ON THE WEAK LIMIT LAW OF THE MAXIMAL UNIFORM \( k \)-SPACING

BY ALEKSANDAR MIJATOVIĆ AND VLADISLAV VYSOTSKY

Abstract

In this paper we give a simple proof of a limit theorem for the length of the largest interval straddling a fixed number of points that are independent and uniformly distributed on a unit interval. The key step in our argument is a classical theorem of Watson on the maxima of \( m \)-dependent stationary stochastic sequences.

Keywords: Maximal \( k \)-spacing; order statistics; uniform distribution

2010 Mathematics Subject Classification: Primary 60F05
Secondary 60G70; 62G30

1. Introduction and the main result

Both the distributional and asymptotic theories of spacings between consecutive order statistics of a sample of independent and identically distributed (i.i.d.) random variables play a central role in classical probability theory and mathematical statistics; see [4], [5, Sections 18–21], and the references therein. A deep understanding of this subject has been achieved over the past decades. In particular, Devroye [3] and Deheuvels [1] gave a very fine description of the almost-sure behaviour (as the sample size tends to infinity) of the maximal spacing between the order statistics of uniform random variables. The laws of iterated logarithms proved in these papers for the maximal spacings were further extended by Deheuvels and Devroye [2] 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, a result of this type has not previously been reported; it is truly surprising that this problem was not even mentioned in [2].

More precisely, let \( U_1, \ldots, U_n \) be independent random variables that are uniformly distributed on \([0, 1]\). Denote by

\[
U_{1:n} \leq \cdots \leq U_{n:n}
\]

their order statistics, which are the elements of \( U_1, \ldots, U_n \) arranged in ascending order, and define \( U_{0:n} := 0 \) and \( U_{n+1:n} := 1 \). 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 \( U_1, \ldots, U_n \). The classical representation of uniform spacings given below in (1.1), 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.

doi:10.1017/apr.2016.52 © Applied Probability Trust 2016
Theorem 1.1. Let \( G \) be a random variable that follows a standard Gumbel distribution. For any integer \( k \geq 1 \),
\[
nM_n^{(k)} - \log n - (k - 1) \log \log n + \log(k - 1)! \xrightarrow{D} G \quad \text{as} \quad n \to \infty.
\]

We will use the following well-known fact: the uniform spacings are represented as
\[
(U_{1:n} - U_{0:n}, \ldots, U_{n+1:n} - U_{n:n}) \overset{D}{=} \left( \frac{X_1}{X_1 + \cdots + X_{n+1}}, \ldots, \frac{X_{n+1}}{X_1 + \cdots + X_{n+1}} \right), \quad (1.1)
\]
where \( X_1, X_2, \ldots \) are independent, standard exponential random variables; moreover, the random vector on the right-hand side is independent of the sum \( X_1 + \cdots + X_{n+1} \); see, e.g. [4, Section 4.1].

To discuss the statement of Theorem 1.1, consider the simplest case in which \( k = 2 \), the largest interval straddling a single uniform point. It is not hard to show that \( A_n := \max_{1 \leq i \leq n} (X_{2i-1} + X_{2i}) \) and \( B_n := \max_{1 \leq i \leq n} (X_{2i} + X_{2i+1}) \), which are the maxima of i.i.d. gamma random variables, satisfy
\[
A_n - \log n - \log \log n \xrightarrow{D} G \quad \text{and} \quad B_n - \log n - \log \log n \xrightarrow{D} G.
\]
The crucial observation is that \( A_n - \log n - \log \log n \) and \( B_n - \log n - \log \log n \) are asymptotically independent. Then
\[
M(n) = \max(A_{\lfloor (n+1)/2 \rfloor}, B_{\lfloor n/2 \rfloor}),
\]
and, hence, the law of large numbers and the continuous mapping theorem imply (see the analogous argument after (2.6) below) that
\[
nM_n^{(2)} - \log(\frac{1}{2}n) - \log \log(\frac{1}{2}n) \xrightarrow{D} \max(G_1, G_2),
\]
where \( G_1 \) and \( G_2 \) are independent random variables with a standard Gumbel distribution. Since \( \max(G_1, G_2) \overset{D}{=} \log 2 + G \), Theorem 1.1 follows in the \( k = 2 \) case.

The asymptotic independence of \( A_n \) and \( B_n \) is nontrivial and somewhat unexpected. Our initial approach to the proof of Theorem 1.1 rested on establishing this property using the specific structure of these random variables. However, once the classical result [6] on the maxima of \( m \)-dependent stationary sequences came to our attention, we understood that our Theorem 1.1 can be established as a direct consequence (the asymptotic independence appears not to be an easy consequence of the result in [6]). We describe this shorter and easier proof in the next section.

2. Proofs

We start by recalling the result from [6]. Random variables \( Y_1, Y_2, \ldots \) are said to be \( m \)-dependent if \( |i - j| > m \) implies that \( Y_i \) and \( Y_j \) are independent.

Theorem 2.1. For any \( m \geq 1 \), let \( Y_1, Y_2, \ldots \) be a strictly stationary sequence of \( m \)-dependent unbounded random variables. Assume that
\[
\lim_{y \to \infty} \max_{1 \leq |i-j| \leq m} \mathbb{P}(Y_j \geq y | Y_i \geq y) = 0. \quad (2.1)
\]
Then, for any positive numbers \( \xi, y_1, y_2, \ldots \) satisfying
\[
\lim_{n \to \infty} n \mathbb{P}(Y_1 > y_n) = \xi, \quad (2.2)
\]
it holds that
\[ \lim_{n \to \infty} \mathbb{P}\left( \max_{1 \leq i \leq n} Y_i \leq y_n \right) = \exp(-\xi). \]

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 [6] makes the more restrictive assumption that \(\xi = n\mathbb{P}(Y_1 > y_n)\) for all \(n \geq 1\), which may even be impossible to satisfy for certain \(\xi\), the presented version easily follows by the monotonicity of distribution functions and the continuity of \(\exp(-\xi)\).

The aim is to apply Theorem 2.1 to the \((k-1)\)-dependent stationary sequence of moving sums
\[ Y_i := \sum_{\ell=i}^{i+k-1} X_\ell, \quad i \geq 1, \quad (2.3) \]
and the numbers
\[ \xi := e^{-x}, \quad y_n := \log n + (k-1) \log \log n - \log(k-1)! + x, \quad (2.4) \]
for any fixed real \(x\).

Note first that the \(Y_i\) are gamma random variables with densities \(f_k\), where \(f_\theta(y) := y^{\theta-1}e^{-y}/\Gamma(\theta)\) for any positive \(y\) and \(\theta\). Then it is straightforward to check using l'Hôpital's rule that
\[ \mathbb{P}(Y_1 > y) \sim \frac{y^{k-1}e^{-y}}{(k-1)!} \quad \text{as} \quad y \to \infty \quad (2.5) \]
(where by ‘\(\sim\)’ we mean that the ratio tends to 1); hence, (2.2) holds by
\[ \lim_{n \to \infty} n\mathbb{P}(Y_1 > y_n) = \lim_{n \to \infty} \frac{ny_n^{k-1}e^{-y_n}}{(k-1)!} = e^{-x} \lim_{n \to \infty} \left( \frac{y_n}{\log n} \right)^{k-1} = \xi. \]

It remains to check that assumption (2.1) holds. For any integer \(1 \leq a \leq k-1\), we have
\[ Y_{a+1} = Y_1 \left( 1 - \frac{X_1 + \cdots + X_a}{X_1 + \cdots + X_k} \right) + X_{k+1} + \cdots + X_{k+a}. \]
Hence,
\[ (Y_1, Y_{a+1}) \overset{D}{=} (Y_1, Y_1(1 - U_{a,k-1}) + Z_a), \]
where the three random variables on the right-hand side are mutually independent and \(Z_a\) has a gamma distribution with density \(f_a\). By (2.5), for any \(\varepsilon > 0\), there exists an \(R > 0\) such that
\[ \mathbb{P}(Y_1 > y + R) \leq \varepsilon \mathbb{P}(Y_1 > y) \quad \text{for all large enough} \quad y. \]
Then (2.1) follows since, for such \(y\),
\[ \mathbb{P}(Y_1 > y, Y_{a+1} > y) \leq \mathbb{P}(y < Y_1 \leq y + R, Y_{a+1} > y) + \mathbb{P}(Y_1 > y + R) \]
\[ \leq \int_y^{y+R} \mathbb{P}(Z_a > y - x(1 - U_{a,k-1})) f_k(x) \, dx + \varepsilon \mathbb{P}(Y_1 > y) \]
\[ \leq (\mathbb{P}(Z_{k-1} > yU_{1,k-1} - R) + \varepsilon) \mathbb{P}(Y_1 > y). \]
Thus we showed that Theorem 2.1 applies to the sequence $Y_1, Y_2, \ldots$ defined in (2.3); hence, combined with (2.4), this implies that
\[
\max_{1 \leq i \leq n+2-k} Y_i - \log n - (k - 1) \log \log n + \log(k - 1)! \xrightarrow{\mathcal{D}} G. \tag{2.6}
\]

Then by (1.1) we find that
\[
\frac{n M^{(k)}_n}{X_1 + \cdots + X_{n+1}} = \max_{1 \leq i \leq n+2-k} Y_i.
\]

Now Theorem 1.1 follows by (2.6), the law of large numbers, the continuous mapping theorem, and the relation
\[
(\log n) \left( \frac{n}{X_1 + \cdots + X_{n+1}} - 1 \right) = \frac{\log n}{\sqrt{n}} \frac{(n - (X_1 + \cdots + X_{n+1}))/\sqrt{n}}{(X_1 + \cdots + X_{n+1})/n} \xrightarrow{\mathcal{D}} 0,
\]
which itself holds by the law of large numbers and the central limit theorem.

Acknowledgements

This paper was written in honour of our dear friend and distinguished colleague Nick Bingham. We are very grateful for Nick’s guidance and help over the years.

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

References

[1] Deheuvels, P. (1982). Strong limiting bounds for maximal uniform spacings. *Ann. Prob.***10*, 1058–1065.

[2] Deheuvels, P. and Devroye, L. (1984). Strong laws for the maximal $k$-spacing when $k \leq c \log n$. *Z. Wahrscheinlichkeitstheor.***66*, 315–334.

[3] Devroye, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Prob.***9*, 860–867.

[4] Pyke, R. (1965). Spacings. *J. R. Statist. Soc. B.* ***27***, 395–449.

[5] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.

[6] Watson, G. S. (1954). Extreme values in samples from $m$-dependent stationary stochastic processes. *Ann. Math. Statist.* ***25***, 798–800.

ALEKSANDAR MIJATOVIĆ, King’s College London
Department of Mathematics, King’s College London, Strand, London WC2R 2LS, UK.
Email address: aleksandar.mijatovic@kcl.ac.uk

VLADISLAV VYSOTSKY, Arizona State University, Imperial College London, and St. Petersburg Department of Steklov Mathematical Institute
School of Mathematical and Statistical Sciences, Arizona State University, PO Box 871804, Tempe, AZ 85287, USA.
Email address: vysotsky@asu.edu