) = p_j(1 - p_j)$, $B_n = \text{Var}(S_n) = \sum_{j=1}^n p_j(1 - p_j)$.*

Assume that the series $\sum_j p_j$ diverges and that $p_j = o(1)$. Then the SLLN, CLT, LIL and LLT hold true.

Proof Let $\epsilon > 0$. Then,

$$\begin{aligned} & \frac{1}{B_n} \sum_{j=1}^n \mathbb{E} \left(|X_j - \mathbb{E} X_j|^2 \cdot \chi \{ |X_j - \mathbb{E} X_j| > \epsilon \text{Var}(S_n)^{1/2} \} \right) \\ &= \frac{1}{B_n} \sum_{j=1}^n \mathbb{E} \left(|X_j - p_j|^2 \cdot \chi \{ |X_j - p_j| > \epsilon \left(\sum_{j=1}^n p_j(1 - p_j) \right)^{1/2} \} \right). \end{aligned}$$

As the series $\sum_j p_j$ diverges and $p_j = o(1)$, the above summands are 0 for all $n \geq n_\epsilon$, say. This implies that

$$\lim_{n \rightarrow \infty} \frac{1}{B_n} \sum_{j=1}^n \mathbb{E} \left(|X_j - \mathbb{E} X_j|^2 \cdot \chi \{ |X_j - p_j| > \epsilon \text{Var}(S_n)^{1/2} \} \right) = 0,$$



for any positive ϵ . Whence the Lindeberg condition is satisfied, and so the CLT holds:

$$\frac{S_n - m_n}{\sqrt{B_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad n \rightarrow \infty. \quad (\text{A.1})$$

Concerning the LIL, Kolmogorov's condition that $X_n = o((B_n/\log \log B_n))^{1/2}$, almost surely, is trivially satisfied. Thus the LIL directly follows from Kolmogorov's theorem. See Petrov [23], Th. 7.1 p. 239. The LIL implies the SLLN, given to the assumptions made. For establishing the LLT, we apply Davis and McDonald's theorem [4, Th. 1.1] which we recall.

Theorem A.3 *Let $\{X_j, j \geq 1\}$ be independent, integer valued random variables with partial sums $S_n = X_1 + \dots + X_n$ and let $f_j(k) = \mathbb{P}\{X_j = k\}$. Also for each j and n , let*

$$q(f_j) = \sum_k [f_j(k) \wedge f_j(k+1)], \quad Q_n = \sum_{j=1}^n q(f_j)$$

and assume that $q(f_j) > 0$ for each $j \geq 1$. Further assume that there exist numbers $b_n > 0, a_n$ such that

$$\lim_{n \rightarrow \infty} b_n = \infty, \quad \limsup_{n \rightarrow \infty} b_n^2 / Q_n < \infty,$$

and

$$\frac{S_n - a_n}{b_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Then

$$\lim_{n \rightarrow \infty} \sup_k \left| b_n \mathbb{P}\{S_n = k\} - \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-a_n)^2}{2b_n^2}} \right| = 0.$$

With the notation used, $f_j(k) = \mathbb{P}\{X_j = k\}$ and so

$$q(f_j) = \sum_k (f_j(k) \wedge f_j(k+1)) = (f_j(1) \wedge f_j(0)) = (p_j \wedge 1 - p_j) > 0.$$

Further $Q_n = \sum_{j=1}^n q(f_j) = \sum_{j=1}^n (p_j \wedge 1 - p_j)$, and $b_n^2 = B_n = \sum_{j=1}^n p_j(1 - p_j)$. Thus

$$\lim_{n \rightarrow \infty} b_n = \infty, \quad \limsup_{n \rightarrow \infty} b_n^2 / Q_n = \limsup_{n \rightarrow \infty} \frac{\sum_{j=1}^n p_j(1 - p_j)}{\sum_{j=1}^n (p_j \wedge 1 - p_j)} < \infty.$$

As moreover the CLT holds, we infer from the above cited result,

$$\lim_{n \rightarrow \infty} \sup_k \left| \sqrt{B_n} \mathbb{P}\{S_n = k\} - \frac{1}{\sqrt{2\pi}} e^{-\frac{(k-m_n)^2}{2B_n}} \right| = 0,$$

namely the LLT holds either. □

A.2 The characteristic function of S_n

Let $\varphi_k(t)$ be the characteristic function of ξ_k , $\varphi_k(t) = \mathbb{E} e^{2i\pi \xi_k t}$. Let also

$$\Phi_n(t) = \mathbb{E} e^{2i\pi t S_n} = \prod_{k=1}^n \varphi_k(t),$$

be the characteristic function of S_n . We prove the following estimate with explicit constants.



Proposition A.4 We have $|\Phi_n(t)| \leq \exp \{-2B_n \sin^2 \pi t\}$. Further

$$\Phi_n(t) = e^{2i\pi t m_n - 2B_n(\pi t)^2 + E_n(t)}, \quad \text{with} \quad |E_n(t)| \leq 12m_n(\pi|t|)^3.$$

In particular,

$$\mathbb{E} e^{iy \frac{S_n - m_n}{\sqrt{B_n}}} = e^{-\frac{y^2}{2} + E_n(y)}, \quad \text{and} \quad |E_n(y)| \leq 2 \frac{\sqrt{\log n} |y|^3}{\sqrt{n}}.$$

For the proof, we use the Lemma below which is inspired by Lemma 3 in Freiman-Pitman [7]. The goal being to obtain an estimate of $\Phi(t)$ with explicit constants, this however requires to base the proof on different calculations. In particular we use the convenient estimate (Lemma 4.14 in Kallenberg [17]),

$$\left| e^{ix} - \sum_{k=0}^n \frac{(ix)^k}{k!} \right| \leq \frac{2|x|^n}{n!} \wedge \frac{|x|^{n+1}}{(n+1)!}, \quad (\text{A.2})$$

valid for any $x \in \mathbb{R}$ and $n \in \mathbb{Z}_+$.

Lemma A.5 Let m be a positive real and p be a real such that $0 < p < 1$. Let β be a random variable defined by $\mathbb{P}\{\beta = 0\} = p$, $\mathbb{P}\{\beta = m\} = 1 - p = q$. Let $\varphi(t) = \mathbb{E} e^{2i\pi t \beta}$. Then we have the following estimates,

- (i) For all real t , $|\varphi(t)| \leq \exp \{-2pq \sin^2 \pi t m\}$
- (ii) If $q|\sin \pi t m| \leq 1/3$, then

$$\varphi(t) = e^{q2i\pi m t - 2qp(\pi m t)^2 + E},$$

$$\text{with } |E| \leq 12q(\pi m|t|)^3.$$

Proof (i) One verifies that $|\varphi(t)|^2 = 1 - 4pq \sin^2 \pi m t$. As moreover $1 - \theta \leq e^{-\theta}$ if $\theta \geq 0$, we obtain $|\varphi(t)|^2 \leq e^{-4pq \sin^2 \pi m t}$.

(ii) Let $|u| \leq u_0 < 1$. From the series expansion of $\log(1+u)$, it follows that

$$\log(1+u) = u - \frac{u^2}{2} + R, \quad |R| \leq |u|^3 \sum_{j=0}^{\infty} \frac{|\theta|^j}{3+j} \leq \frac{|u|^3}{3(1-u_0)}.$$

Then $1+u = \exp\{u - \frac{u^2}{2} + B\}$, with $|B| \leq C_0 |u|^3$ and $C_0 = \frac{1}{3(1-u_0)}$.

Writing that $\varphi(t) = 1 + q(e^{2i\pi m t} - 1) = 1 + u$, where $|u| = 2q|\sin \pi m t| \leq 2q(1 \wedge \pi m|t|)$, we obtain

$$\varphi(t) = 1 + u = e^{q(e^{2i\pi m t} - 1) - \frac{u^2}{2} + B}, \quad |B| \leq 8C_0 q^3 |\sin \pi m t|^3. \quad (\text{A.3})$$

In order to estimate u^2 , we let $A(t) = e^{2i\pi m t} - 1 - 2i\pi m t + \frac{(2\pi m t)^2}{2}$, and write $(e^{2i\pi m t} - 1)^2$ under the form

$$\begin{aligned} A(2t) - 2A(t) - \left\{ -1 - 4i\pi m t + 8(\pi m t)^2 \right\} + 2 \left\{ -1 - 2i\pi m t + \frac{(2\pi m t)^2}{2} \right\} + 1 \\ = A(2t) - 2A(t) - (2\pi m t)^2. \end{aligned}$$

Then $(e^{2i\pi m t} - 1)^2 + (2\pi m t)^2 = A(2t) - 2A(t)$, and so $\frac{u^2}{2} = -2q^2(\pi m t)^2 + \frac{q^2}{2}(A(2t) - 2A(t))$.

Let $u_0 = \frac{2}{3}$ so that $C_0 = 1$. We assumed $q|\sin \pi t m| \leq 1/3$, thus $|u| \leq 2/3$. We consequently get with (A.3),

$$\begin{aligned} \varphi(t) &= e^{q(e^{2i\pi m t} - 1) - \frac{u^2}{2} + B} = e^{q(e^{2i\pi m t} - 1) + 2q^2(\pi m t)^2 - \frac{q^2}{2}(A(2t) - 2A(t)) + B} \\ &= e^{q2i\pi m t - 2qp(\pi m t)^2 + H + B}, \end{aligned} \quad (\text{A.4})$$

with $H = qA(t) - \frac{q^2}{2}(A(2t) - 2A(t))$. By estimate (A.2), letting $\delta(x) = x^2(1 \wedge \frac{|x|}{6})$,

$$|A(t)| \leq \delta(2\pi m|t|). \quad (\text{A.5})$$

Using the rough bound $\delta(x) \leq \frac{|x|^3}{6}$, and $|B| \leq 8q^3(1 \wedge \pi m|t|)^3$, we get

$$|H| + |B| \leq 2q|A(t)| + \frac{q^2}{2}|A(2t)| + |B| \leq \left(\frac{8}{3}q + \frac{4}{3}q^2 + 8q^3\right)(\pi m|t|)^3 \leq 12q(\pi m|t|)^3. \quad (\text{A.6})$$

We conclude by inserting estimate (A.6) into (A.4). $\square$



Proof of Proposition A.4 Here we have $m = 1$, $q = \frac{1}{\log k}$. As $|\varphi_k(t)| \leq \exp \left\{ -2 \left(1 - \frac{1}{\log k} \right) \left(\frac{1}{\log k} \right) \sin^2 \pi t \right\}$, we get

$$|\Phi_n(t)| \leq \exp \left\{ -2 \sum_{k=3}^n \left(1 - \frac{1}{\log k} \right) \left(\frac{1}{\log k} \right) \sin^2 \pi t \right\}.$$

Further condition $q|\sin \pi t m| \leq 1/3$ in Lemma A.5 reduces for $\varphi_k(t)$ to $\frac{1}{\log k} |\sin \pi t| \leq 1/3$. Thus

$$\varphi_k(t) = e^{2i\pi \left(\frac{1}{\log k} \right) t - 2\pi^2 \left(1 - \frac{1}{\log k} \right) \left(\frac{1}{\log k} \right) t^2 + C_k(t)}, \quad (\text{A.7})$$

where $|C_k(t)| \leq \frac{12(\pi|t|)^3}{\log k}$. Recalling that $m_n = \sum_{j=3}^n \frac{1}{\log j}$, $B_n = \sum_{j=3}^n \frac{1}{\log j} \left(1 - \frac{1}{\log j} \right)$, it follows that

$$\mathbb{E} e^{2i\pi t S_n} = \Phi_n(t) = e^{2i\pi t m_n - 2\pi^2 B_n t^2 + D_n(t)}, \quad (\text{A.8})$$

and $|D_n(t)| \leq 12 m_n (\pi|t|)^3$. In particular,

$$\mathbb{E} e^{iy \frac{S_n - m_n}{\sqrt{B_n}}} = e^{-\frac{y^2}{2} + E_n(y)}, \quad (\text{A.9})$$

with $|E_n(y)| \leq \frac{3m_n}{2} \left(\frac{y}{\sqrt{B_n}} \right)^3 \leq 2 \frac{\sqrt{\log n} |y|^3}{\sqrt{n}}$. This achieves the proof. $\square$

One can derive from Proposition A.4 the following value distribution result on divisors of S_n . We omit the proof.

Theorem A.6 Let $\Theta(d, m, B)$ be the elliptic Theta function defined by

$$\Theta(d, m, B) = \sum_{\ell=0}^{\infty} \cos \left(2m\pi \frac{\ell}{d} \right) e^{-\frac{B\pi^2 \ell^2}{2d^2}}.$$

We have

$$\sup_{2 \leq d \leq n} \left| \mathbb{P}\{d|S_n\} - \frac{\Theta(d, m_n, B_n)}{d} \right| = \mathcal{O} \left(\frac{(\log n)^3}{n} \right). \quad (\text{A.10})$$

A.3 Remarks complementary to Cramér's proof

Recall the way (1.9) is established. We provide details, which are necessary for the sequel. Let $c > 0$, and set

$$E_m = \{ \xi_{m+j} = 0, 1 \leq j \leq c(\log m)^2 \}, \quad m \geq 2.$$

Then

$$\mathbb{P}\{E_m\} = \prod_{1 \leq j \leq c(\log m)^2} \left(1 - \frac{1}{\log(m+j)} \right) \asymp \frac{1}{m^c}. \quad (\text{A.11})$$

Introduce the increasing sequence of integers $m_1 = 2$ and $m_{r+1} = m_r + [c(\log m_r)^2] + 1$.

— If $c \leq 1$, the events E_{m_r} being independent, by the second Borel-Cantelli lemma

$$\mathbb{P}\{\limsup_{r \rightarrow \infty} E_{m_r}\} = 1,$$

in other words, almost surely, $S_{m_{r+1}} - S_{m_r} \geq c(\log m_r)^2$, r infinitely often.

— If $c > 1$, by the first Borel-Cantelli lemma, the above probability is 0. Thus with probability one, $S_{m+[c(\log m)^2]} - S_m \leq c(\log m)^2$, m ultimately.

This suffices to imply (1.9), indeed:

(i) If $c > 1$, let $P_{v(m)}$ denote the greatest P_v less than S_m . As with probability one $\xi_{m+j} = 1$, for some $1 \leq j \leq c(\log m)^2$, m ultimately, $P_{v(m+1)}$ cannot be larger than $S_{m+[c(\log m)^2]}$. Thus $P_{v(m+1)} - P_{v(m)} \leq c(\log m)^2$, m ultimately, almost surely. Now by Lemma A.1, $S_m \sim \frac{m}{\log m}$, almost surely, thus $\log S_m \sim \log m$. We deduce



that for any $c' > c$ fixed, $P_{v(m+1)} - P_{v(m)} \leq c' (\log P_{v(m)})^2$, m ultimately, almost surely, which naturally implies that the limsup in (1.9) is less or equal to 1.

In addition, letting $m = m_r$, $r \geq 1$ in the above, we have since $m_{r+1} = m_r + [c(\log m_r)^2] + 1$, that with probability one: for any r large enough, there exists a jump in the interval $[m_r, m_{r+1}]$. Therefore

$$\mathbb{P}\{\exists r_0 : \forall r \geq r_0, \quad P_r \leq m_{r+1} + m_{r_1}\} = 1. \quad (\text{A.12})$$

(ii) If $c \leq 1$, then almost surely $P_{v(m_{r+1})} - P_{v(m_r)} \geq c(\log m_r)^2 \geq c(\log P_{v(m_r)})^2$, r infinitely often. Thus (1.9) is supported by the probability of the set $\limsup_{r \rightarrow \infty} E_m$ (resp. $\limsup_{r \rightarrow \infty} E_{m_r}$) which is trivially 0 (resp. 1) depending on $c > 1$ (resp. $c \leq 1$).

Remark A.7 As (1.9) does not depend on the first random variables ξ_j , it follows that the jumps associated to the truncated sequence $S_n^a = \sum_{j=a}^n \xi_j$, a integer, $n > a$, also satisfy the same asymptotic property. Therefore there is no harm in Cramér's conjecture to consider instead the 'primes' associated to this one.

We now indicate several interesting results complementing (1.9).

(1) Let $N_J = \sum_{r=1}^J \chi_{E_{m_r}}$, $J \geq 1$, and put $B_J = \sum_{r=1}^J \mathbb{P}(E_{m_r})(1 - \mathbb{P}(E_{m_r}))$. By Kolmogorov's LIL,

$$\mathbb{P}\left\{\limsup_{J \rightarrow \infty} \frac{N_J - \sum_{r=1}^J \mathbb{P}(E_{m_r})}{\sqrt{2B_J \log \log B_J}} = 1\right\} = 1. \quad (\text{A.13})$$

(2) Further, by Berry–Esseen inequality [23],

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left\{\frac{N_J - \sum_{r=1}^J \mathbb{P}(E_{m_r})}{\sqrt{B_J}} < x\right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du \right| = \mathcal{O}\left(\frac{1}{\sqrt{\sum_{r=1}^J \mathbb{P}(E_{m_r})}}\right). \quad (\text{A.14})$$

(3) Let $0 < c < 1$. From (A.13), (A.11) and $m_r \asymp cr(\log r)^2$, it follows that with probability one

$$N_J \asymp J^{1-c} (\log J)^{-2c}, \quad \text{for any } J \text{ large enough.} \quad (\text{A.15})$$

Whence it follows that for any J large enough, the interval $[1, cJ(\log J)^2]$ contains at least $C J^{1-c} (\log J)^{-2c}$ 'primes' P_v with large gaps, namely such that $P_{v+1} - P_v \geq c \log^2 P_v$. Therefore the Cramér model also predicts that with probability one, for any J large enough, the number of large gaps between 'primes' in the interval $[1, cJ(\log J)^2]$ is $\Omega_c(J^{1-c} (\log J)^{-2c})$.

A.4 Value distribution of the divisors of the Bernoulli sum

The general problem of estimating the probability $\mathbb{P}\{d|B_n\}$ amounts to the one of estimating the d series

$$\sum_j e^{-\frac{(jd-\rho)^2}{2n}}, \quad (\text{A.16})$$

in which $d \geq 2$ and $0 \leq \rho < d$ are integers. This is clear from the following uniform estimate proved in [32], see also [29], to which we refer for details, and from Poisson's summation formula.

Theorem A.8

$$\sup_{2 \leq d \leq n} \left| \mathbb{P}\{d|B_n\} - \frac{\Theta(d, n)}{d} \right| = \mathcal{O}((\log n)^{5/2} n^{-3/2}),$$

where $\Theta(d, n)$ is the elliptic Theta function

$$\Theta(d, n) = \sum_{\ell \in \mathbb{Z}} e^{in\pi \frac{\ell}{d} - \frac{n\pi^2 \ell^2}{2d^2}},$$



By applying Poisson's summation formula: for $x \in \mathbb{R}$, $0 \leq \delta \leq 1$,

$$\sum_{\ell \in \mathbb{Z}} e^{-(\ell+\delta)^2 \pi x^{-1}} = x^{1/2} \sum_{\ell \in \mathbb{Z}} e^{2i\pi \ell \delta - \ell^2 \pi x}, \quad (\text{A.17})$$

we further get

$$\frac{\Theta(d, n)}{d} = \sqrt{\frac{2}{\pi n}} \sum_{z \equiv 0 \pmod{d}} e^{-\frac{(2z-n)^2}{2n}}. \quad (\text{A.18})$$

The series in (A.18) is of the type given in (A.16). An indication of the behavior of these series when d and n may simultaneously vary, is given by the already sharp estimate,

$$\left| \frac{\Theta(d, n)}{d} - \frac{1}{d} \right| \leq \begin{cases} \frac{C}{d} e^{-\frac{n\pi^2}{2d^2}} & \text{if } d \leq \sqrt{n}, \\ r \frac{C}{\sqrt{n}} & \text{if } \sqrt{n} \leq d \leq n. \end{cases} \quad (\text{A.19})$$

The following estimate is also proved in [32].

Theorem A.9 *There exist a positive real c and positive constants C_0, ζ_0 such that for k large enough we have,*

$$\left| \mathbb{P}\{P^-(B_k) > \zeta\} - \frac{e^{-\gamma}}{\log \zeta} \right| \leq \frac{C_0}{\log^2 \zeta} \quad (\zeta_0 \leq \zeta \leq k^{c/\log \log k}). \quad (\text{A.20})$$

As $\mathbb{P}\{B_n \text{ is prime}\} \leq \mathbb{P}\{P^-(B_n) > \zeta\}$, the following corollary is immediate.

Corollary A.10 *There exists an absolute constant C_1 , such that for all n large enough,*

$$\mathbb{P}\{B_n \text{ is prime}\} \leq C_1 \frac{\log \log n}{c \log n}.$$

The constant c is the same as in Theorem A.9.

By using Borel-Cantelli Lemma it follows that, along subsequences of integers growing at least exponentially, B_n is ultimately *not* prime with probability 1.

The collection of variables $\mathfrak{D} = \{\chi\{d|B_n\}, n \geq 1, d \geq 1\}$ further forms a mixing system, that is the correlation function

$$\Delta((d, n), (\delta, m)) = \mathbb{P}\{d|B_n, \delta|B_m\} - \mathbb{P}\{d|B_n\}\mathbb{P}\{\delta|B_m\}, \quad (\text{A.21})$$

verifies for each d and δ ,

$$\lim_{m \rightarrow \infty} \Delta((d, n), (\delta, m)) = 0. \quad (\text{A.22})$$

The second order theory of this system is thoroughly studied in [29]. Three zones of dependence, weak independence and strong independence can be identified, according to the cases $n < m \leq n + n^c$, $n + n^c \leq m \leq 2n$ and $m \geq 2n$, where $0 < c < 1$. The corresponding correlation estimates are established in [29].

References

1. T. M. Apostol. *Introduction to analytic number theory*. Springer-Verlag, New York-Heidelberg, (1976), Undergraduate Texts in Mathematics.
2. H. Cramér, Some theorems concerning prime numbers, *Acta Arith.* **2**, no. 1, (1936) 23–46.
3. H. Cramér, On the order of magnitude of the difference between consecutive prime numbers, *Arkiv för Mat. o. Fys.* **15**, no. 5, (1920) 1–32.
4. B. Davis and D. MacDonald, An elementary proof of the local central limit theorem, *J. Theoretical Prob.* **8** no. 3, (1995) 695–701.
5. H. M. Edwards, *Riemann's Zeta Function*. Academic Press, New York-London, (1974). Pure and Applied Mathematics, **58**. Reprinted by Dover Publications Inc., Mineola (2001)
6. W. J. Ellison, Differences between prime numbers, *Séminaire Delange-Pisot-Poitou. Théorie des nombres*, tome **15**, no 1 (1973-1974), exp. no 18, 1–10.



7. G. A. Freiman and J. Pitman, Partitions into distinct large parts, *J. Austral. Math. Soc. (Series A)* **57** (1994), 386–416.
8. R. Giuliano and M. Weber, Approximate local limit theorems with effective rate and application to random walks in random scenery, *Bernoulli* **23** (2017) (4B), 3268–3310.
9. A. Granville, Harald Cramér and the Distribution of Prime Numbers, *Scand. Actuarial J.* No. 1, (1995) 12–28.
10. D. R. Heath-Brown, The Differences Between Consecutive Primes. V, arXiv:1906.09555v1, (2019).
11. D. R. Heath-Brown, The number of primes in a short interval, *J. reine angew. Math.* **389** (1988), 22–63.
12. G. Hoheisel, Primzahlprobleme in der Analysis, *Sitz-Ber. preuss. Akad. Wiss.*, t. **33** (1930), 3–11.
13. M. N. Huxley, On the differences between consecutive primes, *Invent. Math.* **15** (1972), 164–170.
14. A. E. Ingham, On the estimation of $N(\sigma, T)$, *Quart. J. Math. (Oxford)* **11** (1940), 291–292.
15. A. E. Ingham, *Distribution of Prime Numbers* (C.U. Press)
16. A. E. Ingham, On the differences between consecutive primes, *Quart. J. Math. (Oxford)* **8** (1937), 255–266.
17. O. Kallenberg, (1997) *Foundations of modern probability theory*, Springer Verlag New-York.
18. E. Landau, Über einige Summen, die von den Nullstellender Riemann'schen Zetafunktion abhängen, *Acta math.* **35** (1912), 271–294.
19. C. MacDiarmid, (1998) Concentration, *Prob. Methods for Algorithmic Discrete Math.*, 195–248.
20. H. L. Montgomery and K. Soundarajan, (2004) Primes in short intervals, *Commun. Math. Phys.* **252**, 589–617.
21. T. R. Nicely, First occurrence prime gaps, web page: <http://www.trnicely.net/gaps/gaplist.html>
22. A. M. Odlyzko, (2000) An improved bound for the Bruijn-Newman constant, *Numerical Algorithms* **25** (1), 293–303.
23. V. V. Petrov, (1975) *Sums of Independent Random Variables*, Ergebnisse der Math. und ihre Grenzgebiete **82**, Springer.
24. J. Pintz, Cramér vs. Cramér, On Cramér's probabilistic model for primes, *Funct. Approx. Comment. Math.* **37**, part 2, (2007) 361–376.
25. I. Richards, (1980) On the normal density of primes in short intervals, *J. Number Th.* **12**, 378–384.
26. J. B. Rosser and L. Schoenfeld, (1962) Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6**(1), 64–94.
27. A. I. Sakhanenko, Estimates in the invariance principle in terms of truncated power moments, *Siberian Math. Journal* **47** no6 (2006), 1113–1127.
28. A. Selberg, On the normal density of primes in small intervals, and the difference between consecutive primes. *Arch. Math. Naturvid.*, **47** (1943), 87–105.
29. M. Weber, Distribution and Correlation Properties of Divisors in the Bernoulli Model for Integers, (2021), *Preprint*, 104 pages.
30. M. Weber, Instants of small amplitude of the Brownian motion and application to the Kubilius model, *Periodica Hungarica Math.* **67** (1), (2013), 95–113.
31. M. Weber, The law of the iterated logarithm for subsequences—Characterizations, *Nagoya Math. J.* **118** (1990), 65–97.
32. M. Weber, Small divisors of Bernoulli sums, *Indag. Math.* **18** No2 (2007), 281–293.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

