A note on the length of maximal arithmetic progressions in random subsets

Yilun Shang

Institute for Cyber Security, University of Texas at San Antonio

San Antonio, TX 78249, USA

email: shylmath@hotmail.com

Abstract

Let $U^{(n)}$ denote the maximal length arithmetic progression in a non-uniform random subset of $\{0,1\}^n$, where 1 appears with probability p_n. By using dependency graph and Stein-Chen method, we show that $U^{(n)} - c_n \ln n$ converges in law to an extreme type distribution with $\ln p_n = -2/c_n$. Similar result holds for $W^{(n)}$, the maximal length aperiodic arithmetic progression (mod n).

MSC 2000: 60C05, 11B25.

Keywords: arithmetic progression, random subset, Stein-Chen method.

1 Introduction

An arithmetic progression is a sequence of numbers such that the difference of any two successive members of the sequence is a constant. A celebrated result of Szemerédi [5] says that any subset of integers of positive upper density contains arbitrarily long arithmetic progressions. The recent work [6] reviews some extremal problems closely related with arithmetic progressions and prime sequences, under the name of the Erdös-Turán conjectures, which are known to be notoriously difficult to solve.

Let $\xi_1, \xi_2, \cdots, \xi_n$ be a uniformly chosen random word in $\{0,1\}^n$ and Ξ_n be the random set consisting elements i such that $\xi_i = 1$. Benjamini et al. [3] studies the length of maximal arithmetic progressions in Ξ_n. Denote by $U^{(n)}$ the maximal length arithmetic progression in Ξ_n and $W^{(n)}$ the maximal length aperiodic arithmetic progression (mod n) in Ξ_n. They show, among others, that the expectation of $U^{(n)}$ and $W^{(n)}$ is roughly $2\ln n/\ln 2$.

In view of the random graph theory [4], a natural extension of [3] is to consider non-uniform random subset of $\{0,1\}^n$, which is the main interest of this note. Let $\xi_i = 1$ with probability p_n and $\xi_i = 0$ with probability $1 - p_n$, where $p_n \in [0,1]$ is a function of n. Following [3], the key to our work is to construct proper dependency graph and apply the Stein-Chen method of Poisson approximation (see e.g. [4, 1]). Our result implies that, in the non-uniform scenarios, the expectation of $U^{(n)}$ and $W^{(n)}$ is roughly $c_n \ln n$, with $\ln p_n = -2/c_n$. Obviously, taking $p_n \equiv 1/2$ and $c_n \equiv 2/\ln 2$, we then recover the main result of Benjamini et al.

The rest of the note is organized as follows. We present the main results in Section 2. Section 3 is devoted to the proofs.

2 Results

Let $\xi_1, \xi_2, \cdots$ be i.i.d. random variables with $P(\xi_i = 1) = p_n$ and $P(\xi_i = 0) = 1 - p_n$. For integers $1 \le s, t \le n$, define

$$W_{s,t}^{(n)} := \max\left\{1 \le k \le n : \xi_s = 0, \prod_{i=1}^{k} \xi_{s+it(\mod n)} = 1\right\}. \tag{1}$$

Therefore, $W_{s,t}^{(n)}$ is the length of the longest arithmetic progression ($\mod n$) in $\{1, 2, \cdots, n\}$ starting at s with difference t. Moreover, set $W^{(n)} = \max_{1\le s,t\le n} W_{s,t}^{(n)}$. Similarly, define

$$U_{s,t}^{(n)} := \max\left\{1 \le k \le \left\lfloor \frac{n-s}{t} \right\rfloor : \xi_s = 0, \prod_{i=1}^{k} \xi_{s+it} = 1\right\}, \tag{2}$$

and $U^{(n)} = \max_{1\le s,t\le n} U_{s,t}^{(n)}$, where $\lfloor a \rfloor$ is the integer part of a.

Theorem 1. *Suppose that* $\ln p_n = -2/c_n$ *and* $\alpha < c_n = o(\ln n)$ *for some* $\alpha > 0$. *Let* $\{x_n\}$ *be a sequence such that* $c_n \ln n + x_n \in \mathbb{Z}$ *for all* n, *and* $\inf_n x_n \ge \beta$ *for some* $\beta \in \mathbb{R}$. *We have*

$$\lim_{n\to\infty} e^{\lambda(x_n)} P(W^{(n)} \le c_n \ln n + x_n) = 1, \tag{3}$$

where $\lambda(x) = p_n^{x+2}$. *In particular,* $W^{(n)}/c_n \ln n$ *converges to* 1 *in probability, as* $n \to \infty$.

Theorem 2. *Suppose that* $\ln p_n = -2/c_n$ *and* $\alpha < c_n = o(\ln n)$ *for some* $\alpha > 0$. *Let* $\{y_n\}$ *be a sequence such that* $c_n \ln n - \ln(2c_n \ln n) + y_n \in \mathbb{Z}$ *for all* n, *and* $\inf_n y_n \ge \beta$ *for some* $\beta \in \mathbb{R}$. *We have*

$$\lim_{n\to\infty} e^{\lambda(y_n)} P(U^{(n)} \le c_n \ln n - \ln(2c_n \ln n) + y_n) = 1, \tag{4}$$

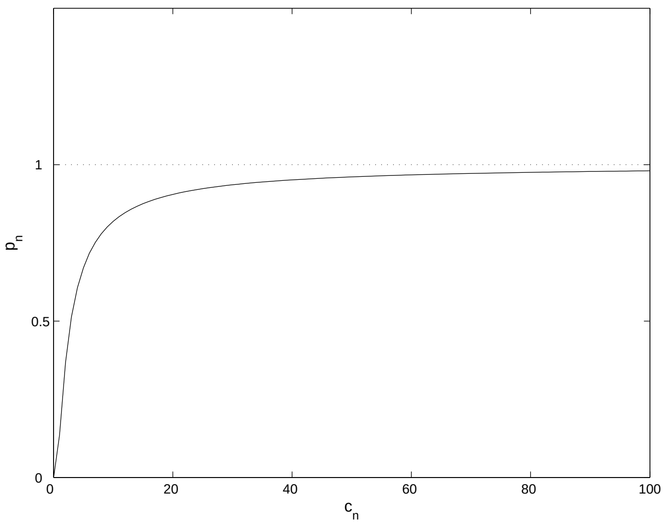

Figure 1: The probability p_n versus c_n.

where $\lambda(x) = p_n^{x+2}$. *In particular,* $U^{(n)}/c_n \ln n$ *converges to* 1 *in probability, as* $n \to \infty$.

The relationship between p_n and c_n is depicted in Fig. 1. We observe that the probability p_n, by our assumptions, should within the regime $e^{-2/\alpha} < p_n = e^{-2/o(\ln n)}$ for $\alpha > 0$. For the case $p_n = o(1)$ (i.e., $c_n = o(1)$), by letting $\alpha \to 0$, we can infer that $W^{(n)} \ll \ln n$ and $U^{(n)} \ll \ln n$ whp.

3 Proofs

In this section, we will only consider Theorem 1 since the proofs are very similar. Theorem 1 will be proved through a series of lemmas by similar reasoning to [3] with some modifications.

For a collection of random variables $\{X_i\}_{i=1}^n$, a graph G of order n is called a dependency graph [4] of $\{X_i\}_{i=1}^n$ if for any vertex i, X_i is independent of the set $\{X_j :$ vertices i and j are not adjacent$\}$. The following is a result of Arratia et al. [2], which is a instrumental version of the Stein-Chen method in numerous probabilistic combinatorial problems [1].

Lemma 1.([2]) *Let* $\{X_i\}_{i=1}^n$ *be* n *Bernoulli random variables with* $EX_i = p_i > 0$. *Let* G *be a dependency graph of* $\{X_i\}_{i=1}^n$. *Set* $S_n = \sum_{i=1}^n X_i$ *and* $\lambda = ES_n = \sum_{i=1}^n p_i$. *Define*

$$B_1(G) = \sum_{i=1}^n \sum_{j:j\sim i} EX_i EX_j \tag{5}$$

and

$$B_2(G) = \sum_{i=1}^{n} \sum_{j \neq i: j \sim i} E(X_i X_j). \tag{6}$$

Let Z be a Poisson random variable with $EZ = \lambda$. For any $A \subset \mathbb{N}$, we have

$$|P(S_n \in A) - P(Z \in A)| \leq B_1(G) + B_2(G). \tag{7}$$

Fix $\varepsilon > 0$ and set $m = \lfloor (c_n + \varepsilon) \ln n \rfloor$. Define the truncated version

$$W_{s,t}'^{(n)} := \max \left\{ 1 \leq k \leq m : \xi_s = 0, \prod_{i=1}^{k} \xi_{s+it (\mod n)} = 1 \right\} \tag{8}$$

and $W'^{(n)} = \max_{1 \leq s,t \leq n} W_{s,t}'^{(n)}$. For $x \in \mathbb{R}$ define the indicator variable

$$I_{s,t}(x) = 1_{\{W_{s,t}'^{(n)} > c_n \ln n + x\}} \quad \text{and} \quad S(x) = \sum_{1 \leq s,t \leq n} I_{s,t}(x). \tag{9}$$

By definition, it is clear that $W'^{(n)} > c_n \ln n + x$ if and only if $S(x) > 0$. Set $A(s,t) = \{s+it\}_{i=0}^{m}$. Fix $x \in \mathbb{R}$ such that $x < \varepsilon \ln n$. Hence, as in [3], we can construct a dependency graph G of random variables $\{I_{s,t}(x)\}_{s,t=1}^{n}$ by setting the vertex set $\{(s,t)\}_{s,t=1}^{n}$ and edges $(s_1,t_1) \sim (s_2,t_2)$ if and only if $A(s_1,t_1) \cap A(s_2,t_2) \neq \emptyset$.

The following combinatorial lemma is useful.

Lemma 2.([3]) *Let $D_{s,t}(k)$ be the number of pairs (s_1,t_1) such that $t \neq t_1$ and $|A(s,t) \cap A(s_1,t_1)| = k$. Then we have*

$$D_{s,t}(k) \leq \begin{cases} (m+1)^2 n, & k = 1 \\ (m+1)^2 m^2, & 2 \leq k \leq \frac{m}{2} + 1 \\ 0, & k > \frac{m}{2} + 1 \end{cases} \tag{10}$$

Recall the definitions (5) and (6). Let

$$B_1(x,G) = \sum_{s_1,t_1} \sum_{\substack{s_2,t_2 \\ (s_1,t_1) \sim (s_2,t_2)}} EI_{s_1,t_1}(x) EI_{s_2,t_2}(x) \tag{11}$$

and

$$B_1(x,G) = \sum_{s_1,t_1} \sum_{\substack{(s_1,t_1) \neq (s_2,t_2) \\ (s_1,t_1) \sim (s_2,t_2)}} E[I_{s_1,t_1}(x) I_{s_2,t_2}(x)]. \tag{12}$$

Lemma 3. *For all $x < \varepsilon \ln n$ and $\delta > 0$, we have*

$$B_1(x, G) + B_2(x, G) = O(p_n^{2(x+1)} n^{\delta-1}). \tag{13}$$

Proof. From (9), we have $EI_{s,t}(x) = P(W'^{(n)}_{s,t} > c_n \ln n + x) \leq p_n^{c_n \ln n + x + 1}$. Since for fixed s and t, the number of pairs (s_1, t_1) such that $|A(s,t) \cap A(s_1, t_1)| = k$ is at most $D_{s,t}(k) + 1$, we obtain by Lemma 2

$$\begin{aligned}
B_1(x, G) &\leq \sum_{s,t} \sum_{k=1}^{m+1} (D_{s,t}(k) + 1) p_n^{2(c_n \ln n + x + 1)} \\
&\leq p_n^{2(x+1)} \cdot \frac{1}{n^4} \sum_{s,t} \left((m+1)^2 n + 1 + \sum_{k=2}^{m/2+1} ((m+1)^2 m^2 + 1) \right) \\
&= p_n^{2(x+1)} \cdot O\left(\frac{m^2 n + m^6}{n^2} \right) \\
&= O(p_n^{2(x+1)} n^{\delta - 1}),
\end{aligned} \tag{14}$$

for all $\delta > 0$, where the last equality holds using the assumption $c_n = o(\ln n)$.

Next, we have $E(I_{s,t}(x) I_{s_1,t_1}(x)) \leq p_n^{2(c_n \ln n + x + 1) - k}$ when $|A(s,t) \cap A(s_1, t_1)| = k$. Therefore, by Lemma 2

$$\begin{aligned}
B_2(x, G) &\leq \sum_{s,t} \sum_{k=1}^{m} D_{s,t}(k) p_n^{2(c_n \ln n + x + 1) - k} \\
&\leq p_n^{2(x+1)} \cdot \frac{1}{n^4} \sum_{s,t} \left(2(m+1)^2 n + (m+1)^2 m^2 \sum_{k=2}^{m/2+1} p_n^{-k} \right).
\end{aligned} \tag{15}$$

Since $c_n > \alpha > 0$, we obtain

$$\sum_{k=2}^{m/2+1} p_n^{-k} = O\big(p_n^{-\frac{m}{2}}\big) = O\big(n^{\frac{c_n + \varepsilon}{c_n}}\big). \tag{16}$$

Combining (15), (16) and the assumption $c_n = o(\ln n)$, we derive

$$\begin{aligned}
B_2(x, G) &= p_n^{2(x+1)} \cdot O\left(\frac{m^2 n + m^4 n^{\frac{c_n + \varepsilon}{c_n}}}{n^2} \right) \\
&= O(p_n^{2(x+1)} n^{\delta - 1})
\end{aligned} \tag{17}$$

for all $\delta > 0$. □

The following lemma is a simplified version of Theorem 1.

Lemma 4. *$W^{(n)}/c_n \ln n$ converges to 1 in probability, as $n \to \infty$; i.e., for any $\delta > 0$,*

$$\lim_{n\to\infty} P\left(\left|\frac{W^{(n)}}{c_n \ln n} - 1\right| > \delta\right) = 0. \tag{18}$$

Proof. Fix $\varepsilon > 0$, we have

$$P(W^{(n)}_{s,t} > (c_n + \varepsilon)\ln n) \leq p_n^{(c_n+\varepsilon)\ln n + 1}. \tag{19}$$

Since $c_n = o(\ln n)$, it follows that

$$P(W^{(n)} > (c_n + \varepsilon)\ln n) \leq n^2 p_n^{(c_n+\varepsilon)\ln n+1} \leq e^{-\frac{2\varepsilon \ln n}{c_n}} \to 0 \tag{20}$$

as $n \to \infty$.

Next, let $x = -\varepsilon \ln n$ and $Z(x)$ be a Poisson random variable with

$$EZ(x) = \lambda(x) = ES(x) = n^2 p_n^{\lfloor c_n \ln n + x + 2\rfloor} \geq e^{\frac{2\varepsilon \ln n - 4}{c_n}}. \tag{21}$$

Note that $\{W^{(n)} \leq (c_n - \varepsilon)\ln n\}$ implies that $\{W'^{(n)} \leq (c_n - \varepsilon)\ln n\}$. By Lemma 1 and Lemma 3,

$$\begin{aligned} P(W^{(n)} \leq (c_n - \varepsilon)\ln n) &\leq P(S(x) = 0) \\ &\leq B_1(x, G) + B_2(x, G) + P(Z(x) = 0) \\ &= O(p_n^{2(x+1)} n^{\delta - 1} + e^{-e^{\frac{2\varepsilon \ln n - 4}{c_n}}}) \to 0, \end{aligned} \tag{22}$$

as $n \to \infty$, for $\delta > 0$ and $\varepsilon < \alpha/5$. Thus, by (20) and (22), it follows that

$$\lim_{n\to\infty} P\left(\left|\frac{W^{(n)}}{c_n \ln n} - 1\right| > \delta\right) = 0. \tag{23}$$

for any $0 < \delta < 1/5$. □

To prove of Theorem 1, we need to further refine the proof of Lemma 4.

Proof of Theorem 1. As in the proof of Lemma 4, let $Z(x)$ be a Poisson random variable with

$$EZ(x) = \lambda(x) = ES(x) = n^2 p_n^{\lfloor c_n \ln n + x + 2\rfloor}. \tag{24}$$

If $c_n \ln n + x \in \mathbb{Z}$, then $\lambda(x) = p_n^{x+2}$. Recall that $W'^{(n)} > c_n \ln n + x$ if and only if $S(x) > 0$. Thus, by Lemma 1 and Lemma 3

$$\begin{aligned} |P(W'^{(n)} > c_n \ln n + x) - P(Z(x) \neq 0)| &= |P(S(x) > 0) - P(Z(x) > 0)| \\ &\leq B_1(x, G) + B_2(x, G) \\ &= O(p_n^{2(x+1)} n^{\delta - 1}). \end{aligned} \tag{25}$$

Note that $x < \varepsilon \ln n$, and then we have

$$\{W^{(n)} > c_n \ln n + x\} = \{W^{(n)} > (c+\varepsilon)\ln n\} \cup \{W'^{(n)} > c_n \ln n + x\}. \tag{26}$$

Hence, by (20), (25) and (26), we obtain

$$\begin{aligned} |P(W^{(n)} \le c_n \ln n + x) - e^{-\lambda(x)}| &= |P(W^{(n)} > c_n \ln n + x) - P(Z(x) \neq 0)| \\ &\le P(W^{(n)} > (c_n+\varepsilon)\ln n) \\ &\quad + |P(W'^{(n)} > c_n \ln n + x) - P(Z(x) \neq 0)| \\ &\le e^{-\frac{2\varepsilon \ln n}{c_n}} + O(p_n^{2(x+1)} n^{\delta-1}), \end{aligned} \tag{27}$$

for $0 < \delta < 1$, where the first item on the right-hand side of (27) tends to 0 as $n \to \infty$.

Let $\{x_n\}$ be a sequence such that $c_n \ln n + x_n \in \mathbb{Z}$ for all n. If $\inf_n x_n \ge \beta \in \mathbb{R}$, then $p_n^{2(x_n+1)} n^{\delta-1} \to 0$ and $e^{\lambda(x_n)}$ is a bounded sequence. Thus, from (27) it follows that

$$\left| e^{\lambda(x_n)} P(W^{(n)} \le c_n \ln n + x_n) - 1 \right| = O\Big(e^{-\frac{2\varepsilon \ln n}{c_n}} + p_n^{2(x_n+1)} n^{\delta-1} \Big) \to 0, \tag{28}$$

as $n \to \infty$. □

References

[1] N. Alon, J. H. Spencer, *The Probabilistic Method*. John Wiley & Sons, New York, 2008

[2] R. Arratia, L. Goldstein, L. Gordon, Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1989) 9–25

[3] I. Benjamini, A. Yadin, O. Zeitouni, Maximal arithmetic progressions in random subsets. *Elect. Comm. in Probab.*, 12(2007) 365–376

[4] S. Janson, T. Łuczak, A. Rucinski, *Random Graphs*. John Wiley & Sons, New York, 2000

[5] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27(1975) 299–345

[6] T. Tao, What is good mathematics? *Bull. Amer. Math. Soc.*, 44(2007) 623–634