Combinatorics, Probability and Computing (2002) 11, 113–127. © 2002 Cambridge University Press
DOI: 10.1017/S0963548301005041 Printed in the United Kingdom

On a Square Packing Problem

S. BOUCHERON and W. FERNANDEZ de la VEGA†

LRI, UMR 8623 CNRS, Bât. 490, Université Paris-Sud,
91405 Orsay cedex, France
(e-mail: bouchero@lri.fr)

Received 19 December 2000; revised 18 June 2001

An instance of the square packing problem consists of n squares with independently, uniformly distributed side-lengths and independently, uniformly distributed locations on the unit d-dimensional torus. A packing is a maximum family of pairwise disjoint squares. The one-dimensional version of the problem is the classical random interval packing problem. This paper deals with the asymptotic behaviour of packings as n tends to infinity while $d = 2$. Coffman, Lueker, Spencer and Winkler recently proved that the average size of packing is $\Theta(n^{d/(d+1)})$. Using partitioning techniques, sub-additivity and concentration of measure arguments, we show first that, after normalization by $n^{2/3}$, the size of two-dimensional square packings tends in probability toward a genuine limit γ. Straightforward concentration arguments show that large fluctuations of order $n^{2/3}$ should have probability vanishing exponentially fast with $n^{2/3}$. Even though γ remains unknown, using a change of measure argument we show that this upper bound on tail probabilities is qualitatively correct.

1. Introduction

This note presents an average-case analysis of a combinatorial optimization problem with a geometric flavour described by Coffman, Lueker, Spencer and Winkler in [4]. An instance of the cube packing problem in the d-dimensional unit torus $T_d = [0,1]^d$ consists of a collection denoted by ω of n (randomly selected) cubes in T_d. The torus is assumed to be oriented. Each cube is obtained by randomly choosing the 'lower left' vertex $x \in T_d$ and a random side-length ℓ according to the uniform distribution in $[0,1]$. Moreover, the cube sides are required to be parallel to the coordinate axes: y belongs to the cube defined by x and ℓ if there exists $z \in [0,\ell]^d$ such that $y = x + z$ where addition is taken modulo 1. Henceforth, the collection will be called a sample. The quantity to be optimized is the size of a sub-family of mutually disjoint cubes. With a slight abuse of language, we

† Research partially supported by Esprit Working Group Rand-APX, and CNRS GDR ALEA and ALCOPHYS.

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

also call this number denoted by $X_n^d(\omega)$, the independence number or packing number of the sample (the term 'packing' is from [4]). As pointed out in [4], cubes cannot be freely translated in T_d. This distinguishes this geometric packing problem from more conventional packing problems.

The one-dimensional version of this problem is a cosmetic variant of the classical random interval packing problem described in [8]. In the latter case, it is known that $n^{-1/2}X_n^1$ converges in probability toward a finite limit $(\sqrt{2/\pi})$ [7]. It is even known that X_n^1 satisfies a central limit theorem and a large deviation principle [3]. But the higher-dimensional versions of the cube packing problem differ significantly from the one-dimensional version: the combinatorial problem is NP-hard as soon as $d = 2$ (see [4] and references therein). Hence there is no hope of understanding the high-dimensional packing problems by analysing the performance of a simple greedy algorithm as in [7, 3]. Nevertheless, it is shown by Coffman, Lueker, Spencer and Winkler in [4] that the order of magnitude of $\mathbb{E}[X_n^d]$ is $n^{d/(d+1)}$

Theorem 1.1 ([4]). *The expectation of the maximum cardinality of a packing of n random cubes in the d-dimensional torus T_d satisfies*

$$\mathbb{E}[X_n^d] \in \Theta(n^{\frac{d}{d+1}}).$$

While the proof of the upper bound relies on a simple covering argument, the lower bound is established through the analysis of a constrained version of the problem. Theorem 1.1 and the proof given in [4] leave open the existence of a limit for the normalized expected independence number $n^{-d/(d+1)}X_n^d$, as n tends to infinity, *i.e.*, of a law of large numbers. Paraphrasing M. Steele on the longest increasing subsequence [9], 'one cannot help but suspect that the ratio $n^{-d/(d+1)}\mathbb{E}[X_n^d]$ has a genuine limit'.

Proving the convergence of $n^{-d/(d+1)}\mathbb{E}[X_n^d]$ toward a finite limit actually suffices to prove that $n^{-d/(d+1)}X_n^d$ converges in probability toward a fixed value. This follows from a basic observation.

As a function of a sample ω of n cubes $(\omega_1, \ldots, \omega_n)$ in $[0,1]^d$, $X_n^d(\omega)$ is a *configuration function* in the sense of Talagrand [10, Section 4]: let $I \subseteq \{1 \ldots n\}$, let the property $p((\omega_i)_{i\in I})$ be satisfied if the cubes $(\omega_i)_{i\in I}$ are pairwise disjoint. Then, if $(\omega_i)_{i\in I}$ satisfies p, so does any subsequence of $(\omega_i)_{i\in J}$, with $J \subseteq I$. $X_n^d(\omega)$ is thus the maximal cardinality of a subsample of ω that satisfies the hereditary property p.

As cubes are drawn independently, generic concentration of measure arguments for configuration functions tell us that the typical deviations of X_n^d should be at most of order $0(\sqrt{\mathbb{E}[X_n^d]})$. Namely, applying results from [10, Section 4] or [2] (Theorem 4.1.1 in [10] actually provides tail bounds for deviations around the median), we obtain the following.

Proposition 1.2. *The independence number of a random sample of n squares in T_d satisfies, for all positive t,*

$$\mathbb{P}\left\{X_n^d > \mathbb{E}[X_n^d] + t\right\} \leqslant \exp\left[-\frac{t^2}{2(\mathbb{E}[X_n^d] + t/3)}\right],$$

$$\mathbb{P}\left\{X_n^d < \mathbb{E}[X_n^d] - t\right\} \leqslant \exp\left[-\frac{t^2}{2\mathbb{E}[X_n^d]}\right].$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

Configuration functions show up frequently in the average-case analysis of combinatorial optimization problems: the longest increasing subsequence in a random permutation [10, 9], the independence number in a random graph [6, 10], and random interval packing [8, 7]. Though laws of large numbers for configuration functions are easily deduced from concentration inequalities provided that the normalized expectations do converge, the situations where general concentration inequalities provide tight results are not fully understood. As concentration of measure arguments for configuration functions are known to provide tight results for interval packings [3] but not for the longest increasing subsequence [5], one naturally wonders whether this remains true for higher-dimensional versions of the problem.

The purpose of this note is to prove the following theorems, which (partially) settle those questions in the two-dimensional case. Henceforth, we assume $d = 2$ and drop the exponent in X_n^d.

Theorem 1.3. *Let $E_n \stackrel{\Delta}{=} \mathbb{E}[X_n]$ denote the expectation of the maximum number X_n of pairwise disjoint squares within n random squares in the two-dimensional torus. Then the sequence $(n^{-2/3}E_n)_{n=1,\ldots}$ converges toward a limit γ.*

The following theorem suggests that the configuration function trick correctly describes the tail behaviour of the square packing number.

Theorem 1.4. *For all $t > 0$, the following hold:*

$$\liminf \frac{1}{n^{2/3}} \log \mathbb{P}\{X_n > n^{2/3}\gamma(1+t)\} > -\infty, \tag{1.1}$$

$$\liminf \frac{1}{n^{2/3}} \log \mathbb{P}\{X_n < n^{2/3}\gamma(1-t)\} > -\infty. \tag{1.2}$$

The paper is organized as follows. In Section 2, Theorem 1.3 is proved. The proof proceeds by showing that E_n satisfies an approximate sub-additivity property (Lemma 2.2). The proof of Lemma 2.2 outlines the fact that, up to some smaller terms, X_n can be considered as the sum of $\Theta(n^{2/3})$ exchangeable random variables. This provides a hint for a change of measure argument that enables us to prove Theorem 1.4 in Section 3.

2. Proof of Theorem 1.3

2.1. A sub-additivity property and its consequences

Let us first note the following almost trivial but useful fact.

Lemma 2.1. *The sequence $E_n = \mathbb{E}[X_n]$ is monotone and sub-additive:*

$$\mathbb{E}[X_m] \leqslant \mathbb{E}[X_{m+n}] \leqslant \mathbb{E}[X_m] + \mathbb{E}[X_n] \leqslant \mathbb{E}[X_m] + n. \qquad \square$$

By Fekete's lemma (see [9]), this implies that E_n/n converges toward a limit. But naive sub-additive arguments provide little help: as E_n does not grow linearly with n, the limit

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

is 0. In the case of the longest increasing subsequence, the analogue of Theorem 1.3 had been established by Hammersley using an elegant embedding into a Poisson process that scales at the right speed (see [9]). We are not aware of a comparable embedding in the case of the cube packing problem. Nevertheless, the proof of Theorem 1.3 also consists of bringing back sub-additive arguments.

The proof of Theorem 1.3 is almost immediate when the following lemma is established.

Lemma 2.2. *There exist an integer m_0, a positive function g, with $m^{-2/3}g(m) \to 0$ as m tends toward infinity, such that, for all $m \geqslant m_0$, $k \geqslant m$,*

$$\mathbb{E}[X_{mk^3}] \leqslant k^2(\mathbb{E}[X_m] + g(m)).$$

The proof of Lemma 2.2 is postponed to the next section. Let us check that Theorem 1.3 is a consequence of Lemma 2.2.

Proof of Theorem 1.3. From [4, Theorem 1], there exist two positive constants γ and ν such that $\limsup_n n^{-2/3}E_n = \nu$ and $\liminf_n n^{-2/3}E_n = \gamma$. Proving Theorem 1.3 boils down to proving that, for any $\epsilon > 0$, $\nu \leqslant \gamma + \epsilon$.

Let m be such that:

- $m > m_0$,
- $m^{-2/3}g(m) \leqslant \gamma\epsilon/2$,
- $m^{-2/3}\mathbb{E}[X_m] \leqslant \gamma(1 + \epsilon/2)$.

For each sufficiently large n, let $k(n)$ be defined by

$$mk(n)^3 \leqslant n < m(k(n) + 1)^3.$$

For all n such that $k(n) > m$, we have, by Lemma 2.2,

$$\mathbb{E}[X_n] \leqslant \mathbb{E}\left[X_{m(k(n)+1)^3}\right] \leqslant (k(n) + 1)^2 (\mathbb{E}[X_m] + g(m)),$$

and hence

$$\frac{\mathbb{E}[X_n]}{n^{2/3}} \leqslant \frac{(k(n) + 1)^2}{k(n)^2} \frac{\mathbb{E}[X_m] + g(m)}{m^{2/3}}.$$

Thus

$$\limsup_n \frac{E_n}{n^{2/3}} \leqslant \gamma(1+\epsilon). \qquad \square$$

2.2. Proof of Lemma 2.2

The proof of Lemma 2.2 relies on several technical results.

For any integer k, let G_k denote the grid defined on the two-dimensional torus by lines with coordinates $x = i/k$ for $0 \leqslant i < k$ and $y = i/k$ for $0 \leqslant i < k$.

Let m be an integer. In this subsection, the following relation between n, m and k is assumed: $n = mk^3$.

Let us agree on the following conventions.

- Let $Y_n(\omega)$ denote the cardinality of the largest packing consisting of squares from ω of

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

side-length smaller than

$$\delta \triangleq m^{1/6}k^{-1} = \frac{\sqrt{m}}{\sqrt[3]{n}},$$

intersecting the grid G_k.

- Let $Z_n(\omega)$ denote the cardinality of the largest packing of squares from ω that do not intersect G_k.

As squares in a packing are non-overlapping and as their surface should not exceed 1, the maximal number of squares with side-length larger than $m^{1/6}k^{-1}$ in a packing is less than $k^2m^{-1/3}$. Hence we have

$$Z_n(\omega) \leqslant X_n(\omega) \leqslant Y_n(\omega) + Z_n(\omega) + k^2m^{-1/3}. \tag{2.1}$$

The following lemma is immediate.

Lemma 2.3. *Let X_{mk^3}, Y_{mk^3} and Z_{mk^3} be defined as above. Then*

$$\mathbb{E}[X_{mk^3}] \leqslant \mathbb{E}[Y_{mk^3}] + \mathbb{E}[Z_{mk^3}] + k^2m^{-1/3}. \qquad \square$$

We will first determine an upper bound on Y_{mk^3}. A packing consisting of squares from ω of side-length smaller than δ intersecting the grid G_k defines a packing consisting of squares from ω of side-length smaller than δ intersecting a meridian $x = i/k$ or a parallel $y = j/k$ for each possible value of i and j. A packing of squares intersecting a given parallel defines a packing of intervals on the unit circle. This interval packing on the unit circle might contain an interval containing the point with zero longitude; if we exclude this interval from the packing, we define an interval packing on the unit interval.

Let $T_n(\omega)$ denote the cardinality of the largest packing consisting of squares from ω with:

- side-length smaller than δ,
- intersecting the line $x = 0$, and not containing the point $(0, 0)$.

All lines play the same role; hence

$$\mathbb{E}[Y_{mk^3}] \leqslant 2k\mathbb{E}[T_{mk^3} + 1]. \tag{2.2}$$

As T_n is the size of a packing of intervals, it can be analysed by the very same techniques that were used for the original random interval packing problem [7, 3].

Lemma 2.4. *Let $k \geqslant m$ and let T_{mk^3} denote the maximum number of disjoint squares with side-length less than $m^{1/6}k^{-1}$ intersecting the line $x = 0$ without containing the point $(0, 0)$ among mk^3 random squares on the unit torus:*

$$\mathbb{E}[T_{mk^3}] \leqslant 4m^{1/3}k.$$

The proof of Lemma 2.4 proceeds in two steps: we first define a Poisson process on the plane in the style of [7] and then show that $\mathbb{E}[T_{mk^3}]$ is well approximated by a functional of this Poisson process. Then we upper-bound the expectation of this functional.

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

Proof. The random interval packing problem we are facing differs from the original random interval packing problem in the following ways: the number of intervals to be considered is random, being binomially distributed with parameters n and $\delta^2/2 - \delta^3/3$ (the corrective term is due to the fact that we exclude small squares containing $0,0$); the probability distribution of the length of the random intervals differs from the original model described in [8].

In other words, the probability that a square with side-length s hits the line $x = 0$ is s. The probability that a random square has side-length smaller than $s \leqslant \delta$ conditionally on the fact that it has side-length less than δ and that it hits the line $x = 0$, is s^2/δ^2.

The Poisson embedding technique described in [7] still performs a flawless job. Recall that $n = mk^3$, and consider the non-homogeneous Poisson process on $[0,1]^2$ with the following intensity:

$$\lambda(x,y) \stackrel{\Delta}{=} \begin{cases} n(y-x), & \text{for} \quad y-\delta < x < y, \\ 0, & \text{otherwise,} \end{cases}$$

where we still have $\delta = m^{1/6}k^{-1}$.

A realization of this process is a random collection σ of points. The number of points in σ belonging to some region $A \subseteq [0,1]^2$ is distributed according to the Poisson law with parameter $\int_A \lambda(x,y)\,dx\,dy$ and the number of points in σ falling into two disjoint regions are independent. A point (x,y) in σ with $0 \leqslant y - x \leqslant \delta$ defines an interval $[x,y]$ with length $y - x < \delta$. Let Po denote the law of the Poisson process. Let $T_n(\sigma)$ denote the maximum cardinality of an interval packing on the unit interval induced by σ.

If the number $M(\sigma)$ of points falling in the region $0 \leqslant y - x \leqslant \delta$ equals the number $N(\omega)$ of squares in ω with side-length less than δ hitting the line $x = 0$ and not containing the point $(0,0)$, the maximum cardinality $T_n(\sigma)$ of an interval packing on $[0,1]$ induced by σ and the maximal cardinality $T_n(\omega)$ of a square packing among the small squares hitting the line $x = 0$ and avoiding $(0,0)$ on the torus are identically distributed.

Note that, if

$$F(\ell) \stackrel{\Delta}{=} \mathbb{E}\left[T_n(\omega) \mid N(\omega) = \ell\right] = \mathbb{E}\left[T_n(\sigma) \mid M(\sigma) = \ell\right],$$

then, by coupling,

$$DF(\ell) \stackrel{\Delta}{=} |F(\ell+1) - F(\ell)| \leqslant 1.$$

Recall that the Monge–Wasserstein distance between two probability distributions over the integers $\mathbb{P}$ and $\mathbb{Q}$ is

$$\sup_{F:\sup_n |DF(n)| \leqslant 1} \left(\mathbb{E}_{\mathbb{P}}[F] - \mathbb{E}_{\mathbb{Q}}[F]\right).$$

Hence,

$$\begin{aligned} &\left|\mathbb{E}_{\mathbb{P}_n}[T_n] - \mathbb{E}_{\mathrm{Po}}[T_n]\right| \\ &\leqslant \left| \sum_{\ell} \mathbb{E}[T_n \mid N(\omega) = \ell] \times \left(\mathbb{P}_n\{N(\omega) = \ell\} - \mathrm{Po}\{M(\sigma) = \ell\}\right)\right| \\ &\leqslant \sup_{G:\sup_\ell |DG(l)| \leqslant 1} \left(\mathbb{E}_{\mathbb{P}_n}\left[G\left(N(\omega)\right)\right] - \mathbb{E}_{\mathrm{Po}}\left[G\left(N(\sigma)\right)\right]\right). \end{aligned}$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

Hence the difference between the expectations of $T_n(\omega)$ under $\mathbb{P}_n$ and of $T_n(\sigma)$ under Po is upper-bounded by the Monge–Wasserstein distance between the laws of $N(\cdot)$ and $M(\cdot)$ under those two probability laws.

Under $\mathbb{P}_n$, $N(\omega)$ is binomially distributed with parameter $n = mk^3$ and success probability $\delta^2/2 - \delta^3/3 = \frac{1}{2}m^{1/3}k^{-2} - \frac{1}{3}m^{1/2}k^{-3}$ and, under Po, $M(\sigma)$ follows a Poisson law with parameter $n(\delta^2/2 - \delta^3/3)$. By standard results in Poisson approximation [1], the Monge–Wasserstein distance for the binomial with parameters n and p, and the Poisson distribution with parameter np, is upper-bounded by $2p\sqrt{np}$. Hence

$$\mathbb{E}_{\mathbb{P}_n}[T_n] - \mathbb{E}_{\mathrm{Po}}[T_n] \leqslant mk^{-3/2}.$$

We will now upper-bound $\mathbb{E}_{\mathrm{Po}}[T_n]$. Let us observe as in [3, 7] that $T_n(\sigma)$ is distributed as the first-passage time of a random walk.

A sequence of random points (x_i, y_i) satisfying $y_i - \delta < x_i < y_i$, and $y_i < x_{i+1}$ with maximum cardinality can be determined by the greedy algorithm described in [7].

The input of the greedy algorithm is a list of couples $(x_i, y_i)_{1 \leqslant i \leqslant \ell}$ satisfying $y_i - \delta < x_i < y_i$, and without loss of generality we may assume that the sequence has been sorted so that $y_i \leqslant y_{i+1}$ for all $i < \ell$. Starting from an empty list $\mathscr{L}$, index $i = 0$, and threshold $h = 0$, the algorithm repeats the following loop: find the smallest index $j, i < j \leqslant \ell$ such that $x_j > h$, add (x_j, y_j) to the list $\mathscr{L}$, update h by assigning y_j to h and i by assigning j to i. The algorithm terminates when no such j can be found.

Because of the definition of the Poisson process, the sequence of second coordinates y_i that are picked by the greedy algorithm is a random walk with drift. As the Poisson process could be extended to $x > 1$, $T_n(\sigma)$ can be analysed as the first-passage time of a random walk with drift though the boundary $y = 1$. Let us call $(W_i)_{i \geqslant 0}$ the sequence of the positions of this random walk.

Such a first-passage time satisfies a law of large numbers and a central limit theorem (see [3] for applications to the original random intervals packing).

The probability that a step $W_{i+1} - W_i$ in the random walk exceeds u is the probability that no point of the Poisson process falls in a region with intensity

$$\begin{aligned} n\frac{u^3}{6} \quad &\text{if} \quad 0 \leqslant u \leqslant \delta, \\ n\left(\frac{\delta^2 u}{2} - \frac{\delta^3}{3}\right) \quad &\text{for} \quad u > \delta. \end{aligned}$$

This probability equals

$$\exp\left(-n\frac{u^3}{6}\right) \quad \text{if} \quad 0 \leqslant u \leqslant \delta, \tag{2.3}$$

$$\exp\left(-n\left(\frac{\delta^2 u}{2} - \frac{\delta^3}{3}\right)\right) \quad \text{for} \quad u > \delta. \tag{2.4}$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

As $\delta = \frac{m^{1/2}}{n^{1/3}}$, the expected step size is thus lower-bounded by

$$\begin{aligned}
&\int_0^\delta \exp\left(-n\frac{u^3}{6}\right) du + \int_\delta^1 \exp\left(-n\left(\frac{\delta^2 u}{2} - \frac{\delta^3}{3}\right)\right) du \\
&\geqslant \int_0^{n^{-1/3}} \exp\left(-n\frac{u^3}{6}\right) du \\
&\geqslant \exp(-1/6)\, n^{-1/3}.
\end{aligned}$$

Now consider the coupled random walk $(W_i')_{i\geqslant 0}$ with steps

$$W'_{i+1} - W'_i \stackrel{\Delta}{=} (W_{i+1} - W_i) \wedge 1.$$

The sequence $W'_i - i\,\mathbb{E}[W'_1 - W'_0]$ is a centred martingale. Let τ denote the stopping time $\min\{i : W'_i \geqslant 1\}$, by a straightforward coupling:

$$\mathbb{E}[T_n(\sigma)] \leqslant \mathbb{E}[\tau].$$

Then, by the optional sampling theorem,

$$0 = \mathbb{E}\left[W'_\tau - \tau\mathbb{E}[W'_1 - W'_0]\right],$$

while $W'_\tau \leqslant 2$, which implies

$$\mathbb{E}[\tau] \leqslant 2\exp(1/6)n^{1/3}.$$

Hence,

$$\mathbb{E}[T_{mk^3}(\omega)] \leqslant 2\exp(1/6)km^{1/3} + mk^{-3/2},$$

which is smaller than $4km^{1/3}$ for $k \geqslant m$. □

Plugging Lemma 2.4 into equation (2.2),

$$\mathbb{E}[Y_{mk^3}] \leqslant 9m^{1/3}k^2. \tag{2.5}$$

Let us now define $X'_n(\omega)$ as the maximal cardinality of a packing of squares from ω that do not intersect the parallel $y = 0$ and the meridian $x = 0$. Then $X'_n(\omega)$ may be regarded as a the maximal cardinality of a packing on the unit square rather than on the unit torus.

Before analysing Z_{mk^3}, let us point out the following relationship between X_n and X'_n.

Lemma 2.5. *Assuming $n = mk^3$ and $k > m$, then*

$$\mathbb{E}[X'_n] \leqslant \mathbb{E}[X_n] \leqslant \mathbb{E}[X'_n] + 9\sqrt[3]{n}.$$

Proof. The first inequality is obvious. The second comes from the ancient results on random interval packings. X'_n and X_n differ at most by twice the cardinality of a packing of random intervals on $[0, 1]$. First observe that such a packing contains at most $\sqrt[3]{n}$ intervals with length greater than $n^{-1/3}$. We may then upper-bound the expected cardinality of a packing of intervals of length smaller than $n^{-1/3}$ on the parallel $y = 0$ or the meridian $x = 0$, by the very same approach that was used in the proof of the preceding lemma. □

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

Henceforth, we will use the following notation.

- For any i, j in $\{0, \ldots, k-1\}$, let $R_{i,j}$ denote the region $i/k \leqslant x < (i+1)/k$ and $j/k \leqslant y < (j+1)/k$.
- Let $N_{i,j}(\omega)$ denote the number of squares in ω with side-length less than $1/k$ whose lower left corner is in region $R_{i,j}$.
- Let $N'_{i,j}$ denote the number of squares in ω that fit into region $R_{i,j}$.
- Let $V_{i,j}(\omega)$ denote the cardinality of a maximum packing among the $N'_{i,j}$ squares fitting into region $R_{i,j}$.

Remember that

$$\mathbb{E}[N_{i,j}] = m \quad \mathbb{E}[N'_{i,j}] = \frac{m}{3},$$

and that $N_{i,j}$ (resp. $N'_{i,j}$) has variance smaller than m (resp. $m/3$).

Remark. Conditionally on $N_{i,j} = l$, $V_{i,j}$ is distributed like X'_l.

Lemma 2.6. $\mathbb{E}[Z_{mk^3}] \leqslant k^2(\mathbb{E}[X_m] + \sqrt{m})$.

Proof. Z_{mk^3} is the sum of k^2 exchangeable random variables $V_{i,j}$. Now,

$$\begin{aligned} \mathbb{E}[V_{i,j}] &= \mathbb{E}_{N_{i,j}}\left[\mathbb{E}[V_l \mid N_{i,j} = l]\right] \\ &\leqslant \mathbb{E}_{N_{i,j}}\left[\mathbb{E}[X'_m + (N_{i,j} - m)^+]\right] \\ &\leqslant \mathbb{E}[X'_m] + \sqrt{m}, \end{aligned}$$

where the last inequality follows from Cauchy–Schwarz inequality. The lemma follows from the first inequality in Lemma 2.5. □

We may now prove Lemma 2.2. Indeed, the lemma follows from Lemma 2.3, inequality (2.5) and Lemma 2.6, choosing $g(m) = \sqrt{m} + 10m^{1/3} + 1$. □

3. Proof of Theorem 1.4

The proof of Theorem 1.3 highlights the fact that, for $n = mk^3$ with large values of k, X_n is well approximated by the sum of k^2 exchangeable and actually weakly dependent random variables. This suggests that the typical fluctuations of X_{mk^3} should be of order k and thus that, in this respect at least, the square packing problem is much closer to the interval packing problem than to the longest increasing subsequence problem.

We may build changes of measure that will be sufficient to prove Theorem 1.4. We customize the pattern of proof usually adopted to prove large deviation lower bounds: for a given $t > 0$ construct a sequence of alternate probability distributions $\mathbb{Q}_n$ for X_n such that the relative entropy,

$$K(\mathbb{Q}_n \mid \mathbb{P}_n) \triangleq \mathbb{E}_{\mathbb{Q}_n}\left[\log \frac{d\mathbb{Q}_n}{d\mathbb{P}_n}\right],$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

is of order $n^{2/3}$, and such that

$$\mathbb{Q}_n\{|X_n - E_n| > \gamma t n^{2/3}\} \to 1.$$

The argument proceeds according to the following folklore lemma.

Lemma 3.1. *Let $\mathbb{P}_n$ and $\mathbb{Q}_n$ denote a sequence of probability measures on the same space such that, for each n, $\mathbb{Q}_n$ is absolutely continuous with respect to $\mathbb{P}_n$. Let A_n denote a sequence of events such that $\mathbb{Q}_n\{A_n\} \to_n 1$. Letting $f(n)$ be defined by*

$$f(n) = K(\mathbb{Q}_n \mid \mathbb{P}_n),$$

then, if $\lim_n f(n) = +\infty$,

$$\liminf_{n\to\infty} \frac{1}{f(n)} \log \mathbb{P}_n\{A\} \geqslant -1.$$

Proof. Elementary properties of relative entropy imply

$$\begin{aligned} K(\mathbb{Q}_n \mid \mathbb{P}_n) &\geqslant \mathbb{Q}_n\{A_n\} \log \frac{\mathbb{Q}_n\{A_n\}}{\mathbb{P}_n\{A_n\}} + \mathbb{Q}_n\{A_n^c\} \log \frac{\mathbb{Q}_n\{A_n^c\}}{\mathbb{P}_n\{A_n^c\}} \\ &\geqslant \mathbb{Q}_n\{A_n\} \log \frac{\mathbb{Q}_n\{A_n\}}{\mathbb{P}_n\{A_n\}} - \frac{1}{e}, \end{aligned}$$

and thus

$$\frac{\mathbb{Q}_n\{A_n\}}{f(n)} \log \mathbb{P}_n\{A_n\} \geqslant -1 - \frac{2}{ef(n)}.$$

The lemma follows by taking n to infinity and taking advantage of $\lim_{n\to\infty} \mathbb{Q}_n\{A_n\} = 1$. □

This approach is made easy by the following observations. X_n is bracketed by two quantities Z_n and $Z_n + Y_n + n^{2/3}/m$. Both Z_n and Y_n are configuration functions in the sense described in Section 1. Hence, by Proposition 1.2, if $\mathbb{Q}_n$ is still a product measure, Z_n and Y_n are again concentrated around their mean. Hereafter, we will be careful to define our alternative measures $\mathbb{Q}_n$ as product measures. Squares will still be selected independently. But the distributions of side-lengths will be twisted according to the goal we are looking for. Remember that squares with side-lengths larger than $n^{-1/3}$ contribute only marginally to large packings and hence to X_n. To induce a large deviation of Z_n from above, we will twist the distribution of side-lengths, slightly favouring small squares. On the other hand, to induce a large deviation from below of Z_n and Y_n, we will discriminate against small squares.

Proof of Theorem 1.4, inequality (1.1). Assume for a moment that $n = mk^3$ where m is fixed and large and k may vary. Recall the notation of Section 2.2: Z_{mk^3} is a sum of k^2 exchangeable random variables, denoted by $V_{i,j}$ with $0 \leqslant i < k$ and $0 \leqslant j < k$, where $V_{i,j}$ denotes the largest packing of disjoint squares included in region $R_{i,j} = \{(x, y) : i/k \leqslant x < (i+1)/k \text{ and } j/k \leqslant y < (j+1)/k\}$. $N_{i,j}$ denotes the number of squares with lower left corner in the region and side-length smaller than $1/k$ (those squares may not fit entirely

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

into the region $R_{i,j}$). Remember:

$$\mathbb{E}\left[V_{i,j} \mid N_{i,j} = \ell\right] = \mathbb{E}\left[X'_\ell\right].$$

The construction of the alternate measure $\mathbb{Q}_n$ follows from the preceding observations. Under $\mathbb{Q}_n$, squares are still placed at random uniformly on the torus but the probability density of side-length is now $g(s)$:

$$g(s) = (1+u) \quad \text{for} \quad s < \frac{1}{k},$$
$$g(s) = \left(1 - \frac{u}{k-1}\right) \quad \text{for} \quad s \geqslant \frac{1}{k},$$

where u is tuned according to the amount of deviation we want to induce on $\mathbb{E}[X_n]$. $\mathbb{Q}_n$ is still a product probability distribution, and Z_{mk^3} remains a configuration function. This implies that, under $\mathbb{Q}_n$, Z_{mk^3} is concentrated around its mean value. Under $\mathbb{Q}_n$, $N_{i,j}$ is a binomial random variable with parameters $n = mk^3$ and $(1+u)/(k^3)$; it is centred around $m(1+u)$ instead of m.

Finally, observe that, conditionally on the value of $N_{i,j}$, the distributions of $V_{i,j}$ under $\mathbb{Q}_n$ and $\mathbb{P}_n$ are identical, and thus

$$\mathbb{E}_{\mathbb{P}_n}[V_{i,j} \mid N_{i,j}] = \mathbb{E}_{\mathbb{Q}_n}[V_{i,j} \mid N_{i,j}]. \tag{3.1}$$

Let ϵ be such that $0 \leqslant \epsilon \leqslant 1/9$, and let m be sufficiently large that the twisting parameter u can be chosen in such a way that

$$\mathbb{E}_{\mathbb{Q}_n}\left[N_{1,1}^{2/3}\right] \geqslant (1+t)(1+5\epsilon)\mathbb{E}_{\mathbb{P}_n}\left[N_{1,1}^{2/3}\right]. \tag{3.2}$$

Furthermore, let m be sufficiently large that, for $l \geqslant \ell_0 \triangleq m/2$,

$$\left|l^{-2/3}\mathbb{E}_{\mathbb{P}_l}[X_l] - \gamma\right| < \gamma\epsilon, \tag{3.3}$$

and so that

$$\gamma\epsilon\ell^{2/3} > 9\sqrt[3]{\ell} \quad \text{and} \quad \ell_0 > 4\log\frac{1}{\epsilon}. \tag{3.4}$$

Furthermore, by Janson's inequality (see [1]), for all $\eta \in [0,1]$,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_n}\left[N_{1,1}^{2/3}\right] &\geqslant [m(1-\eta)]^{2/3}\,\mathbb{P}_n\{N_{1,1} \geqslant m(1-\eta)\} \\ &\geqslant [m(1-\eta)]^{2/3}\left(1 - \exp\left(-m\left[\eta + (1-\eta)\log(1-\eta)\right]\right)\right).\end{aligned} \tag{3.5}$$

Hence we may choose m sufficiently large that

$$\mathbb{E}_{\mathbb{P}_n}\left[N_{1,1}^{2/3}\right](1+\epsilon) \geqslant m^{2/3}. \tag{3.6}$$

Note that Janson's inequality also implies

$$\mathbb{Q}_n\{N_{1,1} \leqslant \ell_0\} \leqslant \exp(-\ell_0/4). \tag{3.7}$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

Under those choices,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}_n}[Z_n] &= k^2 \mathbb{E}_{\mathbb{Q}_n}[V_{1,1}] \\
&= k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\mathbb{E}_{\mathbb{Q}_n}\left[V_{1,1} \mid N_{1,1}\right]\right] \\
&\overset{(a)}{=} k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\mathbb{E}_{\mathbb{P}_n}\left[V_{1,1} \mid N_{1,1}\right]\right] \\
&\overset{(b)}{=} k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\sum_{\ell=1}^{\infty} \mathbb{E}_{\mathbb{P}_\ell}\left[X'_\ell\right] \mathbb{1}_{N_{1,1}=\ell}\right] \\
&\overset{(c)}{\geqslant} k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\sum_{\ell=\ell_0}^{\infty} \mathbb{E}_{\mathbb{P}_\ell}\left[X_\ell - 9\sqrt[3]{\ell}\right] \mathbb{1}_{N_{1,1}=\ell}\right] \\
&\overset{(d)}{\geqslant} k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\sum_{\ell=\ell_0}^{\infty} \mathbb{1}_{N_{1,1}=\ell}\left(\gamma(1-\epsilon)\ell^{2/3} - 9\sqrt[3]{\ell}\right)\right] \\
&\overset{(e)}{\geqslant} k^2 \mathbb{E}_{\mathbb{Q}_n}\left[\mathbb{1}_{N_{1,1}>\ell_0}\, \gamma(1-2\epsilon) N_{1,1}^{2/3}\right] \\
&\overset{(f)}{\geqslant} k^2\gamma(1-2\epsilon)\left((1+t)(1+5\epsilon)\mathbb{E}_{\mathbb{P}_n}[N_{1,1}^{2/3}] - \mathbb{E}_{\mathbb{Q}_n}[N_{1,1}^{2/3}\mathbb{1}_{N_{1,1}<\ell_0}]\right) \\
&\overset{(g)}{\geqslant} (1+t)(1+\epsilon)\gamma k^2 m^{2/3},
\end{aligned}$$

where (a) follows from remark (3.1), (b) follows from the definition of X'_ℓ (see Remark page 121), (c) follows from Lemma 2.5, (d) follows from Theorem 1.3 and assumption (3.3), (e) follows from assumption (3.4), (f) follows from (3.2) and finally (g) follows from (3.7), (3.6) and from assumption (3.4).

Now let $h(x,y) \overset{\Delta}{=} x\log(x/y) + (1-x)\log[(1-x)/(1-y)]$; then

$$\begin{aligned}
\frac{1}{k^2} K(\mathbb{Q}_{mk^3} \mid \mathbb{P}_{mk^3}) &= mk\, h\left(\frac{1+u}{k}, \frac{1}{k}\right) \\
&\to_{k\to\infty} m[(1+u)\log(1+u) - u].
\end{aligned}$$

Now, recalling Proposition 1.2 and the fact that $\mathbb{Q}_n$ is defined as a product measure, combining the lower bound on $\mathbb{E}_{\mathbb{Q}_n}[Z_n]$ and Lemma 3.1, we get that, for $n = mk^3$, m satisfying assumptions (3.3) and (3.4), for $\epsilon < 1/9$,

$$\liminf_{k>\infty} \frac{1}{k^2} \log \mathbb{P}\left\{X_{mk^3} > (1+t)(1+\epsilon/2)\gamma k^2 m^{2/3}\right\} \geqslant -m[(1+u)\log(1+u) - u]. \tag{3.8}$$

Now, for sufficiently large n, let $k(n)$ again be defined by

$$mk(n)^3 \leqslant n < m(k(n)+1)^3.$$

Then

$$\begin{aligned}
&\mathbb{P}_n\left\{X_n > (1+t)\gamma n^{2/3}\right\} \\
&\geqslant \mathbb{P}_{mk(n)^3}\left\{X_{mk(n)^3} > (1+t)\gamma m^{2/3}(k(n)+1)^2\right\} \\
&\geqslant \mathbb{P}_{mk(m)^3}\left\{X_{mk(n)^3} > (1+t)\left(1+\frac{1}{k(n)}\right)^2 \gamma m^{2/3}(k(n))^2\right\},
\end{aligned}$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

and hence, by using the limiting inequality (3.8), we get

$$\liminf_n \frac{1}{n^{2/3}} \log \mathbb{P}\{X_n > (1+t)\gamma n^{2/3}\} \geqslant -m^{1/3}[(1+u)\log(1+u) - u].$$

This establishes inequality (1.1) in Theorem 1.4. □

Let us now look at deviations from below; the proof of inequality (1.2) proceeds in essentially the same way.

Proof of Theorem 1.4, inequality (1.2). Using Lemma 2.3, it is enough to prove that, for sufficiently large m and k, under a suitable product measure $\mathbb{Q}_n$,

$$\mathbb{E}_{\mathbb{Q}_{mk^3}}[Y_{mk^3}] \leqslant \mathbb{E}_{\mathbb{P}_{mk^3}}[Y_{mk^3}] \leqslant \eta\gamma m^{2/3}k^2,$$
$$\mathbb{E}_{\mathbb{Q}_{mk^3}}[Z_{mk^3}] \leqslant (1-t-3\eta)\gamma m^{2/3}k^2.$$

Under $\mathbb{Q}_n$ the density of side-lengths is now

$$g(s) = (1-u) \quad \text{for} \quad s < \frac{1}{k},$$
$$g(s) = \left(1 + \frac{u}{k-1}\right) \quad \text{for} \quad s \geqslant \frac{1}{k}.$$

There exists a probability space defined on pairs of samples (ω, ω') of n random squares, such that ω is distributed according to $\mathbb{P}_n$ and ω' is distributed according to $\mathbb{Q}_n$, and where, for each index i in $\{1 \ldots n\}$, ω_i and ω_i' have the same lower left corner, and where the side-length of ω_i is always smaller than the side-length of ω_i'. Any packing in ω is thus a packing in ω'. This coupling shows that

$$\mathbb{E}_{\mathbb{Q}_{mk^3}}[Y_{mk^3}] \leqslant \mathbb{E}_{\mathbb{P}_{mk^3}}[Y_{mk^3}].$$

If m is sufficiently large,

$$9m^{1/3}k^2 \leqslant \eta\gamma m^{2/3}k^2,$$

and

$$k^2 m^{-1/3} \leqslant \eta\gamma k^2 m^{2/3}.$$

Hence by Lemma 2.3, inequality (2.5), for sufficiently large m, k,

$$\mathbb{E}_{\mathbb{Q}_n}[X_{mk^3}] \leqslant \mathbb{E}_{\mathbb{Q}_n}[Z_{mk^3}] + 2\eta\gamma m^{2/3}k^2.$$

The auxiliary quantity ℓ_0 now satisfies

$$\ell_0 < m(1-u)/2, \tag{3.9}$$

and hence

$$\mathbb{Q}_n\{N_{1,1} < \ell_0\} \leqslant \exp(-\ell_0/4). \tag{3.10}$$

We again assume that m, and hence ℓ_0, is sufficiently large that assumptions (3.3) and (3.4) are satisfied as well as

$$\mathbb{E}_{\mathbb{Q}_n}[N_{1,1}^{2/3}] \geqslant \frac{1}{1+\epsilon}(m(1-u))^{2/3}, \tag{3.11}$$

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

and

$$\ell_0 \exp(-\ell_0/4) \leqslant \frac{\gamma\epsilon}{1+\epsilon}(2\ell_0)^{2/3}. \tag{3.12}$$

Let us further assume that the twisting parameter u is now chosen so that

$$\mathbb{E}_{\mathbb{Q}_n}[N_{1,1}^{2/3}] \leqslant (1-t)(1-3\epsilon)\mathbb{E}_{\mathbb{P}_n}[N_{1,1}^{2/3}]. \tag{3.13}$$

Let us now check that $\mathbb{E}_{\mathbb{Q}_n}[Z_n]$ is small enough.

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}_n}[Z_n] &= k^2\mathbb{E}_{\mathbb{Q}_n}[V_{1,1}] \\
&\overset{(a)}{=} k^2\mathbb{E}_{\mathbb{Q}_n}\left[\mathbb{E}_{\mathbb{P}_n}[V_{1,1} \mid N_{1,1}]\right] \\
&\overset{(b)}{=} k^2\mathbb{E}_{\mathbb{Q}_n}\left[\sum_{\ell\geqslant 1}\mathbb{E}_{\mathbb{P}_\ell}[X'_\ell]\,\mathbb{1}_{N_{1,1}=\ell}\right] \\
&\overset{(c)}{\leqslant} k^2\mathbb{E}_{\mathbb{Q}_n}\left[\sum_{\ell\geqslant 1}\mathbb{E}_{\mathbb{P}_\ell}[X_\ell]\mathbb{1}_{N_{1,1}=\ell}\right] \\
&\overset{(d)}{\leqslant} k^2\mathbb{E}_{\mathbb{Q}_n}\left[\mathbb{1}_{N_{1,1}>\ell_0}\left(\gamma(1+\epsilon)N_{1,1}^{2/3}\right) + \mathbb{1}_{N_{1,1}\leqslant\ell_0}N_{1,1}\right] \\
&\overset{(e)}{\leqslant} k^2\mathbb{E}_{\mathbb{Q}_n}\left[\gamma(1+2\epsilon)N_{1,1}^{2/3}\right] \\
&\overset{(f)}{\leqslant} k^2\gamma(1+2\epsilon)(1-3\epsilon)(1-t)\,\mathbb{E}_{\mathbb{P}_n}\left[N_{1,1}^{2/3}\right] \\
&\overset{(g)}{\leqslant} (1-t)(1-\epsilon)\gamma k^2 m^{2/3},
\end{aligned}$$

where (a) follows from observation (3.1), (b) follows from the definition of X'_ℓ, (c) follows from Lemma 2.5, (d) follows from assumptions (3.9) and (3.3), (e) follows from assumptions (3.10) and (3.11) since they imply $\mathbb{E}_{\mathbb{Q}_n}[N_{1,1}\mathbb{1}_{N_{1,1}\leqslant\ell_0}] \leqslant \ell_0\exp(-\ell_0/4) \leqslant \gamma\epsilon\mathbb{E}_{\mathbb{Q}_n}[N_{1,1}^{2/3}]$, (f) follows from assumption (3.13), and (g) follows from Jensen's inequality. One may choose ϵ so that $(1-t)(1-\epsilon) \leqslant 1-t-3\eta$.

The proof is again completed by observing that as $\mathbb{Q}_n$ is still a product measure, Z_n is still concentrated around its expected value and that the relative entropy between $\mathbb{Q}_n$ and $\mathbb{P}_n$ is again of order $k^2m = m^{1/3}n^{2/3}$. □

4. Conclusion

The two results presented in this note call for some comments. The ingredients used in the derivation, partitioning, sub-additivity, and concentration of measure have been used to investigate the behaviour of the Euclidean travelling salesman (see [9] and references therein). It is quite plausible that the cube packing problem satisfies an analogue of the Beardwood, Halton and Hammersley theorem, in other words that $n^{-d/(d+1)}X_n^d$ converges toward a genuine limit γ_d for each $d \geqslant 1$. But we are not aware of an analogue of the space-filling curve argument to give a proof that would work for all dimensions. The argument we have in mind would instead rely on a (tedious) induction on dimension. Second, establishing limit laws for the square packing problem seems as hard as establishing

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press

a central limit theorem for the travelling salesman. But, while partitioning and space-filling curve arguments do not seem to provide lower bounds on tail probabilities for the travelling salesman, partitioning provides such lower bounds for square packings.

Acknowledgements

The authors would like to thank an anonymous referee for his patience and insightful suggestions.

References

[1] Barbour, A., Holst, L. and Janson, S. (1992) *Poisson Approximation*, Clarendon Press, Oxford.
[2] Boucheron, S., Lugosi, G. and Massart, P. (2000) A sharp concentration inequality with applications. *Random Struct. Alg.* **16** 277–292.
[3] Boucheron, S. and Fernandez de la Vega, W. F. (2001) On the independence number of random interval graphs. *Combin. Probab. Comput.* **10** 385–396.
[4] Coffman, E. G., Lueker G. S., Spencer J. and Winkler, P. (2001) Packing random rectangles. *Prob. Theory Rel. Fields* **120** 585–599.
[5] Deuschel, J.-D. and Zeitouni, O. (1999) On increasing subsequences of i.i.d. samples. *Combin. Probab. Comput.* **8** 247–263.
[6] Frieze, A. (1990) On the length of the longest monotone subsequence in a random permutation. *Ann. Appl. Probab.* **1** 301–305.
[7] Justicz, J., Scheinerman, E. R. and Winkler, P. M. (1990) Random intervals. *Amer. Math. Monthly* **97** 881–889.
[8] Scheinerman, E. R. (1989) Random interval graphs. *Combinatorica* **8** 357–371.
[9] Steele, M. (1997) *Probability Theory and Combinatorial Optimization*, Vol. 69 of *SIAM CBMS–NSF Regional Conference Series in Applied Mathematics*.
[10] Talagrand M. (1995) Concentration of measure and isoperimetric inequalities in product spaces (English). *Publ. Math., Inst. Hautes Etud. Sci.* **81** 73–205.

https://doi.org/10.1017/S0963548301005041 Published online by Cambridge University Press