Electronic Journal of Probability

Electron. J. Probab. **18** (2013), no. 79, 1–21.
ISSN: 1083-6489 DOI: 10.1214/EJP.v18-2018

On the maximal length of arithmetic progressions*

Minzhi Zhao† Huizeng Zhang‡

Abstract

This paper is a continuation of Benjamini, Yadin and Zeitouni's paper [4] on maximal arithmetic progressions in random subsets. In this paper the asymptotic distributions of the maximal arithmetic progressions and arithmetic progressions modulo n relative to an independent Bernoulli sequence with parameter p are given. The errors are estimated by using the Chen-Stein method. Then the almost sure limit behaviour of these statistics is discussed. Our work extends the results in [4] and gives an affirmative answer to the conjecture raised at the end of that paper.

Keywords: arithmetic progression; Bernoulli sequence; limit distribution; Chen-Stein method.
AMS MSC 2010: 60F05, 60C05.
Submitted to EJP on May 10, 2012, final version accepted on Aug 4, 2013.
Supersedes arXiv:1204.1149.

1 Introduction and main results

As T. Tao stated in [15], long arithmetic progressions are very important in number theory. The first major result goes back to the work of van der Waerden in 1927. He proved that if the positive integers are divided into finitely many classes, then at least one of the classes contains arithmetic progressions of arbitrary length. In 1936, Erdös and Turán [7] conjectured that any subset of positive integers whose sum of reciprocals diverges must contain arbitrarily long arithmetic progressions. Roth [12] in 1953 proved that any subset with positive upper density contains an arithmetic progression of length three. Later in 1975, Szemerédi [13] established that such subset contains arbitrarily large arithmetic progressions. Recently, many authors have been interested in the arithmetic progressions in sumsets (see for instance [14]) and in random sets (see for instance [10]).

Although Bernoulli sequences are a rather simple probabilistic model, they are also a source of interesting discoveries. Limit behavior of the maximal length of runs in

*Support: Zhejiang Province education department scientific research project (Grant Y201225728), NSFC (Grants 11001070 and 11101113), ZJNSF (Grant R6090034).

†Zhejiang University, Hangzhou, Zhejiang, China.
E-mail: zhaomz@zju.edu.cn
‡Corresponding author. Hangzhou Normal University, Hangzhou, Zhejiang, China.
E-mail: zhanghz789@163.com

Bernoulli sequence have been studied for a long time by many authors, see e.g. [5], [6],[8], [9] and [11]. In 2007, Benjamini et al. [4] investigated the limit distribution of the maximal length of arithmetic progressions in Bernoulli sequence with $p = 1/2$. In this paper, we extend their result to general $0 < p < 1$, as well as the case when $p \to 0$ as $n \to \infty$, and prove a conjecture in their paper.

Suppose that $\xi_1, \xi_2, \dots$ is a Bernoulli sequence with $P(\xi_i = 1) = p = 1 - q$, where $0 < p < 1$. Let $\Sigma_n = \{1 \le i \le n : \xi_i = 1\}$ be the random subset of $\{1, \dots, n\}$ determined by $\xi_1, \dots, \xi_n$. For any $1 \le a, s \le n$, set

$$U_{a,s}^{(n)} = \max\Big\{1 \le m \le 1 + \Big[\frac{n-a}{s}\Big] : \xi_a = \xi_{a+s} = \cdots = \xi_{a+(m-1)s} = 1\Big\},$$

where $[x]$ denotes the integer part of x. Then $U_{a,s}^{(n)}$ stands for the maximal length of arithmetic progressions in Σ_n starting at a, with difference s. The maximal length of arithmetic progressions relative to $\xi_1, \dots, \xi_n$, denoted by $U^{(n)}$, is defined by

$$U^{(n)} = \max_{1 \le a,s \le n} U_{a,s}^{(n)}.$$

For any $1 \le a, s \le n$, the numbers

$$a, a + s \,(\mathrm{mod}\, n), \dots, a + \Big(\frac{n}{\gcd(s,n)} - 1\Big) s \,(\mathrm{mod}\, n)$$

are different while $a + \frac{n}{\gcd(s,n)} s \,(\mathrm{mod}\, n) = a$, where $\gcd(s,n)$ is the greatest common divisor of s and n. For convenience, when $n|k$, we write $k \,(\mathrm{mod}\, n) = n$. Define

$$W_{a,s}^{(n)} = \max\Big\{1 \le m \le \frac{n}{\gcd(s,n)} : \prod_{i=0}^{m-1} \xi_{a+is \,(\mathrm{mod}\, n)} = 1\Big\},$$

and

$$W^{(n)} = \max_{1 \le a,s \le n} W_{a,s}^{(n)}.$$

We call $W^{(n)}$ the maximal length of arithmetic progressions modulo n relative to $\xi_1, \dots, \xi_n$. Note that $U^{(n)}$ is increasing in n while $W^{(n)}$ is not.

In [4], the authors discussed the limit distribution of $U^{(n)}$ and $W^{(n)}$ when $p = 1/2$. We first extend their results to general p. Set $C = -2/\log p$. In [4], it was proved that as n tends to ∞, $\frac{U^{(n)}}{C \log n} \to 1$ and $\frac{W^{(n)}}{C \log n} \to 1$ in probability. As for almost sure limit, they proved that

$$\lim_{n \to \infty} \frac{U^{(n)}}{C \log n} = 1 \text{ a.s.}$$

and

$$1 = \liminf_{n \to \infty} \frac{W^{(n)}}{C \log n} < \frac{3}{2} \le \limsup_{n \to \infty} \frac{W^{(n)}}{C \log n} \text{ a.s.} \tag{1.1}$$

They conjectured at the end of [4] that

$$\limsup_{n \to \infty} \frac{W^{(n)}}{C \log n} = \frac{3}{2} \text{ a.s.} \tag{1.2}$$

In this paper, like in [4], we use the Chen-Stein method to study the asymptotic distributions of $U^{(n)}$ and $W^{(n)}$ but more carefully. For clarity, let us first introduce the concept of dependency graph (see [1] and [4]) and the Chen-Stein method that will be used in our proof.

Suppose that $\{X_i; i \in V\}$ is a family of random variables indexed by the vertices of a graph $G = (V, E)$. For convenience, we write $i \sim j$ if $(i, j) \in E$, that is if (i, j) is an edge. We call G a dependency graph of $\{X_i; i \in V\}$ if for any $i \in V$, X_i is in independent of $\{X_j; j \not\sim i\}$.

Stein's method was first proposed by Stein in 1972 for normal approximation. Chen, Barbour and others have this method adapted to approximate Poisson distribution (see for instance [3]). The following is a basic Poisson approximation theorem which was proved by Arratia et al. [2] in 1989. See also Theorem 3 of [4].

Theorem 1.1. *Suppose that $\{X_i; i \in V\}$ is a family of Bernoulli random variables with $\mathbb{E}X_i = p_i$ and suppose that $G = (V, E)$ is a dependency graph. Let $W = \sum_{i \in V} X_i$ and Z be a Poisson random variable with mean $\lambda = \sum_{i \in V} p_i$. If $0 < \lambda < \infty$, then*

$$\sup_{A \subseteq \mathbb{Z}} |\mathbb{P}(W \in A) - \mathbb{P}(Z \in A)| \le \frac{1 - e^{-\lambda}}{\lambda} \left[\sum_{i \in V} \sum_{j \sim i} p_i p_j + \sum_{i \in V} \sum_{j \sim i, j \neq i} \mathbb{E}(X_i X_j) \right].$$

To estimate the probability of $U^{(n)} < r$, we write $U^{(n)} \ge r$ as the union of $A_{a,s}$ with $(a, s) \in B_n$, where

$$A_{a,s} = \{\xi_a = \xi_{a+s} = \cdots = \xi_{a+(r-1)s} = 1\} \cap \{a - s \le 0 \text{ or } \xi_{a-s} = 0\}$$

and

$$B_n = \{(a, s) : 1 \le a, s \le n, a + (r - 1)s \le n\}.$$

It follows that

$$\mathbb{P}(U^{(n)} < r) = \mathbb{P}\Big(\sum_{(a,s) \in B_n} \mathbb{1}_{A_{a,s}} = 0 \Big),$$

where $\mathbb{1}_A$ denotes the indicator function of A. Find a dependency graph of the Bernoulli random variables $\{\mathbb{1}_{A_{a,s}} : (a, s) \in B_n\}$ and apply Theorem 1.1, then we will give an estimate of $\mathbb{P}(U^{(n)} < r)$. Similar arguments can be applied to $W^{(n)}$.

In this paper, we use Bachmann-Landau notation to describe the limiting behaviours of two functions. For any functions f and g, if $\lim_{n\to\infty} g(n)/f(n) = 0$, then we write $g(n) = o\big(f(n)\big)$. If $\limsup_{n\to\infty} |g(n)|/|f(n)| < \infty$, then we write $g(n) = O\big(f(n)\big)$. When $f(n) = O\big(g(n)\big)$, we also write $g(n) = \Omega\big(f(n)\big)$. The notation $g(n) = \Theta\big(f(n)\big)$ means that $0 < \liminf_{n\to\infty} |g(n)|/|f(n)| \le \limsup_{n\to\infty} |g(n)|/|f(n)| < \infty$. Set $D = -C/2 = 1/\log p$, $h'_n = C \log n$, $g_n = D \log\log n$ and $h_n = h'_n + g_n$. We have the following distribution approximations.

Theorem 1.2. 1. *For any $0 \le a < 1$, let $L_n = \{x : x \ge a g_n, h_n + x \in \mathbb{Z}\}$. Then we have*

$$\max_{x \in L_n} \Big| \exp\Big(\frac{-q \log p}{4} p^x \Big) \mathbb{P}(U^{(n)} < h_n + x) - 1 \Big| = O\Big(\frac{\log\log n}{\log^{1-a} n} \Big). \tag{1.3}$$

Hence, for any sequence $\{x_n\}$ with $x_n \in L_n$,

$$\lim_{n\to\infty} \exp\Big(\frac{-q \log p}{4} p^{x_n} \Big) \mathbb{P}(U^{(n)} < h_n + x_n) = 1. \tag{1.4}$$

2. *Let $L'_n = \{x : x \ge g_n, h'_n + x \in \mathbb{Z}\}$. Then we have*

$$\max_{x \in L'_n} \Big| \exp\Big(\frac{q}{2} p^x \Big) \mathbb{P}\Big(W^{(n)} < h'_n + x \Big) - 1 \Big| = O\Big(\frac{\log^7 n}{n^{1-\frac{q}{2}}} \Big). \tag{1.5}$$

Hence, for any sequence $\{x_n\}$ with $x_n \in L'_n$, we have

$$\lim_{n\to\infty} \exp\Big(\frac{q}{2} p^{x_n} \Big) \mathbb{P}(W^{(n)} < h'_n + x_n) = 1. \tag{1.6}$$

Note that if one replaces "$\leq$" by "$<$" in (1) of [4], and "$\log(2C\log N)$" by "$\log_2(2C\log N)$ " in (2) of [4], then the results match the corresponding equations (1.6) and (1.4) in our paper.

The following theorem gives the almost sure limits of $U^{(n)}$ and $W^{(n)}$.

Theorem 1.3. *1. As n tends to ∞,*

$$\frac{U^{(n)} - C\log n}{\log\log n} \to D \text{ in probability} \tag{1.7}$$

and

$$\frac{W^{(n)} - C\log n}{\log\log n} \to 0 \text{ in probability.} \tag{1.8}$$

2. For almost every ω,

$$D = \liminf_{n\to\infty} \frac{U^{(n)}(\omega) - C\log n}{\log\log n} < \limsup_{n\to\infty} \frac{U^{(n)}(\omega) - C\log n}{\log\log n} = 0. \tag{1.9}$$

3. The conjecture (1.2) holds and for almost every ω,

$$D = \liminf_{n\to\infty} \frac{W^{(n)}(\omega) - C\log n}{\log\log n} < \limsup_{n\to\infty} \frac{W^{(n)}(\omega) - C\log n}{\log\log n} = \infty. \tag{1.10}$$

Since $U^{(n)}$ is increasing, (1.9) also holds on the subsequence $\{2^n\}$. But (1.10) fails on the subsequence $\{2^n\}$ due to the fact that $W^{(n)}$ is not increasing. The following theorem states the behaviours on the subsequence $\{2^n\}$.

Theorem 1.4. *1. For almost every ω,*

$$0 = \liminf_{n\to\infty} \frac{W^{(2^n)}(\omega) - C\log 2^n}{\log\log 2^n} < \limsup_{n\to\infty} \frac{W^{(2^n)}(\omega) - C\log 2^n}{\log\log 2^n} = -D. \tag{1.11}$$

2. With probability one,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \frac{U^{(2^k)}(\omega) - C\log 2^k}{\log\log 2^k} = D \tag{1.12}$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \frac{W^{(2^k)}(\omega) - C\log 2^k}{\log\log 2^k} = 0. \tag{1.13}$$

Using the fact that $U^{(n)}$ is increasing, we deduce the following Corollary from (1.12).

Corollary 1.5. *For almost every ω,*

$$\lim_{n\to\infty} \frac{1}{\log n\log\log n}\left(\sum_{k=1}^{n} \frac{U^{(k)}(\omega)}{k} + D\log^2 n\right) = D. \tag{1.14}$$

Next, consider the case that the success probability is not necessarily the same. Suppose that $\xi_1^{(n)}, \dots, \xi_n^{(n)}$ are i.i.d. with

$$\mathbb{P}(\xi_i^{(n)} = 1) = p_n = 1 - \mathbb{P}(\xi_i^{(n)} = 0).$$

Use $U^{(n,p_n)}$ and $W^{(n,p_n)}$ to denote the the maximal length of arithmetic progressions or of arithmetic progressions modulo n relative to $\xi_1^{(n)}, \dots, \xi_n^{(n)}$ respectively. The following theorem states an interesting phenomenon that both statistics will eventually be concentrated on a few neighbouring integer values.

Theorem 1.6. *Assume that*

$$\lim_{n\to\infty} p_n = 0, \ \lim_{n\to\infty} np_n = \infty \ \textit{and} \ \lim_{n\to\infty} \frac{2\log n}{-\log p_n} = b \tag{1.15}$$

for some $2 \le b \le \infty$.

1. *If* $b = \infty$, *then*

$$\lim_{n\to\infty} \mathbb{P}\left(U^{(n,p_n)} \in \{k_n, k_n+1\}\right) = 1 \tag{1.16}$$

and

$$\lim_{n\to\infty} \mathbb{P}\left(W^{(n,p_n)} \in \{k'_n, k'_n \pm 1\}\right) = 1, \tag{1.17}$$

where $k_n = \left[\frac{-2\log n + \log\log n}{\log p_n}\right]$ *and* $k'_n = \left[\frac{2\log n}{-\log p_n}\right]$.

2. *If* $b = 2$, *or if* $2 < b < \infty$ *and* b *is not an integer, then*

$$\lim_{n\to\infty} \mathbb{P}(W^{(n,p_n)} = [b]) = \lim_{n\to\infty} \mathbb{P}(U^{(n,p_n)} = [b]) = 1. \tag{1.18}$$

3. *If* $b \ge 3$ *and* b *is an integer, then*

$$\lim_{n\to\infty} \mathbb{P}(W^{(n,p_n)} \in \{b, b-1\}) = \lim_{n\to\infty} \mathbb{P}(U^{(n,p_n)} \in \{b, b-1\}) = 1. \tag{1.19}$$

If in addition $u = \lim_{n\to\infty} n^2 p_n^b \le \infty$ *exists, then*

$$\lim_{n\to\infty} \mathbb{P}(U^{(n,p_n)} = b-1) = e^{-\frac{u}{2(b-1)}} = 1 - \lim_{n\to\infty} P(U^{(n,p_n)} = b) \tag{1.20}$$

and

$$\lim_{n\to\infty} \mathbb{P}(W^{(n,p_n)} = b-1) = e^{-\frac{u}{2}} = 1 - \lim_{n\to\infty} \mathbb{P}(W^{(n,p_n)} = b). \tag{1.21}$$

The rest of the paper is organized as follows. In §2 and §3, we discuss the asymptotic distributions of $U^{(n)}$ and $W^{(n)}$, respectively, and give the proof of Theorem 1.2. The proofs of Theorem 1.3, Theorem 1.4 and Theorem 1.6 are given in §4, §5 and §6 respectively. In §5, we also give the proof of Corollary 1.5.

2 The asymptotic distribution of $U^{(n)}$: Proof of Part 1 of Theorem 1.2

Suppose that $2 \le r \le n$. Let

$$\begin{aligned} B_n = B_n^{(r)} &= \{(a,s) : 1 \le a, s \le n, a + (r-1)s \le n\} \\ &= \left\{(a,s) : 1 \le s \le \left[\frac{n-1}{r-1}\right], 1 \le a \le n-(r-1)s\right\}. \end{aligned}$$

For any $(a,s) \in B_n$, let

$$A_{a,s} = A_{a,s}^{(r)} = \{\xi_a = \xi_{a+s} = \cdots = \xi_{a+(r-1)s} = 1\} \cap \{a - s \le 0 \text{ or } \xi_{a-s} = 0\}.$$

Then

$$\mathbb{P}(U^{(n)} \ge r) = \mathbb{P}\Big(\bigcup\nolimits_{(a,s)\in B_n} A_{a,s}\Big). \tag{2.1}$$

Set

$$B_{a,s} = B_{a,s}^{(r)} = \begin{cases} \{a, a+s, \ldots, a+(r-1)s\}, & \text{if } a \le s; \\ \{a-s, a, \ldots, a+(r-1)s\}, & \text{otherwise.} \end{cases}$$

Let G be the graph with vertex set B_n and edges defined by $(a,s) \sim (b,t)$ if and only if $B_{a,s} \cap B_{b,t} \neq \emptyset$. Then G is a dependency graph of $\{\mathbb{1}_{A_{a,s}} : (a,s) \in B_n\}$. Set $I = I_{n,r} = \sum_{(a,s)\in B_n} \mathbb{P}(A_{a,s})$ and

$$e^{(n,r)} = \sum_{(a,s)\in B_n} \sum_{(b,t)\sim(a,s)} \mathbb{P}(A_{a,s})\mathbb{P}(A_{b,t}) + \sum_{(a,s)\in B_n} \sum_{\substack{(b,t)\sim(a,s),\\ (b,t)\neq(a,s)}} \mathbb{P}(A_{a,s} \cap A_{b,t}).$$

Note that $\mathbb{P}(U^{(n)} < r) = \mathbb{P}\big(\sum_{(a,s)\in B_n} \mathbb{1}_{A_{a,s}} = 0\big)$. By Theorem 1.1, we have

$$\left|\mathbb{P}(U^{(n)} < r) - e^{-I}\right| \le e^{(n,r)}. \tag{2.2}$$

We shall estimate the value I and $e^{(n,r)}$.

Lemma 2.1. *For any* $0 < p < 1$ *and* $2 \le r \le n$*, we have*

$$p^r \frac{(n-r)^2}{2(r-1)} - p^{r+1}\frac{n^2}{2r} \le I \le p^r \frac{n^2}{2(r-1)} - p^{r+1}\frac{(n-r)^2}{2r}. \tag{2.3}$$

Proof. Clearly, $|B_n| = \frac{1}{2}[\frac{n-1}{r-1}]\left(2n - r + 1 - (r-1)[\frac{n-1}{r-1}]\right)$. It implies that

$$\frac{(n-r)^2}{2(r-1)} \le |B_n| \le \frac{n^2}{2(r-1)}. \tag{2.4}$$

Note that $B_n \cap \{(a,s) : a > s\} = \left\{(a,s) : 1 \le s \le \left[\frac{n-1}{r}\right], s < a \le n - (r-1)s\right\}$. We have

$$\frac{(n-r)^2}{2r} \le |B_n \cap \{(a,s) : a > s\}| \le \frac{n^2}{2r}. \tag{2.5}$$

Obviously, $I = p^r|B_n| - p^{r+1}|B_n \cap \{(a,s) : a > s\}|$. This, together with (2.4) and (2.5), gives (2.3). □

Lemma 2.2. *For any* $0 < p < 1$ *and* $2 \le r \le n$*, we have*

$$e^{(n,r)} \le 9\left(n^3 p^{2r-1} + n^2 r^3 p^{\frac{5}{3}r-1} + n^2 p^{\frac{3}{2}r-1}\right).$$

Proof. Let

$$c_1 = |\{(a,s,b,t) \in B_n \times B_n : (a,s) \sim (b,t)\}|,$$
$$c_2 = |\{(a,s,b,t) \in B_n \times B_n : |B_{a,s} \cap B_{b,t}| \ge 2\}|$$

and

$$c_3 = |\{(a,s,b,t) \in B_n \times B_n : (a,s) \sim (b,t), t = 2s \text{ or } s = 2t\}|.$$

Then

$$c_1 \le \left|\left\{(a,s,b,t) : (a,s) \in B_n, 1 \le t \le \left[\frac{n-1}{r-1}\right], b = a + is - jt, \right.\right.$$
$$\left.\left. -1 \le i,j \le r-1\right\}\right| \le 9n^3/2$$

and

$$c_3 \le 2\,|\{(a,s,b,t) : (a,s) \in B_n, t = 2s, b = a + is - jt, -1 \le i,j \le r-1\}|$$
$$= 2\,|\{(a,s,b) : (a,s) \in B_n, b = a + ks, -2r+1 \le k \le r+1\}| \le 7n^2.$$

Let us estimate c_2. Suppose that $|B_{a,s} \cap B_{b,t}| \ge 2$ and x_0 is the minimal number of the set $B_{a,s} \cap B_{b,t}$. Then $x_0 = a + is = b + jt$ for some $-1 \le i,j \le r-1$. If $x \in B_{a,s} \cap B_{b,t}$ and $x > x_0$, then $x = a + i's = b + j't$ for some $-1 \le i', j' \le r-1$. It follows that $x - x_0 = (i'-i)s = (j'-j)t$. Thus $t = sk_1/k_2$ for some $1 \le k_1, k_2 \le r$. In addition, there is a positive integer k such that $i' - i = kt_0, j' - j = ks_0$ and $x - x_0 = kst_0$, where $s_0 = s/\gcd(s,t)$ and $t_0 = t/\gcd(s,t)$. Since $i' - i \le r$, $k \le r/t_0$. Similarly, $k \le r/s_0$. Therefore

$$|B_{a,s} \cap B_{b,t}| \le r/\max(s_0,t_0) + 1.$$

Consequently, $|B_{a,s} \cap B_{b,t}| \le r/3 + 1$ whenever $\max(s_0,t_0) \ge 3$. When $\max(s_0,t_0) = 2$, $|B_{a,s} \cap B_{b,t}| \le r/2 + 1$. Actually, in this case, $s = 2t$ or $t = 2s$. When $\max(s_0,t_0) = 1$, $s = t$. We are now in a position to show that if $a \ne b$ and $B_{a,s} \cap B_{b,s} \ne \emptyset$, then $A_{a,s} \cap A_{b,s} = \emptyset$. Assume that $b > a$ without loss of generality. Since $B_{a,s} \cap B_{b,s} \ne \emptyset$, $a + is = b + js$ for some $-1 \le i,j \le r-1$ and hence $b = a + ks$ for some $1 \le k \le r$. Thus $A_{a,s} \subseteq \{\xi_{a+(k-1)s} = 1\}$ and $A_{b,s} \subseteq \{\xi_{b-s} = 0\} = \{\xi_{a+(k-1)s} = 0\}$. It implies that $A_{a,s} \cap A_{b,s} = \emptyset$ as desired. In view of the discussion above, we have

$$c_2 \le |\{(a,s,b,t) : (a,s) \in B_n, t = sk_1/k_2, b = a + is - jt,$$
$$-1 \le i,j \le r-1, 1 \le k_1,k_2 \le r\}| \le 9n^2r^3/4$$

and

$$e^{(n,r)} \le 2c_1p^{2r-1} + c_2p^{\frac{5r}{3}-1} + c_3p^{\frac{3r}{2}-1} \le 9\big(n^3p^{2r-1} + n^2r^3p^{\frac{5}{3}r-1} + n^2p^{\frac{3}{2}r-1}\big)$$

as desired. □

For any integer $2 \le r \le n$, let

$$\lambda_{n,r} = \lambda_{n,r,p} = \frac{n^2p^r(p+qr)}{2r(r-1)}. \tag{2.6}$$

Lemma 2.3. *For any $0 < p < 1$, we have*

$$\max_{2 \le r \le n} |\mathbb{P}(U^{(n)} < r) - e^{-\lambda_{n,r}}| = O\Big(\frac{\log^4 n \log\log n}{n}\Big). \tag{2.7}$$

Proof. Lemma 2.1 and (2.6) imply that

$$\big|e^{-I_{n,r}} - e^{-\lambda_{n,r}}\big| \le \big|I_{n,r} - \lambda_{n,r}\big| \le 2np^r. \tag{2.8}$$

Let $r_n = \Big[\frac{-2\log n}{\log p} + 2\frac{\log\log n}{\log p} + \frac{\log\log\log n}{2\log p}\Big]$ and $R_n = \Big[\frac{-3\log n}{\log p}\Big]$. By (2.2), (2.8) and Lemma 2.2,

$$\begin{aligned}
&\max\Big\{\Big|\mathbb{P}(U^{(n)} < r) - e^{-\lambda_{n,r}}\Big| : r_n \le r \le R_n\Big\} \\
&\le \max\Big\{\Big|\mathbb{P}(U^{(n)} < r) - e^{-I_{n,r}}\Big| + \big|e^{-I_{n,r}} - e^{-\lambda_{n,r}}\big| : r_n \le r \le R_n\Big\} \\
&\le \max\Big\{e^{(n,r)} + 2np^r : r_n \le r \le R_n\Big\} \\
&\le 9\big(n^3p^{2r_n-1} + n^2R_n^3p^{\frac{5}{3}r_n-1} + n^2p^{\frac{3}{2}r_n-1}\big) + 2np^{r_n} \\
&= O\Big(\frac{\log^4 n \log\log n}{n}\Big).
\end{aligned} \tag{2.9}$$

On the other hand, it is easy to check that $e^{-\lambda_{n,r_n}} = e^{-O(\log n\sqrt{\log\log n})} = o(n^{-1})$ and $1-e^{-\lambda_{n,R_n}} = 1-e^{-O(\frac{1}{n\log n})} = o(n^{-1})$. Note that $\mathbb{P}(U^{(n)} < r)$ and $e^{-\lambda_{n,r}}$ are both increasing functions of r. Hence when $r < r_n$,

$$\begin{aligned}|\mathbb{P}(U^{(n)} < r) - e^{-\lambda_{n,r}}| &\le \mathbb{P}(U^{(n)} < r) + e^{-\lambda_{n,r}} \le \mathbb{P}(U^{(n)} < r_n) + e^{-\lambda_{n,r_n}}\\ &\le |\mathbb{P}(U^{(n)} < r_n) - e^{-\lambda_{n,r_n}}| + 2e^{-\lambda_{n,r_n}} = O\Big(\frac{\log^4 n\log\log n}{n}\Big).\end{aligned}$$

Similarly, when $r > R_n$,

$$\begin{aligned}|\mathbb{P}(U^{(n)} < r) - e^{-\lambda_{n,r}}| &\le 1-\mathbb{P}(U^{(n)} < R_n) + 1 - e^{-\lambda_{n,R_n}}\\ &\le |\mathbb{P}(U^{(n)} < R_n) - e^{-\lambda_{n,R_n}}| + 2(1-e^{-\lambda_{n,R_n}}) = O\Big(\frac{\log^4 n\log\log n}{n}\Big).\end{aligned}$$

This completes the proof of our lemma. □

Proof of Part 1 of Theorem 1.2. Let $r = h_n + x$. For convenience, set

$$\epsilon_{n,x} = \Big|\frac{-q\log p}{4}p^x - \lambda_{n,r}\Big| = p^x\Big|\frac{-q\log p}{4} - \frac{q\log n}{2(r-1)} - \frac{p\log n}{2r(r-1)}\Big|.$$

Then

$$\begin{aligned}&\Big|\exp\Big(\frac{-q\log p}{4}p^x\Big)\mathbb{P}(U^{(n)} < h_n + x) - 1\Big|\\ &\quad\le \exp\Big(\frac{-q\log p}{4}p^x\Big)\Big|\mathbb{P}(U^{(n)} < r) - e^{-\lambda_{n,r}}\Big| + |\exp(\epsilon_{n,x}) - 1|.\end{aligned}\tag{2.10}$$

Clearly, $p^{ag_n} = \log^a n$. Since $a < 1$, we have

$$\exp\Big(\frac{-q\log p}{4}p^{ag_n}\Big) = \exp\Big(\frac{-q\log p}{4\log^{1-a} n}\log n\Big) = o(n^{1/3}).$$

Thus by Lemma 2.3, we have

$$\begin{aligned}&\max_{x\in L_n}\exp\Big(\frac{-q\log p}{4}p^x\Big)\Big|\mathbb{P}\Big(U^{(n)} < r\Big) - e^{-\lambda_{n,r}}\Big|\\ &\quad= o(n^{1/3})O(n^{-1}\log^4 n\log\log n) = o(n^{-1/2}).\end{aligned}\tag{2.11}$$

It is easy to verify that

$$\max_{x>-g_n}\epsilon_{n,x} = p^{-g_n}O(1) = O(\log^{-1} n)\tag{2.12}$$

and

$$\max_{ag_n\le x\le -g_n}\epsilon_{n,x} = p^{ag_n}O\Big(\frac{\log\log n}{\log n}\Big) = O\Big(\frac{\log\log n}{\log^{1-a} n}\Big).\tag{2.13}$$

Now (1.3) follows from (2.10)–(2.13). □

3 The asymptotic distribution of $W^{(n)}$: Proof of Part 2 of Theorem 1.2

Suppose that $2\le r\le n$. For any $1\le a,s\le n$, let

$$\tilde{A}_{a,s} = \tilde{A}_{a,s}^{(n,r)} = \Big\{\xi_a = 0, \prod_{i=1}^{r}\xi_{a+is\,(\mathrm{mod}\,n)} = 1\Big\}.\tag{3.1}$$

Let
$$\tilde{B}_n = \tilde{B}_n^{(r)} = \{(a,s) : 1 \le a \le n, 1 \le s \le [n/2], \gcd(n,s) < n/r\}$$
and
$$A_1 = \bigcup\nolimits_{(a,s)\in\tilde{B}_n} \tilde{A}_{a,s}. \tag{3.2}$$

Set $C_{a,s} = \{a, a+s\,(\mathrm{mod}\, n), a+2s\,(\mathrm{mod}\, n), \dots\}$. Then $C_{a,s} = \big\{a, a+s\,(\mathrm{mod}\, n), \dots, a + (n/\gcd(s,n)-1)\,s\,(\mathrm{mod}\, n)\big\}$ and $|C_{a,s}| = n/\gcd(s,n)$. Set
$$A_2 = \bigcup_{s|n, s\le n/r, 1\le a\le s} \{\xi_i = 1 : i \in C_{a,s}\}. \tag{3.3}$$

Lemma 3.1. *For any* $2 \le r \le n$*, we have*
$$\mathbb{P}(A_1) \le \mathbb{P}(W^{(n)} \ge r) \le \mathbb{P}(A_1) + \mathbb{P}(A_2). \tag{3.4}$$

Proof. Put
$$W_s^{(n)} = \max_{1\le i\le n} W_{i,s}^{(n)},$$
which stands for the maximal length of arithmetic progressions modulo n in Σ_n with difference s. For any $m \ge 0$, $\prod_{i=0}^m \xi_{a+is\,(\mathrm{mod}\, n)} = 1$ if and only if $\prod_{i=0}^m \xi_{b+i(n-s)\,(\mathrm{mod}\, n)} = 1$, where $b = a + ms\,(\mathrm{mod}\, n)$. In addition, $\gcd(s,n) = \gcd(n-s,n)$. Hence $W_s^{(n)} = W_{n-s}^{(n)}$ for all $1 \le s \le n$. Consequently,
$$W^{(n)} = \max_{1\le s\le n} W_s^{(n)} = \max_{1\le s\le [n/2]} W_s^{(n)}.$$

For any $1 \le a, b \le n$, $C_{a,s} \cap C_{b,s} = \emptyset$ or $C_{a,s} = C_{b,s}$. Furthermore, $C_{a,s} = C_{b,s}$ if and only if $b = a + k\cdot\gcd(s,n)$ for some integer k. Thus $\{1,\dots,n\}$ is the disjoint union of $C_{a,s}$ with $1 \le a \le \gcd(s,n)$. It follows that
$$W_s^{(n)} = \max_{1\le a\le \gcd(s,n)} \widetilde{W}_{a,s}^{(n)}.$$
where $\widetilde{W}_{a,s}^{(n)} = \max_{i\in C_{a,s}} W_{i,s}^{(n)}$. Note that $\{\widetilde{W}_{a,s}^{(n)} \ge r\} = \{\xi_i = 1 : i \in C_{a,s}\}$ when $n/\gcd(s,n) = r$, and
$$\Big\{\widetilde{W}_{a,s}^{(n)} \ge r\Big\} = \Big(\bigcup\nolimits_{i\in C_{a,s}} \tilde{A}_{i,s}\Big) \cup \{\xi_i = 1 : i \in C_{a,s}\}$$
provided $n/\gcd(s,n) > r$. We deduce that
$$\Big\{W^{(n)} \ge r\Big\} = \Big(\bigcup\nolimits_{(i,s)\in\tilde{B}_n} \tilde{A}_{i,s}\Big) \cup \Big(\bigcup\nolimits_{(a,s)\in B_n'} \{\xi_i = 1 : i \in C_{a,s}\}\Big),$$
where $B_n' = \{(a,s) : 1 \le a \le \gcd(s,n), 1 \le s \le [n/2], n \ge r\cdot\gcd(s,n)\}$. This, together with the fact that $C_{a,s} = C_{a,\gcd(s,n)}$, yields that
$$\{W^{(n)} \ge r\} = A_1 \cup A_2. \tag{3.5}$$
It implies (3.4) immediately. □

We next show that $\mathbb{P}(A_2)$ is small. Then by Lemma 3.1. $\mathbb{P}(W^{(n)} \ge r)$ is approximately equal to $\mathbb{P}(A_1)$. Set
$$\tilde{B}_{a,s} = \tilde{B}_{a,s}^{(n,r)} = \{a, a+s\,(\mathrm{mod}\, n), \dots, a+rs\,(\mathrm{mod}\, n)\}. \tag{3.6}$$

Define $\tilde{G}$ to be the graph with vertex set $\tilde{B}_n$ and edges defined by $(a,s)\sim(b,t)$ if and only if $\tilde{B}_{a,s}\cap\tilde{B}_{b,t}\neq\emptyset$. Then $\tilde{G}$ is a dependency graph of $\{\mathbb{1}_{\tilde{A}_{a,s}}:(a,s)\in\tilde{B}_n\}$. Put $\tilde{I}=\tilde{I}_{n,r}=\sum_{(a,s)\in\tilde{B}_n}\mathbb{P}(\tilde{A}_{a,s})$ and

$$\tilde{e}^{(n,r)}=\sum_{(a,s)\in\tilde{B}_n}\sum_{(b,t)\sim(a,s)}\mathbb{P}(\tilde{A}_{a,s})\mathbb{P}(\tilde{A}_{b,t})+\sum_{(a,s)\in\tilde{B}_n}\sum_{\substack{(b,t)\sim(a,s),\\(b,t)\neq(a,s)}}\mathbb{P}(\tilde{A}_{a,s}\cap\tilde{A}_{b,t}).$$

Then by Theorem 1.1, we have

$$|\mathbb{P}(A_1^c)-e^{-\tilde{I}}|\leq\tilde{e}^{(n,r)}. \tag{3.7}$$

The estimates of $\mathbb{P}(A_2)$, $\tilde{I}$ and $\tilde{e}^{(n,r)}$ are given in the following two lemmas.

Lemma 3.2. *For any* $0<p<1$ *and* $2\leq r\leq n$*, we have*

$$\mathbb{P}(A_2)\leq\frac{np^r}{qr} \tag{3.8}$$

and

$$\Big(1-\frac{(r+1)^2}{2n}\Big)\frac{qn^2p^r}{2}\leq\tilde{I}\leq\frac{qn^2p^r}{2}. \tag{3.9}$$

Proof. We see at once that

$$\mathbb{P}(A_2)\quad\leq\quad\sum_{s|n,s\leq n/r}sp^{\frac{n}{s}}\leq\sum_{i=r}^{n}\frac{n}{i}p^i\leq\frac{n}{r}\sum_{i=r}^{n}p^i\leq\frac{np^r}{qr}.$$

Clearly, $\big\{1\leq s\leq[\frac{n}{2}]:\gcd(n,s)\geq\frac{n}{r}\big\}\subseteq\big\{s=\frac{n}{i}j:2\leq i\leq r,1\leq j\leq[\frac{i}{2}]\big\}$. It implies that

$$\frac{n^2}{2}\geq|\tilde{B}_n|\geq n\Big(\Big[\frac{n}{2}\Big]-\sum_{i=2}^{r}\Big[\frac{i}{2}\Big]\Big)\geq\frac{n^2}{2}\Big(1-\frac{(r+1)^2}{2n}\Big).$$

Now the fact that $\tilde{I}=|\tilde{B}_n|qp^r$ yields (3.9) immediately. □

Lemma 3.3. *For any* $0<p<1$ *and* $2\leq r\leq n$*, we have*

$$\tilde{e}^{(n,r)}\leq 4\big(n^3r^2p^{2r-1}+n^2r^5p^{\frac{3}{2}r-1}+nr^6p^r\big).$$

Proof. Let $H=\big\{(a,s,b,t)\in\tilde{B}_n\times\tilde{B}_n:(a,s)\sim(b,t)\big\}$, $\tilde{c}_1=|H|$,

$$\tilde{c}_2=\Big|\big\{(a,s,b,t)\in H:|\tilde{B}_{a,s}\cap\tilde{B}_{b,t}|\geq 2\big\}\Big|$$

and

$$\tilde{c}_3=\Big|\big\{(a,s,b,t)\in H:(a,s)\neq(b,t),|\tilde{B}_{a,s}\cap\tilde{B}_{b,t}|>r/2+1,\tilde{A}_{a,s}\cap\tilde{A}_{b,t}\neq\emptyset\big\}\Big|.$$

Then

$$\begin{aligned}\tilde{c}_1\leq&\Big|\big\{(a,s,b,t):(a,s)\in\tilde{B}_n,1\leq t\leq[n/2],b=a+is-jt\,(\mathrm{mod}\,n),\\&0\leq i,j\leq r\big\}\Big|\leq n^3r^2.\end{aligned} \tag{3.10}$$

Suppose that $|\tilde{B}_{a,s}\cap\tilde{B}_{b,t}|\geq 2$. Then there are $0\leq j_1<j_2\leq r$ and $x,y\in\tilde{B}_{a,s}$ such that $b+j_1t\,(\mathrm{mod}\,n)=x$ and $b+j_2t\,(\mathrm{mod}\,n)=y$. Hence $(j_2-j_1)t-kn=y-x$ for some $0\leq k\leq j_2-j_1$ and $b=x-j_1t\,(\mathrm{mod}\,n)$. Therefore

$$\begin{aligned}\tilde{c}_2\leq&\Big|\big\{(a,s,b,t):(a,s)\in\tilde{B}_n,t=(kn+y-x)/i,b=x-jt\,(\mathrm{mod}\,n),\\&1\leq i\leq r,0\leq j,k\leq r,x,y\in\tilde{B}_{a,s}\big\}\Big|\leq 3n^2r^5.\end{aligned} \tag{3.11}$$

To estimate $\tilde{c}_3$, we first prove that if $a \neq b$ and $(a,s) \sim (b,s)$, then $\tilde{A}_{a,s} \cap \tilde{A}_{b,s} = \emptyset$. Set

$$j^* = \min\big\{0 \le j \le r : b + js \,(\mathrm{mod}\, n) \in \tilde{B}_{a,s}\big\}.$$

Then there is $0 \le i \le r$ such that $b+j^*s \equiv a+is \,(\mathrm{mod}\, n)$. It follows that $b+(j^*-1)s \equiv a+(i-1)s \,(\mathrm{mod}\, n)$. By the definition of j^*, $j^*=0$ or $i=0$. If $j^*=0$, then $b = a+is \,(\mathrm{mod}\, n)$. But $i \neq 0$ since $a \neq b$. This gives $\tilde{A}_{a,s} \cap \tilde{A}_{b,s} \subseteq \{\xi_{a+is\,(\mathrm{mod}\, n)} = 1, \xi_b = 0\} = \{\xi_b = 1, \xi_b = 0\} = \emptyset$. Similarly, $\tilde{A}_{a,s} \cap \tilde{A}_{b,s} = \emptyset$ when $i = 0$.

Next suppose that $s \neq t$, $|\tilde{B}_{a,s} \cap \tilde{B}_{b,t}| > r/2+1$ and $|\tilde{B}_{a,s} \cap \tilde{B}_{b,t}| = \{a+i_0 s \,(\mathrm{mod}\, n), \ldots, a+i_k s \,(\mathrm{mod}\, n)\}$ with $0 \le i_0 < i_1 < \cdots < i_k \le r$. Then there is l such that $i_{l+1} - i_l = 1$. It follows that there are $j_1 \neq j_2$ such that $0 \le j_1, j_2 \le r$, $a + i_l s \equiv b + j_1 \,(\mathrm{mod}\, n)$ and $a + i_{l+1} s \equiv b + j_2 \,(\mathrm{mod}\, n)$. Accordingly, $s = it \,(\mathrm{mod}\, n)$ with $i = j_2 - j_1$. If $i = 1$, then $s = t$. If $i = -1$, then $s = n - t$ and hence $s = t = n/2$ by the fact that $1 \le s, t \le n/2$. The contradiction shows that $1 < |i| \le r$. Similarly, $t = js \,(\mathrm{mod}\, n)$ for some $1 < |j| \le r$. It follows that $s = ijs \,(\mathrm{mod}\, n)$, that is $(ij-1)s = vn$ for some $|v| \le r^2/2$. Consequently,

$$\begin{aligned}\tilde{c}_3 \le \Big|\big\{&(a,s,b,t) : s = vn/(ij-1), t = js \,(\mathrm{mod}\, n), b = a + ls - mt \,(\mathrm{mod}\, n),\\ &1 \le a \le n, 1 < |i|, |j| \le r, |v| \le r^2/2, 0 \le l, m \le r\big\}\Big| \le 4nr^6. \end{aligned} \tag{3.12}$$

Thus our result holds by noting that $\tilde{e}^{(n,r)} \le 2\tilde{c}_1 p^{2r-1} + \tilde{c}_2 p^{\frac{3r}{2}-1} + \tilde{c}_3 p^r$. □

For any integer $2 \le r \le n$, let

$$\mu_{n,r} = \mu_{n,r,p} = qn^2 p^r/2. \tag{3.13}$$

Lemma 3.4. *For any* $0 < p < 1$*, we have*

$$\max_{2 \le r \le n} |\mathbb{P}(W^{(n)} < r) - e^{-\mu_{n,r}}| = O\Big(\frac{\log^7 n}{n}\Big). \tag{3.14}$$

Proof. Combining (3.13) with (3.9) gives that

$$|e^{-\tilde{I}_{n,r}} - e^{-\mu_{n,r}}| \le |\tilde{I}_{n,r} - \mu_{n,r}| \le qnp^r r^2. \tag{3.15}$$

Set $r_n = \big[\frac{-2\log n}{\log p} + \frac{\log\log n}{\log p} + \frac{\log 2 - \log q}{\log p} - 1\big]$ and $R_n = \big[\frac{-3\log n}{\log p}\big]$. Lemma 3.3, together with (3.4), (3.7), (3.8) and (3.15), yields that

$$\begin{aligned}&\max\Big\{|\mathbb{P}(W^{(n)} < r) - e^{-\mu_{n,r}}| : r_n \le r \le R_n\Big\}\\ &\le \max\big\{|\mathbb{P}(W^{(n)} < r) - e^{-\tilde{I}_{n,r}}| + |e^{-\tilde{I}_{n,r}} - e^{-\mu_{n,r}}| : r_n \le r \le R_n\big\}\\ &\le \max\big\{\tilde{e}^{(n,r)} + np^r/(qr) + qnp^r r^2 : r_n \le r \le R_n\big\}\\ &= O(n^{-1}\log^7 n). \end{aligned} \tag{3.16}$$

Furthermore, it is easy to check that $e^{-\mu_{n,r_n}} = o(n^{-1})$ and $1 - e^{-\mu_{n,R_n}} = O(n^{-1})$. Therefore (3.14) holds by noting that $\mathbb{P}(W^{(n)} < r)$ and $e^{-\mu_{n,r}}$ are both increasing functions of r. □

Proof of Part 2 of Theorem 1.2. Let $r = h'_n + x$. Then $\mu_{n,r} = qp^x/2$. We conclude from (3.14) that

$$\begin{aligned}&\max_{x \in L'_n} \big|\exp(qp^x/2)\mathbb{P}(W^{(n)} < h'_n + x) - 1\big|\\ &\le \exp\Big(\frac{qp^{g_n}}{2}\Big) O\Big(\frac{\log^7 n}{n}\Big) = O\Big(\frac{\log^7 n}{n^{1-\frac{q}{2}}}\Big)\end{aligned}$$

and completes our proof. □

4 Almost sure limits: Proof of Theorem 1.3

We first list two estimates that will be used in the proof of (1.10) and (1.11). Analysis similar to that in the proof of Lemma 3.3 shows that for any m, n,

$$\Big|\Big\{(a,s,b,t) \in \tilde{B}_m^{(r_m)} \times \tilde{B}_n^{(r_n)} : |\tilde{B}_{a,s}^{(m,r_m)} \cap \tilde{B}_{b,t}^{(n,r_n)}| \geq 1\Big\}\Big| \leq m^2 n r_m r_n \tag{4.1}$$

and

$$\Big|\Big\{(a,s,b,t) \in \tilde{B}_m^{(r_m)} \times \tilde{B}_n^{(r_n)} : |\tilde{B}_{a,s}^{(m,r_m)} \cap \tilde{B}_{b,t}^{(n,r_n)}| \geq 2\Big\}\Big| \leq 3m^2 r_m^2 r_n^3. \tag{4.2}$$

Proof of Part 1 of Theorem 1.3. By (2.1), we have

$$\mathbb{P}(U^{(n)} \geq r) \leq I. \tag{4.3}$$

For any $\varepsilon > 0$, (4.3) and (2.3) imply that

$$\mathbb{P}(U^{(n)} \geq C \log n + (1-\varepsilon) D \log\log n) = O\big(\log^{-\varepsilon} n\big). \tag{4.4}$$

On the other hand, by (2.6) and (2.7),

$$\mathbb{P}(U^{(n)} < C \log n + (1+\varepsilon) D \log\log n) = e^{-\Theta(\log^{\varepsilon} n)} + O\Big(\frac{\log^4 n \log\log n}{n}\Big). \tag{4.5}$$

Hence (1.7) holds.

In view of (3.14), we have

$$\mathbb{P}(W^{(n)} < C \log n + \varepsilon D \log\log n) = e^{-\Theta(\log^{\varepsilon} n)} + O\Big(\frac{\log^7 n}{n}\Big) \to \begin{cases} 1, & \varepsilon < 0; \\ 0, & \varepsilon > 0. \end{cases}$$

From this, (1.8) follows immediately.

Part 2. By (4.5), $\sum_{k=1}^{\infty} \mathbb{P}(U^{(2^k)} < C \log 2^k + (1+\varepsilon) D \log\log 2^k) < \infty$ for any $\varepsilon > 0$. One then deduces from the Borel-Cantelli Lemma that

$$\mathbb{P}(U^{(2^k)} < C \log 2^k + (1+\varepsilon) D \log\log 2^k \quad i.o.) = 0.$$

It follows that for almost every ω, there is $K(\omega)$ such that for $k \geq K(\omega)$,

$$U^{(2^k)}(\omega) \geq C \log 2^k + (1+\varepsilon) D \log\log 2^k. \tag{4.6}$$

If $n > 2^{K(\omega)}$, then $2^k \leq n < 2^{k+1}$ for some $k \geq K(\omega)$. Hence $U^{(2^k)}(\omega) \leq U^{(n)}(\omega) \leq U^{(2^{k+1})}(\omega)$. This, together with (4.6), gives that

$$U^{(n)}(\omega) \geq C \log n - C \log 2 + (1+\varepsilon) D \log\log n.$$

Since $\varepsilon > 0$ was chosen arbitrarily, we have that

$$\liminf_{n\to\infty} \frac{U^{(n)}(\omega) - C \log n}{\log\log n} = D,$$

by considering (1.7).

Let T_k be the maximal length of arithmetic progressions relative to $\xi_{2^{k-1}}, \ldots, \xi_{2^k - 1}$. Then $T_1, T_2, \ldots$ are independent. In addition, T_k have the same distribution as $U^{(2^{k-1})}$. By (2.7),

$$\mathbb{P}(T_k \geq C \log 2^k) = \mathbb{P}(U^{(2^{k-1})} \geq C \log 2^k) \geq 1 - e^{-\lambda_{n,r}} - O\Big(\frac{\log^4 n \log\log n}{n}\Big)$$

with $n = 2^{k-1}$ and $r = [C \log 2^k] + 1$. Note that $\lambda_{n,r} = \Theta(1/\log n)$. Hence

$$\mathbb{P}(T_k \ge C \log 2^k) = \Theta(1/\log n) = \Theta(1/k).$$

It follows that $\sum_{k=1}^{\infty} \mathbb{P}(T_k \ge C \log 2^k) = \infty$. By the Borel-Cantelli Lemma, $\mathbb{P}(T_k \ge C \log 2^k \ \ i.o.) = 1$. Consequently,

$$\limsup_{k\to\infty} \frac{U^{(2^k)} - C \log 2^k}{\log\log 2^k} \ge 0 \tag{4.7}$$

by noting that $U^{(2^k)} \ge T_k$. On the other hand, (4.4) yields that

$$\sum_{k=1}^{\infty} \mathbb{P}(U^{(2^k)} \ge C \log 2^k + (1-\varepsilon) D \log\log 2^k) < \infty$$

whenever $\varepsilon > 1$. Hence

$$\limsup_{k\to\infty} \frac{U^{(2^k)} - C \log 2^k}{\log\log 2^k} \le 0. \tag{4.8}$$

Combining (4.7) with (4.8) we conclude that

$$\limsup_{n\to\infty} \frac{U^{(n)} - C \log n}{\log\log n} = 0$$

by the fact that $U^{(n)}$ is increasing. This completes the proof of (1.9).

Part 3. In view of (1.1), to prove (1.2), it remains to prove that

$$\limsup_{n\to\infty} \frac{W^{(n)}}{C \log n} \le \frac{3}{2} \ a.s.$$

We conclude from (3.2) that

$$\mathbb{P}(A_1) \le \tilde{I}. \tag{4.9}$$

For any $\varepsilon > 0$, by (4.9), (3.8) and (3.9),

$$\mathbb{P}(W^{(n)} > (1+\varepsilon) C \log n) = O(n^{-2\varepsilon}) + O\Big(\frac{n^{-1-2\varepsilon}}{\log n}\Big). \tag{4.10}$$

Hence

$$\sum_{n=1}^{\infty} \mathbb{P}(W^{(n)} > (1+\varepsilon) C \log n) < \infty$$

whenever $\varepsilon > \frac{1}{2}$. Therefore, $\limsup_{n\to\infty} \frac{W^{(n)}}{C \log n} \le \frac{3}{2} \ \ a.s.$ as desired.

By (1.2), (1.9) and the fact that $W^{(n)} \ge U^{(n)}$, to prove (1.10), we only need to show that

$$\liminf_{n\to\infty} \frac{W^{(n)} - C \log n}{\log\log n} \le D \ \ a.s. \tag{4.11}$$

Fix any $0 < \varepsilon < 1$. Let $r_n = [C \log n + \varepsilon D \log\log n]$, $H_n = \{W^{(n)} < r_n\}$ and $X_k = \sum_{n=k+1}^{2k} \mathbb{1}_{H_n}$. Then

$$\mathbb{P}\Big(\bigcup_{n=k+1}^{2k} H_n\Big) = \mathbb{P}(X_k > 0) \ge \frac{(\mathbb{E}X_k)^2}{\mathbb{E}X_k^2} = \frac{\big[\sum_{n=k+1}^{2k} \mathbb{P}(H_n)\big]^2}{\sum_{m,n=k+1}^{2n} \mathbb{P}(H_m H_n)}. \tag{4.12}$$

Obviously, $\mu_{n,r_n} = O(\log^\varepsilon n) = o(\log n/4)$. Together with (3.14), it implies that

$$\sum_{n=k+1}^{2k} \mathbb{P}(H_n) \geq \sum_{n=k+1}^{2k} e^{-\mu_{n,r_n}} - O(\log^7 k) = \Omega(k^{3/4}). \tag{4.13}$$

Set $E_n = \{(a,s) : 1 \leq a \leq n, [\frac{3n}{C\log n}] + 1 \leq s \leq [\frac{n}{2}], \gcd(n,s) < \frac{n}{r_n}\}$. Combining (3.2) with (3.5) yields that

$$\mathbb{P}\big[(H_m H_n)^c\big] \geq \mathbb{P}\Big[\Big(\bigcup\nolimits_{(a,s)\in E_m} \tilde{A}_{a,s}^{(m,r_m)}\Big) \cup \Big(\bigcup\nolimits_{(b,t)\in E_n} \tilde{A}_{b,t}^{(n,r_n)}\Big)\Big].$$

Let V be the graph with vertex set $V_k = \{(a,s,n) : k+1 \leq n \leq 2k, (a,s) \in E_n\}$ and edges defined by $(a,s,m) \sim (b,t,n)$ if and only if $\tilde{B}_{a,s}^{(m,r_m)} \cap \tilde{B}_{b,t}^{(n,r_n)} \neq \emptyset$. Then V is a dependency graph of $\{\mathbb{1}_{\tilde{A}_{a,s}^{(n,r_n)}} : (a,s,n) \in V_k\}$. Write $J_m = \sum_{(a,s)\in E_m} \mathbb{P}(\tilde{A}_{a,s}^{(m,r_m)})$. By using the Stein's method, we have

$$\mathbb{P}(H_m H_n) \leq e^{-J_m - J_n} + \tilde{e}^{(m,r_m)} + \tilde{e}^{(n,r_n)} + 2\tilde{e}^{(m,r_m,n,r_n)}, \tag{4.14}$$

where

$$\tilde{e}^{(m,r_m,n,r_n)} = \sum_{(a,s,m)\sim(b,t,n)} \big[\mathbb{P}\big(\tilde{A}_{a,s}^{(m,r_m)}\big)\mathbb{P}\big(\tilde{A}_{b,t}^{(n,r_n)}\big) + \mathbb{P}\big(\tilde{A}_{a,s}^{(m,r_m)} \cap \tilde{A}_{b,t}^{(n,r_n)}\big)\big].$$

It follows that

$$\sum_{m,n=k+1}^{2k} \mathbb{P}(H_m H_n) \leq \Big(\sum_{n=k+1}^{2k} e^{-J_n}\Big)^2 + \mathcal{L}_k, \tag{4.15}$$

where $\mathcal{L}_k = 2k\sum_{n=k+1}^{2k} \tilde{e}^{(n,r_n)} + 2\sum_{m,n=k+1}^{2k} \tilde{e}^{(m,r_m,n,r_n)}$. One deduces from (3.9) that $J_n \geq \mu_{n,r_n}\Big(1 - \frac{(r_n+1)^2}{2n} - \frac{6}{C\log n}\Big)$. Hence

$$\sum_{n=k+1}^{2k} e^{-J_n} = e^{O(\log^{\varepsilon-1} k)} \sum_{n=k+1}^{2k} e^{-\mu_{n,r_n}} \tag{4.16}$$

by noting that $\mu_{n,r_n} = O(\log^\varepsilon n)$. If we have showed that $\mathcal{L}_k = O(k\log^8 k)$, then $\lim_{k\to\infty} \mathbb{P}(\bigcup_{n=k+1}^{2k} H_n) = 1$ and hence (4.11) holds, in view of (4.12)–(4.16).

We are now in a position to show that $\mathcal{L}_k = O(k\log^8 k)$. By Lemma 3.3,

$$k\sum_{n=k+1}^{2k} \tilde{e}^{(n,r_n)} = O(k\log^7 k).$$

Fix any $(a,s,m) \in V_k$, define $\gamma(a,s,m)$ to be the set of all triples $(b,t,n) \in V_k$ such that $|\tilde{B}_{b,t}^{(n,r_n)} \cap \tilde{B}_{a,s}^{(m,r_m)}| > C\log(2k)/2$. Suppose that k is sufficiently large and $(b,t,n) \in \gamma(a,s,m)$. Let $i_l = \min\{j \geq 0 : b + jt > ln\}$, $v = \max\{l : i_l \leq r_m\}$ and $Z_l = \{b + it - ln : i_l \leq i \leq \min(i_{l+1} - 1, r_n)\}$. Then $\tilde{B}_{b,t}^{(n,r_n)}$ is the disjoint union of Z_i with $0 \leq i \leq v$. Since $[\frac{3n}{C\log n}] + 1 \leq t \leq [\frac{n}{2}]$, $|Z_i| \leq C\log n/3 + 1$ for all $i \leq v$, and $|Z_i| \geq 2$ for $0 < i < v$. Thus $|\tilde{B}_{b,t}^{(n,r_n)}| = r_n + 1 \geq 2(v-1)$. It follows that $v < C\log(2k)/2 - 3$. Since $|\tilde{B}_{b,t}^{(n,r_n)} \cap \tilde{B}_{a,s}^{(m,r_m)}| > C\log(2k)/2$ and $|Z_i| \leq C\log n/3 + 1$, there are l and w such that $w \neq l$, $|Z_l \cap \tilde{B}_{a,s}^{(m,r_m)}| \geq 2$ and $|Z_w \cap \tilde{B}_{a,s}^{(m,r_m)}| \geq 1$. That is to say, there are $0 \leq i, j, \ell \leq r_n$

and $x, y, z \in \tilde{B}_{a,s,m}$ such that $b + it - ln = x$, $b + jt - ln = y$ and $b + \ell t - wn = z$. This leads to $(j - i)t = y - x$ and $(w - l)n = (\ell - i)t + x - z$. Therefore

$$|\gamma(a,s,m)| \leq \Big|\big\{(b,t,n) : t = (y-x)/i, n = (\ell t + x - z)/j, b + wt \,(\mathrm{mod}\, n) = x,\\ |i|, |j|, |\ell|, |w| \leq C \log 2k, i, j \neq 0, x, y, z \in \tilde{B}_{a,s,m}\big\}\Big| = O\big(\log^7 k\big).$$

Combining (4.1) with (4.2), one then deduces that

$$\sum_{m,n=k+1}^{2k} \tilde{e}^{(m,r_m,n,r_n)} \leq \sum_{m,n=k+1}^{2k} \Big(2p^{r_m+r_n-1} m^2 n r_m r_n + 3p^{r_m+r_n-C\log(2k)/2} m^2 r_m^2 r_n^3\Big)\\ + \sum_{(a,s,m)\in V_k} p^{r_m} |\gamma(a,s,m)| = O\big(k \log^8 k\big)$$

as desired. This completes the proof of (1.10). □

5 Behaviour of certain subsequences: Proof of Theorem 1.4

We first give some estimates that will be used in the proof of Theorem 1.4. Analysis similar to that in the proof of Lemma 2.2 shows that for any m, n and $2 \leq r_m \leq r_n$, with $H = \{(a,s,b,t) : (a,s) \in B_m^{(r_m)}, (b,t) \in B_n^{(r_n)}\}$,

$$|\{(a,s,b,t) \in H : |B_{a,s}^{(r_m)} \cap B_{b,t}^{(r_n)}| \geq 1\}| \leq 9m^2 n/2 \tag{5.1}$$

and

$$|\{(a,s,b,t) \in H : |B_{a,s}^{(r_m)} \cap B_{b,t}^{(r_n)}| \geq 2\}| \leq 9m^2 r_m r_n^2/4. \tag{5.2}$$

We can also show that if $|B_{a,s}^{(r_m)} \cap B_{b,t}^{(r_n)}| > r_n/2+1$ and $A_{a,s}^{(r_m)} \cap A_{b,t}^{(r_n)} \neq \emptyset$, then $(b,t) = (a,s)$. Therefore when $2 \leq r_m \leq r_n$, we have

$$\Big|\big\{(a,s,b,t) \in H : |B_{a,s}^{(r_m)} \cap B_{b,t}^{(r_n)}| > r_n/2 + 1, A_{a,s}^{(r_m)} \cap A_{b,t}^{(r_n)} \neq \emptyset\big\}\Big| \leq m^2/r_m. \tag{5.3}$$

Lemma 5.1. *Suppose that* $(a,s) \in \tilde{B}_m^{(r_m)}$ *and* $(b,t) \in \tilde{B}_n^{(r_n)}$. *If* $n \geq 2m$, $r_n \geq 36$ *and* $r_n t > 3n$, *then* $\big|\tilde{B}_{a,s}^{(m,r_m)} \cap \tilde{B}_{b,t}^{(n,r_n)}\big| < 3r_n/4$.

Proof. If $1 \leq x \leq m$, $A = \{x, x+t, \ldots, x+kt\} \subseteq \{1, \ldots, n\}$ and $x + (k+1)t > n$, then $k \geq 1$ and $x + t(k+1)/2 > n/2 \geq m$. Thus $|A \cap \{1, \ldots, m\}| \leq (k+1)/2$ when k is odd, or $|A \cap \{1, \ldots, m\}| \leq k/2 + 1$ when k is even. Hence $|A \cap \{1, \ldots, m\}|/|A| \leq 2/3$.

Since $t > 3n/r_n$, there are $h \geq 3$ and $0 \leq i_1 < \cdots < i_h < r_n$ such that $b + (i_1+1)t > n \geq b + i_1 t, \ldots, b + (i_h+1)t > hn \geq b + i_h t$ and $b + r_n t \leq (h+1)n$. Set $i_0 = -1$, $i_{h+1} = r_n$ and

$$H_j = \big\{b + (i_j+1)t \,(\mathrm{mod}\, n), b + (i_j+2)t \,(\mathrm{mod}\, n), \ldots, b + i_{j+1}t \,(\mathrm{mod}\, n)\big\}.$$

According to the above discussion, we have $|H_j \cap \{1, \ldots, m\}| \leq 2|H_j|/3$ provided $0 \leq j \leq h-1$, and $v = |H_h \cap \{1, \ldots, m\}| \leq m/t + 1$. Clearly, $\tilde{B}_{b,t}^{(n,r_n)} = \bigcup_{j=0}^{h} H_j$ and $\tilde{B}_{a,s}^{(m,r_m)} \subseteq \{1, \ldots, m\}$. It implies that

$$\frac{\big|\tilde{B}_{a,s}^{(m,r_m)} \cap \tilde{B}_{b,t}^{(n,r_n)}\big|}{r_n} \leq \frac{\frac{2}{3}\sum_{j=0}^{h-1} |H_j| + v}{r_n} \leq \frac{2(i_h + 1 + v)}{3r_n} + \frac{v}{3r_n}\\ \leq \frac{2}{3} + \frac{m}{3r_n t} + \frac{1}{r_n} < \frac{3}{4}$$

as claimed. □

Before come to the proof of Theorem 1.4, we give some equivalent statements of (1.12) and (1.13). The proof of these equivalent statements will use the following Lemma.

Lemma 5.2. *Suppose that*

$$b_n > 0,\ \sum_n b_n = \infty,\ \lim_{n\to\infty} \frac{\sum_{k=1}^n b_k}{n \max_{1\le k\le n} b_k} = 1 \tag{5.4}$$

and

$$a \le \liminf_{n\to\infty} \frac{a_n}{b_n} \le \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty. \tag{5.5}$$

Then

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n a_k}{\sum_{k=1}^n b_k} = a \tag{5.6}$$

if and only if for any $c > a$*,*

$$\lim_{n\to\infty} \frac{|\{1 \le k \le n : a_k/b_k < c\}|}{n} = 1. \tag{5.7}$$

Proof. We first prove the sufficiency. By (5.5), there is d such that $a_k/b_k < d$ for all k. For any $c > a$, let $A_n = \{1 \le k \le n : a_k/b_k \ge c\}$. Then (5.7) implies that $\lim_{n\to\infty} |A_n|/n = 0$. Hence

$$\frac{\sum_{k=1}^n a_k}{\sum_{k=1}^n b_k} \le c + \frac{(d-c)\sum_{k\in A_n} b_k}{\sum_{k=1}^n b_k} \le c + \frac{(d-c)|A_n| \max_{1\le k\le n} b_k}{\sum_{k=1}^n b_k} \to c.$$

Since this is true for any $c > a$, we have that $\limsup_{n\to\infty} \sum_{k=1}^n a_k/\sum_{k=1}^n b_k \le a$. This, together with (5.4) and (5.5), gives (5.6).

We next show the necessity. By (5.5), for any $\varepsilon > 0$, there is K such that $a_k/b_k > a-\varepsilon$ for all $k \ge K$. For any $c > a$, let $B_n = \{K < k \le n : a_k/b_k < c\}$. Then for $n > K$,

$$\begin{aligned}\frac{\sum_{k=1}^n a_k}{\sum_{k=1}^n b_k} &\ge c + \frac{\sum_{k=1}^K (a_k - cb_k) + (a-\varepsilon-c)\sum_{k\in B_n} b_k}{\sum_{k=1}^n b_k} \\ &\ge c + \frac{\sum_{k=1}^K (a_k - cb_k) + (a-\varepsilon-c)|B_n| \max_{1\le k\le n} b_k}{\sum_{k=1}^n b_k}.\end{aligned}$$

Letting $n \to \infty$, we get that

$$\liminf_{n\to\infty} \frac{|B_n| \max_{1\le k\le n} b_k}{\sum_{k=1}^n b_k} \ge \frac{c-a}{c+\varepsilon-a}.$$

The arbitrary of $\varepsilon > 0$, together with (5.4), implies (5.7) and completes our proof. ☐

Suppose that (5.5) holds. In addition, suppose that $b_n c_n > 0$, $\sum_n b_n c_n = \infty$ and $\lim_{n\to\infty} \sum_{k=1}^n b_k c_k/(n \max_{1\le k\le n} b_k c_k) = 1$. Lemma 5.2 shows that $\lim_{n\to\infty} \frac{\sum_{k=1}^n a_k c_k}{\sum_{k=1}^n b_k c_k} = a$ if and only if (5.7) holds for all $c > a$. Particularly, by letting $c_n = 1/b_n$, we see that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \frac{a_k}{b_k} = a$ if and only if (5.7) holds for all $c > a$. Thus if (5.4) and (5.5) hold, then $\lim_{n\to\infty} \sum_{k=1}^n a_k/\sum_{k=1}^n b_k = a$ if and only if $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \frac{a_k}{b_k} = a$, and also if and only if (5.7) holds for all $c > a$.

Proposition 5.3. *If (1.9) holds, then (1.12) holds if and only if*

$$\lim_{n\to\infty} \frac{1}{n\log n}\left(\sum_{k=1}^{n} U^{(2^k)}(\omega) + \frac{\log 2}{\log p} n^2\right) = D, \tag{5.8}$$

and also if and only if for any $1 > \varepsilon > 0$*,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\left\{U^{(2^k)} < C\log 2^k + D(1-\varepsilon)\log\log 2^k\right\}}(\omega) = 1. \tag{5.9}$$

Proof. Choose $a_n = U^{(2^n)}(\omega) - C\log 2^n$ and $b_n = \log\log\max(2^n, 4)$. Then (5.5) holds with $a = D$. It is easy to check that

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} U^{(2^k)}(\omega) + \frac{\log 2}{\log p} n^2 + \frac{\log 2}{\log p} n$$

and

$$\sum_{k=1}^{n} b_k = \sum_{k=1}^{n} \log k + n\log\log 2 + \log 2.$$

Since $n\log n - n \le \sum_{k=1}^{n}\log k \le n\log n$, $\lim_{n\to\infty}\sum_{k=1}^{n} b_k/(n\log n) = 1$ and (5.4) holds. We also conclude that $\lim_{n\to\infty}\sum_{k=1}^{n} a_k/\sum_{k=1}^{n} b_k = D$ if and only if (5.8) holds. By Lemma 5.2, our result holds. □

The similar conclusion can be drawn for the sequence $\{W^{(2^k)}(\omega)\}$.

Proposition 5.4. *If (1.11) holds, then (1.13) holds if and only if*

$$\lim_{n\to\infty} \frac{1}{n\log n}\left(\sum_{k=1}^{n} W^{(2^k)}(\omega) + \frac{\log 2}{\log p} n^2\right) = 0, \tag{5.10}$$

and also if and only if for any $\varepsilon > 0$*,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\left\{W^{(2^k)} < C\log 2^k - \varepsilon D\log\log 2^k\right\}}(\omega) = 1. \tag{5.11}$$

Proof of Part 1 of Theorem 1.4. By (3.14), for any $0 < \varepsilon < 1$,

$$\mathbb{P}\left(W^{(2^n)} < C\log 2^n + \varepsilon D\log\log 2^n\right) = e^{-\Theta(n^{\varepsilon})} + O(2^{-n}n^7)$$

and

$$\mathbb{P}\left(W^{(2^n)} > C\log 2^n - (1+\varepsilon)D\log\log 2^n\right) = \Theta\big(n^{-(1+\varepsilon)}\big).$$

By the Borel-Cantelli Lemma,

$$\liminf_{n\to\infty} \frac{W^{(2^n)} - C\log 2^n}{\log\log 2^n} \ge 0 \text{ and } \limsup_{n\to\infty} \frac{W^{(2^n)} - C\log 2^n}{\log\log 2^n} \le -D.$$

In view of (1.8), it remains to show that

$$\limsup_{n\to\infty} \frac{W^{(2^n)} - C\log 2^n}{\log\log 2^n} \ge -D. \tag{5.12}$$

For any $0 < \varepsilon < 1$, let $r_n = [C\log n - \varepsilon D\log\log n]$ and

$$F_n = \{(a, s) : 1 \le a \le n, 3n/r_n < s \le [n/2], \gcd(n, s) < n/r_n\}.$$

To show (5.12), it suffices to show that $\lim_{n\to\infty}\mathbb{P}\big(\bigcup_{n=k+1}^{2k}\{W^{(2^n)}\geq r_{2^n}\}\big)=1$. What is left is to show that $\lim_{n\to\infty}\big(\bigcup_{(a,s,m)\in G_k}\tilde{A}_{a,s}^{(m,r_m)}\big)=1$, where $G_k=\{(a,s,2^n):k+1\leq n\leq 2k,(a,s)\in F_{2^n}\}$. Let $X_k=\sum_{(a,s,m)\in G_k}\mathbb{1}_{\tilde{A}_{a,s}^{(m,r_m)}}$. It is easy to verify that $\mathbb{E}X_k=\Theta(k^{1-\varepsilon})$. By Lemma 3.3, Lemma 5.1, (4.1) and (4.2), we have

$$\begin{aligned}\mathrm{Var}(X_k)\leq&\sum_{(a,s,m)\in G_k}\mathbb{P}\big(\tilde{A}_{a,s}^{(m,r_m)}\big)+2\sum_{n=k+1}^{2k}\tilde{e}^{(2^n,r_{2^n})}\\&+2\sum_{k+1\leq m<n\leq 2k}\big(p^{r_{2^m}+r_{2^n}-1}2^{2m}2^n r_{2^m}r_{2^n}+3p^{r_{2^m}+\frac{1}{4}r_{2^n}}2^{2m}r_{2^m}^2r_{2^n}^3\big)\\=&O(k^{1-\varepsilon}).\end{aligned}$$

Consequently,

$$\mathbb{P}\Big(\bigcup\nolimits_{(a,s,m)\in G_k}\tilde{A}_{a,s}^{(m,r_m)}\Big)=\mathbb{P}(X_k>0)\geq\frac{(\mathbb{E}X_k)^2}{\mathrm{Var}(X_k)+(\mathbb{E}X_k)^2}\to 1,$$

and complete the proof of (1.11).

Part 2. We now come to prove (1.12). By Proposition 5.3, we only need to show (5.9). Let $r_k=\big[C\log 2^k+D(1-\varepsilon)\log\log 2^k\big]$, $V_n=\big\{(a,s,k):1\leq k\leq n,(a,s)\in B_{2^k}^{(r_k)}\big\}$ and $\Lambda(n)=\sum_{(a,s,k)\in V_n}\mathbb{1}_{A_{a,s}^{(r_k)}}$. Then it suffices to show that $\lim_{n\to\infty}\Lambda(n)/n=0$ a.s. Clearly, $\mathbb{E}\Lambda(n)=\sum_{k=1}^n I_{2^k,r_k}=O(n^{1-\varepsilon})$ and $\mathrm{Var}[\Lambda(n)]$ is less than the sum of $p^{r_k+r_m-|B_{a,s}^{(r_k)}\cap B_{b,t}^{(r_m)}|}$ with $(a,s,k),(b,t,m)\in V_n$, $B_{a,s}^{(r_k)}\cap B_{b,t}^{(r_m)}\neq\emptyset$ and $A_{a,s}^{(r_k)}\cap A_{b,t}^{(r_m)}\neq\emptyset$. By (5.1)–(5.3),

$$\begin{aligned}\mathrm{Var}[\Lambda(n)]=&\sum_{1\leq i\leq j\leq n}O\big(2^{-j}ij+2^{-j}i^2j^3+2^{2i-2j}j^{1-\varepsilon}i^{-1}\big)\\=&O\Big(\sum_{j=1}^{\infty}2^{-j}j^6\Big)+O\Big(\sum_{j=1}^{n}j^{-\varepsilon}\sum_{k=0}^{j-1}2^{-2k}j(j-k)^{-1}\Big)\\=&O(1)+\sum_{k=0}^{\infty}2^{-2k}(k+1)O\Big(\sum_{j=1}^{n}j^{-\varepsilon}\Big)=O(n^{1-\varepsilon}).\end{aligned}$$

Then by the Tchebychev inequality, for any $\delta>0$,

$$\sum_{n=1}^{\infty}\mathbb{P}(|\Lambda_n/n-\mathbb{E}(\Lambda_n/n)|>\delta)=\sum_{n=1}^{\infty}O\Big(\frac{1}{n^{1+\varepsilon}\delta^2}\Big)<\infty.$$

The Borel-Cantelli Lemma yields that $\Lambda_n/n\to 0$ $a.s.$. Hence (5.9) holds as desired.

As to (1.13), by Proposition 5.4, we need only to show (5.11). The proof is completed by showing that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{(a,s)\in F_{2^k}}\mathbb{1}_{\tilde{A}_{a,s}^{(2^k,r_{2^k})}}=0\ a.s.\tag{5.13}$$

and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{H_k}=0\ a.s.,\tag{5.14}$$

where $H_k=\{W^{(2^k)}\geq r_{2^k}\}\setminus\big(\bigcup_{(a,s)\in F_{2^k}}\tilde{A}_{a,s}^{(2^k,r_{2^k})}\big)$. Analysis similar to that in the proof (5.9) shows (5.13). Note that $\mathbb{P}(H_k)=O(k^{-1-\varepsilon})$. By the Borel-Cantelli Lemma, $\lim_{k\to\infty}\mathbb{1}_{H_k}=0$ $a.s.$ which gives (5.14) and completes our proof. □

Proof of Corollary 1.5. By Proposition 5.3, (5.8) holds. Let $c_n = [\log n/\log 2]$. Then $2^{c_n} \le n < 2^{c_n+1}$. For any integers $1 \le a \le b$,

$$\log \frac{b+1}{a} = \int_a^{b+1} \frac{1}{x}\,dx \le \sum_{i=a}^{b} \frac{1}{i} \le \int_{a-1}^{b} \frac{1}{x}\,dx = \log \frac{b}{a-1}.$$

Thus

$$\sum_{k=1}^{n} \frac{U^{(k)}}{k} \ge \sum_{i=0}^{c_n-1} U^{(2^i)} \sum_{j=2^i}^{2^{i+1}-1} \frac{1}{j} \ge \log 2 \sum_{i=0}^{c_n-1} U^{(2^i)}$$

and

$$\sum_{k=2}^{n} \frac{U^{(k)}}{k} \le \sum_{i=0}^{c_n} U^{(2^{i+1})} \sum_{j=2^i+1}^{2^{i+1}} \frac{1}{j} \le \log 2 \sum_{i=1}^{c_n+1} U^{(2^i)}.$$

We conclude from (5.8) that

$$\lim_{n\to\infty} \frac{1}{c_n \log c_n} \left(\frac{1}{\log 2} \sum_{k=1}^{n} \frac{U^{(k)}}{k} + \frac{\log 2}{\log p} c_n^2 \right) = D \ a.s.$$

It follows (1.14) immediately. □

6 The asymptotic distributions when $p_n = o(1)$: Proof of Theorem 1.6

Proof of Part 1 of Theorem 1.6. Set $q_n = 1 - p_n$. Clearly, $p_n^{k_n+1} \le n^{-2}\log n \le p_n^{k_n}$. Similar to (2.9) shows that there is a constant $c > 0$ such that

$$\max_{0\le r\le 2} \left| \mathbb{P}\left(U^{(n,p_n)} < k_n + r\right) - e^{-\lambda_{n,k_n+r,p_n}} \right| \le c p_n^{-3} n^{-1} \log^2 n. \tag{6.1}$$

Since $\lim_{n\to\infty} \frac{2\log n}{-\log p_n} = \infty$, $\frac{\log n}{-\log p_n} \ge 10$ and hence $p_n^{-1} \le n^{0.1}$ for sufficiently large n. This, together with (6.1), implies that

$$\lim_{n\to\infty} \max_{0\le r\le 2} \left| \mathbb{P}\left(U^{(n,p_n)} < k_n + r\right) - e^{-\lambda_{n,k_n+r,p_n}} \right| = 0. \tag{6.2}$$

In addition,

$$\lim_{n\to\infty} \lambda_{n,k_n,p_n} \ge \lim_{n\to\infty} \frac{q_n \log n}{2(k_n-1)} = \lim_{n\to\infty} \frac{-\log p_n}{4} = \infty \tag{6.3}$$

by noting that $p_n \to 0$. Similarly,

$$\lim_{n\to\infty} \lambda_{n,k_n+2,p_n} \le \lim_{n\to\infty} \left(\frac{p_n^2 \log^2 p_n}{8\log n} - \frac{p_n \log p_n}{4} \right) = 0. \tag{6.4}$$

Therefore (1.16) holds by (6.2)–(6.4). In the same manner we can prove (1.17).

Part 2 and Part 3. Since $\lim_{n\to\infty} np_n = \infty$, $\lim_{n\to\infty} \mathbb{P}\big(\sum_{i=1}^{n} \xi_i^{(n)} \ge 2\big) = 1$ and hence

$$\lim_{n\to\infty} \mathbb{P}\left(U^{(n,p_n)} \ge 2\right) = 1. \tag{6.5}$$

Choose an $0 < \varepsilon < 0.1$ such that $[b-\varepsilon, b) \cup (b, b+\varepsilon)$ contains no integers. There is N such that $(b-\varepsilon)/2 < -\log n/\log p_n < (b+\varepsilon)/2$ for all $n > N$. It follows that for all $n > N$,

$$p_n^{\frac{-b+\varepsilon}{2}} < n < p_n^{\frac{-b-\varepsilon}{2}}. \tag{6.6}$$

By (4.9), (6.6) and Lemma 3.2, when $n > N$, we have

$$\mathbb{P}\left(W^{(n,p_n)} \geq [b]+1\right) \leq \frac{p_n^{[b]+1-b/2-\varepsilon/2}}{q_n\left([b]+1\right)} + \frac{q_n p_n^{[b]+1-b-\varepsilon}}{2} \to 0. \tag{6.7}$$

Suppose that $r < b \leq r+1$ where r is a positive integer. Applying (6.6) gives that $n^2 p_n^r \geq p_n^{r-(b-\varepsilon)} \to \infty$. Let $X_n = \sum_{(a,s)\in B_n^{(r)}} \mathbb{1}_{A_{a,s}^{(r)}}$. By Lemma 2.1 and Lemma 2.2, $\mathbb{E}X_n = \Theta(n^2 p_n^r)$ and $\mathrm{Var}(X_n) \leq \mathbb{E}X_n + O(n^3 p_n^{2r-1}) = o\big((\mathbb{E}X_n)^2\big)$. Consequently,

$$\mathbb{P}(U^{(n,p_n)} \geq r) = \mathbb{P}(X_n > 0) \geq \frac{(\mathbb{E}X_n)^2}{\mathrm{Var}(X_n) + (\mathbb{E}X_n)^2} \to 1. \tag{6.8}$$

Similar to (2.9) and (3.16), by using (6.6), we can show that

$$\lim_{n\to\infty} |\mathbb{P}(U^{(n,p_n)} < b) - e^{-\lambda_{n,b,p_n}}| = \lim_{n\to\infty} |\mathbb{P}(W^{(n,p_n)} < b) - e^{-\mu_{n,b,p_n}}| = 0 \tag{6.9}$$

when b is an integer satisfying $b \geq 3$. Furthermore, if $u = \lim_{n\to\infty} n^2 p_n^b \leq \infty$ exists, then

$$\lim_{n\to\infty} \mathbb{P}(U^{((n,p_n)} < b) = \lim_{n\to\infty} e^{-\lambda_{n,b,p_n}} = e^{-\frac{u}{2(b-1)}} \tag{6.10}$$

and

$$\lim_{n\to\infty} \mathbb{P}(W^{(n,p_n)} < b) = \lim_{n\to\infty} e^{-\mu_{n,b,p_n}} = e^{-\frac{u}{2}}. \tag{6.11}$$

Thus our result holds by (6.5),(6.7),(6.8),(6.10), (6.11) and by noting that $W^{(n,p_n)} \geq U^{(n,p_n)}$. □

References

[1] Alon, N. and Spencer, J. H.: The probabilistic method. *John Wiley & Sons*, New York, 2008. xviii+352 pp. MR-2437651

[2] Arratia, R., Goldstein, L. and Gordon, L.: Two moments suffice for poisson approximations: the chen-stein method. *Ann. Probab.* **17**, (1989), no. 1, 9–25. MR-0972770

[3] Barbour, A. D. and Chen, L.: An introduction to Stein's method. *Lecture Notes Series* **4**, Institute for Mathematical Sciences, National University of Singapore, 2005. xii+225 pp. MR-2235447

[4] Benjamini, I., Yadin, A. and Zeitouni, O.: Maximal arithmetic progressions in random subsets. *Elec. Comm. in Probab.* **12**, (2007), 365–376. MR-2350574

[5] Erdös, P. and Rényi, A.: On a new law of large numbers. *J. Analyse Math.* **23**, (1970), 103–111. MR-0272026

[6] Erdös, P. and Revesz, P.: On the length of the longest head-run. *Colloq. Math. Soc. Janos Bolyai* **16**, (1977), 219–228. MR-0478304

[7] Erdös, P. and Turán, P.: On some sequences of integers. *J. London Math. Soc.* **11**, (1936), 261–264. MR-1574918

[8] Gordon, L., Schilling, M. F. and Waterman, M. S.: An extreme value theory for long head runs. *Probability Theory and Related Fields* **72**, (1986), no. 2, 279–287. MR-0836278

[9] József, T.: Limit theorems for the longest run. *Annales Mathematicae et Informaticae* **36**, (2009), 133–141. MR-2580909

[10] Kohayakawa,Y., Łuczak, T. and Rödl, V.: Arithmetic progressions of length three in subsets of a random set. *Acta Arith.* **75**, (1996), no. 2, 133–163. MR-1379396

[11] Móri, T. F.: The a.s. limit distribution of the longest head run. *Can. J. Math.* **45**, (1993), no. 6, 1245–1262. MR-1247545

[12] Roth, K. F.: On certain sets of integers. *J. London Math. Soc.* **28**, (1953), 104–109. MR-0051853

[13] Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27**, (1975), 199–245. MR-0369312

[14] Szemerédi, E. and Vu, V. H.: Finite and infinite arithmetic progressions in sumsets. *Annals of Mathematics* **163** , (2006), no. 1, 1–35. MR-2195131

[15] Tao, T.: What is good mathematics? *Bulletin of the American Mathematical Society* **44**, (2007), no. 4, 623-634. MR-2338369

Acknowledgments. The authors wish to express their thanks to the referee for his/her constructive suggestions and many helpful comments which have led to much improvement on the paper. The authors would like to thank Prof. Jiangang Ying and Prof. Ofer Zeitouni for their useful suggestions. They also thank Ruinan Lu for critical reading of the manuscript.