

ON THE WEAK LIMIT LAW OF THE MAXIMAL UNIFORM k -SPACING

ALEKSANDAR MIJATOVIĆ[†] AND VLADISLAV VYSOTSKY[‡]

ABSTRACT. This paper gives a simple proof of a limit theorem for the length of the largest interval straddling a fixed number of i.i.d. points uniformly distributed on a unit interval. The key step in our argument is a classical theorem of [Watson, 1954] on the maxima of m -dependent stationary stochastic sequences.

1. INTRODUCTION AND THE MAIN RESULT

Both the distributional and asymptotic theories of spacings between consecutive order statistics of a sample of i.i.d. random variables play a central role in classical probability theory and mathematical statistics, see [Pyke, 1965], [Shorack and Wellner, 1986, Sec. 18–21] and the references therein. A deep understanding of this subject has been achieved over the past decades. In particular, [Devroye, 1981, Deheuvels, 1982] give a very fine description of the almost sure behaviour (as the sample size tends to infinity) of the maximal spacing between the ordered statistics of uniform random variables. The laws of iterated logarithms proved in these papers for the maximal spacings are further extended in [Deheuvels and Devroye, 1984] to analogous statements on the maximum of k consecutive spacings (called k -spacings).

In this note we prove a weak limit theorem for the maximal k -spacings. To the best of our knowledge, no result of this type was available in the past; it is truly surprising that this problem was not even mentioned in [Deheuvels and Devroye, 1984].

More precisely, let U_1, \dots, U_n be i.i.d. random variables that are uniformly distributed on $[0, 1]$. Denote by

$$U_{1:n} \leq \dots \leq U_{n:n}$$

2000 *Mathematics Subject Classification.* 60F05, 60G70.

Key words and phrases. maximal k -spacing, order statistics, uniform distribution.

AM is supported in part by the Humboldt Foundation Research Fellowship GRO/1151787 STP. The work of VV is supported by Marie Curie IIF Grant 628803 by European Commission and supported in part by Grant 13-01-00256 by RFBR.

their order statistics, which are the elements of U_1, \dots, U_n arranged in the ascending order, and define $U_{0:n} := 0$, $U_{(n+1):n} := 1$. The maximal spacing $M_n^{(1)} := \max_{0 \leq i \leq n} (U_{(i+1):n} - U_{i:n})$ is the length of the longest interval containing no points of the sample U_1, \dots, U_n . The classical representation of the uniform spacings given below in (1), which relates $M_n^{(1)}$ to the maximum of i.i.d. exponential random variables, together with the law of large numbers easily yields

$$nM_n^{(1)} - \log n \xrightarrow{d} G,$$

where G follows a standard Gumbel distribution $\mathbb{P}(G \leq x) = \exp(-e^{-x})$, $x \in \mathbb{R}$.

We study an analogous weak limit of the *maximal k -spacing*, that is the length of the largest open subinterval of $[0, 1]$ that contains $k - 1$ uniform points:

$$M_n^{(k)} := \max_{0 \leq i \leq n+1-k} (U_{(i+k):n} - U_{(i):n}).$$

Our main result is as follows.

Theorem 1. *Let G be a random variable that follows a standard Gumbel distribution. For any integer $k \geq 1$, it holds*

$$nM_n^{(k)} - \log n - (k-1) \log \log n + \log(k-1)! \xrightarrow{d} G \quad \text{as } n \rightarrow \infty.$$

We will use the following well-known fact: the uniform spacings are represented as

$$(U_{1:n} - U_{0:n}, \dots, U_{n:n} - U_{(n-1):n}) \stackrel{d}{=} \left(\frac{X_1}{X_1 + \dots + X_{n+1}}, \dots, \frac{X_n}{X_1 + \dots + X_{n+1}} \right), \quad (1)$$

where X_1, X_2, \dots are i.i.d. standard exponential random variables; moreover, the random vector on the right-hand side is independent of the sum $X_1 + \dots + X_{n+1}$, see e.g. [Pyke, 1965, Sec. 4.1].

To discuss the statement of Theorem 1, consider the simplest case that $k = 2$ on the largest interval straddling a single uniform point. It is not hard to show that $A_n := \max_{1 \leq i \leq n} (X_{2i-1} + X_{2i})$, $B_n := \max_{1 \leq i \leq n} (X_{2i} + X_{2i+1})$, which are maxima of i.i.d. gamma random variables, satisfy $A_n - \log n - \log \log n \xrightarrow{d} G$ and $B_n - \log n - \log \log n \xrightarrow{d} G$.

The crucial observation is that $A_n - \log n - \log \log n$ and $B_n - \log n - \log \log n$ are *asymptotically independent*. Then

$$M_n^{(2)} = \frac{\max(A_{\lfloor n/2 \rfloor}, B_{\lfloor (n-1)/2 \rfloor})}{X_1 + \dots + X_{n+1}},$$

and hence the law of large numbers and the continuous mapping theorem imply¹

$$nM_n^{(2)} - \log(n/2) - \log \log(n/2) \xrightarrow{d} \max(G_1, G_2),$$

where G_1 and G_2 are i.i.d. random variables with a standard Gumbel distribution. Since $\max(G_1, G_2) \stackrel{d}{=} \log 2 + G$, Theorem 1 follows in the case that $k = 2$.

The asymptotic independence of A_n and B_n is non-trivial and somewhat unexpected. Our initial approach to the proof of Theorem 1 rested on establishing this property using the specific structure of these random variables. However, once the classical result [Watson, 1954] on the maxima of m -dependent stationary sequences came to our attention, we understood that our Theorem 1 can be established as a direct consequence². We describe this shorter and easier proof in the next section.

2. PROOFS

We start by recalling the result from [Watson, 1954]. Random variables Y_1, Y_2, \dots are said to be m -dependent if $|i - j| > m$ implies that Y_i and Y_j are independent.

Theorem 2. *For any $m \geq 1$, let Y_1, Y_2, \dots be a strictly stationary sequence of m -dependent unbounded random variables. Assume that*

$$\lim_{y \rightarrow \infty} \max_{1 \leq |i-j| \leq m} \mathbb{P}(Y_j > y | Y_i > y) = 0. \quad (2)$$

Then for any positive numbers ξ, y_1, y_2, \dots satisfying

$$\lim_{n \rightarrow \infty} n\mathbb{P}(Y_1 > y_n) = \xi, \quad (3)$$

it holds

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\max_{1 \leq i \leq n} Y_i \leq y_n \right) = \exp(-\xi).$$

The theorem says that the maximum of m -dependent stationary random variables has the same weak limit as the maximum of an i.i.d. sequence with the same common distribution. Although the actual theorem of [Watson, 1954] makes a more restrictive assumption $\xi = n\mathbb{P}(Y_1 > y_n)$ for all $n \geq 1$, which may even be impossible to satisfy for certain ξ , the presented version easily follows by the monotonicity of distribution functions and the continuity of $\exp(-\xi)$.

¹See the analogous argument after (7) below.

²The asymptotic independence appears not to be an easy consequence of the result in [Watson, 1954].

The aim is to apply Theorem 2 to the $(k-1)$ -dependent stationary sequence of moving sums

$$Y_i := \sum_{\ell=i}^{i+k-1} X_\ell, \quad i \geq 1, \quad (4)$$

and the numbers

$$\xi := e^{-x}, \quad y_n := \log n + (k-1) \log \log n - \log(k-1)! + x \quad (5)$$

for any fixed real x .

Note first that Y_i are gamma random variables with densities f_k , where $f_\theta(y) := y^{\theta-1}e^{-y}/\Gamma(\theta)$ for any positive y and θ . Then it is straightforward to check using L'Hopital's rule that

$$\mathbb{P}(Y_1 > y) \sim \frac{y^{k-1}e^{-y}}{(k-1)!}, \quad y \rightarrow \infty \quad (6)$$

(where by \sim we mean that the ratio tends to 1), hence (3) holds by

$$\lim_{n \rightarrow \infty} n\mathbb{P}(Y_1 > y_n) = \lim_{n \rightarrow \infty} \frac{ny_n^{k-1}e^{-y_n}}{(k-1)!} = e^{-x} \lim_{n \rightarrow \infty} \left(\frac{y_n}{\log n} \right)^{k-1} = \xi.$$

It remains to check the assumption (2). For any integer $1 \leq a \leq k-1$, we have

$$Y_{a+1} = Y_1 \left(1 - \frac{X_1 + \dots + X_a}{X_1 + \dots + X_k} \right) + (X_{k+1} + \dots + X_{k+a}).$$

Hence

$$(Y_1, Y_{a+1}) \stackrel{d}{=} (Y_1, Y_1(1 - U_{a:(k-1)}) + Z_a),$$

where the three random variables in the r.h.s. are mutually independent and Z_a has a gamma distribution with density f_a . By (6), for any $\varepsilon > 0$ there exists an $R > 0$ such that

$$\mathbb{P}(Y_1 > y + R) \leq \varepsilon \mathbb{P}(Y_1 > y) \quad \text{for all } y \text{ large enough.}$$

Then (2) follows as for such y ,

$$\begin{aligned} \mathbb{P}(Y_i > y, Y_{a+i} > y) &\leq \mathbb{P}(y < Y_1 \leq y + R, Y_{a+1} > y) + \mathbb{P}(Y_1 > y + R) \\ &\leq \int_y^{y+R} \mathbb{P}(Z_a > y - x(1 - U_{a:(k-1)})) f_k(x) dx + \varepsilon \mathbb{P}(Y_1 > y) \\ &\leq (\mathbb{P}(Z_{k-1} > yU_{1:(k-1)} - R) + \varepsilon) \cdot \mathbb{P}(Y_1 > y). \end{aligned}$$

Thus we showed that Theorem 2 applies to the sequence Y_1, Y_2, \dots defined in (4), hence combined with (5) this implies

$$\max_{1 \leq i \leq n+1-k} Y_i - \log n - (k-1) \log \log n + \log(k-1)! \xrightarrow{d} G. \quad (7)$$

Then by (1), we find

$$nM_n^{(k)} = \frac{n}{X_1 + \cdots + X_{n+1}} \max_{1 \leq i \leq n+1-k} Y_i.$$

Now Theorem 1 follows by (7), the law of large numbers, the continuous mapping theorem, and the relation

$$\log n \left(\frac{n}{X_1 + \cdots + X_{n+1}} - 1 \right) = \frac{\log n}{\sqrt{n}} \cdot \frac{(n - (X_1 + \cdots + X_{n+1}))/\sqrt{n}}{(X_1 + \cdots + X_{n+1})/n} \xrightarrow{d} 0,$$

which itself holds by the law of large numbers, and the central limit theorem.

REFERENCES

- [Deheuvels, 1982] Deheuvels, P. (1982). Strong limiting bounds for maximal uniform spacings. *Ann. Probab.*, 10(4):1058–1065.
- [Deheuvels and Devroye, 1984] Deheuvels, P. and Devroye, L. (1984). Strong laws for the maximal k -spacing when $k \leq c \log n$. *Z. Wahrsch. Verw. Gebiete*, 66(3):315–334.
- [Devroye, 1981] Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Probab.*, 9(5):860–867.
- [Pyke, 1965] Pyke, R. (1965). Spacings. (With discussion.). *J. Roy. Statist. Soc. Ser. B*, 27:395–449.
- [Shorack and Wellner, 1986] Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York.
- [Watson, 1954] Watson, G. S. (1954). Extreme values in samples from m -dependent stationary stochastic processes. *Ann. Math. Statistics*, 25:798–800.

ARIZONA STATE UNIVERSITY[‡], IMPERIAL COLLEGE LONDON^{†,‡}, ST.PETERSBURG DEPARTMENT OF STEKLOV MATHEMATICAL INSTITUTE[‡]