Average-Case Analysis of Rectangle Packings

E.G. Coffman, Jr.[1], George S. Lueker[2], Joel Spencer[3], and Peter M. Winkler[4]

[1] New Jersey Institute of Technology, Newark, NJ 07102
[2] University of California, Irvine, Irvine, CA 92697
[3] New York University, New York, NY 10003
[4] Bell Labs, Lucent Technologies, Murray Hill, NJ 07974

Abstract. We study the average-case behavior of algorithms for finding a maximal disjoint subset of a given set of rectangles. In the probability model, a random rectangle is the product of two independent random intervals, each being the interval between two points drawn uniformly at random from $[0, 1]$. We have proved that the expected cardinality of a maximal disjoint subset of n random rectangles has the tight asymptotic bound $\Theta(n^{1/2})$. Although tight bounds for the problem generalized to $d > 2$ dimensions remain an open problem, we have been able to show that $\Omega(n^{1/2})$ and $O((n \log^{d-1} n)^{1/2})$ are asymptotic lower and upper bounds. In addition, we can prove that $\Theta(n^{d/(d+1)})$ is a tight asymptotic bound for the case of random cubes.

1 Introduction

We estimate the expected cardinality of a maximal disjoint subset of n rectangles chosen at random in the unit square. We say that such a subset is a *packing* of the n rectangles, and stress that a rectangle is specified by its position as well as its sides; it can not be freely moved to any position such as in strip packing or two-dimensional bin packing (see [2] and the references therein for the probabilistic analysis of algorithms for these problems). A random rectangle is the product of two independent random intervals on the coordinate axes; each random interval in turn is the interval between two independent random draws from a distribution G on $[0, 1]$.

This problem is an immediate generalization of the one-dimensional problem of packing random intervals [3]. And it generalizes in an obvious way to packing random rectangles (boxes) in $d > 2$ dimensions into the d-dimensional unit cube, where each such box is determined by $2d$ independent random draws from $[0, 1]$, two for every dimension. A later section also studies the case of random cubes in $d \geq 2$ dimensions. For this case, to eliminate irritating boundary effects that do not influence asymptotic behavior, we wrap around the dimensions of the unit cube to form a toroid. In terms of an arbitrarily chosen origin, a random cube is then determined by $d+1$ random variables, the first d locating the vertex closest to the origin, and the last giving the size of the cube, and hence the coordinates of the remaining $2^d - 1$ vertices. Each random variable is again an independent random draw from the distribution G.

G. Gonnet, D. Panario, and A. Viola (Eds.): LATIN 2000, LNCS 1776, pp. 292–297, 2000.

Applications of our model appear in jointly scheduling multiple resources, where customers require specific “intervals” of a resource or they require a resource for specific intervals of time. An example of the former is a linear communication network and an example of the latter is a reservation system. In a linear network, we have a set S of call requests, each specifying a pair of endpoints (calling parties) that define an interval of the network. If we suppose also that each request gives a future time interval to be reserved for the call, then a call request is a rectangle in the two dimensions of space and time. In an unnormalized and perhaps discretized form, we can pose our problem of finding the expected value of the number of requests in S that can be accommodated.

The complexity issue for the combinatorial version of our problem is easily settled. Consider the two-dimensional case, and in particular a collection of equal size squares. In the associated intersection graph there is a vertex for each square and an edge between two vertices if and only if the corresponding squares overlap. Then our packing problem specialized to equal size squares becomes the problem of finding maximal independent sets in intersection graphs. It is easy to verify that this problem is NP-complete. For example, one can use the approach in [1] which was applied to equal size circles; the approach is equally applicable to equal size squares. We conclude that for any fixed $d \geq 2$, our problem of finding maximal disjoint subsets of rectangles is NP-complete, even for the special case of equal size cubes. As a final note, we point out that, in contrast to higher dimensions, the one-dimensional (interval) problem has a polynomial-time solution [3].

Let S_n be a given set of random boxes, and let C_n be the maximum cardinality of any set of mutually disjoint boxes taken from S_n. After preliminaries in the next section, Section 3 proves that, in the case of cubes in $d \geq 2$ dimensions, $\mathsf{E}[C_n] = \Theta(n^{d/(d+1)})$, and Section 4 proves that, in the case of boxes in d dimensions, $\mathsf{E}[C_n] = \Omega(n^{1/2})$ and $\mathsf{E}[C_n] = O((n \log^{d-1} n)^{1/2})$. Section 5 contains our strongest result, which strengthens the above bounds for $d = 2$ by presenting a $O(n^{1/2})$ tight upper bound. We sketch a proof that relies on a similar result for a reduced, discretized version of the two dimensional problem.

2 Preliminaries

We restrict the packing problem to continuous endpoint distributions G. Within this class, our results are independent of G, because the relevant intersection properties of G depend only on the relative ordering of the points that determine the intervals in each dimension. Thus, for simplicity, we assume hereafter that G is the uniform distribution on $[0, 1]$.

It is also easily verified that we can Poissonize the problem without affecting our results. In this version, the number of rectangles is a Poisson distributed random variable T_n with mean n, and we let $C(n)$ denote the number packed in a maximal disjoint subset. We will continue to parenthesize arguments in the notation of the Poissonized model so as to distinguish quantities like C_n in the model where the number of rectangles to pack is fixed at n.

Let $X_1, \ldots, X_n$ be i.i.d. with a distribution F concentrated on $[0,1]$. We assume that F is *regularly varying* at 0 in that it is strictly increasing and that, for some $\xi > 0$, some constants $K, K' \in (0,1)$, and all $x \in (0,\xi)$, it satisfies $\frac{F(x/2)}{F(x)} \in [K, K']$. For $(s_n \in (0,1]; n \geq 1)$ a given sequence, let $N_n(F, s_n)$ be the maximum number of the X_i that can be chosen such that their sum is at most ns_n on average. Equivalently, in terms of expected values, N_n is such that the sum of the smallest N_n of the X_i is bounded by ns_n, but the sum of the smallest $N_n + 1$ of the X_i exceeds ns_n.

Standard techniques along with a variant of Bernstein's Theorem suffice to prove the following technical lemma.

Lemma 1. *With F and $(s_n, n \geq 1)$ as above, let x_n be the solution to*

$$s_n = \int_0^{x_n} x dF(x), \tag{1}$$

and assume the s_n are such that $\lim_{n\to\infty} x_n = 0$. Then if $\lim_{n\to\infty} nF(x_n) = \infty$ and $nF(x_n) = \Omega(\log^2 s_n^{-1})$, we have

$$\mathsf{E}[N_n(F, s_n)] \sim nF(x_n)\,. \tag{2}$$

3 Random Cubes

The optimum packing of random cubes is readily analyzed. We work with a d-dimensional unit cube, and allow (toroidal) wrapping in all axes. The n cubes are generated independently as follows: First a vertex $(v_1, v_2, \ldots, v_d)$ is generated by drawing each c_i independently from the uniform distribution on $[0,1]$. Then one more value w is drawn independently, again uniformly from $[0,1]$. The cube generated is

$$[v_1, v_1 + w) \times [v_2, v_2 + w) \times \cdots \times [v_d, v_d + w),$$

where each coordinate is taken modulo 1. In this set-up, we have the following result.

Theorem 1. *The expected cardinality of a maximum packing of n random cubes is $\Theta(n^{d/(d+1)})$.*

Proof: For the lower bound consider the following simple heuristic. Subdivide the cube into c^{-d} cells with sides

$$c = \alpha n^{-1/(d+1)},$$

where α is a parameter that may be chosen to optimize constants. For each cell $\mathcal{C}$, if there are any generated cubes contained in $\mathcal{C}$, include one of these in the packing. Clearly, all of the cubes packed are nonoverlapping.

One can now show that the probability that a generated cube fits into a particular cell $\mathcal{C}$ is $c^{d+1}/(d+1)$, and so the probability that $\mathcal{C}$ remains empty after generating all n cubes is

$$\left(1 - \frac{c^{d+1}}{d+1}\right)^n = \left(1 - \frac{\alpha^{d+1}}{n(d+1)}\right)^n \sim \exp\left(-\frac{\alpha^{d+1}}{d+1}\right) .$$

Since the number of cells is $1/c^d = n^{d/(d+1)}/\alpha^d$, the expectation of the total number of cubes packed is

$$\alpha^{-d}\left(1 - \exp\left(-\frac{\alpha^{d+1}}{d+1}\right)\right) n^{d/(d+1)} ,$$

which gives the desired lower bound.

The upper bound is based on the simple observation that the sum of the volumes of the packed cubes is at most 1. First we consider the probability distribution of the volume of a single generated cube. The side of this cube is a uniform random variable U over $[0,1]$. Thus the probability that its volume is bounded by z is

$$F(z) = \Pr\left\{U^d \leq z\right\} = \Pr\left\{U \leq z^{1/d}\right\} = z^{1/d} .$$

Then applying Lemma 1 with $s_n = 1/n$, $x_n = ((d+1)/n)^{d/(d+1)}$, and $F(x_n) = ((d+1)/n)^{1/(d+1)}$, we conclude that the expected number of cubes selected before their total volume exceeds 1 is asymptotic to $(d+1)^{1/(d+1)} n^{d/(d+1)}$, which gives the desired matching upper bound. ∎

4 Bounds for $d \geq 2$ Dimensional Boxes

Let $\mathcal{H}_d$ denote the unit hypercube in $d \geq 1$ dimensions. The approach of the last section can also be used to prove asymptotic bounds for the case of random boxes in $\mathcal{H}_d$.

Theorem 2. *Fix d and draw n boxes independently and uniformly at random from $\mathcal{H}_d$. The maximum number that can be packed is asymptotically bounded from below by $\Omega(\sqrt{n})$ and from above by $O(\sqrt{n \ln^{d-1} n})$.*

Proof sketch: The lower bound argument is the same as that for cubes, except that $\mathcal{H}_d$ is partitioned into cells with sides on the order of $1/n^{2d}$. It is easy to verify that, on average, there is a constant fraction of the $n^{1/2}$ cells in which each cell wholly contains at least one of the given rectangles.

To apply Lemma 1 in a proof of the upper bound, one first conducts an asymptotic analysis of the distribution F_d, the volume of a d-dimensional box, which shows that

$$dF_d(x) \sim \frac{2^d}{(d-1)!} \ln^{d-1} x^{-1}.$$

Then, with $s_n = 1/n$, we obtain

$$x_n \sim \sqrt{(d-1)!/(n \ln^{d-1} n)} \quad \text{and} \quad F_d(x_n) \sim 2\sqrt{\ln^{d-1} n/((d-1)!n)}.$$

which together with Lemma 1 yields the desired upper bound. ∎

5 Tight Bound for $d = 2$

Closing the gaps left by the bounds on $\mathsf{E}[C_n]$ for $d \geq 3$ remains an interesting open problem. However, one can show that the lower bound for $d \geq 2$ is tight, i.e., $\mathsf{E}[C_n] = \Theta(n^{1/2})$. To outline the proof of the $O(n^{1/2})$ bound, we first introduce the following reduced, discretized version. A *canonical* interval is an interval that, for some $i \geq 0$, has length 2^{-i} and has a left endpoint at some multiple of 2^{-i}. A canonical rectangle is the product of two canonical intervals. In the reduced, rectangle-packing problem, a Poissonized model of canonical rectangles is assumed in which the number of rectangles of area a is Poisson distributed with mean λa^2, independently for each possible a. Let $C^*(\lambda)$ denote the cardinality of a maximum packing for an instance of the reduced problem with parameter λ.

Note that there are $i+1$ shapes possible for a rectangle of area 2^{-i}, and that for each of these shapes there are 2^i canonical rectangles. The mean number of each of these is $\lambda/2^{2i}$. Thus, the total number $T(\lambda)$ of rectangles in the reduced problem with parameter λ is Poisson distributed with mean

$$\sum_{i=0}^{\infty}(i+1)2^i(\lambda 2^{-2i}) = \lambda \sum_{i=0}^{\infty}(i+1)2^{-i} = 4\lambda\,. \tag{3}$$

To convert an instance of the original problem to an instance of the reduced problem, we proceed as follows. It can be seen that any interval in $\mathcal{H}_1$ contains either one or two canonical intervals of maximal length. Let the *canonical subinterval* I' of an interval I be the maximal canonical interval in I, if only one exists, and one such interval chosen randomly otherwise. A straightforward analysis shows that a canonical subinterval $I = [k2^{-i}, (k+1)2^{-i})$ has probability 0 if it touches a boundary of $\mathcal{H}_1$, and has probability $\frac{3}{2}2^{-2i}$, otherwise. The *canonical subrectangle* R' of a rectangle R is defined by applying the above separately to both coordinates. Extending the calculations to rectangles, we get $\frac{9}{4}a^2$ as the probability of a canonical subrectangle R of area a, if R does not touch the boundary of $\mathcal{H}_2$, and 0 otherwise. Now consider a random family of rectangles $\{R_i\}$, of which a maximum of $C(n)$ can be packed in $\mathcal{H}_2$. This family generates a random family of canonical subrectangles $\{R_i'\}$. The maximum number $C'(n)$ of the R_i' that can be packed trivially satisfies $C(n) \leq C'(n)$. Since the number of each canonical subrectangle of area a that does not touch a boundary is Poisson distributed with mean $9na^2/4$, we see that an equivalent way to generate a random family $\{R_i'\}$ is simply to remove from a random instance of the reduced problem with parameter $9n/4$ all those rectangles touching a boundary. It follows easily that $\mathsf{E}C(n) \leq \mathsf{E}C'(n) \leq \mathsf{E}C^*(9n/4)$ so if we can prove that $\mathsf{E}C^*(9n/4)$ or more simply $\mathsf{E}C^*(n)$, has the $O(n^{1/2})$ upper bound, then we are done.

The following observations bring out the key recursive structure of maximal packings in the reduced problem. Let Z_1 be the maximum number of rectangles that can be packed if we disallow packings that use rectangles spanning the height of the square. Define Z_2 similarly when packings that use rectangles spanning the width of the square are disallowed. By symmetry, Z_1 and Z_2 have the same distribution, although they may not be independent. To find this distribution, we begin by noting that (i) a rectangle spanning the width of $\mathcal{H}_2$ and one spanning the height of $\mathcal{H}_2$ must intersect and hence can not coexist in a packing; (ii) rectangles spanning the height of $\mathcal{H}_2$ are the only rectangles crossing the horizontal line separating the top and bottom halves of $\mathcal{H}_2$ and rectangles spanning the width of $\mathcal{H}_2$ are the only ones crossing the vertical line separating the left and right halves of $\mathcal{H}_2$. It follows that, if a maximum cardinality packing is not just a single 1×1 square, then it consists of a pair of disjoint maximum cardinality packings, one in the bottom half and one in the top half of $\mathcal{H}_2$, or a similar pair of subpackings, one in the left half and one in the right half of $\mathcal{H}_2$. After rescaling, these subpackings become solutions to our original problem on $\mathcal{H}_2$ with the new parameter λ times the square of half the area of $\mathcal{H}_2$, i.e., $\lambda/4$. We conclude that Z_1 and Z_2 are distributed as the sum of two independent samples of $C^*(\lambda/4)$, and that

$$C^*(\lambda) \leq Z_0 + \max(Z_1, Z_2)\,,$$

where Z_0 is the indicator function of the event that the entire square is one of the given rectangles. Note that Z_0 is independent of Z_1 and Z_2.

To exploit the above recursion, it is convenient to work in terms of the generating function, $S(\lambda) := \mathsf{E}e^{\alpha C^*(\lambda)}$. One can show that $S(\lambda) \leq 2e^{\alpha}\left(S(\lambda/4)\right)^2$, and that a solution to this relation along with the inequality $\mathsf{E}[C^*(\lambda)] \leq \alpha^{-1} \ln \mathsf{E}[e^{\alpha C_*(\lambda)}]$ yields the desired bound, $\mathsf{E}[C^*(\lambda)] = O(n^{1/2})$.

Acknowledgment

In the early stages of this research, we had useful discussions with Richard Weber, which we gratefully acknowledge.

References

1. Clark, B. N., Colburn, C. J., and Johnson, D. S., "Unit Disk Graphs," *Discrete Mathematics*, **86**(1990), 165-177.
2. Coffman, E. G., Jr. and Lueker, G. S., *An Introduction to the Probabilistic Analysis of Packing and Partitioning Algorithms*, Wiley & Sons, New York, 1991.
3. Justicz, J., Scheinermann, E. R., and Winkler, P. M., "Random Intervals," *Amer. Math. Monthly*, **97**(1990), 881-889.