

ON THE DISTRIBUTION OF THE MAXIMUM OF RANDOM VARIABLES

BY JANOS GALAMBOS

Temple University

For a wide class of (dependent) random variables X_1, X_2, \dots, X_n , a limit law is proved for the maximum, with suitable normalization, of X_1, X_2, \dots, X_n . The results are more general in two aspects than the ones obtained earlier by several authors, namely, the stationarity of the X 's is not assumed and secondly, the assumptions on the dependence of the X 's are weaker than those occurring in previous papers. A generalization of the method of inclusion and exclusion is one of the main tools.

1. The results and related works. Let X_1, X_2, \dots, X_n be random variables on a given probability space and put $Z_n = \max(X_1, X_2, \dots, X_n)$ and $F_j(x) = P(X_j < x)$.

We introduce the following combinatorial concepts. Let $H = \{1, 2, \dots, n\}$, and let E be a subset of the set of ordered pairs of distinct elements of H . The pair $G = (H, E)$ is called a finite graph. The elements of H are the vertices and those of E are the edges of G . Let N_E denote the number of elements of E . Let further $H_k^{(0)}$ be the set of those ordered sets (i_1, i_2, \dots, i_k) of k distinct elements of H , which contain no pairs (i_s, i_t) belonging to E , and $H_k^{(1)}$ is the set of those ordered sets (i_1, i_2, \dots, i_k) of k distinct elements of H , which contain exactly one pair $(i_s, i_t) \in E$.

In this note we prove the following

THEOREM. Let G be a finite graph, and let $c_n = c_n(a)$ be a sequence of real numbers for which

$$(1) \quad \lim_{n \rightarrow +\infty} \sum_{j=1}^n (1 - F_j(c_n)) = a.$$

Assume that each of the following properties is satisfied.

(i) There is a finite K such that for all n and j

$$n(1 - F_j(c_n)) \leq K.$$

$$(ii) \quad \lim_{n \rightarrow +\infty} N_E \max_{(i,j) \in E} P(X_i \geq c_n, X_j \geq c_n) = 0$$

and

$$N_E = o(n^2).$$

(iii) For any set $(i_1, i_2, \dots, i_k) \in H_k^{(0)}$

$$P(X_{i_1} \geq c_n, \dots, X_{i_k} \geq c_n) = \prod_{j=1}^k (1 - F_j(c_n)) + r(i_1, \dots, i_k)$$

Received January 21, 1971, revised October 3, 1971.

with

$$\lim_{n \rightarrow +\infty} \sum_{(i_1, \dots, i_k) \in H_k^{(0)}} r(i_1, \dots, i_k) = 0.$$

(iv) There is a real sequence d_k such that for any set $(i_1, \dots, i_k) \in H_k^{(1)}$

$$P(X_{i_1} \geq c_n, \dots, X_{i_k} \geq c_n) \leq d_k P(X_{i_s} \geq c_n, X_{i_t} \geq c_n) \prod_{j=1; j \neq s, t}^k P(X_{i_j} \geq c_n)$$

where $(i_s, i_t) \in E$.

Then

$$(2) \quad \lim_{n \rightarrow +\infty} P(Z_n < c_n) = e^{-a}.$$

Without attempting to review the results in this direction, we shall analyze our theorem in terms of previous results. For a detailed review and bibliography, the reader is referred to the book by David (1970). Let first the X 's be independent. Then we can choose E to be empty and therefore our assumptions (ii)—(iv) are automatically satisfied, and the theorem becomes well known. Let now X_1, X_2, \dots, X_n be from an m -dependent stationary stochastic process. In this case, evidently, (1) implies (i). Define E as the set of all (i, j) for which $|i - j| \leq m$ and $i \neq j$. Thus $r(i_1, \dots, i_k) = 0$ in (iii), and (iv) is also satisfied with $d_k = 1$. Since $N_E = nm + O(1)$, our condition (ii), in view of (1), is equivalent to the assumption of Watson (1954), namely, that

$$\lim_{n \rightarrow +\infty} \frac{\max_{|i-j| \leq m} P(X_i \geq c_n, X_j \geq c_n)}{P(X_1 \geq c_n)} = 0.$$

It is worthwhile to look at the conditions of our Theorem in more detail. The idea behind the conditions is that, if we have a set of random variables, "almost all" of which are "almost independent", then a restriction on their bivariate distributions already guarantees that their maximum behaves as if they were completely independent. As a matter of fact, if the almost independence, expressed in (iii), is not satisfied with X_i and X_j , then include (i, j) in E ; otherwise do not. Condition (ii) requires that a positive percentage of all pairs (X_i, X_j) cannot be in the exceptional set, but there is no other restriction on the set E . I wish to point out two facts about assumption (iv). First of all, that it imposes restrictions only on the elements of $H_k^{(1)}$, and secondly, that a much weaker assumption is made than the independence of (X_{i_s}, X_{i_t}) and the rest of the random variables in question. Our model thus includes the mixing sequences of random variables by constructing E in the same way as in the case of m -dependence. As a matter of fact, the essential difference between m -dependence and mixing is that, in the first case, X_i and X_j are independent if $|i - j| > m$, while in the second case they are "almost" independent for m large. For the asymptotic distribution of the maximum of random variables from a mixing sequence, see Loynes (1965). Our result is thus more general than these previous ones, since their assumptions on the dependence are

much stronger. In addition, stationarity is always assumed, except for the case of independent random variables, Juncosa (1949). For a stationary sequence, with stronger assumptions than (iii) and (iv), a general theorem is proved in Galambos (1970a).

It is easy to construct several different kinds of examples for stationary sequences of random variables to which the present result applies but previous ones do not. The following example plays a significant role in a problem of probabilistic number theory, see Galambos (1970b) and a paper to appear [5]. The infinite sequence X_1, X_2, \dots of random variables is identically distributed with distribution function $1 - e^{-x}$. For their joint distribution we have

$$P(X_{i_1} > x, X_{i_2} > x, \dots, X_{i_k} > x) = [1 + D_k(\frac{x}{\theta})^{i_1}] \exp(-kx),$$

where $1 \leq i_1 < i_2 < \dots < i_k$ are positive integers and D_k is a constant, bounded over the possible choices of the subscripts i_j . Note the exponent i_1 in the error term. This shows that X_1 and X_m do not become "almost independent" whatever the value of m be, hence the sequence X_i is not mixing (therefore not m -dependent either). Our conditions are, however, satisfied by choosing E as the set of all (i, j) , $i < j$, with $1 \leq i < \log n$, say (any function can be taken for $\log n$ which is of smaller order than n). The result of Galambos (1970a) does not contain this example as a special case.

2. Proof of Theorem. We shall apply a graph sieve theorem (a generalization of the method of inclusion and exclusion) of Rényi (1961) which will be stated as Lemma 1. By the notations of the previous section, let $H_k = H_k^{(0)} \cup H_k^{(1)}$. Let A_1, A_2, \dots, A_n be a sequence of events and let B denote the event that none of the A 's occurs. Let $S_{1,0} = S_{2,0} = 1$ and for $k \geq 1$, put

$$(3) \quad S_{i,k} = \sum P(A_{i_1} A_{i_2} \dots A_{i_k}), \quad i = 1 \text{ or } 2,$$

the summation being extended over $(i_1, i_2, \dots, i_k) \in H_k$ if $i = 1$ and k is even or if $i = 2$ and k is odd, and over $(i_1, i_2, \dots, i_k) \in H_k^{(0)}$ if $i = 1$ and k is odd or if $i = 2$ and k is even. Then we have

LEMMA 1. For any integer $s \geq 0$,

$$(4) \quad \sum_{k=0}^{2s+1} (-1)^k S_{2,k} \leq P(B) \leq \sum_{k=0}^{2s} (-1)^k S_{1,k}.$$

Since the proof of this lemma is only available in Hungarian, I give the major steps in it. The details are similar (and simpler) than those in the proof of a generalization of Lemma 1 in Galambos (1966).

By the method of indicators of Loève (1942) (see also Loève's book on probability theory), it is sufficient to prove (4) when all the A 's are replaced by their indicator variables. Let h denote the number of those indicators which are equal to 1. (4) is proved by induction over h . For $h = 0$, both inequalities in (4) evidently hold. For $h > 0$, (4) reduces to the following combinatorial

inequalities. First of all, note that, without loss of generality, we may assume that the indicators of A_1, \dots, A_h are 1, and those of the rest are 0. Let G_h denote the graph obtained from G by considering the vertices $1, 2, \dots, h$ only. Let $H(h) = \{1, 2, \dots, h\}$ and let $N_1(G_h; j, m)$ be the number of those subsets of $H(h)$, the number of elements of which is at most m and is an odd number; further, the number of pairs of the elements of these subsets which belong to E is at most j . $N_2(G_h; j, m)$ is defined in the same way, only the number of elements of the subsets considered is required to be an even number. The empty set is considered as a set with even number of elements. (4) now becomes

$$(4a) \quad N_2(G_h; 1, 2s) - N_1(G_h; 0, 2s - 1) \geq 0$$

and

$$(4b) \quad N_2(G_h; 0, 2s) - N_1(G_h; 1, 2s + 1) \leq 0.$$

To prove (4a, b) by induction over h , assume that they hold for any graph with at most h vertices and consider G_{h+1} . We shall adopt the usual terminology that i and j are connected with an edge if $(i, j) \in E$. Observe that those subsets of $H(h + 1)$ which contain at most one edge of G_{h+1} , belong to one of the following three categories: (a) those which do not contain the vertex $h + 1$ (b) those which contain $h + 1$ and one element is connected with $h + 1$ and (c) those which contain $h + 1$ and none of their elements is connected with $h + 1$. Denoting by G^* that subgraph of G_h the vertices of which are not connected with $h + 1$, we have by definition

$$N_2(G_{h+1}; 1, 2s) = N_2(G_h; 1, 2s) + N_1(G^*; 1, 2s - 1) + R_1$$

and

$$N_1(G_{h+1}; 0, 2s - 1) = N_1(G_h; 0, 2s - 1) + N_2(G^*; 0, 2s - 2)$$

where R_1 is the number of those subsets which correspond to the category (b). Since the number of vertices of G_h and of G^* is at most h , the assumption of induction is applicable to them; therefore by subtracting the two equations above, (4a) easily follows. The proof of (4b) is exactly the same.

From Lemma 1 we easily get

LEMMA 2. Let A_1, A_2, \dots, A_n be a sequence of events and suppose that for any fixed $k \geq 0$

$$(5) \quad \lim_{n \rightarrow +\infty} S_{i,k} = a^k/k!, \quad i = 1, 2.$$

Then

$$(6) \quad \lim_{n \rightarrow +\infty} P(B) = e^{-a}.$$

PROOF. Fix s in (4) and let $n \rightarrow +\infty$. We thus have from (4) and (5)

$$\sum_{k=0}^{2s-1} (-1)^k a^k/k! \leq \liminf_{n \rightarrow +\infty} P(B) \leq \limsup_{n \rightarrow +\infty} P(B) \leq \sum_{k=0}^{2s} (-1)^k a^k/k!.$$

Letting $s \rightarrow +\infty$, we obtain the statement of the lemma. Hence the proof is complete.

We can now turn to the proof of the Theorem. Put $A_j = \{X_j \geq c_n\}$. Thus by Lemma 2 it suffices to show that (5) applies. By the definition of $S_{i,k}$, for $k \geq 1$,

$$|S_{1,k} - S_{2,k}| = \sum_{(i_1, \dots, i_k) \in H_k^{(1)}} P(A_{i_1} A_{i_2} \dots A_{i_k})$$

and thus by assumption (iv)

$$(7) \quad |S_{1,k} - S_{2,k}| \leq d_k N_E \max_{(i,j) \in E} P(A_i A_j) \sum_{(i_1, \dots, i_{k-2}) \in H_{k-2}^{(0)}} P(A_{i_1}) \dots P(A_{i_{k-2}}) \\ \leq d_k N_E \max_{(i,j) \in E} P(A_i A_j) [\sum_{t=1}^n P(A_t)]^{k-2} \rightarrow 0 \quad (n \rightarrow +\infty)$$

where the last step follows from (1) and (ii).

On the other hand, for $i = 1$ and k odd or for $i = 2$ and k even, in view of (iii) and (3),

$$S_{i,k} = \sum_{(i_1, \dots, i_k) \in H_k^{(0)}} P(A_{i_1}) \dots P(A_{i_k}) + o(1) \\ = \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k}) - R_k + o(1),$$

where

$$R_k = \sum P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

the summation being extended over the vectors (i_1, i_2, \dots, i_k) for which at least one pair $(i_s, i_t) \in E$. Evidently

$$R_k \leq N_E [\max_{1 \leq j \leq n} P(A_j)]^2 [\sum_{j=1}^n P(A_j)]^{k-2}$$

which tends to zero by (i) and (ii). Thus, putting

$$(8) \quad T_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k}),$$

we have that for $i = 1, 2$ and for $k \geq 1$,

$$(9) \quad S_{i,k} = T_k + o(1), \quad (n \rightarrow +\infty).$$

Note that $U_k = (1/k)T_1 T_{k-1}$ contains all terms of T_k and that any term of U_k is majorized by $[\max_{1 \leq j \leq n} P(A_j)]^k$. We thus have

$$(10) \quad 0 \leq (1/k)T_1 T_{k-1} - T_k \leq \left[\frac{n}{k} \binom{n}{k-1} - \binom{n}{k} \right] [\max_{1 \leq j \leq n} P(A_j)]^k,$$

which tends to zero by (i) and by the fact that for fixed k ,

$$\frac{n}{k} \binom{n}{k-1} - \binom{n}{k} = \frac{n}{k} \left[\binom{n}{k-1} - \binom{n-1}{k-1} \right] = \frac{n}{k} \binom{n-1}{k-2} = O(n^{k-1}).$$

Hence (9), (10) and (1) yield by induction that (5) holds for all $k \geq 1$ and for $i = 1, 2$, which terminates the proof of the Theorem.

3. A concluding remark. Let Z_t^* denote the t th largest among X_1, X_2, \dots, X_n

(i.e. $Z_n^* = Z_n$). Under the same conditions as those assumed in the Theorem, limit law can be obtained for Z_{n-1}^* in the same way that the Theorem is proved. This is made possible by the inequalities obtained in Galambos (1966) for $P(B_t)$, where B_t denotes the event that at least t out of the events A_1, A_2, \dots, A_n occur. For Z_{n-m}^* with $m \geq 2$, however, the inequalities of [2] require restriction at least on the trivariate distributions of those X 's, for which the "almost independence" is not satisfied. Recent unpublished improvements on the inequalities of [2] reduce it now to $m \geq 4$.

Evidently the Theorem yields conditions for Z_1^* to have limiting distribution. Also, by the technique of the present paper, the limit of the joint distribution of (Z_1^*, Z_n) can be evaluated.

REFERENCES

- [1] DAVID, H. A. (1970). *Order Statistics*. Wiley, New York.
- [2] GALAMBOS, J. (1966). On the sieve methods in probability theory I. *Studia Sci. Math. Hungar.* **1** 39-50.
- [3] GALAMBOS, J. (1970a). On the sieve methods in probability theory II. *Ghana J. Sci.* **10** 11-15.
- [4] GALAMBOS, J. (1970b). The ergodic properties of the denominators in the Oppenheim expansion of real numbers into infinite series of rationals. *Quart. J. Math. Oxford Ser. 2* **21** 177-191.
- [5] GALAMBOS, J. Further ergodic results on the Oppenheim series. *Quart. J. Math. Oxford Ser. 2* To appear.
- [6] JUNCOSA, M. L. (1949). On the distribution of the minimum in a sequence of mutually independent random variables. *Duke J. Math.* **16** 609-618.
- [7] LOÈVE, M. (1942). Sur les systemes d'evenements. *Ann. Univ. Lyon Sect. A.* **5** 55-74.
- [8] LOYNES, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.* **36** 993-999.
- [9] RÉNYI, A. (1961). A general method to prove theorems of probability and some of its applications (in Hungarian). *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* **11** 79-105.
- [10] WATSON, G. S. (1954). Extreme values in samples from m -dependent stationary stochastic processes. *Ann. Math. Statist.* **25** 798-800.