The Annals of Applied Probability
1996, Vol. 6, No. 2, 572–576

PACKING RANDOM INTERVALS[1]

BY WANSOO T. RHEE AND MICHEL TALAGRAND

Ohio State University

A *packing* of a collection of subintervals of $[0, 1]$ is a pairwise disjoint subcollection of the intervals; its *wasted space* is the measure of the set of points not covered by the packing.

Consider n random intervals, $I_1, \dots, I_n$, chosen by selecting endpoints independently from the uniform distribution. We strengthen and simplify the results of Coffman, Poonen and Winkler, and we show that, for some universal constant K and for each $t \geq 1$, with probability greater than or equal to $1 - 1/n^t$, there is a packing with wasted space less than or equal to $Kt(\log n)^2/n$.

1. Introduction. Packing problems are of fundamental importance, in particular, in computer science [1]. If $I_1, \dots, I_n$ are intervals contained in $[0, 1]$, a packing of these intervals in $[0, 1]$ is a disjoint subcollection of these intervals. The wasted space is the length of the part of $[0, 1]$ that is not covered by this subcollection. The optimal packing minimizes wasted spaces. (So, in other words, one tries to cover as much as possible of $[0, 1]$ using disjoint intervals of the given family.) Coffman, Poonen and Winkler [2] prove the remarkable fact that an optimal packing of n random intervals $I_1, \dots, I_n$ in $[0, 1]$ wastes a space W_n of order $(\log n)^2/n$. They prove that, for every $\varepsilon > 0$, $W_n \geq (1/8 - \varepsilon)(\log n)^2/n$ with a probability that goes to 1 as $n \to \infty$, and they prove that $E(W_n) \leq K(\log n)^2/n$ for some universal constant K. In this paper, we strengthen this latter result and we prove the following theorem.

THEOREM 1. *For all* $1 \leq t \leq n/(\log n)^3$, *we have*

$$P\left(W_n \geq Kt\frac{(\log n)^2}{n}\right) \leq \left(\frac{1}{n}\right)^t, \tag{1}$$

where K *does not depend on* n.

While the proof of [2] uses the second moment method via a rather delicate computation, our approach is considerably more straightforward.

2. Poissonization. Consider $D = \{(x, y) \in [0, 1]^2;\ x \leq y\}$. There is an obvious bijection between the points of D and the subintervals of $[0, 1]$.

Received September 1995; revised January 1996.

[1]This research was supported in part by an NSF grant.

AMS 1991 *subject classifications*. 60G55, 90B05.

Key words and phrases. Wasted space, packing, random intervals, asymptotic estimate.

Through this bijection, the intervals $I_1, \dots, I_n$ appear as n points uniformly distributed over D. Following a well-known scheme, let us consider a homogeneous Poisson point process of uniform intensity λ. The process generates a finite subset Π of D, that is, a finite collection of intervals $I_1, \dots, I_r$ of $[0, 1]$. We denote by V_λ the space wasted by an optimal packing of $[0, 1]$ by these intervals. We claim that, for each n,

$$P(W_n \geq a) \leq P(V_\lambda \geq a) + P(\operatorname{card} \Pi \geq n). \tag{2}$$

Indeed, if we condition with respect to the cardinal of $\Pi = m$, V_λ is distributed like W_m and (considering only the first m intervals) $P(W_n \geq a) \leq P(W_m \geq a)$ when $n \geq m$.

Now we fix $\lambda = n/2$, so that (as is well known)

$$P(\operatorname{card} \Pi \geq n) \leq \exp\left(-\frac{n}{K}\right).$$

(There, as well as in the rest of the paper, K denotes a universal constant, not necessarily the same at each occurrence.) Thus it suffices to prove that, for $1 \leq t \leq n/(\log n)^3$,

$$P\left(V_\lambda \geq Kt\frac{(\log n)^2}{n}\right) \leq \left(\frac{1}{n}\right)^t, \tag{3}$$

where K does not depend on n.

3. The idea. Consider a parameter u and $q = \lfloor n/(u \log n) \rfloor$. We divide $[0, 1]$ into the q consecutive intervals $[k/q, (k+1)/q[$, $0 \leq k < q$, which for simplicity we will call *atoms*. For an atom A, we denote by A^- (resp. A^+) the atom to the left (right) when it exists.

We set $m_1 = \lfloor \log n \rfloor$. We divide the set of atoms into $3m_1 + 2$ consecutive blocks. The first and the last blocks consist, respectively, of the first m_1 and the last m_1 atoms. The $q - 2m_1$ atoms left are divided into $3m_1$ blocks $B_2, \dots, B_{3m_1+1}$ of consecutive atoms, each of them containing either $p = \lfloor (q - 2m_1)/3m_1 \rfloor$ or $p + 1 = \lceil (q - 2m_1)/3m_1 \rceil$ atoms. Now, when $u \leq n/3(\log n)^2$ (and n is larger than some fixed integer n_0), we have $p \geq n/4u(\log n)^2$.

We now define, by induction over k, the set of atoms contained in the block B_k that are *alive*. For $k = 1$, all the atoms contained in B_k are alive. Now we say that an atom A contained in B_{k+1} is alive if there is an atom A_0 contained in B_k that is alive and such that, among the intervals $I_1, \dots, I_r$, we can find one with endpoints in A_0^+ and A.

Thus an atom A of B_{k+1} is alive if, among $I_1, \dots, I_r$, we can find intervals $J_1, \dots, J_k$ with the following properties:

(4) The left endpoint of J_1 belongs to B_1.

(5) For $2 \leq l \leq k$, the right endpoint of J_{l-1} and the left endpoint of J_l belongs to two consecutive atoms of B_l.

What this means is that we have succeeded in constructing a "partial packing" $J_1, \dots, J_k$ of $[0,1]$ starting at 0 and up to A. This partial packing is efficient in the sense that the following occurs:

(6) At most $m_1/q \le 2u(\log n)^2/n$ space is not covered at the left of the left endpoint of J_1.

(7) The gap between J_l and J_{l+1} is at most $2/q$.

Suppose now that we can prove the following for $u \ge K$, $u \le n/(\log n)^3$:

(8) With probability at least $1 - 3\exp(-u\log n/K)$, the number M_k of *live* intervals in B_k satisfies $M_k \ge \min\{3^{k-1}m_1, 2p/3\}$ for each $k \le 2m_1$.

Then we observe that $3^{k-1}m_1 \ge n$ for $k \ge \log n$; thus $M_k \ge 2p/3$ for $k \ge \log n$. This implies that, if we consider a block B near the center of $[0,1]$, at least (approximately) $2/3$ of its atoms contain the right endpoint of the last interval of a partial efficient packing [in the sense that (6) and (7) hold]. Now we could have constructed these partial efficient packings starting at the right of $[0,1]$ rather than the left, and again at least $2/3$ (approximately) of the atoms of B would contain the left endpoint of the last (starting from the right) interval of such a packing. So we can find an atom J of B that contains the end of a partial efficient packing starting from the left, while J^+ contains the end of a partial efficient packing starting from the right. The union of these two packings wastes at most

$$4u\frac{(\log n)^2}{n} + \frac{3m_1}{q} \le Ku\frac{(\log n)^2}{n}.$$

Combining this with (8) shows that, for $u \le n/(\log n)^3$,

$$P\left(V_\lambda \ge Ku\frac{(\log n)^2}{n}\right) \le 4\left(\frac{1}{n}\right)^{u/K}.$$

This implies (1).

4. End of the proof. Let us denote by b_k the right endpoint of B_k, and set $\Pi_k = \Pi \cap ([0, b_k] \times [0,1])$. We observe that the set of live atoms of B_k depends on Π_{k-1} only. First, we show that, to prove (8), it suffices to prove the following:

$$\begin{aligned} P(M_{k+1} \ge \min\{3M_k, 2p/3\} | \Pi_{k-1}) \\ \ge 1 - \exp(-u\min\{3M_k, 2p/3\}/K). \end{aligned} \tag{9}$$

Indeed, using (9), we get, setting $n_k = \min\{3^{k-1}m_1, 2p/3\}$, that

$$P(M_{k+1} \ge n_{k+1}) \ge P(M_k \ge n_k) - \exp(-n_k u/K),$$

so that

$$P(\forall\, k \le 2m_1,\, M_k \ge n_k) \ge 1 - \sum_{l \le 2m_1} \exp(-n_l u/K)$$

and the latter sum is bounded by $\exp(-um_1/K)$ if $u \le N/\log N$. Thus the only task left is to prove (9).

There is a set $\mathscr{A}$ of atoms of B_k such that $\operatorname{card}\mathscr{A} \ge M_k - 1 \ge M_k/2$ and that, whenever $A \in \mathscr{A}$, A^- is alive. Consider for an atom B of B_{k+1} the random variable δ_{AB} that is equal to 1 if $\Pi \cap (A \times B) \ne \varnothing$, and is 0 otherwise. The set of atoms of B_{k+1} that are alive is

$$\mathscr{B} = \{B \in B_{k+1} \colon \exists\, A \in \mathscr{A},\ \delta_{AB} = 1\}.$$

Conditionally on Π_{k-1}, the random variables δ_{AB} are independent and

$$P(\delta_{AB} = 0) = \exp(-\lambda \operatorname{Area}(A \times B)) \le \exp\bigl(-u^2(\log n)^2/(2n)\bigr).$$

Thus

$$\forall\, B \in B_{k+1},$$
$$P(B \notin \mathscr{B} | \Pi_{k-1}) \le \exp\bigl(-u^2(\log n)^2 M_k/(4n)\bigr) \le \exp(-uM_k/(16p)) := \tau.$$

Thus, conditionally on Π_{k-1}, the number of live atoms in B_{k+1} dominates the number $H(p, 1-\tau)$ of outcomes of a sequence of p independent Bernoulli trials, each with probability $1-\tau$ of success. We need the following lemma which is a very weak form of the Chernoff bounds [3].

Lemma 1. *For some universal constant K_0, we have*

$$a \ge 3/4 \Rightarrow P(H(p,a) \le 2p/3) \le \exp(-p/K_0), \tag{10}$$

$$P(H(p,a) \le ap/2) \le \exp(-ap/K_0). \tag{11}$$

We finish the proof. For clarity, we distinguish two cases.

Case 1. In this case $uM_k/(16p) \ge 2$. Then $1 - \tau \ge 3/4$, and the conclusion holds by (10).

Case 2. In this case $uM_k/(16p) \le 2$. Here, we use that $1 - \exp(-x) \ge x/4$ for $x \le 2$ to get $1 - \tau \ge uM_k/(64p)$. We use (11) with $a = uM_k/(64p)$, so that $ap/2 \ge 3M_k$ provided $u \ge 384$. Also, $ap = uM_k/64$, and $3M_k = \min(3M_k, 2p/3)$ since $M_k \le 32p/u$. This completes the proof. □

REFERENCES

[1] Coffman, E. G., Jr. and Lueker, G. S. (1991). *Probabilistic Analysis of Packing and Partitioning Algorithms*. Wiley, New York.

[2] Coffman, E. G., Jr., Poonen, B. and Winkler, P. Packing random intervals. *Probab. Theory Related Fields*. To appear.

[3] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

Department of Management Sciences
Ohio State University
1775 College Road
Columbus, Ohio 43210
E-mail: rhee.1@osu.edu

Department of Mathematics
Ohio State University
231 West 18th Avenue
Columbus, Ohio 43210
and
Equipe d'Analyse–Tom 46
U.A. au CNRS
Universite Paris VI
4 Place Jussieu
75230 Paris Cedex 05
France
E-mail: talagrand.1@osu.edu