

## LIMIT PROPERTIES OF RANDOM VARIABLES ASSOCIATED WITH A PARTIAL ORDERING OF $R^d$

BY J. MICHAEL STEELE

*University of British Columbia*

A limit theorem is established for the length of the longest chain of random values in  $R^d$  with respect to a partial ordering. The result is applied to a question raised by T. Robertson and F. T. Wright concerning the generalized empirical distribution function associated with the class of lower layers.

**Introduction.** There are two distinct roots to the work presented in this paper: one in mathematical statistics, and the other in combinatorial geometry. Since the geometrical origin is historically prior, as well as the more succinct, we will consider it first.

In 1935, Erdős and Szekeres [3] proved that any sequence of  $n$  distinct real numbers has a subsequence of length at least  $n^{1/2}$  which is either monotone increasing or monotone decreasing. S. Ulam later posed a probabilistic version of the result of Erdős and Szekeres and made a conjecture which we will now describe.

Let  $X_i = (\zeta_i^{(1)}, \zeta_i^{(2)})$ ,  $i = 1, 2, \dots$  be a sequence of i.i.d. random variables which are uniformly distributed in  $[0, 1] \times [0, 1]$ , and let  $l(n)$  be equal to the largest integer  $k$  such that there is a subsequence  $i_1 < i_2 < \dots, i_k \leq n$  such that

$$\zeta_{i_1}^{(1)} \leq \zeta_{i_2}^{(1)} \leq \dots \leq \zeta_{i_k}^{(1)} \quad \text{and} \quad \zeta_{i_1}^{(2)} \leq \zeta_{i_2}^{(2)} \leq \dots \leq \zeta_{i_k}^{(2)}.$$

In a form equivalent to our description, Ulam conjectured that the sequence of random variables  $l(n)/n^{1/2}$ ,  $n = 1, 2, \dots$  converges in probability to a nonzero constant  $c$ . This conjecture was proved by Hammersley [4], and his estimate for  $c$  was sharpened by Kingman [7]. A very nice addition to this work was made in a comment by Kesten (see [7, page 903], and [5]) who proved that  $l(n)/n^{1/2}$  converges with probability one.

In the present paper we present a natural extension of Ulam's problem to  $R^d$ ,  $d \geq 2$ , and provide results on the limit behavior of the analogue to  $l(n)$ . These results are given in the first section of this paper and are presented in a form which makes them immediately applicable to a question in mathematical statistics raised by Richardson and Wright [10].

To state their problem we need to introduce the concept of a "lower layer." If  $x = (\zeta^{(1)}, \zeta^{(2)}, \dots, \zeta^{(d)})$  and  $y = (\tilde{\zeta}^{(1)}, \tilde{\zeta}^{(2)}, \dots, \tilde{\zeta}^{(d)})$  are two points in  $R^d$ , we write  $x \ll y$  if  $\zeta^{(i)} \leq \tilde{\zeta}^{(i)}$  for  $i = 1, 2, \dots, d$ . A Borel subset  $A$  of  $R^d$  is called a

---

Received November 20, 1975.

AMS 1970 classification numbers. 60C05, 60F15, 60K99.

Key words and phrases. Monotone subsequences, lower layers, partial ordering, discrepancy functions, subadditive processes.

lower layer if for any  $y \in A$  we have  $\{x: x \ll y\} \subset A$ . We will denote this class of subsets by  $\mathcal{L}$ . To check ones understanding of this definition, he should note that in  $R^2$  a lower layer corresponds to the set of points which lie below the graph of a monotone decreasing function. Also we note that for any  $d \geq 1$ , the class of sets  $(-\infty, \zeta^{(1)}] \times (-\infty, \zeta^{(2)}] \times \dots \times (-\infty, \zeta^{(d)}]$  are a small subclass of the class of lower layers, which we will denote by  $Q$ .

We denote by  $1_A$  the characteristic function of the set  $A$  and we denote by  $X_i, i = 1, 2, 3 \dots$  a sequence of i.i.d. random variables which take values in  $R^d$ . Next we define random variables  $\pi_n$  and  $\tilde{\pi}_n$  by

$$\pi_n = \sup_{A \in \mathcal{L}} \left| \frac{1}{n} \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A) \right|$$

and

$$\tilde{\pi}_n = \sup_{A \in Q} \left| \frac{1}{n} \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A) \right|.$$

Of course,  $\tilde{\pi}_n$  is the classical Kolmogorov–Smirnov statistic and describes the discrepancy between the empirical distribution and the theoretical distribution function. For the generalized discrepancy  $\pi_n$ , Blum [1] proved that  $\pi_n \rightarrow 0$  with probability one provided the  $X_i$  have a density (see also Dehardt [2]). Now one of the deepest results on the classical discrepancy  $\tilde{\pi}_n$  is Kiefer’s law of the iterated logarithm [6], and the question raised by Robertson and Wright [10] is whether a similar result holds for  $\pi_n$ . That is, does one have

$$\lim_{n \rightarrow \infty} \frac{(2n)^{\frac{1}{2}} \pi_n}{(\log \log n)^{\frac{1}{2}}} = 1 \quad \text{with probability 1?}$$

The answer to this question for  $d \geq 3$  is proved in the second section of this paper to be NO. The results provided there show that the rate of growth of  $\pi_n$  is much greater than that of  $\tilde{\pi}_n$ . The question of the possibility of a law of the iterated logarithm for  $\pi_n$  in  $R^2$  is still open.

In the final section of this paper we offer two easily stated conjectures whose solution would be of interest, either from the viewpoint of combinatorial geometry or mathematical statistics.

**1. The probability of long chains.** The section is concerned primarily with chains of points in  $R^d$  with respect to the partial ordering  $\ll$  defined by  $(\zeta_1, \zeta_2, \dots, \zeta_d) \ll (\zeta'_1, \zeta'_2, \dots, \zeta'_d)$  if and only if  $\zeta_i \leq \zeta'_i$  for  $i = 1, 2, \dots, d$ . A set which is totally ordered by  $\ll$  is called a *chain*, and a set such that no two of its elements are comparable is called an *antichain*. The only fact we require from the theory of partial orderings is the Dilworth decomposition theorem:

*If a partially ordered set  $P$  possesses no antichain of cardinal  $m + 1$ , then it can be expressed as the union of  $m$  chains.*

(For a brief and elegant proof of Dilworth’s theorem see Tverberg [13].)

To gain a quick view of the power of Dilworth's theorem one should note that applied to  $(R^2, \ll)$  it immediately implies the result of Erdős and Szekeres mentioned in the introduction.

Now if  $\{y_1, y_2, \dots, y_n\}$  is a set of points in  $R_d$  we write  $\Lambda_d^+(y_1, y_2, \dots, y_n)$  for the cardinal of the largest subset of  $\{y_1, y_2, \dots, y_n\}$  which is a chain. Similarly  $\Lambda_d^-(y_1, y_2, \dots, y_n)$  is the cardinal of the largest antichain of  $\{y_1, y_2, \dots, y_n\}$ . If  $X_i$   $i = 1, 2, \dots$  are independent random variables which are uniformly distributed in  $[0, 1]^d$  we write

$$\Lambda_d^+(n) = \Lambda_d^+(X_1, X_2, \dots, X_n)$$

and

$$\Lambda_d^-(n) = \Lambda_d^-(X_1, X_2, \dots, X_n).$$

These random variables have direct relevance to a variety of problems in probability and statistics. The remainder of this section develops inequalities for these variables and establishes results on their limit behavior.

**THEOREM 1. (Recursive inequalities).** For  $d \geq 3$  and any positive integers  $n, \alpha$  and  $\beta$  we have

$$(1.1) \quad P(\Lambda_d^+(n) \geq \beta) - P(\Lambda_{d-1}^+(n) \geq \alpha) \leq nP(\Lambda_2^+(\alpha) \geq \beta)$$

and

$$(1.2) \quad P(\Lambda_d^+(n) \leq \beta) - P(\Lambda_{d-1}^+(n) \leq \alpha) \leq nP(\Lambda_2^+(\alpha) \leq \beta).$$

**PROOF.** The proof depends upon constructing a set of functions which will be used to carry estimates from  $d - 1$  dimensions to estimates for  $d$  dimensions. We let  $y_1, y_2, \dots, y_n$  be a set of points in  $[0, 1]^{d-1}$  such that no two of the  $y_i$  have equal first coordinate. By Dilworth's theorem we can decompose the set  $\{y_1, y_2, \dots, y_n\}$  into  $\Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$  chains. We denote these chains by  $C_i$ ,  $i = 1, 2, \dots, \Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$  and assume that as the index increases the first coordinate of the least element of the chain increases. Now we can define a set of  $\Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$  functions  $\lambda_i(y_1, y_2, \dots, y_n; x_1, x_2, \dots, x_n)$  where the  $x_i$  are distinct points in  $[0, 1]$ . The  $\lambda_i$  are defined as follows:

$\lambda_i(y_1, y_2, \dots, y_n; x_1, x_2, \dots, x_n)$  is the cardinality of the largest subset  $B_i \subset C_i$  such that for any two elements  $y_{j_1}, y_{j_2}$  of  $B_i$  such that  $y_{j_1} \ll y_{j_2}$  we have  $x_{j_1} < x_{j_2}$ .

If  $z_1, z_2, \dots, z_n$  are i.i.d. random variables with uniform distribution in  $[0, 1]$  we write

$$\lambda_i(y, z) = \lambda_i(y_1, \dots, y_n; z_1, z_2, \dots, z_n).$$

We note that  $\lambda_i(y, z)$  is a random variable which has a distribution that depends only on the cardinality of the chain  $C_i$ . Moreover, we note

$$(1.3) \quad \text{If } \Lambda_{d-1}^+(y_1, y_2, \dots, y_n) \leq \alpha \text{ we have} \\ P(\lambda_i(y_1, y_2, \dots, y_n; z_1, z_2, \dots, z_n) \geq \beta) \leq P(\Lambda_2^+(\alpha) \geq \beta).$$

The second basic observation is that

$$(1.4) \quad P(\Lambda_d^+(n) \geq \beta) = \int_{[0,1]^{n(d-1)}} P(\max_i \lambda_i \geq \beta) dy$$

where the maximum is taken over  $i = 1, 2, \dots, \Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$ .

These two observations follow directly from the definition of the  $\lambda_i$ , and with these observations under our belt the proof of the theorem proceeds by computation. To begin we have

$$(1.5) \quad P(\Lambda_d^+(n) \geq \beta) \leq \int_A P(\max_i \lambda_i \geq \beta) dy + P(\Lambda_{d-1}^+(n) \geq \alpha)$$

where  $A = \{y: \Lambda_{d-1}^+(y_1, y_2, \dots, y_n) \leq \alpha\}$ , and we have used the fact that  $\int_{A^c} P(\max_i \lambda_i \geq \beta) dy \leq \int_{A^c} dy = P(\Lambda_{d-1}^+(n) > \alpha)$ . Next we have

$$(1.6) \quad \int_A P(\max_i \lambda_i \geq \beta) dy \leq \int_A \sum_i P(\lambda_i \geq \beta) dy$$

where the sum over  $i$  is for  $i$  from 1 to  $\Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$ . Since  $\Lambda_{d-1}^+(y_1, y_2, \dots, y_n) \leq \alpha$  on  $A$  we have for  $y \in A$  that

$$(1.7) \quad P(\lambda_i(y_1, y_2, \dots, y_n; z_1, z_2, \dots, z_n) \geq \beta) \leq P(\Lambda_2^+(\alpha) \geq \beta)$$

by the application of inequality (1.3). This shows that (1.6) reduces to

$$(1.8) \quad \int_A P(\max_i \lambda \geq \beta) dy \leq nP(\Lambda_2^+(\alpha) \geq \beta)$$

where we have also applied the trivial fact that for all  $y$ ,  $\Lambda_{d-1}^-(y_1, y_2, \dots, y_n) \leq n$ . Now by applying inequality (1.8) to the basic (1.5) we obtain

$$(1.9) \quad P(\Lambda_d^+(n) \geq \beta) \leq nP(\Lambda_2^+(\alpha) \geq \beta) + P(\Lambda_{d-1}^+(n) \geq \alpha)$$

and the first inequality of the theorem is proved.

To prove the second half we note that

$$(1.10) \quad P(\Lambda_d^+(n) \leq \beta) \leq P(\Lambda_{d-1}^+(n) \leq \alpha) + \int_B P(\max_i \lambda_i \leq \beta) dy$$

where  $B = \{y: \Lambda_{d-1}^+(y_1, y_2, \dots, y_n) > \alpha\} \subset [0, 1]^{n(d-1)}$  and  $i = 1, 2, \dots, \Lambda_{d-1}^-(y_1, y_2, \dots, y_n)$ . Further we have

$$(1.11) \quad \int_B P(\max_i \lambda_i \leq \beta) dy \leq \int_B \sum_i P(\lambda_i \leq \beta) dy \leq nP(\Lambda_2^+(\alpha) \leq \beta),$$

by the same proof as that of (1.8). When this last inequality is applied to (1.10) the inequality of the second half of the theorem is proved.

To make use of the recursive inequalities of Theorem 1, both tails of the distributions of  $\Lambda_2^+(n)$  must be estimated. We obtain such estimates with the following results.

LEMMA 1. For all positive integers  $n$ , we have

$$(1.12) \quad P(\Lambda_2^+(n) \geq 8(n)^{\frac{1}{2}}) \leq \gamma^{n^{\frac{1}{2}}}$$

and

$$(1.13) \quad P(\Lambda_2^+(n) \leq n^{\frac{1}{2}}/8) \leq \gamma^{n^{\frac{1}{2}}}$$

where  $\gamma = (e/4)^4 < 1$ .

PROOF. The method for proving the first inequality is a slight variant of a method due to Hammersley [4] which we include for completeness. Let  $X_i = (X_i^{(1)}, X_i^{(2)})$   $i = 1, 2, \dots, n$  be i.i.d. random variables which are uniformly distributed in  $[0, 1]^2$ . Next let  $\hat{X}_i = X_j^{(2)}$  if  $X_j^{(1)}$  is the  $i$ th largest element of the set  $\{X_1^{(1)}, X_2^{(1)}, \dots, X_r^{(1)}\}$ . Now choose  $r$  so that  $2r < n$  and let  $\nu$  be the number of sequences  $i_1 < i_2 < \dots < i_r \leq n$  such that  $\hat{X}_{i_1} < \hat{X}_{i_2} < \dots < \hat{X}_{i_r}$ . By the uniform distribution of the  $\hat{X}_i$  we have that

$$(1.14) \quad E(\nu) = \sum_{i_1 < i_2 < \dots < i_r} P(\hat{X}_{i_1} > \hat{X}_{i_2} \dots > \hat{X}_{i_r}) = \binom{n}{r} (r!)^{-1}.$$

But if  $\Lambda_2^+(X_1, X_2, \dots, X_n) \geq 2r$  then  $\nu \geq \binom{2r}{r}$  so by Chebyshev's inequality and (1.14) we have

$$(1.15) \quad P(\Lambda_2^+(X_1, X_2, \dots, X_n) \geq 2r) \leq \binom{n}{r} (r! \binom{2r}{r})^{-1}.$$

By Stirling's formula applied to (1.15) and by letting  $r = 4[n^{\frac{1}{2}}]$  we obtain inequality (1.12) of the lemma.

For the second half of the lemma, we first observe that Dilworth's theorem implies that  $\Lambda_2^+(n)\Lambda_2^-(n) \geq n$ . This shows that  $\Lambda_2^-(X_1, X_2, \dots, X_n) \leq n^{\frac{1}{2}}/8$  implies  $\Lambda_2^+(X_1, X_2, \dots, X_n) \geq 8(n)^{\frac{1}{2}}$  so by the first half of the lemma we obtain

$$(1.16) \quad P(\Lambda_2^-(X_1, X_2, \dots, X_n) \leq n^{\frac{1}{2}}/8) \leq \gamma^{n^{\frac{1}{2}}}.$$

But the distribution of  $\Lambda_2^-(X_1, X_2, \dots, X_n)$  is the same as the distribution of  $\Lambda_2^+(X_1, X_2, \dots, X_n)$  so (1.16) implies the second half of the lemma.

The inequalities of Theorem 1 and Lemma 1 can now be applied to yield a results on the limit behavior of  $\Lambda_d^+(n)$ .

THEOREM 2. For  $d \geq 2$  we have with probability one that

$$(1.17) \quad \lim_{n \rightarrow \infty} \log \Lambda_d^+(n) / \log n = 2^{1-d}.$$

PROOF. We first consider  $d \geq 3$ . For any sequence of positive integers  $\beta_1, \beta_2, \dots, \beta_d$  we have

$$(1.18) \quad P(\Lambda_k^+(n) \geq \beta_k) - P(\Lambda_{k-1}^+(n) \geq \beta_{k-1}) \leq nP(\Lambda_2^+(\beta_{k-1}) \geq \beta_k)$$

by Theorem 1. On adding these inequalities for  $k = 3, 4, \dots, d$  we obtain

$$(1.19) \quad P(\Lambda_d^+(n) \geq \beta_d) - P(\Lambda_2^+(n) \geq \beta_2) \leq n \sum_{k=3}^d P(\Lambda_2^+(\beta_{k-1}) \geq \beta_k).$$

Now by choosing  $\beta_2 \geq 8(n)^{\frac{1}{2}}$  and  $\beta_k \geq 8(\beta_{k-1})^{\frac{1}{2}}$  we have by Lemma 1 that  $P(\Lambda_2^+(\beta_{k-1}) \geq \beta_k) \leq \gamma^{(\beta_{k-1})^{\frac{1}{2}}}$  for  $k = 3, 4, \dots, d$ . So (1.19) provides the estimate

$$P(\Lambda_d^+(n) \geq \beta_d) \leq \gamma^{n^{\frac{1}{2}}} + n \sum_{k=3}^d \gamma^{(\beta_{k-1})^{\frac{1}{2}}}$$

and consequently

$$(1.20) \quad \sum_{n=1}^{\infty} P(\Lambda_d^+(n) \geq 2^8 n^{2^{1-d}}) < \infty.$$

The same procedure based on the second inequalities of Theorem 1 and Lemma 1 provide

$$(1.21) \quad \sum_{n=1}^{\infty} P(\Lambda_d^+(n) \leq 2^{-8} n^{2^{1-d}}) < \infty.$$

The Borel–Cantelli lemma thus gives us that  $2^{-8} \leq \liminf \Lambda_d^+(n)/n^{2^1-d} \leq \limsup \Lambda_d^+(n)/n^{2^1-d} \leq 2^8$  with probability one. The result (1.17) of Theorem 2 now follows by taking logarithms and dividing by  $\log n$ . For the case  $d = 2$  we can go directly to (1.19) from Lemma 1 and thus obtain the analogous result to that obtained for  $d \geq 3$ , but as noted in the introduction more precise results have already been obtained for  $d = 2$ .

COROLLARY 1. For  $d \geq 2$  we have

$$(1.22) \quad \liminf \log_{n \rightarrow \infty} \Lambda_d^-(n)/\log n \geq 1 - 2^{1-d}$$

with probability one.

PROOF. By Dilworth’s theorem we have  $\log \Lambda_d^+(n) + \log \Lambda_d^-(n) \geq \log n$  so (1.22) is an immediate consequence of Theorem 1.

**2. Application to discrepancy functions.** We will now focus our attention on the discrepancy function for the class of lower layers defined in the introduction. The next result shows in a strong way that the law of the iterated logarithm is quite far from the nature of the  $\pi_n$  in  $R^d$ ,  $d \geq 3$ .

THEOREM 3. Let  $X_i$   $i = 1, 2, \dots$  be i.i.d. uniformly distributed random variables in  $[0, 1]^d$ ,  $d \geq 2$ . For the generalized discrepancy  $\pi_n = \sup_{A \in L} |(1/n) \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A)|$  where  $L$  is the class of lower layers we have

$$(2.1) \quad \pi_n \geq \frac{1}{2n} \Lambda_d^-(n)$$

and consequently

$$(2.2) \quad \liminf_{n \rightarrow \infty} n^{2^1-d} \pi_n \geq 2^{-9} \quad \text{a.s.}$$

PROOF. For any  $A \in L$  we can define for  $0 < \theta < 1$  another element of  $L$  by  $\theta A = \{y : y = \theta x, x \in A\}$ . We note that  $\lim_{\theta \rightarrow 1} P(X_1 \in \theta A) = P(X_1 \in A)$ . Further, if we choose  $A$  such that  $A$  has piecewise linear boundary and such that  $\Lambda_d^-(n)$  of the elements of  $\{X_1(w), X_2(w), \dots, X_n(w)\}$  are boundary points of  $A$  then we have

$$(2.3) \quad \sum_{i=1}^n 1_A(X_i) - \sum_{i=1}^n 1_{\theta A}(X_i) \geq \Lambda_d^-(n).$$

By (2.3) we immediately obtain that

$$(2.4) \quad n\pi_n \geq 2^{-1}\Lambda_d^-(n)$$

Thus the first half of the theorem is proved. Finally (2.2) follows from (2.1), (1.20) and (1.21).

The preceding result shows that for  $d \geq 3$  there is no log log law for  $\pi_n$ . This can also be obtained by a completely different method as has recently been proved by W. Schmidt [12, page 387]. If  $\lambda$  is used to denote Lebesgue measure in  $R^d$ , and  $x_1, x_2, \dots, x_n$  is any sequence of points in  $R^d$ , W. Schmidt has proved that

$$(2.5) \quad \sup_{A \in L} \left| \frac{1}{n} \sum_{i=1}^n 1_A(X_i) - \lambda(A) \right| \geq c_d n^{-1/d}$$

where  $c_d$  is a constant depending only on  $d$ . The rate of growth of  $n\pi_n$  given by Schmidt's result is not as great as that provided by (2.2) but, of course, that is more than to be expected since his result pertains to any sequence of points  $x_1, x_2, \dots$  no matter how favorably chosen. In the case of  $d = 2$  both Schmidt's result and the result of (2.2) are compatible with the log log law.

Our next inequality provides an upper bound for  $\pi_n$  in the case of  $R^2$ . This result was suggested by an inequality of Hlawka (see [8, 229-233]) for the class of convex subsets of  $R^d$ . Letting  $\bar{\pi}_n = \sup_{A \in \mathcal{C}} |(1/n) \sum_{i=1}^n 1_A(x_i) - \lambda(A)|$  and  $\mathcal{C}$  the class of convex subsets of  $[0, 1]^d$ , Hlawka proved essentially that  $\bar{\pi}_n \leq \mathcal{C}_d(\bar{\pi}_n)^{1/d}$ . It would be futile for us to seek to prove that  $\pi_n \leq \mathcal{C}_d'(\bar{\pi}_n)^{1/d}$  for all  $d \geq 2$  since by Kiefer's theorem  $\bar{\pi}_n = O((n^{-1} \log \log n)^{1/2})$  whereas by (2.2) we know that  $\pi_n$  is not  $O((\log \log n)/n)^{1/2d}$  if  $d \geq 5$ . We therefore provide the analog for the class of lower layers of Hlawka's inequality only in the case  $d = 2$ .

**THEOREM 4.** For  $R^d$  with  $d = 2$  we have the inequality

$$(2.6) \quad \bar{\pi}_n \leq \pi_n \leq 4(\bar{\pi}_n)^{1/2}.$$

**PROOF.** The first inequality of (2.6) is, of course, valid for  $d \geq 2$  and is trivial. To prove the second inequality we first define

$$(2.7) \quad n\bar{\pi}_n' = \sup_{A \in \mathcal{R}} |\sum_{i=1}^n 1_A(X_i) - P(X_1 \in A)|$$

where  $\mathcal{R}$  is the class of all rectangles of the form  $[0, \gamma] \times [j/r, (j+1)/r]$ . Since any such rectangle can be written as  $[0, \gamma] \times [0, \beta] \setminus [0, \gamma] \times [0, \beta]$  we have  $\bar{\pi}_n' \leq 2\bar{\pi}_n$ . Next let  $r$  be any positive integer and suppose  $A$  is a lower layer. It is easy to prove (but tedious to write down) that for any lower layer  $A$  there are polygons  $P_1, P_2$  with the following properties:

- (i) Each  $P_j$  is the union of  $r$  rectangles of the form  $[0, \alpha] \times [j/r, (j+1)/r]$ ,
- (ii)  $P_1 \subset A \subset P_2$ , and
- (iii)  $P(X_1 \in P_2 \setminus P_1) = 1/r$ .

From these properties we have that

$$\max_{j=1,2} \left| \frac{1}{n} \sum_{i=1}^n 1_{P_j}(X_i) - P(X_1 \in P_j) \right| \leq r\bar{\pi}_n' \leq 2r\bar{\pi}_n.$$

Further by (iii) we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n 1_{P_1}(X_i) - P(X_1 \in P_1) + P(X_1 \in P_1) - P(X_1 \in A) \\ & \leq \frac{1}{n} \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A) \\ & \leq \frac{1}{n} \sum_{i=1}^n 1_{P_2}(X_i) - P(X_1 \in P_2) + P(X_1 \in P_2) - P(X_1 \in A). \end{aligned}$$

This shows as a consequence of (iii) that

$$\left| \frac{1}{n} \sum_{i=1}^n 1_A(X_i) - P(X_1 \in A) \right| \leq 2r\bar{\pi}_n + \frac{1}{r}.$$

Since  $A$  was arbitrary we have proved that

$$(2.8) \quad \pi_n \leq 2r\bar{\pi}_n + \frac{1}{r} \quad \text{for all } r.$$

By taking  $r = [\bar{\pi}_n^{-1}]$  the inequality (2.6) is established.

At this point a note should be made about the generality of the results given here. The assumption that the  $X_i$  are uniformly distributed has been used through this paper only to avoid making repeated reductions to this case. Since the class of lower layers is preserved under mappings of  $R^d$  which are monotone in each coordinate, the study of arbitrary  $X_i$  with a density can be reduced to the case of the uniform distribution (see e.g. Rosenblatt [11]). Indeed, all of the results of this paper are valid under the assumption that the  $X_i$  have a density.

**3. Two conjectures.** In this final section we offer two conjectures.

CONJECTURE 1.

$$\lim_{n \rightarrow \infty} \Lambda_d^+(n)/n^{2-d} = c < \infty.$$

CONJECTURE 2.

$$\lim_{n \rightarrow \infty} n^{\frac{1}{2}} \pi_n / (\log \log n)^{\frac{1}{2}} = \infty \quad \text{a.s.} \quad \text{when } d = 2.$$

The first conjecture is, of course, a stronger version of our Theorem 2 and is the “one rightful heir” of the beautiful theorem of Hammersley–Kingman–Kesten which solves Ulam’s problem. The second conjecture would complete the answer to the question raised by Richardson and Wright. One should note that Philipp [9] has proved the log log law in  $R^2$  for the convex sets; so, although the second conjecture is not as central as the first, there is still, perhaps, something subtle to uncover.

#### REFERENCES

- [1] BLUM, J. (1955). On convergence of empirical distribution functions. *Ann. Math. Statist.* **26** 527–529.
- [2] DEHARDT, J. (1971). Generalization of the Glivenko–Cantelli theorem. *Ann. Math. Statist.* **42** 2050–2055.
- [3] ERDÖS, P. and SZEKERES, GY. (1935). A combinatorial problem in geometry. *Composito* **2** 463–470.
- [4] HAMMERSLEY, J. M. (1972). A few seedlings of research. *Proc. Sixth Berkeley Symp. Math. Math. Statist. Prob.* 345–394, Univ. of California Press.
- [5] HAMMERSLEY, J. M. (1974). Postulates for subadditive processes. *Ann. Probability* **2** 652–680.
- [6] KIEFER, J. (1961). On the large deviation of the empiric d.f. of vector chance variables and a law of the iterated logarithm. *Pacific J. Math.* **11** 649–660.
- [7] KINGMAN, J. F. C. (1973). Subadditive ergodic theory. *Ann. Probability* **1** 883–909.
- [8] KUIPERS, L. and NIEDERREITER, H. (1974). *Uniform Distribution of Sequences*. Wiley, Toronto.
- [9] PHILIPP, W. (1973). Empirical distribution functions and uniform distribution mod 1. *Diophantine Approximation and Its Applications* (C. F. Osgood, ed.). Academic Press, New York.
- [10] ROBERTSON, T. and WRIGHT, F. T. (1974). On the maximum likelihood estimation of stochastically ordered random variates. *Ann. Statist.* **2** 528–534.