2010

Packing Rates Of Measures And A Conjecture For The Packing Density Of 2413

C. B. Presutti

Walter Stromquist
Swarthmore College

Recommended Citation

C. B. Presutti and Walter Stromquist. (2010). "Packing Rates Of Measures And A Conjecture For The Packing Density Of 2413". *Permutation Patterns: St. Andrews 2007*. Volume 376, 287-316. DOI: 10.1017/CBO9780511902499.015
https://works.swarthmore.edu/fac-math-stat/230

Packing rates of measures and a conjecture for the packing density of 2413

Cathleen Battiste Presutti

Department of Mathematics
Ohio University - Lancaster
Lancaster, Ohio 43130 USA

Walter Stromquist

Department of Mathematics and Statistics
Swarthmore College
Swarthmore, Pennsylvania 19081 USA

Abstract

We give a new lower bound of 0.10472422757673209041 for the packing density of 2413, justify it by a construction, and conjecture that this value is actually equal to the packing density. Along the way we define the packing rate of a permutation with respect to a measure, and show that maximizing the packing rate of a pattern over all measures gives the packing density of the pattern.

In this paper we consider the packing density of the pattern 2413. This pattern is significant because it is not layered, and because up to

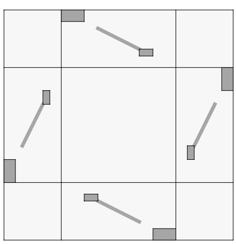


Fig. 1. The conjecture is based on this measure, μ_2

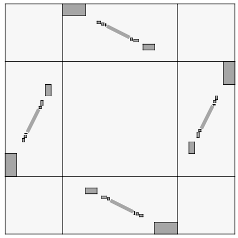


Fig. 2. The measure μ_∞

symmetry it is the smallest nontrivial pattern that is simple in the sense of [3]. We conjecture that its packing density is given by

$$\delta(2413) = 0.10472422757673209041\ldots,$$

and we show by a construction that this value is a lower bound. It is slightly larger than the lower bound of $0.10425\ldots$ given in [4].

We leap ahead briefly to describe the construction. Figure 1 describes a probability distribution on the unit square. Probability is concentrated on the dark shaded rectangles and the dark shaded segments. The distribution is described below. It is not uniform along the segments or in the rectangles; in fact, the rectangles are "recursion bubbles," meaning that each of them is a scaled-down replica of the entire figure. To construct a permutation of size n (for n large) with a large number of occurrences of the pattern 2413, we select n points independently from this distribution, and treat them as the graph of a permutation. In the limit of large n, with probability one, the packing density of 2413 in the resulting permutations approaches the value given above for $\delta(2413)$. (Figure 1 is not drawn to scale. If it were, the smaller recursion boxes would be too small to see.)

Permutations constructed in this way tend to consist of an initial increasing sequence, then two interleaved increasing sequences (one of high values, one of low values), then a final increasing sequence. Figure 9, below, shows the graph of one such permutation (with $n = 8$) and Figure 11, a prototype of Figure 1, suggests the more general pattern.

The basic definitions related to packing densities are reviewed in Section 1.

Our principal technique is to reinterpret packing densities in the language of measures. By a measure we mean a probability distribution on the unit square. In Section 2 we define the packing rate of a pattern with respect to a measure, and define the packing rate $\delta'(\pi)$ of a pattern π as the supremum of its packing rates over all measures. Our main result, in Section 4, is that the packing rate of a pattern is equal to its packing density, $\delta'(\pi) = \delta(\pi)$. Finding the packing density of a pattern is then a matter of finding an optimal measure for the pattern.

We return to the packing density of 2413 in Section 5. The language of measures allows us to bring to bear the techniques of analysis, including the calculus of variations and extensive calculations involving integrals of probability distribution functions. In Sections 6 to 8 we define four-segment measures, and by extensive calculation we find the optimal measure for 2413 within this class. In Sections 9 to 11 we improve the measure slightly by the use of recursion bubbles. We conjecture that the optimal measure is the one we call μ_2, which is the measure that is illustrated in Figure 1, on which the conjecture is based.

Figure 2 illustrates an attractive alternative called μ_∞, in which there is an infinite sequence of recursion bubbles at each end of each segment. We do not believe that μ_∞ is optimal, but we cannot rule it out.

1 Packing densities

Let $\pi \in S_m$. A sequence $x_1, \ldots, x_m$ *has the order type of* π if, for all i and j, $x_i < x_j \Leftrightarrow \pi_i < \pi_j$. This requires at least that the terms x_i be distinct. If $\sigma \in S_n$ then an *occurrence of* π *in* σ is an m-term subsequence of σ that has the order type of π. The number $\nu(\pi, \sigma)$ of such occurrences is called the *packing number of* π *in* σ, and the ratio

$$\delta(\pi, \sigma) = \frac{\nu(\pi, \sigma)}{\binom{n}{m}} \tag{1}$$

is called the *packing density of* π *in* σ. Clearly $0 \leq \delta(\pi, \sigma) \leq 1$. In this context π is called a *pattern*. We always assume that $\pi \in S_m$, $\sigma \in S_n$, and $n \geq m \geq 1$.

For a fixed pattern π we are concerned with finding permutations $\sigma \in S_n$ that maximize the packing density, especially in the limit as $n \to \infty$. Write

$$\delta(\pi, n) = \max_{\sigma \in S_n} \delta(\pi, \sigma). \tag{2}$$

If σ realizes this maximum—that is, if $\delta(\pi, \sigma) = \delta(\pi, n)$—then σ is called an *optimizer* (or "optimizing permutation") of size n for π. The *packing density of* π is

$$\delta(\pi) = \lim_{n \to \infty} \delta(\pi, n). \tag{3}$$

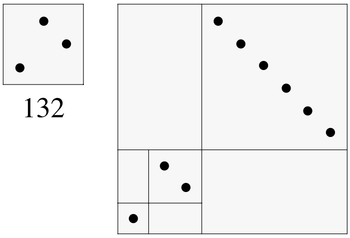


Fig. 3. $\nu(132, 132987654) = 46$

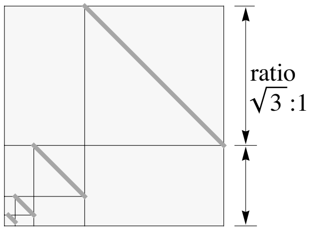


Fig. 4. A measure for 132

Galvin showed that the sequence $\{\delta(\pi, n)\}$ is non-increasing, so its limit always exists.

Equivalently we could define the packing density as the largest number D for which there is a sequence of permutations $\sigma_1, \sigma_2, \ldots$ of increasing size with

$$D = \lim_{i \to \infty} \delta(\pi, \sigma_i). \tag{4}$$

Such a sequence with $D = \delta(\pi)$ is called an *optimizing sequence* for π, and the permutations σ_i are called (collectively) *near-optimizers*. They do not need actually to be optimizers; they need only be close enough to give the right limit.

As an example consider the pattern $\pi = 132$. If $\sigma = 132987654$ then $\nu(\pi, \sigma) = 46$ and $\delta(\pi, \sigma) = 46/84 \approx 0.548$. This turns out to be the unique optimizer of size 9 for π, so $\delta(\pi, 9) = 46/84$ as well. These permutations are illustrated in Figure 3. The shape of σ suggests a recursive construction of near-optimizers for larger n. In fact, as is well known, this construction does produce an optimizing sequence for 132, whose packing density turns out to be $\delta(132) = 2\sqrt{3} - 3 \approx 0.464$. (Even for $\pi = 132$ it is not so easy to find optimizers for particular values of

n. Rounding issues arise and many possibilities need to be considered. Near-optimizers are easier.)

A simpler illustration of the near-optimizers for 132 is in Figure 4. The points in the graph of σ line up along the diagonal lines in the figure, and are distributed uniformly by length (to the extent possible for any particular value of n).

Figure 4 could be understood as simply a guide to the imagination. We prefer to give it a more formal meaning: we interpret pictures like this as defining probability measures on the unit square. In the next sections we will clarify this interpretation and show how it relates to packing densities.

2 Packing rates for measures

We consider probability measures μ on the unit square $S = [0,1]\times[0,1] \subseteq \mathbf{R}^2$. In this section we define $\delta'(\pi,\mu)$, the packing rate of a pattern π with respect to a measure μ. The packing rate of the pattern, $\delta'(\pi)$, is the supremum of the rates $\delta'(\pi,\mu)$ over all measures μ.

Recall that a *measure* μ on S assigns a non-negative value $\mu(A)$ to each Borel set $A \subseteq S$ in such a way that $\mu(\cup A_i) = \sum(\mu(A_i))$ whenever $\{A_i\}$ is a finite or countable sequence of pairwise disjoint sets. It is a *probability measure* if $\mu(S) = 1$. Borel sets are subsets of S that can obtained from closed rectangles in finitely many steps, each step being a complementation, a union or intersection of finitely many sets, or a union or intersection of countably many sets. In this paper all of the sets we encounter are Borel sets and "measure" always means probability measure.

A measure can be interpreted as a guide for selecting points randomly from S. When we say that a point is selected "according to μ" we mean that the probability that the point is in any set A is $\mu(A)$. Our plan is to pick m points independently according to μ and look at the order type of the resulting configuration.

Suppose that an m-tuple of points in S has no repeated x coordinates and no repeated y coordinates. We say that it *has the order type of* π if, when the points are arranged in order of increasing x coordinates, their y coordinates form a sequence with the order type of π. See Figure 5. (The order of the points in the m-tuple does not affect the order type.) An m-tuple that has a repeated x coordinate or a repeated y coordinate is called *degenerate* and has no order type.

An *order-preserving transformation* of S is a map of the form $(x, y) \mapsto$

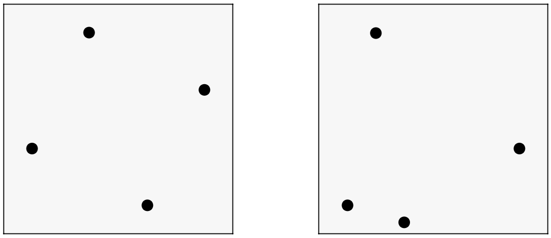


Fig. 5. Graph of 2413, and a 4-tuple with the order type of 2413

$(f(x), g(y))$ where each of f and g is an order-preserving bijection of $[0, 1]$. Transformations of this type preserve the order type of any m-tuple. Given any two m-tuples with the same order type, we can find an order-preserving transformation of S that maps one onto the other.

Let $\pi \in S_m$ be a pattern and let μ be a measure. The *packing rate* of π with respect to μ is the probability, if m points are selected independently according to μ, that they have the order type of π. We denote the packing rate by $\delta'(\pi, \mu)$.

More precisely: Let $\mu^m = \mu \times \cdots \times \mu$ be the product measure on S^m, and let $C_\pi \subseteq S^m$ contain all m-tuples of points that have order type π. Then the packing rate is

$$\delta'(\pi, \mu) = \mu^m(C_\pi). \tag{5}$$

The notation δ' for packing rates is temporary. After we relate packing rates to packing densities in Theorem 4.1 we will replace δ' with δ everywhere.

Examples of packing rates.

(i) Let μ be the uniform measure on S (Figure 6). Then all order types are equally likely. We have $\delta'(123, \mu) = 1/6$ and in general

$$\delta'(\pi, \mu) = \frac{1}{m!} \tag{6}$$

if π has size m.

(ii) Let μ be concentrated on the main diagonal of S (Figure 7). Then

$$\delta'(123, \mu) = 1 \tag{7}$$

and $\delta'(\pi, \mu) = 0$ for any other pattern π of size 3. It isn't necessary that μ be uniform on the diagonal as long as single points have zero probability.

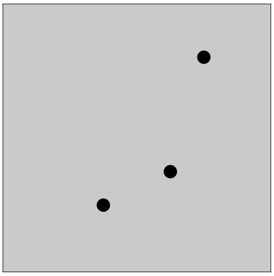


Fig. 6. uniform

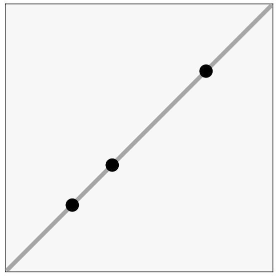


Fig. 7. diagonal

(iii) Let μ be concentrated along countably many diagonal segments as shown in Figure 4, with probability being proportional to length. Then

$$\delta'(132, \mu) = 2\sqrt{3} - 3 \approx 0.464. \tag{8}$$

This is equal to the packing density of 132. We call this an "optimal measure" for 132, because no other measure gives a higher packing rate.

(iv) **A challenge.** Let μ be uniform on a disk in S (Figure 8). Then what is $\delta'(123, \mu)$? (We don't know!)

(v) **Template measures.** Let $\tau \in S_k$ be a permutation and form a measure μ_τ as shown in Figure 9. The measure is concentrated uniformly on the union of k small squares arranged like the graph of τ. Then μ_τ is called the *template measure* corresponding to τ. We will use template measures in the proof of Theorem 4.1 and we will have more to say about them in Section 5. For now, consider the packing rate of a pattern π with respect to its own template measure μ_π. If m points are selected from m different cells in the template—an event that occurs with probability $m!/m^m$—then they are guaranteed to form an occurrence of τ. Therefore,

$$\delta'(\pi, \mu_\pi) \geq \frac{m!}{m^m}. \tag{9}$$

The *packing rate of* π is the supremum of $\delta'(\pi, \mu)$:

$$\delta'(\pi) = \sup_\mu \delta'(\pi, \mu), \tag{10}$$

the supremum being taken over all probability measures μ on S. An

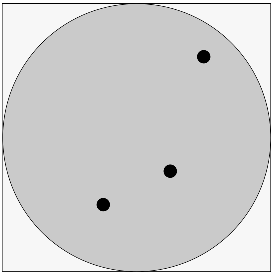


Fig. 8. disk

optimal measure for π (or "optimizer" when we are considering only measures) is a measure μ that achieves the supremum.

The example of the template measure shows that $\delta'(\pi) \geq m!/m^m$ for any $\pi \in S_m$.

3 Limits of measures

Is there an optimal measure for every pattern? That is, is the supremum in (10) really a maximum? The answer is yes, and we can prove it by forming a limit of of a sequence of measures whose packing rates approach the supremum. But the proof requires care for two reasons:

- We need a suitable definition for the limit of a sequence of measures; and
- Limits of measures do not always respect packing rates.

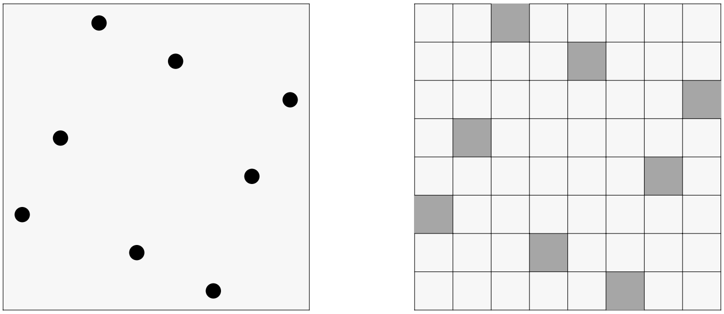


Fig. 9. $\tau = 35827146$ and the template measure μ_τ

As a cautionary example consider the measures μ_j defined as in Figure 10. Each μ_j is concentrated (uniformly) on the square $[1/(j+1), 1/j]^2 \subseteq S$ for $j = 1, 2, \ldots$. The only candidate for a limiting measure is the measure μ defined by a point mass at the origin. Then for each j we have $\delta'(123, \mu_j) = 1/6$, but $\delta'(123, \mu) = 0$.

The trouble, of course, is that the limit measure allows degenerate m-tuples. We need to identify circumstances in which this does not occur.

First we define limits of measures. We say that μ is the limit of a sequence of measures $\{\mu_j\}$,

$$\mu = \lim \mu_j,$$

if

$$\int_S f(p) d\mu(p) = \lim_{j\to\infty} \int_S f(p) d\mu_j(p) \tag{11}$$

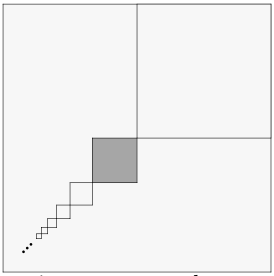
for every continuous function $f : S \to \mathbf{R}$. With this definition the limit of a sequence is unique (if it exists) and the probability measures on S form a compact topological space. This means that from any sequence of measures $\{\nu_i\}$ we can select a sequence $\{\mu_j\}$ that has a limit measure.

It is not generally true that $\mu(A) = \lim \mu_j(A)$ for an arbitrary set A. (Consider $A = \{(0,0)\}$ in the cautionary example.) But it can be shown from the definition that if A is a closed set,

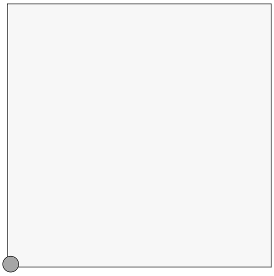
$$\mu(A) \geq \limsup_{j\to\infty} \mu_j(A). \tag{12}$$

and if B is an open set,

$$\mu(B) \leq \liminf_{j\to\infty} \mu_j(B). \tag{13}$$



μ_j is concentrated on $[1/(j+1), 1/j]^2$.



μ is a point mass at the origin.

Fig. 10. Limiting measures do not always respect packing rates

Well-behaved measures. We now identify some special classes of measures. For this paper, a measure μ is *smooth* if $\mu(A) = 0$ whenever A is a vertical or horizontal line. If points are selected independently according to a smooth measure, the probability of their forming a degenerate configuration is zero.

A measure μ is *normalized* if its projection onto each axis is the uniform measure; that is, if

$$\mu([0,a] \times [0,1]) = \mu([0,1] \times [0,a]) = a \tag{14}$$

for every $a \in [0,1]$. Every normalized measure is smooth. Better, as we now show, the limit of a sequence of normalized measures is necessarily normalized, and limits of normalized measures respect packing rates.

Lemma 3.1. *If $\mu = \lim_{j\to\infty} \mu_j$ and each μ_j is normalized, then μ is also normalized and for any pattern π,*

$$\delta'(\pi, \mu) = \lim_{j\to\infty} \delta'(\pi, \mu_j).$$

Proof. To see that μ is normalized note that

$$\mu([0,a] \times [0,1]) \geq \limsup_{j\to\infty} \mu_j([0,a] \times [0,1]) \tag{15}$$

because this set is closed. Every term in the right-hand sequence is a, so $\mu([0,a] \times [0,1]) \geq a$. If $\epsilon > 0$ then

$$\begin{aligned} \mu\left([0,a] \times [0,1]\right) &\leq \mu\left([0,a+\epsilon) \times [0,1]\right) \\ &\leq \liminf_{j\to\infty} \mu_j\left([0,a+\epsilon) \times [0,1]\right) \end{aligned} \tag{16}$$

so $\mu([0,a] \times [0,1]) < a + \epsilon$. Since this holds for every $\epsilon > 0$ it follows that

$$\mu([0,a] \times [0,1]) = a. \tag{17}$$

The same is true in the other dimension, so μ is normalized.

To see that these limits respect packing rates, note that according to any of the normalized measures μ or μ_j, the boundary of each C_π has measure zero. Therefore

$$\begin{aligned}
\delta'(\pi,\mu) &= \mu^m(C_\pi) \\
&= \mu^m(\text{int } C_\pi) \\
&\leq \liminf \mu_j^m(\text{int } C_\pi) \\
&= \liminf \delta'(\pi,\mu_j) \\
&\leq \limsup \delta'(\pi,\mu_j) \\
&= \limsup \mu_j^m(\text{cl } C_\pi) \\
&\leq \mu^m(\text{cl } C_\pi) \\
&= \mu^m(C_\pi) \\
&= \delta'(\pi,\mu). \qquad (18)
\end{aligned}$$

This is enough to force $\lim \delta'(\pi,\mu_j) = \delta'(\pi,\mu)$. □

Lemma 3.2. *If μ is any measure on S then there is a normalized measure $\tilde{\mu}$ such that $\delta'(\pi,\tilde{\mu}) \geq \delta'(\pi,\mu)$ for every pattern π.*

We call $\tilde{\mu}$ a *normalization* of μ. It is unique if μ is smooth. We conclude from Lemma 3.2 that if we want to maximize $\delta'(\pi,\nu)$ it suffices to look among normalized measures ν.

Proof. If μ is smooth we can find an order-preserving transformation of S that maps μ to a normalized measure. More precisely, we can define $\tilde{\mu}$ by

$$\tilde{\mu}([0,a]\times[0,b]) = \mu([0,x]\times[0,y]) \qquad (19)$$

for every pair (a,b), where x is the least value for which

$$\mu([0,x]\times[0,1]) = a \qquad (20)$$

and y is the least value for which

$$\mu([0,1]\times[0,y]) = b. \qquad (21)$$

The values specified in (19) are enough to determine $\tilde{\mu}$ on all Borel sets. Now $\delta'(\pi,\tilde{\mu}) = \delta'(\pi,\mu)$ for all patterns π.

If μ is not smooth, we apply (19) whenever x and y are uniquely determined, and then extend $\tilde{\mu}$ arbitrarily to a normalized measure on S. The implied mapping is not a bijection, and some m-tuples which have no order type at all (because an x coordinate is repeated or a y coordinate is repeated) may be mapped to m-tuples that do have order types. So, the probability of an order type π arising may be greater under $\tilde{\mu}$ than under μ, and we may have $\delta'(\pi, \tilde{\mu}) > \delta'(\pi, \mu)$. □

The next theorem says that every pattern has an optimal measure.

Theorem 3.3. *For every pattern π there is a normalized measure μ for which $\delta'(\pi, \mu) = \sup_\nu \delta'(\pi, \nu)$, the supremum being taken over all measures ν on S.*

Proof. Let $D = \sup_\nu \delta'(\pi, \nu)$. Find a sequence $\{\nu_i\}$ of measures such that

$$\lim_{i\to\infty} \delta'(\pi, \nu_i) = D. \tag{22}$$

Replacing each ν_i with its normalization does not change the limit (it might increase some values of $\delta'(\pi, \nu_i)$, but not above D) so we might as well assume that each ν_i is normalized. We can find among them a subsequence $\{\mu_j\}$ of $\{\nu_i\}$ that converges to a limit measure μ. Then μ is necessarily normalized and we have

$$\delta'(\pi, \mu) = \lim_{j\to\infty} \delta'(\pi, \mu_j) = D \tag{23}$$

as required. □

4 Packing rates are packing densities

Theorem 4.1. *For every pattern π,*

$$\delta'(\pi) = \delta(\pi).$$

Proof. First we show that $\delta(\pi) \geq \delta'(\pi)$.

More generally, if μ is any measure we show that there exists a sequence of permutations $\{\sigma_i\}$ such that $\lim_{i\to\infty} \delta(\pi, \sigma_i) \geq \delta'(\pi, \mu)$. Since this result holds for every measure μ, including the optimal measure, it follows that $\delta(\pi) \geq \delta'(\pi)$.

First suppose that a permutation σ of length n is chosen randomly according to μ—that is, n points are selected independently according to μ, and σ is their order type. Suppose then that we select an m-element subsequence from σ, also randomly. Then we might has well have chosen m-element subsequence directly according to μ, so the probability that it is an occurrence of π is exactly $\delta'(\pi, \mu)$. This means that the expected value of $\delta(\pi, \sigma)$ is at least $\delta'(\pi, \mu)$.

It follows that (for each n) there exists at least one specific permutation σ such that $\delta(\pi,\sigma) \geq \delta'(\pi,\mu)$. From the sequence of these permutations select a subsequence $\{\sigma_i\}$ for which the limit $\lim_{i\to\infty} \delta(\pi,\sigma_i) = D$ exists; then $\delta(\pi) \geq D \geq \delta'(\pi,\mu)$. Since this result holds for every measure μ, including the optimal measure, it follows that $\delta(\pi) \geq \delta'(\pi)$.

Next we show that $\delta'(\pi) \geq \delta(\pi)$.

Let $\{\sigma_i\}$ be a sequence of permutations of increasing size satisfying $\lim \delta(\pi,\sigma_i) = \delta(\pi)$. From each σ_i construct a template measure ν_i as defined above. Suppose that each σ_i has size n_i.

Select an m-tuple according to ν_i. For it to have the order type of π it suffices that

- the points come from m different boxes, and
- the boxes correspond to an occurrence of π in σ_i.

The probability of the first event is $\frac{n_i!/(n_i-m)!}{n_i^m}$ and then the conditional probability of the second event is $\delta(\pi,\sigma_i)$. Therefore

$$\delta'(\pi,\nu_i) \geq \frac{n_i!/(n_i-m)!}{n_i^m}\delta(\pi,\sigma_i). \tag{24}$$

Each ν_i is normalized, so there is a subsequence $\{\mu_j\}$ with a limit μ that is also normalized, and

$$\begin{aligned}
\delta'(\pi,\mu) &= \lim_{j\to\infty} \delta'(\pi,\mu_j) \\
&= \lim_{i\to\infty}\left(\frac{n_i!/(n_i-m)!}{n_i^m}\right)\lim_{i\to\infty}\delta(\pi,\sigma_i) \\
&= \delta(\pi).
\end{aligned} \tag{25}$$

It follows that $\delta'(\pi) \geq \delta(\pi)$. □

We now abandon the notation δ' in favor of δ in all uses. The packing rate of π with respect to a measure μ is $\delta(\pi,\mu)$, and the packing density is $\delta(\pi)$ whether it arises from a sequence of permutations or an optimal measure.

Open question: Is the normalized optimal measure for π unique?

5 The packing density of 2413

With the language of measures now firmly in place, we come to the packing density of 2413. In this section we summarize the existing lower bounds on $\delta(2413)$, all of which have been obtained using template measures in one form or another.

Recursive template measures. Let $\tau \in S_k$. In Section 2 we defined the template measure μ_τ to be uniform on k small squares arranged like the graph of τ (Figure 9). One way for a pattern π to occur in μ_τ is for m points to be chosen from different squares which happen to correspond to an occurrence of π in τ. The probability of this event is given in equation (24); in the current context it is

$$\frac{k!/(k-m)!}{k^m}\delta(\pi,\tau).$$

If we substitute $\delta(\pi,\tau) = \nu(\pi,\tau)/\binom{k}{m}$ we get the equivalent form

$$\frac{m!}{k^m}\nu(\pi,\tau). \tag{26}$$

This construction can be refined by modifying the measure within each small square to be a reduced-scale copy of the measure on S itself. We call the result the *recursive template measure* corresponding to τ. (From now on we will reserve the notation μ_τ for the recursive template.) In a recursive template there is another good way for occurrences of a pattern π to occur: if m points are drawn from the same square (probability k/k^m) then they form an occurrence of π with probability $\delta(\pi,\mu_\tau)$. Combining these two ways gives

$$\delta(\pi,\mu_\tau) \geq \frac{m!}{k^m}\nu(\pi,\tau) + \frac{k}{k^m}\delta(\pi,\mu_\tau) \tag{27}$$

which can be solved to give

$$\delta(\pi,\mu_\tau) \geq \frac{(m!)\nu(\pi,\tau)}{k^m - k}. \tag{28}$$

This formula appears in [4]. It is an inequality because there may be yet other ways for a pattern to occur (although this is not the case when $\pi = 2413$).

For any pattern π we can use this construction with $\tau = \pi$ (whence $\nu(\pi,\tau) = 1$) to obtain

$$\delta(\pi,\mu_\pi) \geq \frac{m!}{m^m - m} \tag{29}$$

which gives a lower bound for the packing density of any pattern of size m. In the case of $\pi = 2413$, we have $m = 4$ and the bound is

$$\delta(2413) \geq 2/21 \approx 0.0952. \tag{30}$$

In [2] the same construction was applied using $\tau = 35827146$ (Figure 9) to obtain

$$\delta(2413) \geq 51/511 \approx 0.099804. \tag{31}$$

This follows from equation (28) with $k = 8$, $m = 4$, and $\nu(2413, 35827146) = 17$.

Warren [5] used this construction with $k = 12$ and $\tau = 5\ 4\ 7\ 12\ 11\ 3\ 10\ 2\ 1\ 6\ 9\ 8$ to obtain

$$\delta(2413) \geq 16/157 \approx 0.101911. \tag{32}$$

Weighted templates. The above construction can be improved in another way. We can alter the probabilities allocated to the small squares in the template. The probabilities (weights) can be assigned arbitrarily, as long as they add to 1.

Presutti [4] uses a template based on the permutation

$$579(11)(16)4(15)3(14)2(13)168(10)(12)$$

(with $m = 16$) and optimizes weights using Mathematica to obtain

$$\delta(2413) \geq 0.10425098006897487 4, \tag{33}$$

which is the best lower bound that has appeared.

Empirical results. Other researchers have used empirical methods to find optimizing permutations σ for 2413, including cases with large n. Michael Albert, Nik Ruskuc, and Imre Leader found optimizers and near-optimizers for large values of n, and were able to use them to establish lower bounds greater than 51/511. Albert and Vince Vatter (separately) used simulated annealing to find additional examples [1]. The optimizers seem to have a consistent form. If σ is one of these optimizers, then generally σ consists of...

- An initial, increasing segment, consisting of middle-range values;
- A segment with two interleaved decreasing sequences, one with high values and one with low values; and
- A terminal, increasing segment, with middle-range values overlapping those of the initial segment.

In effect, the points in the graph of σ seem to be lining up along the segments illustrated in Figure 11, below. The optimizers also have some local complications corresponding to the "recursion bubbles" that we introduce in Section 9.

The template permutations used above are all of this form. For example, 35827146 (illustrated in Figure 9) consists of an initial increasing segment 35, two interleaved decreasing sequences 8_7 (with high values) and 2_1 (with low values) and a terminal increasing segment 46. The templates used by Warren (size 12) and Presutti (size 16) are also of this form. (Well, actually, Warren's template is not of this form, and does not even have four-fold symmetry. But the slightly modified template $\tau = 457(12)(11)3(10)21689$ is of the above form and has exactly the same 2413-occurrences as Warren's template. None of these examples is large enough to show recursion bubbles.)

This form of the optimizers motivates the definition of a "four-segment measure" in the next section.

6 Four-segment measures

We define a class of measures that offer good packing rates for 2413.

A *symmetrical four-segment measure* (SFS measure) is a measure on S which is

- symmetrical with respect to four-fold rotations of S, and
- concentrated on the line segment from $(1/4, 1/4)$ to $(3/4, 0)$ and the three segments obtained from it by rotations of S.

That means that the measure is concentrated on the four segments illustrated in Figure 11.

There is no requirement that the measures be uniform on the segments. In fact, the measures in this class differ precisely in their distributions along the segments. Because of symmetry, each is determined by the distribution along the bottom segment.

We give a name to this distribution. Let μ be a four-segment measure and let $t \in [0, 1]$. Then let $F(t)$ be the probability, given that a point is on the bottom segment, that it is in the leftmost fraction t of the segment.

One way to make this definition precise is to write

$$F(t) = 4\mu([1/4, 1/4 + t/2] \times [0, 1/4]) \tag{34}$$

for $t \in [0, 1]$.

(This is an awkward formula, mainly because—for convenience in later calculations—we have chosen to make F have domain $[0, 1]$ and range $[0, 1]$. This forces a mismatch of coordinates. While the argument t

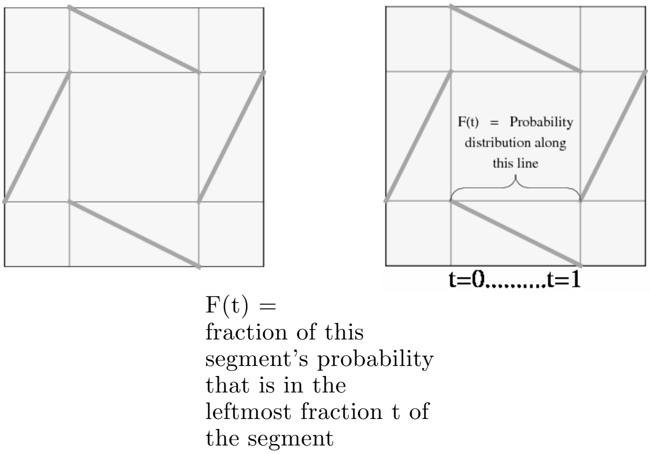


Fig. 11. The segments on which four-segment measures are concentrated

runs from 0 to 1, the coordinate x runs from $1/4$ to $3/4$. One unit of t corresponding to $1/2$ unit of x.)

Now any SFS measure is completely determined by the function F. Like any cumulative distribution function on $[0,1]$, F is non-decreasing and satisfies $F(1) = 1$. It is not necessarily differentiable.

An SFS measure is not *a priori* smooth or normalized. Since we are concerned with maximizing packing rates we can limit our attention to smooth measures, but we can't usually normalize an SFS measure without bending the segments. We can, however, partially normalize the measure by requiring that, projected onto the x axis, the measure be uniform on $[1/4, 3/4]$. To preserve symmetry we then do the same thing for the y axis. The geometry of the four segments is such that the two operations do not interfere with each other. This process does not alter the packing rate for 2413 or any other pattern, so we might as well limit our attention to SFS measures that are partially normalized in this sense.

This assumption gives us the formula

$$F(t) + (1 - F(1-t)) = 2t \tag{35}$$

for every $t \in [0,1]$. We call this the *normalization identity,* and we assume that this relationship holds for every SFS measure we consider. This relationship forces the measure to be smooth and forces the distribution F to be continuous. It also implies that the slope of F is bounded between 0 and 2. (More precisely, since the graph of F need not always have a slope, it implies that every difference quotient

$$\frac{F(b) - F(a)}{b - a}$$

is in the interval $[0, 2]$.) On intervals where F is flat, all of the probability is in the upper segment; on intervals where the graph of F has slope 2, all of the probability is on the lower segment. On other intervals there is probability on both segments (which accounts for interleaved sequences in the empirical optimizing permutations).

We now have a class of measures with which to proceed. We make the following non-conjecture:

Non-Conjecture 6.1. The optimal measure for 2413 is a symmetrical four-segment measure determined by a distribution function F satisfying (35).

We call this a non-conjecture because we will prove in Section 9 that it is false. In that section we will show that adding recursion bubbles to the best SFS measure increases the packing rate; hence, the best SFS measure isn't optimal. Still, it's a starting point. Our plan in the next two sections is to find, with proof, the SFS measure that optimizes the packing density of 2413 among SFS measures. Then we can begin to improve that measure with recursion bubbles.

7 The packing rate for an SFS measure

In this section we give a formula for the packing rate of 2413 with respect to a symmetrical four-segment measure. Most of the rest of the section consists of the proof, which involves heavy calculation. Theorem 7.2, at the end of the section, gives an alternative formula.

Theorem 7.1. *Let μ be the symmetrical four-segment measure determined by a distribution function F satisfying the normalization identity (35). Then the packing rate of* 2413 *with respect to μ is given by*

$$\delta(2413, \mu) = \frac{5}{32} + \frac{3}{4}\left(\int_{t=0}^{1} F(t)\,dt\right)^2 + \int_{t=0}^{1} \left(\left(\frac{3}{4}t - \frac{9}{8}\right) F(t)^2 - \frac{1}{4}F(t)^3\right) dt. \quad (36)$$

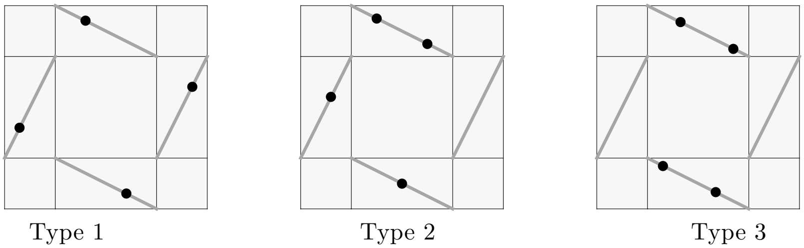


Fig. 12. Three types of 2413-occurrences

Proof. For full generality we will use the notation of Stieltjes integrals. Recall that the integral

$$\int_{t=a}^{b} H(t)dF(t) \tag{37}$$

gives the mean of $H(t)$ when t is drawn from the probability distribution defined by F. If F has a derivative f then the integral can be understood as

$$\int_{t=a}^{b} H(t)dF(t) = \int_{t=a}^{b} H(t)f(t)dt. \tag{38}$$

More generally the integral is defined using Riemann-like sums:

$$\int_{t=a}^{b} H(t)dF(t) = \lim \sum_{i=0}^{n} H(x_i^*)\left(F(x_i) - F(x_{i-1})\right) \tag{39}$$

where x_i^* is an arbitrary point in $[x_{i-1}, x_i]$ and the limit is over partitions $a \leq x_0 < x_1 < \cdots < x_n = b$ with decreasing mesh size. The formula for integration by parts is

$$\int H(t)dF(t) = H(t)F(t) - \int F(t)dH(t) \tag{40}$$

or

$$\int_{t=a}^{b} H(t)dF(t) = H(t)F(t)\big|_{t=a}^{b} - \int_{t=a}^{b} F(t)dH(t). \tag{41}$$

With these tools in hand we begin the evaluation of $\delta(2413, \mu)$. There are three ways for a 4-tuple of points selected according to μ to be an occurrence of 2413: Types 1, 2, and 3, each illustrated in Figure 12.

The 4-tuple is an occurrence of Type 1 if

- One point is chosen from each segment (probability 3/32);
- The top point is to the left of the bottom point; and

- The right point is above the left point.

Let X be the probability (given that there is one point on each segment) that the top point is to the left of the bottom point. Then X is given by

$$X = \int_{t=0}^{1} (1 - F(1-t))dF(t). \tag{42}$$

Using the identity (35) and integration by parts we obtain

$$\begin{aligned} X &= \int_{t=0}^{1} (2t - F(t))dF(t) \\ &= 2\int_{t=0}^{1} tdF(t) - \int_{t=0}^{1} F(t)dF(t) \\ &= 2\left(1 - \int_{t=0}^{1} F(t)dt\right) - \frac{1}{2}\, F(t)^2\big|_{t=0}^{1} \\ &= 2\left(1 - \int_{t=0}^{1} F(t)dt\right) - \frac{1}{2} \\ &= \frac{3}{2} - 2\int_{t=0}^{1} F(t)dt. \end{aligned} \tag{43}$$

As an example, consider the measure μ determined by a uniform distribution along the four segments; that is, consider the case of $F(t) = t$. In this case we would expect to find $X = 1/2$, and that is indeed the value given by equation (43).

By symmetry the probability that the right point is above the left point is also X, so the probability of a Type 1 occurrence is

$$\frac{3}{32}X^2 = \frac{27}{128} + \frac{3}{8}\left(\int_{t=0}^{1} F(t)dt\right)^2 - \frac{9}{16}\int_{t=0}^{1} F(t)dt. \tag{44}$$

The 4-tuple is an occurrence of Type 2 if

- Two points are on the top segment and one each are on the left and bottom segments (probability 3/64 before allowing for rotations);
- The point on the bottom segment is horizontally between the points on the top segment.

Let Y be the probability, given that the points are on the correct segments, that the bottom point is horizontally between the top points. Then Y is given by

$$Y = 2\int_{t=0}^{1} F(1-t)(1 - F(1-t))dF(t). \tag{45}$$

The initial factor 2 appears because the two points on the top segment can occur in either order. Substituting $1-t$ for t makes this

$$Y = 2\int_{t=0}^{1} F(t)\,(1-F(t))\,dF(1-t). \tag{46}$$

Differentiating (35) gives

$$dF(1-t) = 2dt - dF(t), \tag{47}$$

and substituting this into our expression gives

$$Y = 4\int_{t=0}^{1} F(t)\,(1-F(t))\,dt - 2\int_{t=0}^{1} F(t)\,(1-F(t))\,dF(t). \tag{48}$$

The second integral is $\left(\frac{1}{2}F(t)^2 - \frac{1}{3}F(t)^3\right)\Big|_{t=0}^{1} = \frac{1}{6}$, so we can rewrite (48) as

$$Y = -\frac{1}{3} + \int_{t=0}^{1} \left(4F(t)dt - 4F(t)^2\right) dt. \tag{49}$$

For example, if $F(t) = t$, then $Y = \frac{1}{3}$ as we would expect. Now the probability of a Type 2 occurrence (including a factor of 4 to account for the rotations) is

$$4 \cdot \frac{3}{64}Y = -\frac{1}{16} + \int_{t=0}^{1} \left(\frac{3}{4}F(t)dt - \frac{3}{4}F(t)^2\right) dt. \tag{50}$$

A 4-tuple is a Type 3 occurrence if

- Two points are from the top segment and two from the bottom segment (probability 3/128 before allowing for rotations); and
- From the left, the four points have the order bottom, top, bottom, top.

Let Z be the probability, given that the points are on the correct segments, that they have the correct order. Then Z is given by

$$Z = 4\int_{t=0}^{1}\int_{s=1-t}^{1} F(1-t)F(1-s)dF(s)dF(t). \tag{51}$$

In this formulation t represents the distance from the left end of the bottom segment to the rightmost bottom point, and s represents the distance from the right end of the top segment to the leftmost of the top points. These must satisfy $s > 1-t$, hence the limits of the integrals. The order requirement is that one bottom point is left of s, one top point is at s, one bottom point is at t, and one top point is right of t; hence factors of $F(1-t)$, $dF(s)$, $F(1-s)$, $dF(t)$ respectively. The initial factor of 4 is there because the top points can exchange roles and the bottom points can exchange roles.

Write this expression as an iterated integral:

$$Z = 4\int_{t=0}^{1} F(1-t)\left[\int_{s=1-t}^{1} F(1-s)dF(s)\right] dF(t). \tag{52}$$

To evaluate the inner integral we first substitute $1-s$ for s, again using the identity $dF(1-s) = 2ds - dF(s)$:

$$\begin{aligned}
\int_{s=1-t}^{1} F(1-s)dF(s) &= \int_{s=0}^{t} F(s)dF(1-s) \\
&= \int_{s=0}^{t} F(s)(2ds - dF(s)) \\
&= 2\int_{s=0}^{t} F(s)ds - \int_{s=0}^{t} F(s)dF(s) \\
&= 2\int_{s=0}^{t} F(s)ds - \left(\frac{1}{2}F(s)^2\right)\Big|_{s=0}^{t} \\
&= 2G(t) - \frac{1}{2}F(t)^2
\end{aligned}$$

where $G(t) = \int_{s=0}^{t} F(s)ds$. Substituting this into the main integral gives

$$\begin{aligned}
Z &= 4\int_{t=0}^{1} F(1-t)\left(2G(t) - \frac{1}{2}F(t)^2\right) dF(t) \\
&= 4\int_{t=0}^{1} (1-2t+F(t))\left(2G(t) - \frac{1}{2}F(t)^2\right) dF(t).
\end{aligned}$$

Expanding this integral into six terms and integrating (usually by parts) gives

$$\begin{aligned} Z &= 8\int_{t=0}^{1} G(t)dF(t) - 16\int_{t=0}^{1} tG(t)dF(t) + 8\int_{t=0}^{1} F(t)G(t)dF(t) \\ &\quad -2\int_{t=0}^{1} F(t)^2dF(t) + 4\int_{t=0}^{1} tF(t)^2dF(t) - 2\int_{t=0}^{1} F(t)^3dF(t) \\ &= 8\left(\int_{t=0}^{1} F(t)dt - \int_{t=0}^{1} F(t)^2dt\right) \\ &\quad -16\left(\int_{t=0}^{1} F(t)dt - \frac{1}{2}\left(\int_{t=0}^{1} F(t)dt\right)^2 - \int_{t=0}^{1} tF(t)^2dt\right) \\ &\quad +8\left(\frac{1}{2}\int_{t=0}^{1} F(t)dt - \frac{1}{2}\int_{t=0}^{1} F(t)^3dt\right) \\ &\quad -2\left(\frac{1}{3}\right) + 4\left(\frac{1}{3} - \frac{1}{3}\int_{t=0}^{1} F(t)^3dt\right) - 2\left(\frac{1}{4}\right) \\ &= \frac{1}{6} + 8\left(\int_{t=0}^{1} F(t)dt\right)^2 \\ &\quad + \int_{t=0}^{1}\left(-4F(t) - 8F(t)^2 - \frac{16}{3}F(t)^3 + 16tF(t)^2\right)dt. \end{aligned}$$

When $F(t) = t$ this is $1/6$ as we would expect. The probability of a Type-3 occurrence (multiplying by 2 to account for the rotation) is

$$\begin{aligned} 2\cdot\frac{3}{128}Z &= \frac{1}{128} + \frac{3}{8}\left(\int_{t=0}^{1} F(t)dt\right)^2 \\ &\quad + \int_{t=0}^{1}\left(-\frac{3}{16}F(t) - \frac{3}{8}F(t)^2 - \frac{1}{4}F(t)^3 + \frac{3}{4}tF(t)^2\right)dt. \end{aligned} \quad (53)$$

Combining the probabilities for the three types, we have

$$\begin{aligned} \delta(2413,\mu) &= \frac{5}{32} + \frac{3}{4}\left(\int_{t=0}^{1} F(t)dt\right)^2 \\ &\quad + \int_{t=0}^{1}\left(\left(\frac{3}{4}t - \frac{9}{8}\right)F(t)^2 - \frac{1}{4}F(t)^3\right)dt \end{aligned} \quad (54)$$

as required. □

We aren't free to choose $F(t)$ arbitrarily for all $t \in [0,1]$. We may choose $F(t)$ on $[0,1/2]$ subject to certain constraints, but then the values on $[1/2,1]$ are forced on us by the normalization identity (35). It is helpful, therefore, to have an alternative to Theorem 7.1 in which the integrals are limited to the interval $[0,1/2]$.

Theorem 7.2. *Let μ be the symmetrical four-segment measure determined by a distribution function F satisfying (35). Then the packing rate of* 2413 *with respect to μ is given by*

$$\delta(2413,\mu) = \frac{3}{32} + 3\left(\int_{t=0}^{1/2} F(t)dt\right)^2 + \int_{t=0}^{1/2}\left[\left(3t-\frac{3}{4}\right)F(t) + \left(\frac{3}{2}t-\frac{9}{4}\right)F(t)^2 - \frac{1}{2}F(t)^3\right]dt. \quad (55)$$

Proof. Divide each of the integrals in Theorem 7.1 into two integrals, substitute $1-t$ for t in the second integral, simplify using the normalization identity, and recombine the integrals. For example:

$$\begin{aligned}\int_{t=0}^{1} F(t)dt &= \int_{t=0}^{1/2} F(t)dt + \int_{t=1/2}^{1} F(t)dt \\ &= \int_{t=0}^{1/2} F(t)dt + \int_{t=0}^{1/2} F(1-t)dt \\ &= \int_{t=0}^{1/2} F(t)dt + \int_{t=0}^{1/2} (1-2t+F(t))dt \\ &= 2\int_{t=0}^{1/2} F(t)dt + \int_{t=0}^{1/2} (1-2t)dt \\ &= 2\int_{t=0}^{1/2} F(t)dt + \frac{1}{4} \qquad (56)\end{aligned}$$

The other integral in (36) can be restated in the same way, and the results can be combined to give (55). We leave the calculation to the reader. (Actually we invite the reader to leave the calculation to us. This is a good time to thank the volunteer referees who make mathematical publication possible.) □

8 The optimal SFS measure

In this section we use the calculus of variations to find a measure that maximizes $\delta(2413,\mu)$ among symmetrical four-segment measures μ.

Define a functional Φ by

$$\Phi[F] = \frac{3}{32} + 3\left(\int_{t=0}^{1/2} F(t)dt\right)^2 + \int_{t=0}^{1/2} \left(\left(3t - \frac{3}{4}\right) F(t) + \left(\frac{3}{2}t - \frac{9}{4}\right) F(t)^2 - \frac{1}{2}F(t)^3\right) dt \quad (57)$$

when F is defined on the interval $[0, 1/2]$. This is the formula from Theorem 7.2, which says that $\delta(\pi, \mu) = \Phi[F]$ when μ is the SFS measure determined by F. To find the optimal SFS measure, we need to maximize $\Phi[F]$ subject to certain constraints on F.

We are free to choose any distribution F provided that $F(0) = 0$, F is non-decreasing, and F satisfies the normalization identity (35). Equivalently: We can choose $F(t)$ arbitrarily on the interval $0 \le t \le 1/2$ subject to two constraints:

- $F(0) = 0$, and
- The difference quotients of F satisfy

$$0 \le \frac{F(t) - F(s)}{t - s} \le 2 \quad (58)$$

whenever $0 \le s < t \le 1/2$.

Then F can be extended to all of $[0, 1]$ using equation (35), and equation (58) is automatically satisfied on the entire interval. These requirements also force F to be continuous and nondecreasing on $[0, 1]$ and to satisfy $F(1) = 1$.

We say that F is *unconstrained* at t if, in some neighborhood of t, the difference quotients are bounded away from 0 and 2. Otherwise, F is *constrained* at t. The easiest way for F to be constrained at t is for the graph of F to have slope 0 or 2 on an interval containing t, but F can also be constrained at t (for example) if t is a limit point of such intervals. If F is unconstrained at t, we are free to make positive or negative adjustments to F in a neighborhood of t in an attempt to maximize $\delta(2413, \mu)$.

(When F is constrained at t, the corresponding measure has probability only on one segment—on top when $F(t) = 0$, and on the bottom when F has slope 2. When F is unconstrained, there is probability on both segments.)

Theorem 8.1. *Let* $J = \int_0^{1/2} F(t)dt$*. If* F *maximizes* $\delta(2413, \mu)$ *subject* *to the above requirements, then $F(t)$ must be given by*

$$F(t) = \sqrt{\left(t - \frac{1}{2}\right)^2 + \frac{3}{2} + 4J} + \left(t - \frac{3}{2}\right) \tag{59}$$

whenever F is unconstrained at t.

This is a local requirement. We will prove it first, then extend it to a global description of F in the next theorem.

Proof. Let H be any function with a continuous derivative on $[0, 1/2]$ and satisfying $H(0) = 0$ and $H(t) = 0$ whenever F is constrained at t. Then F may be altered by adding or subtracting a small multiple of H. It follows that the derivative of

$$\Phi[F + \epsilon H]$$

with respect to ϵ must be zero at $\epsilon = 0$. Compute:

$$\begin{aligned}
&\lim_{\epsilon \to 0} \frac{\Phi[F + \epsilon H] - \Phi[F]}{\epsilon} \\
&= \int_{t=0}^{1} \left(\left(6J + 3t - \frac{3}{4}\right) + \left(3t - \frac{9}{2}\right) F(t) - \frac{3}{2} F(t)^2 \right) H(t)\, dt. \quad (60)
\end{aligned}$$

This expression must be zero for F to be optimal, for any suitable H. When F is unconstrained at t we can choose H to be positive in a small neighborhood of t. Therefore we must have

$$\left(6J + 3t - \frac{3}{4}\right) + \left(3t - \frac{9}{2}\right) F(t) - \frac{3}{2} F(t)^2 = 0 \tag{61}$$

whenever $F(t)$ is unconstrained.

This can be solved uniquely for $F(t)$ (since $F(t) \geq 0$) giving

$$F(t) = \sqrt{\left(t - \frac{1}{2}\right)^2 + \frac{3}{2} + 4J} + \left(t - \frac{3}{2}\right). \tag{62}$$

as required. □

That is a local result. To understand the behavior of F globally, we must know when F is unconstrained. If $J < 1/8$ (which is the case for all plausible F) we can check that equation (59) never gives a slope greater than 2, but that it sometimes does give negative values for F.

In fact, equation (59) gives $F(t) \leq 0$ whenever $t \leq 1/4 - 2J$. Write $t^* = 1/4 - 2J$. This means that for the optimal F we must have $F(t) = 0$

when $t \le t^*$, and $F(t)$ given by the formula when $t^* \le t \le 1/2$. To summarize:

$$F(t) = \begin{cases} 0 & \text{when } t \le t^* \\ \sqrt{\left(t - \frac{1}{2}\right)^2 + \frac{3}{2} + 4J} + \left(t - \frac{3}{2}\right) & \text{when } t^* \le t \le \frac{1}{2}. \end{cases} \tag{63}$$

This formula for F is circular because it makes F depend on its own integral J. In fact, there is only one value of J that makes the formula consistent, which we will call J^*, and only one corresponding value of t^*. We turn to Mathematica for a numerical integral and solution:

$$J^* \approx 0.05110454191162339225 \tag{64}$$

$$t^* = \frac{1}{4} - 2J^* \approx 0.14779091617675321550 \tag{65}$$

(This calculation is almost analytic. If K is the smallest positive solution of $K \ln K = K - 5/2$ then the above values are given by $J^* = (K - 3/2)/4$ and $t^* = 1 - K/2$.)

Extending F to $[0,1]$ using (35) leaves the formula (59) unchanged. For $t > 1 - t^*$ it gives $F(t) = 2t - 1$. We have proved:

Theorem 8.2. *There is a unique distribution F that maximizes $\Phi[F]$ for four-segment measures. If J^* and t^* are chosen as above with approximate values given by (64) and (65), then $J^* = \int_0^{1/2} F(t)dt$ and F is given on $[0,1]$ by*

$$F(t) = \begin{cases} 0 & \textit{when } 0 \le t \le t^* \\ \sqrt{\left(t - \frac{1}{2}\right)^2 + \frac{3}{2} + 4J} + \left(t - \frac{3}{2}\right) & \textit{when } t^* \le t \le 1 - t^* \\ 2t - 1 & \textit{when } 1 - t^* \le t \le 1. \end{cases} \tag{66}$$

□

Figure 13 is a graph of F. The function is convex, has $F(1/2) \approx 0.30553$ and $F'(1/2) = 1$, and has no derivative at t^* or $1 - t^*$. We call the corresponding measure μ. Its packing density is calculated from (59):

$$\delta(2413, \mu) = \Phi[F] \approx 0.10472339512772223636. \tag{67}$$

This number is a new lower bound for the packing density $\delta(2413)$, and we have proven that it is the best packing rate possible using a symmetrical four-segment measure.

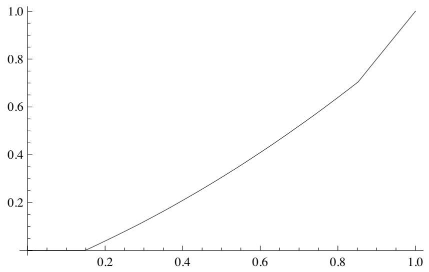


Fig. 13. Graph of F for μ or μ_1

9 The first recursion bubble

Having found the optimal four-segment measure, we now improve it using recursion.

The measure μ defined at the end of the last section is determined by a function F with $F(t) = 0$ for $t \in [0, t^*]$. This corresponds to the part of S with $x \in [1/4, 1/4 + t^*/2]$. In this region all of the probability is concentrated on the upper segment, in the rectangle $R = [1/4, 1/4 + t^*/2] \times [1 - t^*/4, 1]$. The probability itself is $\mu(R) = t^*/2$.

There is no probability above, below, or to the left or right of this rectangle, so it is easy to check that no occurrence of 2413 includes more than one point from this rectangle. Therefore nothing is lost by rearranging probability within the rectangle.

Define μ_1 recursively by $\mu_1 = \mu$ except on the rectangle R and its rotated images, in which μ_1 is a reduced-scale image of μ itself. (It makes no difference that the transformation between S and R is not aspect-preserving.) We call the four altered rectangles *recursion bubbles.* (We have seen recursion bubbles before, in recursive templates and in Figure 3.)

Now we have the same 2413 occurrences as before, plus additional occurrences when all four points fall within one of the recursion bubbles (probability $4(t^*/2)^4$) and happen to form a 2413 occurrence within the bubble. Therefore

$$\delta(2413, \mu_1) = \Phi[F] + 4\left(\frac{t^*}{2}\right)^4 \delta(2413, \mu_1) \tag{68}$$

or, solving,

$$\delta(2413, \mu_1) = \frac{\Phi[F]}{1 - 4\left(\frac{t^*}{2}\right)^4}. \tag{69}$$

Using the values of $\Phi[F]$ and t^* from the previous section we obtain

$$\delta(2413, \mu_1) \approx 0.10473588696991414716\ldots, \tag{70}$$

a new lower bound for $\delta(2413)$. The increase due to the recursion bubble appears in the fifth decimal place.

10 The second recursion bubble

Shouldn't the recursion bubble be bigger?

The measure μ was optimal in the absence of recursion. With recursion, there is a greater advantage to selecting points in the recursion box than there was before. At the margin, shouldn't that shift the optimum configuration in the direction of a larger bubble?

So, let's increase the size of the bubble. We can't do that in isolation, because it would cause F to be inconsistent with (59) immediately to the right of the bubble. We must allow F to increase with slope 2 until it catches up with the formula. This creates a small region in which all of the probability is on the lower segment, so we might as well turn it into a second recursion bubble.

The resulting measure is the same as an SFS with this distribution:

$$F(t) = \begin{cases} 0 & \text{when } t \leq t_1 \\ 2(t - t_1) & \text{when } t_1 \leq t \leq t_2 \\ \sqrt{\left(t - \frac{1}{2}\right)^2 + \frac{3}{2} + 4J} + \left(t - \frac{3}{2}\right) & \text{when } t_2 \leq t \leq 1 - t_2 \\ 1 - 2t_1 & \text{when } 1 - t_2 \leq t \leq 1 - t_1 \\ 2t - 1 & \text{when } 1 - t_1 \leq t. \end{cases}$$

We have extended F to $[0, 1]$ using the normalization identity. There are now two recursion bubbles on each segment, one corresponding to the interval $[0, t_1]$ (probability $(t_1/2)^2$ for each box) and one corresponding to the interval $[t_1, t_2]$ (probability $((t_2 - t_1)/2)^2$ for each box). Both t_1 and t_2 are parameters that we can choose, along with J, subject to the requirement that F be continuous at t_2 and have integral J on $[0, 1/2]$.

We optimize t_1 and t_2 by naked calculation:

$$\begin{aligned} t_1 &= 0.14861089461296151506\ldots \\ t_2 &= 0.14909030676438411460\ldots \end{aligned}$$

and if μ_2 is the measure with these revisions, we get

$$\delta(2413, \mu_2) = \frac{\Phi[F]}{1 - 4\left(\frac{t_1}{2}\right)^4 - 4\left(\frac{t_2 - t_1}{2}\right)^4} \approx 0.10473602526603545023\ldots.$$

The improvement over μ_1 is about 10^{-7}. This is the best lower bound we have found for $\delta(2413)$.

Conjecture 10.1. The measure μ_2 is optimal for 2413 and the packing density of 2413 is $\delta(2413) = 0.10473602526603545023\ldots$.

The measure μ_2 is illustrated in Figure 1.

How much is proof and how much is conjecture? We conjectured that the optimal measure would be related to a four-segment measure, and we proved that the optimal four-segment measure is given by Theorem 8.2. We conjectured that adding two recursion bubbles would make this optimal, and calculated the best location of the recursion bubbles by brute force. Hence, gaps remain before we can be sure that μ_2 is optimal.

11 More bubbles

An alternative possibility is that the recursion bubbles continue to multiply, alternating between the top and bottom segment and reaching a limit point before the center of the segment. We cannot calculate a positive contribution even for the third box, which may just mean that it is too small to be found by our methods. A measure μ_∞ with an infinite sequence of recursion blocks is illustrated in Figure 2.

References

[1] M. H. Albert. personal communication. 2007.

[2] M. H. Albert, M. D. Atkinson, C. C. Handley, D. A. Holton, and W. Stromquist. On packing densities of permutations. *Electron. J. Combin.*, 9(1):Research Paper 5, 20 pp., 2002.

[3] R. Brignall. A survey of simple permutations. In *this volume*, pages 41–65.

[4] C. B. Presutti. Determining lower bounds for packing densities of non-layered patterns using weighted templates. *Electron. J. Combin.*, 15(1):Research paper 50, 10, 2008.

[5] D. Warren. *Optimizing the packing behavior of layered permutation patterns.* PhD thesis, University of Florida, 2005.