A Variational Principle for Permutations

Richard Kenyon* Daniel Král'† Charles Radin‡
Peter Winkler§

Abstract

We define an entropy function for scaling limits of permutations, called permutons, and prove that under appropriate circumstances, both the shape and number of large permutations with given constraints are determined by maximizing entropy over permutons with those constraints. We also describe a useful equivalent version of permutons using a recursive construction.

This variational principle is used to study permutations with one or two fixed pattern densities. In particular, we compute (sometimes directly, sometimes numerically) the maximizing permutons with fixed density of 12 patterns or of fixed 123 density or both; with fixed 12 density and sum of 123 and 132 densities; and with finally with fixed 123 and 321 densities. In the last case we study a particular phase transition.

*Department of Mathematics, Brown University, Providence, RI 02912; rkenyon@math.brown.edu

†Mathematics Institute, DIMAP and Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK; d.kral@warwick.ac.uk

‡Department of Mathematics, University of Texas, Austin, TX 78712; radin@math.utexas.edu

§Department of Mathematics, Dartmouth College, Hanover, New Hampshire 03755; Peter.Winkler@Dartmouth.edu

1 Introduction

Suppose something is known about the permutation $\pi \in S_n$ for n large, e.g. the relative fraction of "321" patterns (double inversions) in π. How many permutations are there with that relative fraction? What do they look like? In this work we approach such problems by showing that the size of many classes of permutations can be estimated by maximizing a certain function over limit objects called permutons. Furthermore when—as is usually the case—the maximizing permuton is unique, properties of "most" permutations in the class can then be deduced from it.

To a permutation $\pi \in S_n$ one can associate a probability measure γ_π on $[0,1]^2$ as follows. Divide $[0,1]^2$ into an $n \times n$ grid of squares of size $1/n \times 1/n$. Define the density of γ_π on the square in the ith row and jth column to be the constant n if $\pi(i) = j$ and 0 otherwise. In other words, γ_π is a geometric representation of the permutation matrix of π.

Define a *permuton* to be a probability measure γ on $[0,1]^2$ with uniform marginals:

$$\gamma([a,b] \times [0,1]) = b - a = \gamma([0,1] \times [a,b]), \quad \text{for all } 0 \le a \le b \le 1. \tag{1}$$

Note that γ_π is a permuton for any permutation $\pi \in S_n$. Permutons were introduced in [12, 13] with a different but equivalent definition; the measure theoretic view of large permutations can be traced to [21] and was used in [10] as an analytic representation of permutation limits equivalent to that used in [12, 13]; the term "permuton" first appeared, we believe, in [10].

Let Γ be the space of permutons. There is a natural topology on Γ defined by the metric $d_\square$ given by $d_\square(\gamma_1, \gamma_2) = \max |\gamma_1(R) - \gamma_2(R)|$, where R ranges over aligned rectangles in $[0,1]^2$. This topology is the same as that given by the L^∞ metric on the cumulative distribution functions $G_i(x,y) = \gamma_i([0,x] \times [0,y])$. It is also the weak topology on probability measures.

We say that a sequence of permutations π_n with $\pi_n \in S_n$ *converges* as $n \to \infty$ if the associated permutons converge in the above sense. More generally, let ν_n be a probability measure on S_n. Associated to ν_n is the permuton that is the ν_n-expectation of the permuton of π as π ranges over S_n. We can then consider permuton convergence of a sequence of probability measures ν_n on S_n as $n \to \infty$.

A *pattern* τ is a permutation in S_k for some fixed k. Given a permuton γ, the *pattern density* of τ in γ, denoted $\rho_\tau(\gamma)$, is by definition the probability that, when k points are selected independently from γ and their

x-coordinates are ordered, the permutation induced by their y-coordinates is τ. For example, for γ with density $g(x,y)dx\,dy$, the density of pattern $12 \in S_2$ in γ is

$$\rho_{12}(\gamma) = 2\int_{x_1<x_2\in[0,1]}\int_{y_1<y_2\in[0,1]} g(x_1,y_1)g(x_2,y_2)dx_1dy_1dx_2dy_2. \tag{2}$$

It follows from results of [12, 13] that two permutons are equal if they have the same pattern densities (for all k).

1.1 Results

Our main result, Theorem 1 below, is a variational principle for permutons: it describes explicitly how many large permutations lie 'near' a given permuton. The statement is essentially that the number of permutations in S_n lying near a permuton γ is

$$n!e^{(H(\gamma)+o(1))n}, \tag{3}$$

where $H(\gamma)$ is the "permuton entropy" (defined below).

A corollary to Theorem 1 is Theorem 3, which describes both the number and (when uniqueness holds) limit shape of permutations in which a finite number of pattern densities have been fixed. The theorem states that the number of such permutations is determined by the entropy maximized over the set of permuton(s) having those fixed pattern densities.

Only in special cases have we been able to prove uniqueness of the entropy maximizers, although we suspect that the maximizers are unique for generic densities.

One can describe a permuton via a family of *insertion measures* $\{\mu_t\}_{t\in[0,1]}$, which is analgous to building a permutation by inductively inserting one element at a time into a growing list: for each $i \in [n]$ one inserts i into a random location in the permuted list of the first $i-1$ elements. This point of view is used to describe explicitly the entropy maximizing permutons with fixed densities of patterns of type $**\cdots*i$ (here each $*$ represents an element not exceeding the length of the pattern, so that $**2$ represents the patterns 132 and 312). We prove that for this family of patterns the maximizing permutons are analytic, and the entropy function as a function of the constraints is analytic and strictly concave.

The most basic example, the entropy-maximizing permuton for a fixed density ρ_{12} of 12 patterns, has the cumulative distribution function (CDF)

$$G(x,y) = \frac{1}{r}\log\left(1 + \frac{(e^{rx}-1)(e^{ry}-1)}{e^r - 1}\right), \tag{4}$$

where r is an explicit function of ρ_{12}. See Figure 1.

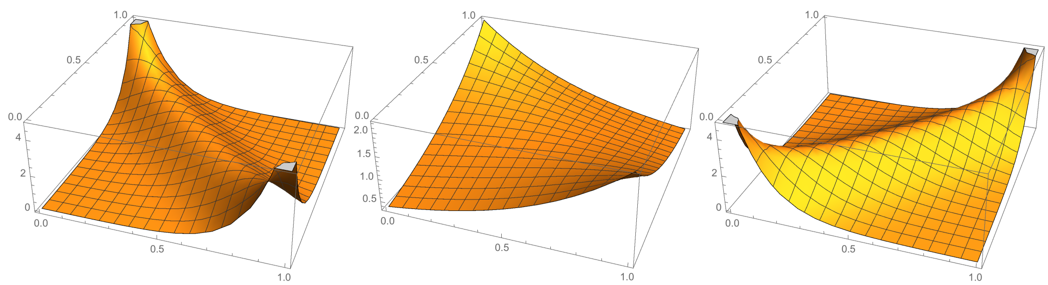

Figure 1: The permuton with fixed density ρ of pattern 12, shown for $\rho = .2, .4, .8$.

While maximizing permutons can be shown to satisfy certain explicit PDEs (see Section 9), they can also exhibit a very diverse set of behaviors. Even in one of the simplest cases, that of fixed density of the two patterns 12 and 123, the variety of shapes of permutons (and therefore of the approximating permutations) is remarkable: see Figure 7. In this case we prove that the feasible region of densities is the so-called "scalloped triangle" of Razborov [25, 26] which also describes the space of feasible densities for edges and triangles in the graphon model. (Here by *feasible region* we mean the closure of the set of achievable densities).

Another example which has been studied recently [8, 14, 15] is the case of the two patterns 123 and 321. In this case we describe a phase transition in the feasible region, where the maximizing permuton changes abruptly.

The variational principle can be applied directly to analyze classes of permutations which are defined by constraints that are continuous in the permuton topology. For constraints that are not continuous, for example the number of cycles of a fixed size, one can analyze an analogous "weak" characteristic, which is continuous, by applying the characteristic to patterns. For example, while the number of fixed points of a permuton is not well-defined, we can compute the expected number of fixed points for the permutation in S_n obtained by choosing n points independently from the permuton, and analyze the large-n limit of this quantity.

1.2 Analogies with graphons

For those who are familiar with variational principles for dense graphs [7, 6, 22, 23], we note the following differences between the graph case and the permutation case (see [19] for background on graph asymptotics):

1. Although permutons serve the same purpose for permutations that graphons serve for graphs, and (being defined on $[0,1]^2$) are superficially similar, they are measures (not symmetric functions) and represent permutations in a different way. (One *can* associate a graphon with a limit of permutations, via comparability graphs of two-dimensional posets, but these have trivial entropy in the Chatterjee-Varadhan sense [7] and we do not consider them here.)

2. The classes of constrained (dense) graphs considered in [7] have size about e^{cn^2}, n being the number of vertices and the (nonnegative) constant c being the target of study. Classes of permutations in S_n are of course of size at most $n! \sim e^{n(\log n-1)}$ but the constrained ones we consider here have size of order not $e^{cn\log n}$ for $c \in (0,1)$, as one might at first expect, but instead $e^{n\log n-n+cn}$ where $c \in [-\infty, 0]$ is the quantity of interest.

3. The "entropy" function, i.e., the function of the limit structure to be maximized, is bounded for graphons but unbounded for permutons. This complicates the analysis significantly for permutations.

4. The limit structures that do maximize the entropy function tend, in the graph case, to be combinatorial objects: step-graphons corresponding to what Radin, Ren and Sadun call "multipodal" graphs [24]. In contrast, maximizing permutons at interior points of feasible regions seem always to be smooth measures with analytic densities. Although they are more complicated than maximizing graphons, these limit objects are more suitable for classical variational analysis, e.g., differential equations of the Euler-Lagrange type.

2 Variational principle

For convenience, we denote the unit square $[0,1]^2$ by Q.

Let γ be a permuton with density g defined almost everywhere. We compute the *permutation entropy* $H(\gamma)$ of γ as follows:

$$H(\gamma) = \int_Q -g(x,y)\log g(x,y)\,dx\,dy \tag{5}$$

where "$0\log 0$" is taken as zero. Then H is finite whenever g is bounded (and sometimes when it is not). In particular for any $\sigma \in S_n$, we have $H(\gamma_\sigma) = -n(n\log n/n^2) = -\log n$ and therefore $H(\gamma_\sigma) \to -\infty$ for any sequence of increasingly large permutations even though $H(\lim \gamma_\sigma)$ may be finite. Note that H is zero on the uniform permuton (where $g(x,y) \equiv 1$) and negative (sometimes $-\infty$) on all other permutons, since the function $z\log z$ is concave downward. If γ has no density, we define $H(\gamma) = -\infty$.

Theorem 1 (The Large Deviations Principle)**.** *Let Λ be a set of permutons, Λ_n the set of permutations $\pi \in S_n$ with $\gamma_\pi \in \Lambda$. Then:*

1. *If Λ is closed,*
$$\lim_{n\to\infty} \frac{1}{n}\log\frac{|\Lambda_n|}{n!} \le \sup_{\gamma\in\Lambda} H(\gamma); \tag{6}$$
2. *If Λ is open,*
$$\lim_{n\to\infty} \frac{1}{n}\log\frac{|\Lambda_n|}{n!} \ge \sup_{\gamma\in\Lambda} H(\gamma). \tag{7}$$

Since we will be approximating H by Riemann sums, it is useful to define, for any permuton γ and any positive integer m, an approximating "step-permuton" γ^m as follows. Fixing m, denote by Q_{ij} the half-open square $((i-1)/m, i/m] \times ((j-1)/m, j/m]$; for each $1 \le i,j \le m$, we want γ^m to be uniform on Q_{ij} with $\gamma^m(Q_{ij}) = \gamma(Q_{ij})$. In terms of the density g^m of γ^m, we have $g^m(x,y) = m^2\gamma(Q_{ij})$ for all $(x,y) \in Q_{ij}$.

In order to prove Theorem 1 we will need the following result.

Theorem 2. *For any permuton γ, $\lim_{m\to\infty} H(\gamma^m) = H(\gamma)$, with $H(\gamma^m)$ diverging downward when $H(\gamma) = -\infty$.*

And finally we make a connection with our applications to large constrained permutations. Let us fix some finite set $\mathcal{P} = \{\pi_1, \ldots, \pi_k\}$ of patterns, that is, permutations of various sizes. Let $\alpha = (\alpha_1, \ldots, \alpha_k)$ be a vector of desired pattern densities. We then define two sets of permutons:

$$\Lambda^{\alpha,\epsilon} = \{\gamma \in \Gamma \,|\, |\rho_{\pi_j}(\gamma) - \alpha_j| < \epsilon \text{ for each } 1 \leq j \leq k\} \tag{8}$$

and

$$\Lambda^{\alpha} = \{\gamma \in \Gamma \,|\, \rho_{\pi_j}(\gamma) = \alpha_j \text{ for each } 1 \leq j \leq k\}. \tag{9}$$

With that notation, and the understanding that $\Lambda_n^{\alpha,\epsilon} = \Lambda^{\alpha,\epsilon} \cap S_n$, we have:

Theorem 3.

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log \frac{|\Lambda_n^{\alpha,\epsilon}|}{n!} = \max_{\gamma \in \Lambda^\alpha} H(\gamma).$$

The value $\max_{\gamma \in \Lambda^\alpha} H(\gamma)$ (which is guaranteed by the theorem to exist, but may be $-\infty$) will be called the *constrained entropy* and denoted by $s(\alpha)$. In the next two sections, we will prove Theorems 2, 1 and 3 (in that order); in subsequent sections, we will study the properties of maximizing permutons and use them to compute properties of some classes of permutations defined by pattern densities.

3 Proof of Theorem 2

In what follows we will, in order to increase readability, we write

$$\int_0^1 \int_0^1 -g(x,y) \log g(x,y) dx dy \tag{10}$$

as just $\int_Q -g \log g$. Also for the sake of readability, we will for this section only state results in terms of $g \log g$ rather than $-g \log g$; this avoids clutter caused by a multitude of absolute values and negations. Eventually, however, we will need to deal with an entropy function $H(\gamma) = \int_Q -g \log g$ that takes values in $[-\infty, 0]$.

Define

$$g_{ij} = m^2 \gamma(Q_{ij}). \tag{11}$$

We wish to show that the Riemann sum

$$\frac{1}{m^2} \sum_{0 \le i,j \le m} g_{ij} \log g_{ij}, \tag{12}$$

which we denote by $R_m(\gamma)$, approaches $\int_Q g \log g$ when γ is absolutely continuous with respect to Lebesgue measure, i.e., when the density g exists a.e., and otherwise diverges to ∞. There are thus three cases:

1. g exists and $\int_Q g \log g < \infty$;
2. g exists but $g \log g$ is not integrable, i.e., its integral is ∞;
3. γ is singular.

Let $A(t) = \{(x, y) \in Q : \ g(x, y) \log g(x, y) > t\}$.

In the first case, we have that $\limsup \int_{A(t)} g \log g = 0$, and since $g \log g \ge t$ on $A(t)$, we have $\limsup |A(t)|t = 0$ where $|A|$ denotes the Lebesgue measure of $A \subset Q$. (We need not concern ourselves with large negative values, since the function $x \log x$ is bounded below by $-1/e$.)

In the second case, we have the opposite, i.e., for some $\epsilon > 0$ and any s there is a $t > s$ with $t|A(t)| > \epsilon$.

In the third case, we have a set $A \subset Q$ with $\gamma(A) > 0$ but $|A| = 0$.

In the proof that follows we don't actually care that γ has uniform marginals, or that it is normalized to have $\gamma(Q) = 1$. Thus we restate Theorem 2 in greater generality:

Theorem 4. *Let γ be a finite measure on $Q = [0, 1]^2$ and $R_m = R_m(\gamma)$. Then:*

1. *If γ is absolutely continuous with density g, and $g \log g$ is integrable, then $\lim_{m\to\infty} R_m = \int_Q g \log g$.*
2. *If γ is absolutely continuous with density g, and $g \log g$ is not integrable, then $\lim_{m\to\infty} R_m = \infty$.*
3. *If γ is singular, then $\lim_{m\to\infty} R_m = \infty$.*

Proof. We begin with the first case, where we need to show that for any $\epsilon > 0$, there is an m_0 such that for $m \ge m_0$,

$$\int_Q g \log g - \frac{1}{m^2} \sum_{i,j=0}^{m} g_{ij} \log g_{ij} < \epsilon \ . \tag{13}$$

Note that since $x \log x$ is convex, the quantity on the left cannot be negative.

Lemma 5. *Let $\epsilon > 0$ be fixed and small. Then there are $\delta > 0$ and s with the following properties:*

1. $|\int_{A(s)} g \log g| < \delta^2/4$;
2. $|A(s)| < \delta^2/4$;
3. *for any* $u, v \in [0, s+1]$*, if* $|u - v| < \delta$ *then* $|u \log u - v \log v| < \epsilon/4$;
4. *for any* $B \subset Q$*, if* $|B| < 2\delta$ *then* $\int_B |g \log g| < \epsilon/4$.

Proof. By Lebesgue integrability of $g \log g$, we can immediately choose δ_0 such that any $\delta < \delta_0$ will satisfy the fourth property.

We now choose s_1 so that $\int_{A(s_1)} g \log g < \epsilon^2/4$, and $t|A(t)| < 1$ for all $t \geq s_1$. Since $[0, s_1]$ is compact we may choose $\delta_1 < \delta_0$ such that for any $u, v \in [0, s_1 + 1]$, $|u - v| < \delta_1$ implies $|u \log u - v \log v| < \epsilon/4$. We are done if $|A(s_1)| < \delta_1$ but since δ_1 depends on s_1, it is not immediately clear that we can achieve this. However, we know that since $\frac{d}{du} u \log u = 1 + \log u$, the dependence of δ_1 on s_1 is only logarithmic, while $|A(s_1)|$ descends at least as fast as $1/s_1$.

So we take $k = \lceil \log(\delta/2)/\log(\epsilon/2) \rceil$ and let $\delta = \delta_1/k$, $s = s_1^k$. Then $u, v \in [0, s+1]$ and $|u - v| < \delta$ implies $|u \log u - v \log v| < \epsilon/4$, and

$$\int_{A(s)} g \log g \leq \left(\int_{A(s_1)} g \log g \right)^k < (\epsilon/2)^{2 \log(\delta/2)/\log(\epsilon/2)} = \delta^2/4 \tag{14}$$

as desired. Since $u \log u > u > 1$ for $u > e$, we get $|A(s)| < \delta^2/4$ as a by-product. □

Henceforth s and δ will be fixed, satisfying the conditions of Lemma 5. Since g is measurable we can find a subset $C \subset Q$ with $|C| = |A(s)| < \delta^2/4$ such that g, and thus also $g \log g$, is continuous on $Q \setminus C$. Since $\int_B g \log g$ is maximized by $B = A(s)$ for sets B with $B = |A(s)|$, $|\int_C g \log g| < \int_{A(s)} g \log g$, so $|\int_{A(s) \cup C} g \log g| < \delta^2/2$. We can then find an open set A containing $A(s) \cup C$ with $|A|$ and $\int_A g \log g$ both bounded by δ^2.

We now invoke the Tietze Extension Theorem to choose a continuous $f : Q \to \mathbb{R}$ with $f(x, y) = g(x, y)$ on $Q \setminus A$, and $f \log f < s$ on all of Q.

Since f is continuous and bounded, f and $f \log f$ are Riemann integrable. Let f_{ij} be the mean value of f over Q_{ij}, i.e.,

$$f_{ij} = m^2 \int_{Q_{ij}} f \ . \tag{15}$$

Since, on any Q_{ij}, $\inf f \log f \le f_{ij} \log f_{ij} < \sup f \log f$, we can choose m_0 such that $m \ge m_0$ implies

$$\left| \int_Q f \log f - \frac{1}{m^2} \sum_{ij} f_{ij} \log f_{ij} \right| < \epsilon/4 \ . \tag{16}$$

We already have

$$\begin{aligned} |\int_Q g \log g \ &- \ \int_Q f \log f| = |\int_A g \log g - \int_A f \log f| \\ &\le \ |\int_A g \log g| - |\int_A f \log f| < 2\delta^2 << \epsilon/4. \end{aligned} \tag{17}$$

Thus, to get (13) from (16) and (17), it suffices to bound

$$\left| \frac{1}{m^2} \sum_{ij} g_{ij} \log g_{ij} - \frac{1}{m^2} \sum_{ij} f_{ij} \log f_{ij} \right| \tag{18}$$

by $\epsilon/2$.

Fixing $m > m_0$, call the pair (i, j), and its corresponding square Q_{ij}, "good" if $|Q_{ij} \cap A| < \delta/(2m^2)$. The number of bad (i.e., non-good) squares cannot exceed $2\delta m^2$, else $|A| > 2\delta m^2 \delta/(2m^2) = \delta^2$.

For the good squares, we have

$$|g_{ij} - f_{ij}| = m^2 \left| \int_{Q_{ij} \cap A} (g - f) \right| \le m^2 \left| \int_{Q_{ij} \cap A} 2g \right| \le 2(\delta/2) = \delta \tag{19}$$

with $f_{ij} \le s$, thus f_{ij} and g_{ij} both in $[0, s+1]$. It follows that

$$|g_{ij} \log g_{ij} - f_{ij} \log f_{ij}| < \epsilon/4 \tag{20}$$

and therefore the "good" part of the Riemann sum discrepancy, namely

$$\frac{1}{m^2} \left| \sum_{\text{good } ij} (g_{ij} \log g_{ij} - f_{ij} \log f_{ij}) \right| \ , \tag{21}$$

is itself bounded by $\epsilon/4$.

Let Q' be the union of the bad squares, so $|Q'| < m^2 2\delta/(2m^2) = 2\delta$; then by (16) and convexity of $u \log u$,

$$\frac{1}{m^2}\left|\sum_{\text{bad } ij} g_{ij}\log g_{ij} - f_{ij}\log f_{ij}\right| < 2\left|\int_{Q'} g\log g\right| < 2(\epsilon/8) = \epsilon/4 \qquad (22)$$

and we are done with the case where $g\log g$ is integrable.

Suppose g exists but $g\log g$ is not integrable; we wish to show that for any M, there is an m_1 such that $m \geq m_1$ implies $\frac{1}{m^2}\sum g_{ij}\log g_{ij} > M$.

For $t \geq 1$, define the function g^t by $g^t(x,y) = g(x,y)$ when $g(x,y) \leq t$, i.e. when $(x,y) \notin A(t)$, otherwise $g^t(x,y) = 0$. Then $\int_Q g^t \log g^t \to \infty$ as $t \to \infty$, so we may take t so that $\int_Q g^t \log g^t \geq M+1$. Let γ^t be the (finite) measure on Q for which g^t is the density. Since g^t is bounded (by t), $g^t \log g^t$ is integrable and we may apply the first part of Theorem 4 to get an m_1 so that $m \geq m_1$ implies that $R^m(\gamma^t) > M$.

Since $t \geq 1$, $g\log g \geq g^t \log g^t$ everywhere and hence, for every m, $R_m(\gamma^t) \leq R_m(\gamma)$. It follows that $R_m(\gamma) > M$ for $m \geq m_1$ and this case is done.

Finally, suppose γ is singular and let A be a set of Lebesgue measure zero for which $\gamma(A) = a > 0$.

Lemma 6. *For any $\epsilon > 0$ there is an m_2 such that $m > m_2$ implies that there are ϵm^2 squares of the $m \times m$ grid that cover at least half the γ-measure of A.*

Proof. Note first that if B is an open disk in Q of radius at most δ, then for $m > 1/(2\delta)$, then we can cover B with cells of an $m \times m$ grid of total area at most $64\delta^2$. The reason is that such a disk cannot contain more than $\lceil 2\delta/(1/m)\rceil^2 < (4\delta m)^2$ grid vertices, each of which can be a corner of at most four cells that intersect the disk. Thus, rather conservatively, the total area of the cells that intersect the disk is bounded by $(4/m^2)\cdot(4\delta m)^2 = 64\delta^2$. It follows that as long as a disk has radius at least $1/(2m)$, it costs at most a factor of $64/\pi$ to cover it with grid cells.

Now cover A with open disks of area summing to at most $\pi\epsilon/64$. Let b_n be the sum of the gamma-measures of the disks of radius in the half-open interval $[1/(2n), 1/(2n-2))$, so that $\sum_n^\infty b_n \geq a$, and take m_2 so that the partial sum $\sum_n^{m_2} b_n$ exceeds $a/2$, to get the desired result. □

Let M be given and use Lemma 6 to find m_2 such that for any $m \geq m_2$, there is a set $I \subset \{1, \ldots, m\}^2$ of size at most δm^2 such that $\gamma(\bigcup_I Q_{ij}) > a/2$, where $Q_{ij} = ((i-1)/m, i/m] \times ((j-1)/m, j/m]$ as before and δ is a small positive quantity depending on M and a, to be specified later. Then

$$\begin{aligned} R_m(\gamma) &= \sum_{ij} \frac{1}{m^2} g_{ij} \log g_{ij} \\ &\geq -1/e + \frac{1}{m^2} \delta m^2 \bar{g} \log \bar{g} = -1/e + \delta \bar{g} \log \bar{g} \end{aligned} \tag{23}$$

where $\bar{g}$ is the mean value of g_{ij} over $(i, j) \in I$, the last inequality following from the convexity of $u \log u$. The $-1/e$ term is needed to account for possible negative values of $g \log g$.

But $\sum_I g_{ij} = m^2 \gamma(\bigcup_I Q_{ij}) > m^2 a/2$, so $\bar{g} > (m^2 a/2)/(\delta m^2) = a/(2\delta)$. Consequently

$$R_m(\gamma) > -\frac{1}{e} + \delta \frac{a}{2\delta} \log \frac{a}{2\delta} = -\frac{1}{e} + \frac{a}{2} \log \frac{a}{2\delta} \ . \tag{24}$$

Taking

$$\delta = \frac{a}{2} \exp\left(-2 \frac{\left(M + \frac{1}{e}\right)}{a} \right) \tag{25}$$

gives $R_m(\gamma) > M$ as required, and the proof of Theorem 2 is complete. □

4 Proof of Theorem 1

We begin with a simple lemma.

Lemma 7. *The function $H : \Gamma \to \mathbb{R}$ is upper semicontinuous.*

Proof. Let $\gamma_1, \gamma_2, \ldots$ be a sequence of permutons approaching the permuton γ (in the $d_\square$-topology); we need to show that $H(\gamma) \geq \limsup H(\gamma_n)$.

If $H(\gamma)$ is finite, fix $\epsilon > 0$ and take m large enough so that $|H(\gamma^m) - H(\gamma)| < \epsilon$; then since $H(\gamma_n^m) \geq H(\gamma_n)$ by concavity,

$$\limsup_n H(\gamma_n) \leq \limsup_n H(\gamma_n^m) = H(\gamma^m) < \epsilon + H(\gamma) \tag{26}$$

and since this holds for any $\epsilon > 0$, the claimed inequality follows.

If $H(\gamma) = -\infty$, fix $t < 0$ and take m so large that $H(\gamma^m) < t$. Then

$$\limsup_n H(\gamma_n) \leq \limsup_n H(\gamma_n^m) = H(\gamma^m) < t \tag{27}$$

for all t, so $\limsup_n H(\gamma_n^m) \to -\infty$ as desired. $\square$

Let $B(\gamma, \epsilon) = \{\gamma' | d_\square(\gamma, \gamma') \leq \epsilon\}$ be the (closed) ball in Γ of radius $\epsilon > 0$ centered at the permuton γ, and let $B_n(\gamma, \epsilon)$ be the set of permutations $\pi \in S_n$ with $\gamma_\pi \in B(\gamma, \epsilon)$.

Lemma 8. *For any permuton γ, $\lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \frac{1}{n} \log(|B_n(\gamma, \epsilon)|/n!)$ exists and equals $H(\gamma)$.*

Proof. Suppose $H(\gamma)$ is finite. It suffices to produce two sets of permutations, $U \subset B_n(\gamma, \epsilon)$ and $V \supset B_n(\gamma, \epsilon)$, each of size

$$\exp\left(n \log n - n + n(H(\gamma) + o(\epsilon^0)) + o(n)\right) \tag{28}$$

where by $o(\epsilon^0)$ we mean a function of ϵ (depending on γ) which approaches 0 as $\epsilon \to 0$. (The usual notation here would be $o(1)$; we use $o(\epsilon^0)$ here and later to make it clear that the relevant variable is ϵ and not, e.g., n.)

To define U, fix $m > 5/\epsilon$ so that $|H(\gamma^m) - H(\gamma)| < \epsilon$ and let n be a multiple of m with $n > m^3/\epsilon$. Choose integers $n_{i,j}$, $1 \leq i, j \leq m$, so that:

1. $\sum_{i=1}^n n_{i,j} = n/m$ for each j;
2. $\sum_{j=1}^n n_{i,j} = n/m$ for each i; and
3. $|n_{i,j} - n\gamma(Q_{ij})| < 1$ for every i, j.

The existence of such a rounding of the matrix $\{n\gamma(Q_{ij})\}_{i,j}$ is guaranteed by Baranyai's rounding lemma [1].

Let U be the set of permutations $\pi \in S_n$ with exactly $n_{i,j}$ points in the square Q_{ij}, that is, $|\{i : (i/n, \pi(i)/n) \in Q_{ij}\}| = n_{i,j}$, for every $1 \leq i, j \leq m$. We show first that U is indeed contained in $B_n(\gamma, \epsilon)$. Let $R = [a, b] \times [c, d]$ be a rectangle in $[0, 1]^2$. R will contain all Q_{ij} for $i_0 < i < i_1$ and $j_0 < j < j_1$ for suitable i_0, i_1, j_0 and j_1, and by construction the γ_π-measure of the union of those rectangles will differ from its γ-measure by less than $m^2/n < \epsilon/m$. The squares cut by R are contained in the union of two rows and two columns of width $1/m$, and hence, by the construction of π and the uniformity of the marginals of γ, cannot contribute more than $4/m < 4\epsilon/5$ to the difference in measures. Thus, finally, $d_\square(\gamma_\pi, \gamma) < \epsilon/m + 4\epsilon/5 < \epsilon$.

Now we must show that $|U|$ is close to the claimed size

$$\exp\big(n \log n - n - H(\gamma)n\big) \tag{29}$$

We construct $\pi \in U$ in two phases of m steps each. In step i of Phase I, we decide for each k, $(i-1)n/m < k \leq in/m$, which of the m y-intervals $\pi(k)$ should lie in. There are

$$\binom{n/m}{n_{i,1}, n_{i,2}, \ldots, n_{i,m}} = \exp\big((n/m)h_i + o(n/m)\big) \tag{30}$$

ways to do this, where $h_i = -\sum_{j=1}^m (n_{i,j}/(n/m)) \log(n_{i,j}/(n/m))$ is the entropy of the probability distribution $n_{i,\cdot}/(n/m)$.

Thus, the number of ways to accomplish Phase I is

$$\begin{aligned}
\exp\big(o(n) \;+\; (n/m)\sum_i h_i\big) &= \exp\big(o(n) - \sum_{i,j} n_{i,j} \log(n_{i,j}/(n/m))\big) \\
&= \exp\left(o(n) - \sum_{i,j} n_{i,j}(\log(n_{i,j}/n) + \log m)\right) \\
&= \exp\left(o(n) - n\log m - \sum_{i,j} n_{i,j} \log \gamma(Q_{ij})\right) \\
&= \exp\left(o(n) - n\log m - n\sum_{i,j} \gamma(Q_{ij}) \log \gamma(Q_{ij})\right) \;.
\end{aligned} \tag{31}$$

Recalling that the value taken by the density g^m of γ^m on the points of Q_{ij} is $m^2\gamma(Q_{ij})$, we have that

$$\begin{aligned}
H(\gamma^m) &= \sum_{i,j} \frac{1}{m^2}\big(-m^2\gamma(Q_{ij}) \log(m^2\gamma(Q_{ij}))\big) \\
&= -\sum_{i,j} \gamma(Q_{ij})(\log \gamma(Q_{ij}) + 2\log m) \\
&= -\sum_{i,j} \gamma(Q_{ij})(\log \gamma(Q_{ij}) + 2\log m) \\
&= -2\log m - \sum_{i,j} \gamma(Q_{ij}) \log \gamma(Q_{ij}) \;.
\end{aligned} \tag{32}$$

Therefore we can rewrite the number of ways to do Phase I as

$$\exp\big(n\log m + nH(\gamma^m) + o(n)\big). \tag{33}$$

In Phase II we choose a permutation $\pi_j \in S_{n/m}$ for each j, $1 \le j \le m$, and order the y-coordinates of the n/m points (taken left to right) in row j according to π_j. Together with Phase I this determines π uniquely, and the number of ways to accomplish Phase II is

$$\begin{aligned}(n/m)!^m &= \Big(\exp\Big(\frac{n}{m}\log\frac{n}{m} - \frac{n}{m} + o(n/m)\Big)\Big)^m \\ &= \exp\big(n\log n - n - n\log m + o(n)\big)\end{aligned} \tag{34}$$

so that in total,

$$\begin{aligned}|U| &\ge \exp\big(n\log m + nH(\gamma^m) + o(n)\big)\exp\big(n\log n - n - n\log m + o(n)\big) \\ &= \exp\big(n\log n - n + nH(\gamma^m) + o(n)\big)\end{aligned} \tag{35}$$

which, since $|H(\gamma) - H(\gamma^m)| < \epsilon$, does the job.

We now proceed to the other bound, which involves similar calculations in a somewhat different context. To define the required set $V \supset B_n(\gamma,\epsilon)$ of permutations we must allow a wide range for the number of points of π that fall in each square Q_{ij}— wide enough so that a violation causes Q_{ij} itself to witness $d_\square(\gamma_\pi,\gamma) > \epsilon$, thus guaranteeing that if $\pi \notin V$ then $\pi \notin B_n(\gamma,\epsilon)$.

To do this we take m large, $\epsilon < 1/m^4$, and $n > 1/\epsilon^2$. We define V to be the set of permutations $\pi \in S_n$ for which the number of points $(k/n, \pi(k)/n)$ falling in Q_{ij} lies in the range $[n(\gamma(Q_{ij}) - \sqrt{\epsilon}), n(\gamma(Q_{ij}) + \sqrt{\epsilon})]$. Then, as promised, if $\pi \notin V$ we have a rectangle $R = Q_{ij}$ with $|\gamma(R) - \gamma_\pi(R)| > \sqrt{\epsilon}/m^2 > \epsilon$.

It remains only to bound $|V|$. Here a preliminary phase is needed in which the exact count of points in each square Q_{ij} is determined; since the range for each $n_{i,j}$ is of size $2n\sqrt{\epsilon}$, there are at most $(2n\sqrt{\epsilon})^{m^2} = \exp\big(m^2\log(2n\sqrt{\epsilon})\big)$ ways to do this, a negligible factor since $m^2\log(n\sqrt{\epsilon}) = o(n)$. For Phase I we must assume the $n_{i,j}$ are chosen to maximize each h_i but since the entropy function h is continuous, the penalty shrinks with ϵ. Counting as before, we deduce that here the number of ways to accomplish Phase I is bounded by

$$\exp\big(n\log m + n(H(\gamma^m) + o(\epsilon^0)) + o(n)\big) = \exp\big(n\log m + n(H(\gamma) + o(\epsilon^0)) + o(n)\big). \tag{36}$$

The computation for Phase II is exactly as before and the conclusion is that

$$\begin{aligned} |V| &\leq \exp\left(n \log m - n + n(H(\gamma) + o(\epsilon^0)) + o(n)\right) \\ &\quad \cdot \exp\left(n \log n - n - n \log m + o(n)\right) \\ &= \exp\left(n \log n - n + nH(\gamma) + o(n)\right) \end{aligned} \tag{37}$$

proving the lemma in the case where $H(\gamma) > -\infty$.

If $H(\gamma) > -\infty$, we need only the upper bound provided by the set V. Fix $t < 0$ with the idea of showing that $\frac{1}{n} \log \frac{|B_n(\gamma,\epsilon_\gamma)|}{n!} < t$. Define V as above, first insuring that m is large enough so that $H(\gamma^m) < t-1$. Then the number of ways to accomplish Phase I is bounded by

$$\exp\left(n \log m + n(H(\gamma^m) + o(\epsilon^0)) + o(n)\right) < \exp\left(n \log m + n(t-1+o(\epsilon^0)) + o(n)\right) \tag{38}$$

and consequently $|V|$ is bounded above by

$$\exp\left(n \log n - n + n(t-1) + o(n)\right) < \exp\left(n \log n - n + nt\right) . \tag{39}$$

□

We are finally in a position to prove Theorem 1. If our set Λ of permutons is closed, then, since Γ is compact, so is Λ. Let $\delta > 0$ with the idea of showing that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \leq H(\mu) + \delta \tag{40}$$

for some $\mu \in \Lambda$. If not, for each $\gamma \in \Lambda$ we may, on account of Lemma 8, choose ϵ_γ and n_γ so that $\frac{1}{n} \log \frac{|B_n(\gamma,\epsilon_\gamma)|}{n!} < H(\gamma) + \delta/2$ for all $n \geq n_\gamma$. Since a finite number of these balls cover Λ, we have too few permutations in Λ_n for large enough n, and a contradiction has been reached.

If Λ is open, we again let $\delta > 0$, this time with the idea of showing that

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{|\Lambda_n|}{n!} \geq H(\mu) - \delta . \tag{41}$$

To do this we find a permuton $\mu \in \Lambda$ with

$$H(\mu) > \sup_{\gamma\in\Lambda} H(\gamma) - \delta/2 , \tag{42}$$

and choose $\epsilon > 0$ and n_0 so that $B_n(\mu,\epsilon) \subset \Lambda$ and $\frac{1}{n} \log \left(\frac{|B_n(\mu,\epsilon)|}{n!}\right) > H(\mu) - \delta/2$ for $n \geq n_0$.

This concludes the proof of Theorem 1.

5 Proof of Theorem 3

In Theorem 3, the set $\Lambda^{\alpha,\epsilon}$ of permutons under consideration consists of those for which certain pattern densities are close to values in the vector α. Note first that since the density function $\rho(\pi,\cdot)$ is continuous in the topology of Γ, λ^α is closed and by compactness $H(\gamma)$ takes a maximum value on Λ^α.

Again by continuity of $\rho(\pi,\cdot)$, $\Lambda^{\alpha,\epsilon}$ is an open set and we have from the second statement of Theorem 1 that for any ϵ,

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{|\Lambda_n^{\alpha,\epsilon}|}{n!} \geq \max_{\gamma\in\Lambda^{\alpha,\epsilon}} H(\gamma) \geq \max_{\gamma\in\Lambda^\alpha} H(\gamma) \tag{43}$$

from which we deduce that

$$\lim_{\epsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\log\frac{|\Lambda_n^{\alpha,\epsilon}|}{n!} \geq \max_{\gamma\in\Lambda^\alpha} H(\gamma). \tag{44}$$

To get the reverse inequality, fix a $\gamma\in\Lambda^\alpha$ maximizing $H(\gamma)$. Let $\delta>0$; since H is upper semi-continuous and Λ^α is closed, we can find an $\epsilon'>0$ such that no permuton γ' within distance ϵ' of Λ^α has $H(\gamma')>H(\gamma)+\delta$. But again since $\rho(\pi,\cdot)$ is continuous, for small enough ϵ, every $\gamma'\in\Lambda^{\alpha,\epsilon}$ is indeed within distance ϵ' of Λ^α. Let Λ' be the (closed) set of permutons γ' satisfying $\rho(\pi_j,\gamma')\leq\epsilon$; then, using the first statement of Theorem 1, we have thus

$$\lim_{n\to\infty}\frac{1}{n}\log\frac{|\Lambda'_n|}{n!} \leq H(\gamma)+\delta \tag{45}$$

and since such a statement holds for arbitrary $\delta>0$, the result follows.

Theorem 3 puts us in a position to try to describe and enumerate permutations with some given pattern densities. It does not, of course, guarantee that there is just one $\gamma\in\Lambda^\alpha$ that maximizes $H(\gamma)$, nor that there is one with finite entropy. As we shall see it seems to be the case that interior points in feasible regions for pattern densities do have permutons with finite entropy, and *usually* just one "winner." Points on the boundary of a feasible region (e.g., pattern-avoiding permutations) often have only singular permutons, and since the latter always have entropy $-\infty$, Theorem 3 will not be of direct use there.

6 Insertion densities

A permuton γ can be described in terms of its *insertion measures*. This is a family of probability measures $\{\nu_x\}_{x\in[0,1]}$, with measure ν_x supported on $[0,x]$. This is a continuum version of the process of building a permutation on $[n]$ by, for each i, inserting i at a random location in the permutation formed from $\{1,\dots,i-1\}$. Any permutation measure can be built this way. We describe here how a permuton measure can be built from a family of insertion measures, and conversely, how every permuton arises from a family of independent insertions.

We first describe how to reconstruct the insertion measures from the permuton γ. Let $Y_x \in [0,1]$ be the random variable with law $\gamma|_{\{x\}\times[0,1]}$. Let $Z_x \in [0,x]$ be the random variable (with law ν_x) giving the location of the insertion of x (at time x), and let $F(x,\cdot)$ be its CDF. Then

$$F(x,y) = \Pr(Z_x < y) = \Pr(Y_x < \tilde{y}) = G_x(x,\tilde{y}) \tag{46}$$

where $\tilde{y}$ is defined by $G(x,\tilde{y}) = y$.

More succinctly, we have

$$F(x, G(x,\tilde{y})) = G_x(x,\tilde{y}). \tag{47}$$

Conversely, given the insertion measures, equation (47) is a differential equation for G. Concretely, after we insert x_0 at location $X(x_0) = Z_{x_0}$, the image flows under future insertions according to the (deterministic) evolution

$$\frac{d}{dx}X(x) = F_x(X(x)), \quad X(x_0) = Z_{x_0}. \tag{48}$$

If we let $\Psi_{[x,1]}$ denote the flow up until time 1, then the permuton is the push-forward under Ψ of ν_x:

$$\gamma_t = (\Psi_{[x,1]})_*(\nu_x). \tag{49}$$

See an example in the next section.

Another way to see this correspondence is as follows. Project the graph of G in $\mathbb{R}^3$ onto the xz-plane; the image of the curves $G([0,1]\times\{\tilde{y}\})$ are the flow lines of the vector field (48). The divergence of the flow lines at (x,y) is $f(x,y)$, the density associated with $F(x,y)$.

The permuton entropy can be computed from the entropy of the insertion measures as follows.

Lemma 9.
$$H(\gamma) = \int_0^1 \int_0^x -f(x,y)\log(xf(x,y))dy\,dx. \tag{50}$$

Proof. Differentiating (47) with respect to $\tilde{y}$ gives

$$f(x, G(x,\tilde{y}))G_y(x,\tilde{y}) = g(x,\tilde{y}). \tag{51}$$

Thus the RHS of (50) becomes

$$\int_0^1 \int_0^x -\frac{g(x,\tilde{y})}{G_y(x,\tilde{y})} \log \frac{xg(x,\tilde{y})}{G_y(x,\tilde{y})}\, dy\, dx. \tag{52}$$

Substituting $y = G(x,\tilde{y})$ with $dy = G_y(x,\tilde{y})d\tilde{y}$ we have

$$\begin{aligned}\int_0^1 \int_0^1 -g(x,\tilde{y}) \log \frac{xg(x,\tilde{y})}{G_y(x,\tilde{y})}\, d\tilde{y}\, dx = H(\gamma) \quad - \quad \int_0^1 \int_0^1 g(x,\tilde{y}) \log x \, d\tilde{y}\, dx \\ + \int_0^1 \int_0^1 g(x,\tilde{y}) \log G_y(x,\tilde{y})\, d\tilde{y}\, dx. \end{aligned} \tag{53}$$

Integrating over $\tilde{y}$ the first integral on the RHS is

$$\int_0^1 -\log x \, dx = 1, \tag{54}$$

while the second one is

$$\int_0^1 \int_0^1 \frac{\partial}{\partial x}(G_y \log G_y - G_y) = -1, \tag{55}$$

since $G(1,y) = y$ and $G(0,y) = 0$. So those two integrals cancel. $\square$

7 Inversions

The number of occurrences $k(\pi)$ of the pattern 12 in a permutation of S_n has a simple generating function:

$$\sum_{\pi \in S_n} x^{k(\pi)} = \prod_{j=1}^{n} (1 + x + \cdots + x^j) = \sum_{i=0}^{\binom{n}{2}} C_i x^i. \tag{56}$$

One can see this by building up a permutation by insertions: when i is inserted into the list of $\{1,\dots,i-1\}$, the number of 12 patterns created is exactly one less than the position of i in that list.

Theorem 3 suggests that to sample a permutation with a fixed density $\rho \in [0,1]$ of occurrences of pattern 12, we should choose x in the above expression so that the monomial $C_{[\rho n^2/2]} x^{[\rho n^2/2]}$ is the maximal one, and then use the insertion probability measures which are (truncated) geometric random variables with rate x.

Here x is determined as a function of ρ by Legendre duality (see below for an exact formula). Let r be defined by $e^{-r} = x$. In the limit of large n, the truncated geometric insertion densities converge to truncated exponential densities

$$f(x,y) = \frac{re^{-ry}}{1-e^{-rx}} 1_{[0,x]}(y). \tag{57}$$

We can reconstruct the permuton from these insertion densities as follows. Note that the CDF of the insertion measure is

$$F(x,y) = \frac{1-e^{-ry}}{1-e^{-rx}}. \tag{58}$$

We need to solve the ODE (47), which in this case (to simplify notation we changed $\tilde{y}$ to y) is

$$\frac{1-e^{-rG(x,y)}}{1-e^{-rx}} = \frac{dG(x,y)}{dx}. \tag{59}$$

This can be rewritten as

$$\frac{dx}{1-e^{-rx}} = \frac{dG}{1-e^{-rG(x,y)}}. \tag{60}$$

Integrating both sides and solving for G gives the permuton

$$G(x,y) = \frac{1}{r}\log\left(1 + \frac{(e^{rx}-1)(e^{ry}-1)}{e^r-1}\right) \tag{61}$$

which has density

$$g(x,y) = \frac{r(1-e^{-r})}{(e^{r(1-x-y)/2} - e^{r(x-y-1)/2} - e^{r(y-x-1)/2} + e^{r(x+y-1)/2})^2}. \tag{62}$$

See Figure 1 for some examples for varying ρ.

The permuton entropy of this permuton is obtained from (50), and as a function of r it is, using the dilogarithm,

$$H(r) = -\frac{2\text{Li}_2\left(e^r\right)}{r} + \frac{\pi^2}{3r} - 2\log\left(1-e^r\right) + \log\left(e^r-1\right) - \log(r) + 2. \quad (63)$$

The density of inversions ρ is the integral of the expectation of f:

$$\rho(r) = \frac{r\left(r - 2\log\left(1-e^r\right)+2\right) - 2\text{Li}_2\left(e^r\right)}{r^2} + \frac{\pi^2}{3r^2}; \quad (64)$$

see Figure 2 for the inversion density as a function of r.

See Figure 3 for the entropy as a function of the inversion density.

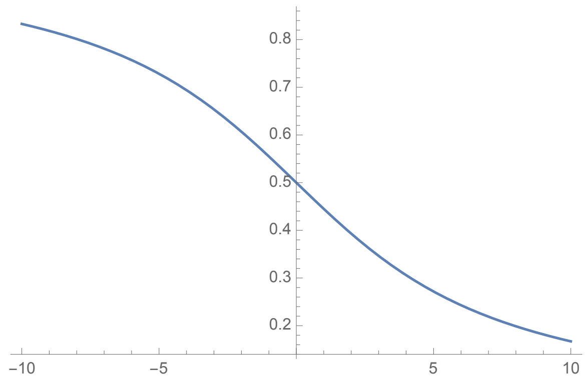

Figure 2: Inversion density as function of r.

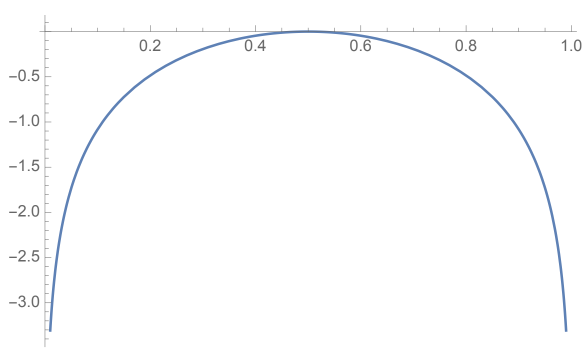

Figure 3: Entropy as function of inversion density.

8 Star models

Equation (56) gives the generating function for patterns 12, i.e. occurrences of pattern 12. For a permutation π let $k_1 = k_1(\pi)$ be the number of 12 patterns. Let k_2 be the number of '$**3$' patterns, that is, patterns of the form 123 or 213.

A similar argument to the above shows that the joint generating function for k_1 and k_2 is

$$\sum_{k_1=0}^{\binom{n}{2}} C_{k_1,k_2} x^{k_1} y^{k_2} = \prod_{j=1}^{n} \left(\sum_i x^i y^{i(i-1)/2} \right). \tag{65}$$

More generally, letting k_3 be the number of patterns $**2$, that is, 132 or 312, and k_4 be the number of $**1$ patterns, that is, 231 or 321. The joint generating function for these four types of patterns is

$$\sum_{k_1=0}^{\binom{n}{2}} C_{k_1,k_2,k_3,k_4} x^{k_1} y^{k_2} z^{k_3} w^{k_4} = \prod_{j=1}^{n} \left(\sum_{i=0}^{j} x^i y^{i(i-1)/2} z^{i(j-i)} w^{(j-i)(j-i-1)/2} \right). \tag{66}$$

One can similarly write down the joint generating function for all patterns of the type $** \ldots *i$, with a string of some number of stars followed by some i in $[n]$. (Note that with this notation, 12 patterns are $*2$ patterns.) These constitute a significant generalization of the Mallows model; see [27] for background.

The term star model was suggested by analogy with subgraph densities of graphons. The pattern $1***$, for example, corresponds to the subgraph $K_{1,3}$.

8.1 The $*2/**3$ model

By way of illustration, let us consider the simplest case of $*2$ (that is, 12) and $**3$.

Theorem 10. *The feasible region for (ρ_{*2}, ρ_{**3}) is the region bounded below by the parametrized curve*

$$(2t - t^2, 3t^2 - 2t^3)_{t \in [0,1]} \tag{67}$$

and above by the parametrized curve

$$(1-t^2, 1-t^3)_{t\in[0,1]}. \tag{68}$$

One can show that the permutons on the boundaries are unique and supported on line segments of slopes ± 1, and are as indicated in Figure 4.

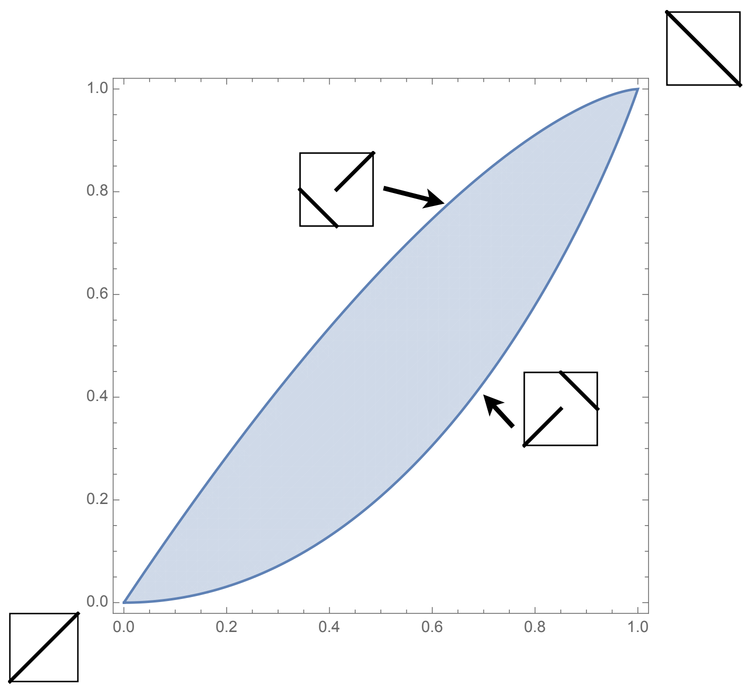

Figure 4: Feasible region for (ρ_{*2}, ρ_{**3}).

Proof. While this can be proved directly from the generating function (65), we give a simpler proof using the insertion density procedure. During the insertion process let $I_{12}(x)$ be the fractional number of 12 patterns in the partial permutation constructed up to time x, that is, $I_{12}(x) = \int_0^x Y_t\,dt$, where Y_t is the random variable giving the location of the insertion of t. By the law of large numbers we can replace Y_t here by its mean value. Let $I_{**3}(x)$ likewise be the fraction of $**3$ patterns created by time x. We have

$$I_{**3}(x) = \int_0^x Y_t^2/2\,dt. \tag{69}$$

Let us fix $\rho_{12} = I_{12}(1)$. To maximize $I_{**3}(x)$, we need to maximize

$$\int_0^1 (I'_{12}(t))^2\,dt \quad \text{subject to} \quad \int_0^1 I'_{12}(t)\,dt = \rho_{12}. \tag{70}$$

This is achieved by making $I'_{12}(t)$ either zero or maximal. Since $I'_{12}(t) \leq t$, we can achieve this by inserting points at the beginning for as long as possible and then inserting points at the end, that is, $Y_t = 0$ up to $t = a$ and then $Y_t = t$ for $t \in [a, 1]$. The resulting permuton is then as shown in the figure: on the square $[0, a]^2$ it is a descending diagonal and on the square $[a, 1] \times [a, 1]$ it is an ascending diagonal.

Likewise to minimize the above integral (70) we need to make the derivatives $I'_{12}(t)$ as equal as possible. Since $I'_{12}(t) \leq t$, this involves setting $I'_{12}(t) = t$ up to $t = a$ and then having it constant after that. The resulting permuton is then as shown in the figure: on the square $[0, a]^2$ it is an ascending diagonal and on the square $[a, 1] \times [a, 1]$ it is a descending diagonal.

A short calculation now yields the algebraic form of the boundary curves.

□

Using the insertion density procedure outlined earlier, we see that the permuton as a function of x, y has an explicit analytic density (which cannot be written in terms of elementary functions, however). The permutons for various values of (ρ_{*2}, ρ_{**3}) are shown in Figure 5.

The entropy $s(\rho_{*2}, \rho_{**3})$ is plotted in Figure 6. It is strictly concave (see Theorem 11 below) and achieves its maximal value, zero, precisely at the point $1/2, 1/3$, the uniform measure.

8.2 Concavity and analyticity of entropy for star models

Theorem 11. *For a star model with a finite number of densities $\rho_1, \ldots, \rho_k$ of patterns $\tau_1 \ldots, \tau_k$ respectively, the feasible region is convex and the entropy $H(\rho_1, \ldots, \rho_k)$ is strictly concave and analytic on the feasible region.*

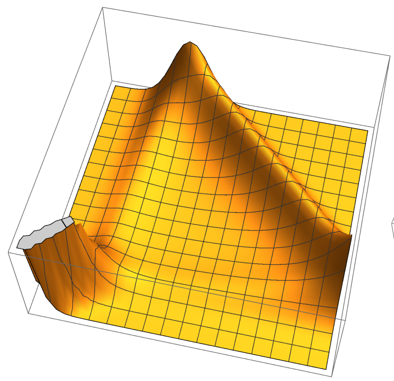

Figure 5: Permutons with $(\rho_{*2}, \rho_{**3}) = (.5, .2)$, and $(.5, .53)$ respectively.

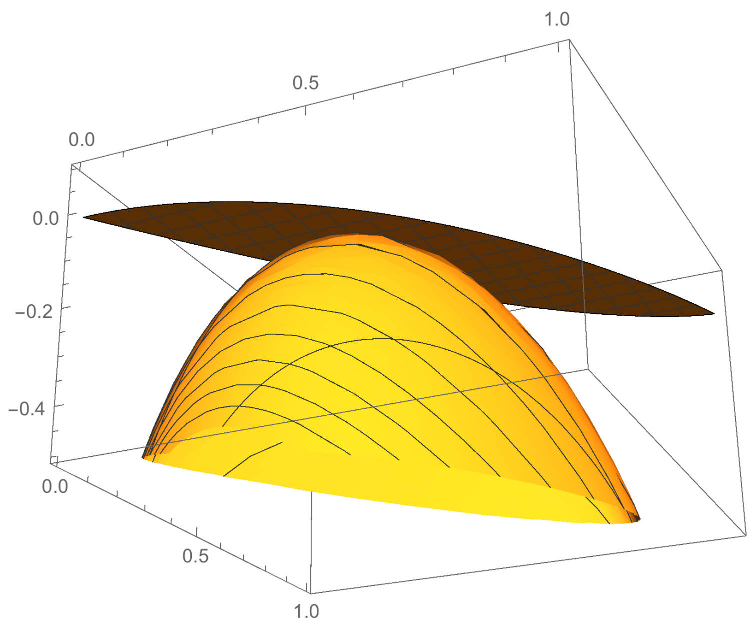

Figure 6: The entropy function on the parameter space for ρ_{12}, ρ_{**3}.

Proof. The generating function for permutations of n counting patterns τ_i is

$$Z_n(x_1,\dots,x_k) = \sum_{\pi\in S_n} x_1^{n_1}\dots x_k^{n_k} \tag{71}$$

where $n_i = n_i(\pi)$ is the number of occurrences of pattern τ_i in π. As discussed above, Z_n can be written as a product generalizing (66). Write $x_i = e^{a_i}$. Then the product expression for Z_n is

$$Z_n = \prod_{j=1}^{n}\sum_{i=0}^{j} e^{p(i,j)}, \tag{72}$$

where $p(i,j)$ is a polynomial in i and j with coefficients that are linear in the a_i. For large n it is necessary to normalize the a_i by an appropriate power of n: write

$$x_i = e^{a_i} = e^{\alpha_i/n^{k_i-1}} \tag{73}$$

where k_i is the length of the pattern τ_i.

Writing $i/n = t$ and $j/n = x$, the expression for $\log Z_n$ is then a Riemann sum, once normalized: In the limit $n\to\infty$ the "normalized free energy" F is

$$F := \lim_{n\to\infty}\frac{1}{n}(\log Z_n - \log n!) = \int_0^1 \left[\log\int_0^x e^{\tilde{p}(t,x)}\,dt\right]\,dx \tag{74}$$

where $\tilde{p}(t,x) = p(nt,nx) + o(1)$ is a polynomial in t and x, independent of n, with coefficients which are linear functions of the α_i. Explicitly we have

$$\tilde{p}(t,x) = \sum_{i=1}^{k}\alpha_i\frac{t^{r_i}(x-t)^{s_i}}{r_i!s_i!} \tag{75}$$

where $r_i + s_i = k_i - 1$ and, if $\tau_i = *\dots*\ell_i$ then $s_i = k_i - \ell_i$.

We now show that F is concave as a function of the α_i, by computing its Hessian matrix. We have

$$\frac{\partial F}{\partial\alpha_i} = \int_0^1 \frac{\int_0^x t^{r_i}(x-t)^{s_i}e^{p(t,x)}dt}{\int_0^x e^{p(t,x)}dt}dx = \int_0^1 \langle t^{r_i}(x-t)^{s_i}\rangle dx \tag{76}$$

where $\langle\cdot\rangle$ is the expectation with respect to the measure on $[0,x]$ with (unnormalized) density $e^{p(t,x)}$.

Differentiating a second time we have

$$\begin{aligned}\frac{\partial^2 F}{\partial\alpha_j\partial\alpha_i} &= \int_0^1 \langle t^{r_i+r_j}(x-t)^{s_i+s_j}\rangle - \langle t^{r_i}(x-t)^{s_i}\rangle\langle t^{r_j}(x-t)^{s_j}\rangle dx \\ &= \int_0^1 \mathrm{Cov}(t^{r_i}(x-t)^{s_i}, t^{r_j}(x-t)^{s_j})\, dx \qquad (77)\end{aligned}$$

where Cov is the covariance. By independence of the variables for different values of x, the integral of the covariance is the covariance of the integrals:

$$= \mathrm{Cov}(T^{r_i}(x-T)^{s_i}, T^{r_j}(x-T)^{s_j}) \qquad (78)$$

where

$$T = \int_0^1 \frac{\int_0^x t e^{p(t,x)} dt}{\int_0^x e^{p(t,x)} dt}\, dx. \qquad (79)$$

Thus we see that the Hessian matrix is the covariance matrix of a set of random variables with no linear dependencies, and so is positive definite. This completes the proof of strict concavity of the free energy F.

Since Z_n is the (unnormalized) probability generating function, the vector of densities as a function of the $\{\alpha_i\}$ is obtained for each n by the gradient of the logarithm

$$(\rho_1, \ldots, \rho_k) = \frac{1}{n}\nabla \log Z_n(\alpha_1, \ldots, \alpha_n). \qquad (80)$$

In the limit we can replace $\nabla \log Z_n$ by ∇F; by strict concavity of F its gradient is injective, and surjective onto the interior of the feasible region. The entropy function H is the Legendre dual of the free energy F, that is,

$$H(\rho_1, \ldots, \rho_k) = \max_{\{\alpha_i\}}\{F(\alpha_1, \ldots, \alpha_k) - \sum \alpha_i \rho_i\}. \qquad (81)$$

Since F is analytic, so is H. □

9 PDEs for permutons

For permutations with constraints on patterns of length 3 (or less) one can write explicit PDEs for the maximizers. Let us first redo the case of 12-patterns, which we already worked out by another method in Section 7.

9.1 Patterns 12

The density of patterns 12 is given in (2). Consider the problem of maximizing $H(\gamma)$ subject to the constraint $I_{12}(\gamma) = \rho$. This involves finding a solution to the Euler-Lagrange equation

$$dH + \alpha\, dI_{12} = 0 \tag{82}$$

for some constant α, for all variations $g \mapsto g + \epsilon h$ fixing the marginals.

Given points $(a_1, b_1), (a_2, b_2) \in [0,1]^2$ we can consider the change in H and I_{12} when we remove an infinitesimal mass δ from (a_1, b_1) and (a_2, b_2) and add it to locations (a_1, b_2) and (a_2, b_1). (Note that two measures with the same marginals are connected by convolutions of such operations.) The change in H to first order under such an operation is δ times (letting $S_0(p) := -p\log p$)

$$- S_0'(g(a_1,b_1)) - S_0'(g(a_2,b_2)) + S_0'(g(a_1,b_2)) + S_0'(g(a_2,b_1)) \\ = \log \frac{g(a_1,b_1)g(a_2,b_2)}{g(a_1,b_2)g(a_2,b_1)}. \tag{83}$$

The change in I_{12} to first order is δ times

$$\sum_{i,j=1}^{2} (-1)^{i+j} \left(\int_{a_i<x_2} \int_{y_2<b_j} g(x_2,y_2)dx_2\, dy_2 + \int_{x_1<a_i} \int_{b_j<y_1} g(x_1,y_1)dx_1\, dy_1 \right) \\ = 2\sum_{i,j=1}^{2} (-1)^{i+j} \int_{a_i<x_2} \int_{y_2<b_j} g(x_2,y_2)dx_2\, dy_2. \tag{84}$$

Differentiating (82) with respect to $a = a_1$ and $b = b_1$, we find

$$\frac{\partial}{\partial a}\frac{\partial}{\partial b} \log g(a,b) + 2\alpha g(a,b) = 0. \tag{85}$$

One can check that the formula (62) satisfies this PDE.

9.2 Patterns 123

The density of patterns 123 is

$$I_{123}(\gamma) = 6 \int_{x_1<x_2<x_3,\ y_1<y_2<y_3} g(x_1,y_1)g(x_2,y_2)g(x_3,y_3)dx_1\, dx_2\, dx_3\, dy_1\, dy_2\, dy_3. \tag{86}$$

Under a similar perturbation as above the change in I_{123} to first order is δ times

$$dI_{123} = \sum_{i,j=1}^{2} (-1)^{i+j} \left(\int_{a_i<x_2<x_3,\ b_j<y_2<y_3} g(x_2,y_2)g(x_3,y_3)dx_2\,dx_3\,dy_2\,dy_3 \right.$$
$$+ \int_{x_1<a_i<x_3,\ y_1<b_j<y_3} g(x_1,y_1)g(x_3,y_3)dx_1\,dx_3\,dy_1\,dy_3$$
$$\left. + \int_{x_1<x_2<a_i,\ y_1<y_2<b_j} g(x_1,y_1)g(x_2,y_2)dx_1\,dx_2\,dy_1\,dy_2 \right). \quad (87)$$

The middle integral here is a product

$$\int_{x_1<a_i,\ y_1<b_j} g(x_1,y_1)dx_1\,dy_1 \int_{a_i<x_3,\ b_j<y_3} g(x_3,y_3)dx_3\,dy_3$$
$$= G(a_i,b_j)(1-a_i-b_j+G(a_i,b_j)). \quad (88)$$

Differentiating each of these three integrals with respect to both $a=a_1$ and $b=b_1$ (then only the $i=j=1$ term survives) gives, for the first integral

$$g(a,b)\int_{a<x_3,\ b<y_3} g(x_3,y_3)dx_3\,dy_3 = g(a,b)(1-a-b+G(a,b)), \quad (89)$$

for the second integral

$$g(a,b)(1-a-b+2G(a,b)) + G_x(a,b)(-1+G_y(a,b))$$
$$+ G_y(a,b)(-1+G_x(a,b)), \quad (90)$$

and the third integral

$$g(a,b)\int_{x_1<a,\ b<y_1} g(x_1,y_1)dx_1\,dy_1 = g(a,b)G(a,b). \quad (91)$$

Summing, we get (changing a,b to x,y)

$$(dI_{123})_{xy} = 2G_{xy}(1-x-y+2G) + 2G_xG_y - G_x - G_y. \quad (92)$$

Thus the Euler-Lagrange equation is

$$(\log G_{xy})_{xy} + \alpha\big(2G_{xy}(1-x-y+2G) + 2G_xG_y - G_x - G_y\big) = 0. \quad (93)$$

This simplifies somewhat if we define $K(x,y) = 2G(x,y) - x - y + 1$. Then

$$(\log K_{xy})_{xy} + \frac{\alpha}{2}\,(2K_{xy}K + K_xK_y - 1) = 0. \tag{94}$$

In a similar manner we can find a PDE for the permuton with fixed densities of other patterns of length 3. In fact one can proceed similarly for longer patterns, getting systems of PDEs, but the complexity grows with the length.

10 The $12/123$ model

When we fix the density of patterns 12 and 123, the feasible region has a complicated structure, see Figure 7.

Theorem 12. *The feasible region for ρ_{12} versus ρ_{123} is the same as the feasible region of edges and triangles in the graphon model.*

Proof. Let $\mathcal{R}$ denote the feasible region for pairs $\big(\rho_{12}(\gamma), \rho_{123}(\gamma)\big)$ consisting of the 12 density and 123 density of a permuton (equivalently, for the closure of the set of such pairs for finite permutations).

Each permutation $\pi \in S_n$ determines a (two-dimensional) poset P_π on $\{1,\ldots,n\}$ given by $i \prec j$ in P_π iff $i < j$ and $\pi_i < \pi_j$. The *comparability graph* $G(P)$ of a poset P links two points if they are comparable in P, that is, $x \sim y$ if $x \prec y$ or $y \prec x$. Then $i \sim j$ in $G(P_\pi)$ precisely when $\{i,j\}$ constitutes an incidence of the pattern 12, and $i \sim j \sim k \sim i$ when $\{i,j,k\}$ constitutes an incidence of the pattern 123. Thus the 12 density of π is equal to the edge density of $G(P_\pi)$, and the 123 density of π is the triangle density of $G(P_\pi)$—that is, the probability that three random vertices induce the complete graph K_3. This correspondence extends perfectly to limit objects, equating 12 and 123 densities of permutons to edge densities and triangle densities of graphons.

The feasible region for edge and triangle densities of graphs (now, for graphons) has been studied for many years and was finally determined by Razborov [25]; we call it the "scalloped triangle" $\mathcal{T}$. It follows from the above discussion that the feasibility region $\mathcal{R}$ we seek for permutons is a subset of $\mathcal{T}$, and it remains only to prove that $\mathcal{R}$ is all of $\mathcal{T}$. In fact we can realize $\mathcal{T}$ using only a rather simple two-parameter family of permutons.

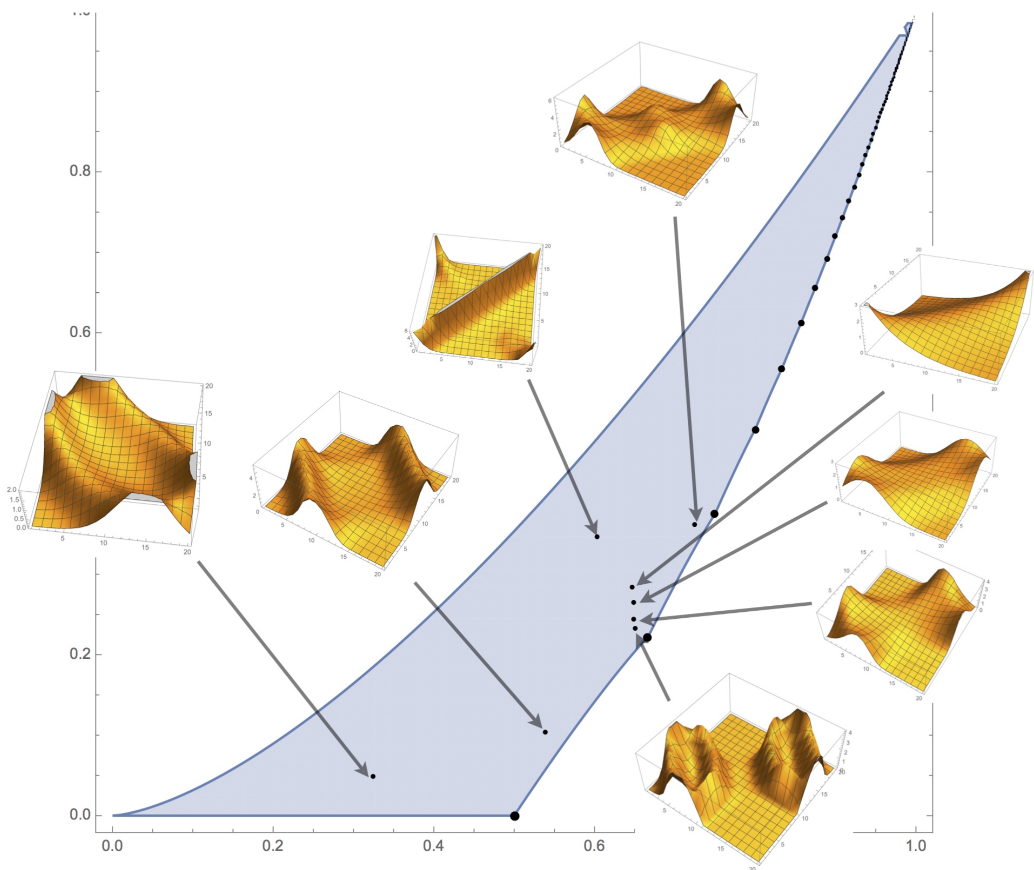

Figure 7: The feasible region for ρ_{12} versus ρ_{123}, with corresponding permutons at selected points.

Let reals a, b satisfy $0 < a \leq 1$ and $0 < b \leq a/2$, and set $k := \lfloor a/b \rfloor$. Let us denote by $\gamma_{a,b}$ the permuton consisting of the following diagonal line segments, all of equal density:

1. The segment $y = 1 - x$, for $0 \leq x \leq 1{-}a$;
2. The k segments $y = (2j{-}1)b{-}1{+}a{-}x$ for $1{-}a{+}(j{-}1)b < x \leq 1{-}a{+}jb$, for each $j = 1, 2, \ldots, k$;
3. The remaining, rightmost segment $y = 1{+}kb{-}x$, for $1{-}a{+}kb < x \leq 1$.

(See Fig. 8 below.)

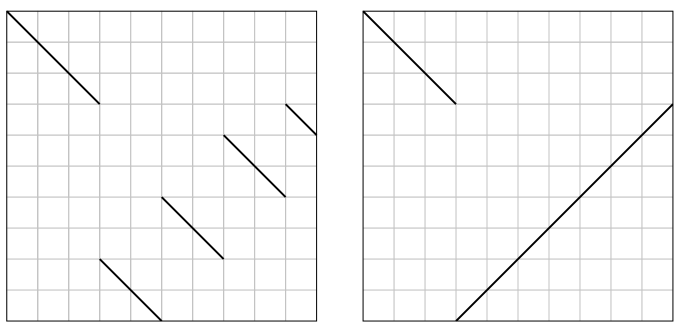

Figure 8: Support of the permutons $\gamma_{.7,.2}$ and $\gamma_{.7,0}$.

We interpret $\gamma_{a,0}$ as the permuton containing the segment $y = 1 - x$, for $0 \leq x \leq 1{-}a$, and the positive-slope diagonal from $(1{-}a, 0)$ to $(1, 1{-}a)$; finally, $\gamma_{0,0}$ is just the reverse diagonal from $(0, 1)$ to $(1, 0)$. These interpretations are consistent in the sense that $\rho_{12}(\gamma_{a,b})$ and $\rho_{123}(\gamma_{a,b})$ are continuous functions of a and b on the triangle $0 \leq a \leq 1$, $0 \leq b \leq a/2$. (In fact, $\gamma_{a,b}$ is itself continuous in the topology of Γ, so all pattern densities are continuous.)

It remains only to check that the comparability graphons corresponding to these permutons match extremal graphs in [25] as follows:

- $\gamma_{a,0}$ maps to the upper left boundary of $\mathcal{T}$, with $\gamma_{0,0}$ going to the lower left corner while $\gamma_{1,0}$ goes to the top;
- $\gamma_{a,a/2}$ goes to the bottom line, with $\gamma_{1,1/2}$ going to the lower right corner;
- For $1/(k+2) \leq b \leq 1/(k+1)$, $\gamma_{1,b}$ goes to the kth lowest scallop, with $\gamma_{1,1/(k+1)}$ going to the bottom cusp of the scallop and $\gamma_{1,1/(k+2)}$ to the top.

It follows that $(a,b) \mapsto \big(\rho_{12}(\gamma_{a,b}), \rho_{123}(\gamma_{a,b})\big)$ maps the triangle $0 \leq a \leq 1$, $0 \leq b \leq a/2$ onto all of $\mathcal{T}$, proving the theorem. □

It may be prudent to remark at this point that while the feasible region for 12 versus 123 density of permutons is the same as that for edge and triangle density of graphs, the *topography* of the corresponding entropy functions within this region is entirely different. In the graph case the entropy landscape is studied in [22, 23, 24]; one of its features is a ridge along the "Erdős-Rényi" curve (where triangle density is the 3/2 power of edge density). There is a sharp drop-off below this line, which represents the very high entropy graphs constructed by choosing edges independently with constant probability. The graphons that maximize entropy at each point of the feasible region all appear to be very combinatorial in nature: each has a partition of its vertices into finitely many classes, with constant edge density between any two classes and within any class, and is thus described by a finite list of real parameters.

The permuton topography features a different high curve, representing the permutons (discussed above) that maximize entropy for a fixed 12 density. Moreover, the permutons that maximize entropy at interior points of the region appear, as in other regions discussed above, always to be analytic.

We do not know explicitly the maximizing permutons (although they satisfy an explicit PDE, see Section 9) or the entropy function.

11 123/321 case

The feasible region for fixed densities ρ_{123} versus ρ_{321} is the same as the feasible region $\mathcal{B}$ for triangle density $x = d(K_3, G)$ versus anti-triangle density $y = d(\bar{K}_3, G)$ of graphons [15]. Let C be the line segment $x + y = \frac{1}{4}$ for $0 \leq x \leq \frac{1}{4}$, D the x-axis from $x = \frac{1}{4}$ to $x = 1$, and E the y-axis from $y = \frac{1}{4}$ to $y = 1$. Let F_1 be the curve given parametrically by $(x,y) = (t^3, (1-t)^3 + 3t(1-t)^2)$,

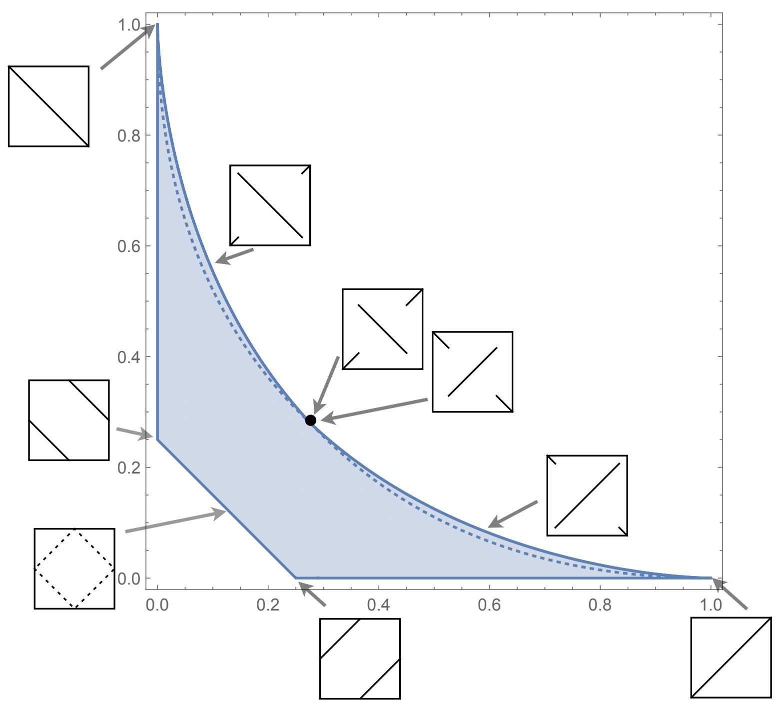

Figure 9: The feasible region for ρ_{123}, ρ_{321}. It is bounded above by the parametrized curves $(1-3t^2+2t^3, t^3)$ and $(t^3, 1-3t^2+2t^3)$ which intersect at $(x,y) = (.278..., .278...)$. The lower boundaries consist of the axes and the line $x+y = 1/4$.

for $0 \le t \le 1$, and F_2 its symmetric twin $(x,y) = ((1-t)^3 + 3t(1-t)^2, t^3)$. Then $\mathcal{B}$ is the union of the area bounded by C, D, E and F_1 and the area bounded by C, D, E and F_2.

The curves F_1 and F_2 cross at a concave "dimple" (r,r) where $r = s^3 = (1-s)^3 + 3s(1-s)^2)$, with $s \sim .653$ and $r \sim .278$; see Fig. 9 below.

To see that $\mathcal{B}$ is also the feasible region for 123 versus 321 density of permutons, an argument much like the one above for 12 density versus 123 can be (and was, by [8]) given. Permutons realizing various boundary points are illustrated in Fig. 9; they correspond to the extremal graphons described in [15]. The rest are filled in by parameterization and a topological argument.

Of note for both graphons and permutons is the double solution at the dimple. These solutions are significantly different, as evidenced by the fact that their edge-densities (12 densities, for the permutons) differ. This multiplicity of solutions, if there are no permutons bridging the gap, suggests a phase transition in the entropy-optimal permuton in the interior of $\mathcal{B}$ in a neighborhood of the dimple. In fact, we can use a stability theorem from [14] to show that the phenomenon is real.

For $0 \le p \le 1$, let $M_p = \max\big((1-p^{1/3})^3 + 3p^{1/3}(1-p^{1/3})^2, (1-q)^{1/3}\big)$ where q is the unique real root of $q^3 + 3q^2(1-q) = p$.

Theorem 13 (special case of Theorems 1.1 and 1.2 of [14])**.** *For any $\epsilon > 0$ there is a $\delta > 0$ and an N such that for any n-vertex graph G with $n > N$ and $d(\bar{K}_3, G) > p$ and $|d(K_3,G) - M_p| < \delta$ is ϵ-close to a graph H on n vertices consisting of a clique and isolated vertices, or an independent set whose vertices are adjacent to all other vertices of the graph. Here $M_p :=$ $\max\big((1-p^{1/3})^3 + 3p^{1/3}(1-p^{1/3})^2, (1-q)^{1/3}\big)$ where q is the unique real root of $q^3 + 3q^2(1-q) = p$; that is, M_p is the largest possible value of $d(K_3,G)$ given $d(\bar{K}_3, G) = p$.*

(Two n-vertex graphs are ϵ-close if one can be made isomorphic to the other by adding or deleting at most $\epsilon \cdot \binom{n}{2}$ edges.)

From Theorem 13 we conclude:

Lemma 14. *There is a neighborhood of the point (r,r) in the feasible region for patterns 123 and 321 within which no permuton has 12-density near $\frac{1}{2}$.*

Proof. There are in fact many permutons representing the dimple (r,r), but only two classes if we consider permutons with isomorphic comparability graphs to be equivalent. The class that came from the curve F_1 has 12 density $s^2 \sim .426$, the other $1 - s^2 \sim .574$. (Interestingly, the other end of the F_1 curve—represented uniquely by the identity permuton—had 12 density 1, while the F_2 class "began" at 12 density 0. Thus, the 12 densities crossed on the way in from the corners of $\mathcal{B}$.)

Taking, say, $\epsilon = .07$ in Lemma 14, we find a $\delta > 0$ with the stated property. Let $\delta' = (M_{r-\delta} - r)/2$, so that if $p = r - \delta'$, we get $|M_p - \rho_{123}(\gamma)| < \delta$ as required by the hypothesis of Theorem 13 (noting that $\rho_{123}(\gamma)$ is the triangle density of the comparability graph corresponding to γ). We conclude that any permuton γ for which $(\rho_{123}(\gamma), \rho_{321}(\gamma))$ lies in the square $[r - \delta', r] \times [r - \delta', r]$ has 12-density within .07 of either .426 or .574, thus outside the range $[.496, .504]$. □

Acknowledgments

The second author would like to thank Roman Glebov, Andrzej Grzesik and Jan Volec for discussions on pattern densities in permutons.

The authors gratefully acknowledge the support of the National Science Foundation's Institute for Computational and Experimental Research in Mathematics, at Brown University, where this research was conducted. This work was also partially supported by NSF grants DMS-1208191, DMS-1208941 and DMS-0901475, the Simons Foundation, and by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 259385.

References

[1] Zs. Baranyai, On the factorization of the complete uniform hypergraph. In *Infinite and finite sets* (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, pages 91–108. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[2] D. Bevan, Growth rates of permutation grid classes, tours on graphs, and the spectral radius, ArXiv:1302.2037v4 (22 Oct 2013).

[3] D. Bevan, Permutations avoiding 1324 and patterns in Lukasiewicz paths, ArXiv:1406.2890v2 (8 Jan 2015).

[4] N. Bhatnagar, and R. Peled, Lengths of monotone subsequences in a Mallows permutation, ArXiv:1306.3674v2 (21 Mar 2014).

[5] G. Brightwell and N. Georgiou, Continuum limits for classical sequential growth models, *Rand. Struc. Algor.* **36** (2010) 218–250.

[6] S. Chatterjee and P. Diaconis, Estimating and understanding exponential random graph models, *Ann. Stat.* **41** (2013) 2428-2461.

[7] S. Chatterjee and S.R.S. Varadhan, The large deviation principle for the Erdős-Rényi random graph, *Eur. J. Combin.* **32** (2011) 1000–1017.

[8] S. Elizalde and M. Noy, The regions obtained by pairs of densities of patterns of length 3, preprint, 2015.

[9] G.B. Folland, *Real analysis* (Second edition), John Wiley & Sons Inc., New York, 1999, pp. xvi+386; ISBN 0-471-31716-0.

[10] R. Glebov, A. Grzesik, T. Klimošová and D. Král', Finitely forcible graphons and permutons, *J. Comb. Thy B* **110** (2015) 112–135.

[11] J. Hladký, A. Máthé, V. Patel and O. Pikhurko, Poset limits can be totally ordered, ArXiv:1211.2473v2 (4 Nov 2013).

[12] C. Hoppen, Y. Kohayakawa, C.G. Moreira, B. Ráth and R.M. Sampaio, Limits of permutation sequences through permutation regularity, Arxiv:1106.1663v1 (8 Jun 2011).

[13] C. Hoppen, Y. Kohayakawa, C.G. Moreira, and R.M. Sampaio, Limits of permutation sequences, *J. Combin. Theory B* **103** (2013) 93–113.

[14] H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov, On the densities of cliques and independent sets in graphs, ArXiv:1211.4532v2 (8 Dec 2013).

[15] H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov, On the 3-local profiles of graphs, ArXiv:1211.3106v2 (8 Dec 2013).

[16] S. Janson, Poset limits and exchangeable random posets, *Combinatorica* **31** (2011) 529–563.

[17] R. Kenyon, C. Radin, K. Ren and L. Sadun, Multipodal structure and phase transitions in large constrained graphs, ArXiv:1405.0599v2 (9 Sep 2014).

[18] D. Král' and O. Pikhurko, Quasirandom permutations are characterized by 4-point densities, *Geom. Funct. Anal.* **23** (2013) 570–579.

[19] L. Lovász, *Large Networks and Graph Limits*, AMS, Providence RI 2012.

[20] S. Minor and I. Pak, The shape of random pattern-avoiding permutations, *Adv. Appl. Math.* **55** (2014), 86–130.

[21] C.B. Presutti and W.R. Stromquist, Packing rates of measures and a conjecture for the packing density of 2413, *London Math. Soc. Lecture Notes* **376** (2010), 3–40.

[22] R. Radin and L. Sadun, Phase transitions in a complex network, *J. Phys. A: Math. Theor.* **46** (2013) 305002.

[23] R. Radin and L. Sadun, Singularities in the entropy of asymptotically large simple graphs, *J. Stat. Phys.* **158**(2015) 853–865.

[24] C. Radin, K. Ren and L. Sadun, The asymptotics of large constrained graphs, *J. Phys. A: Math. Theor.* **47** (2014) 175001.

[25] A. Razborov, On the minimal density of triangles in graphs, *Combinatorics, Probability and Computing* **17** (2008), 603–618.

[26] A. Razborov, Flag algebras, *J. Symbolic Logic* **72** (2007), 1239–1282.

[27] S. Starr, Thermodynamic limit for the Mallows model on S_n, *J. Math. Phys.* **50** (2009), 095208.