High-dimensional permutations

Nati Linial

Nogafest, Tel Aviv, January '16
What are high dimensional permutations?

A permutation can be encoded by means of a permutation matrix. As we all know, this is an $n \times n$ array of zeros and ones in which every line contains exactly one 1-entry. A line here means either a row or a column.
A permutation can be encoded by means of a permutation matrix.
A permutation can be encoded by means of a permutation matrix. As we all know, this is an $n \times n$ array of zeros and ones in which every line contains exactly one 1-entry.
A permutation can be encoded by means of a permutation matrix. As we all know, this is an $n \times n$ array of zeros and ones in which every line contains exactly one 1-entry. A line here means either a row or a column.
A notion of high dimensional permutations

This suggests the following definition of a $d$-dimensional permutation on $[n]$. 

A line is a set of $n$ entries in the array that are obtained by fixing $d$ out of the $d+1$ coordinates and the letting the remaining coordinate take all values from 1 to $n$. 

Whereas a matrix has two kinds of lines, namely rows and columns, now there are $d+1$ kinds of lines.
This suggests the following definition of a \( d \)-dimensional permutation on \([n]\).
It is an array \([n] \times [n] \times \ldots \times [n] = [n]^{d+1}\) (with \(d + 1\) factors) of zeros and ones in which every line contains exactly one 1-entry.
This suggests the following definition of a $d$-dimensional permutation on $[n]$. It is an array $[n] \times [n] \times \ldots \times [n] = [n]^{d+1}$ (with $d + 1$ factors) of zeros and ones in which every line contains exactly one 1-entry. Whereas a matrix has two kinds of lines, namely rows and columns, now there are $d + 1$ kinds of lines.
This suggests the following definition of a $d$-dimensional permutation on $[n]$. It is an array $[n] \times [n] \times \ldots \times [n] = [n]^{d+1}$ (with $d + 1$ factors) of zeros and ones in which every line contains exactly one 1-entry.

Whereas a matrix has two kinds of lines, namely rows and columns, now there are $d + 1$ kinds of lines.

A line is a set of $n$ entries in the array that are obtained by fixing $d$ out of the $d + 1$ coordinates and the letting the remaining coordinate take all values from 1 to $n$. 
The case $d = 2$. A familiar face?

According to our definition, a 2-dimensional permutation on $[n]$ is an $[n] \times [n] \times [n]$ array of zeros and ones in which every row, every column, and every shaft contains exactly one 1-entry.
The case $d = 2$. A familiar face?

According to our definition, a 2-dimensional permutation on $[n]$ is an $[n] \times [n] \times [n]$ array of zeros and ones in which every row, every column, and every shaft contains exactly one 1-entry. An equivalent description can be achieved by using a topographical map of this terrain.
The two-dimensional case

Rather that an \([n] \times [n] \times [n]\) array of zeros and ones we can now consider an \([n] \times [n]\) array with entries from \([n]\), as follows:

The \((i, j)\) entry in this array is \(k\) where \(k\) is the “height above the ground” of the unique 1-entry in the shaft \((i, j, \ast)\).

It is easily verified that the defining condition is that in this array every row and every column contains every entry \(n \geq i \geq 1\) exactly once.

In other words: Two-dimensional permutations are synonymous with Latin Squares.
The two-dimensional case

Rather that an $[n] \times [n] \times [n]$ array of zeros and ones we can now consider an $[n] \times [n]$ array with entries from $[n]$, as follows: The $(i, j)$ entry in this array is $k$ where $k$ is the "height above the ground" of the unique 1-entry in the shaft $(i, j, \ast)$. 

In other words: Two-dimensional permutations are synonymous with Latin Squares.
Rather that an \([n] \times [n] \times [n]\) array of zeros and ones we can now consider an \([n] \times [n]\) array with entries from \([n]\), as follows: The \((i,j)\) entry in this array is \(k\) where \(k\) is the "height above the ground" of the unique 1-entry in the shaft \((i,j,\ast)\). It is easily verified that the defining condition is that in this array every row and every column contains every entry \(n \geq i \geq 1\) exactly once.
The two-dimensional case

Rather that an \([n] \times [n] \times [n]\) array of zeros and ones we can now consider an \([n] \times [n]\) array with entries from \([n]\), as follows: The \((i,j)\) entry in this array is \(k\) where \(k\) is the "height above the ground" of the unique 1-entry in the shaft \((i,j,*)\).

It is easily verified that the defining condition is that in this array every row and every column contains every entry \(n \geq i \geq 1\) exactly once.

In other words: Two-dimensional permutations are synonymous with Latin Squares.
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations.

▶ Can we determine (or estimate) the number of $d$-dimensional permutations?

▶ Can we generate them randomly and efficiently and describe their typical behavior?

▶ Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?

▶ Of Erd˝ os-Szekeres?

▶ Of the solution to Ulam’s Problem?

▶ Discrepancy questions in this territory?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

▶ Can we determine (or estimate) the number of $d$-dimensional permutations?
▶ Can we generate them randomly and efficiently and describe their typical behavior?
▶ Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
▶ Of Erdős-Szekeres?
▶ Of the solution to Ulam’s Problem?
▶ Discrepancy questions in this territory?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of \( d \)-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
- Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
- Of Erdős–Szekeres?
- Of the solution to Ulam’s Problem?
- Discrepancy questions in this territory?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of $d$-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of $d$-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
- Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of $d$-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
- Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
- Of Erdős-Szekeres?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of $d$-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
- Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
- Of Erdős-Szekeres? Of the solution to Ulam’s Problem?
Where do we go from here?

We seek high-dimensional counterparts of known phenomena in ("classical" = "one-dimensional") permutations. Specifically, we ask:

- Can we determine (or estimate) the number of $d$-dimensional permutations?
- Can we generate them randomly and efficiently and describe their typical behavior?
- Is there an analog of the Birkhoff von-Neumann Theorem on doubly stochastic matrices?
- Of Erdős-Szekeres? Of the solution to Ulam’s Problem?
- Discrepancy questions in this territory?
As we all know (Stirling’s formula)

\[ n! = \left( (1 + o(1)) \frac{n}{e} \right)^n \]
As we all know (Stirling’s formula)

\[ n! = \left( (1 + o(1)) \frac{n}{e} \right)^n \]

As van Lint and Wilson showed, the number of order-\(n\) Latin squares is

\[ |\mathcal{L}_n| = \left( (1 + o(1)) \frac{n}{e^2} \right)^{n^2} \]
So, let us conjecture

**Conjecture**

The number of \(d\)-dimensional permutations on \([n]\) is

\[
|S_n^d| = \left( (1 + o(1))\frac{n}{e^d} \right)^{n^d}
\]
and what we actually know

At present we can only prove the upper bound

**Theorem (NL, Zur Luria ’14)**

*The number of $d$-dimensional permutations on $[n]$ is*

$$|S^d_n| \leq \left( (1 + o(1)) \frac{n}{e^d} \right)^{n^d}$$
How van Lint and Wilson enumerated Latin Squares

Recall that the **permanent** of a square matrix is a "determinant without signs".

\[
\text{per}(A) = \sum_{\sigma \in S_n} \prod a_{i,\sigma(i)}
\]
This is a curious and fascinating mathematical object. E.g.
This is a curious and fascinating mathematical object. E.g.

- It counts perfect matchings in bipartite graphs.
This is a curious and fascinating mathematical object. E.g.

- It counts perfect matchings in bipartite graphs.
- In other words, it counts the generalized diagonals included in a 0/1 matrix.
This is a curious and fascinating mathematical object. E.g.

- It counts perfect matchings in bipartite graphs.
- In other words, it counts the **generalized diagonals** included in a 0/1 matrix.
- It is \#-\textit{P}-hard to calculate the permanent exactly, even for a 0/1 matrix.
This is a curious and fascinating mathematical object. E.g.

- It counts perfect matchings in bipartite graphs.
- In other words, it counts the generalized diagonals included in a 0/1 matrix.
- It is \( \#-P \)-hard to calculate the permanent exactly, even for a 0/1 matrix.
- On the other hand there is an efficient approximation scheme for permanents of nonnegative matrices.
A lower bound on the permanent

Since the permanent is so mysterious and hard to compute, it makes sense to seek bounds on it. We say that $A$ is a **doubly stochastic** matrix provided that

- Its entries are nonnegative.
- The sum of entries in every row is 1.
- The sum of entries in every column is 1.
Since the permanent is so mysterious and hard to compute, it makes sense to seek bounds on it. We say that $A$ is a **doubly stochastic** matrix provided that

- Its entries are nonnegative.
Since the permanent is so mysterious and hard to compute, it makes sense to seek bounds on it. We say that $A$ is a **doubly stochastic** matrix provided that

- Its entries are nonnegative.
- The sum of entries in every row is 1.
A lower bound on the permanent

Since the permanent is so mysterious and hard to compute, it makes sense to seek bounds on it. We say that $A$ is a **doubly stochastic** matrix provided that

- Its entries are nonnegative.
- The sum of entries in every row is 1.
- The sum of entries in every column is 1.
By the marriage theorem, a doubly stochastic matrix has a **positive** permanent.
By the marriage theorem, a doubly stochastic matrix has a positive permanent. The set of doubly stochastic matrices is a convex polytope. The permanent is a continuous function, so: What is \( \min \per A \) over \( n \times n \) doubly-stochastic matrices?

As conjectured by van der Waerden in the 20's and proved over 50 years later, in the minimizing matrix all entries are \( \frac{1}{n} \).

Theorem (Falikman; Egorichev ’80-81)

The permanent of every \( n \times n \) doubly stochastic matrix is \( \geq \frac{n!}{n^n} \).
By the marriage theorem, a doubly stochastic matrix has a **positive** permanent. The set of doubly stochastic matrices is a **convex polytope**. The permanent is a continuous function, so: What is 
\[
\min \per A \text{ over } n \times n \text{ doubly-stochastic matrices?}
\]
As conjectured by van der Waerden in the 20’s and proved over 50 years later, in the minimizing matrix all entries are \( \frac{1}{n} \).

**Theorem (Falikman; Egorichev ’80-81)**

*The permanent of every \( n \times n \) doubly stochastic matrix is \( \geq \frac{n!}{n^n} \).*
The following was conjectured by Minc

**Theorem (Brégman ’73)**

Let $A$ be an $n \times n$ 0/1 matrix with $r_i$ ones in the $i$-th row $i = 1, \ldots, n$. Then $\text{per } A \leq \prod_i (r_i!)^{1/r_i}$. The bound is tight.
The proof can be viewed as an extension of the Minc-Brégman theorem. In fact, our work uses ideas from subsequent papers of Schrijver and Radhakrishnan.
How we proved the upper bound on the number of $d$-dimensional permutations

The proof can be viewed as an extension of the Minc-Brégman theorem. In fact, our work uses ideas from subsequent papers of Schrijver and Radhakrishnan. This gave us an upper bound on the number of $d$-dimensional permutations.
How we proved the upper bound on the number of \(d\)-dimensional permutations

The proof can be viewed as an extension of the Minc-Brégman theorem. In fact, our work uses ideas from subsequent papers of Schrijver and Radhakrishnan.
This gave us an upper bound on the number of \(d\)-dimensional permutations.
What about a matching lower bound?
How we proved the upper bound on the number of $d$-dimensional permutations

The proof can be viewed as an extension of the Minc-Brégman theorem. In fact, our work uses ideas from subsequent papers of Schrijver and Radhakrishnan. This gave us an upper bound on the number of $d$-dimensional permutations.

What about a matching lower bound? We don’t have it (yet....), but there is a reason.
The (obvious) analog of the van der Waerden conjecture fails in higher dimension

It is an easy consequence of the marriage theorem that if in a 0/1 matrix $A$, all row sums and all column sums equal $k \geq 1$, then $\text{per } A > 0$. 
The (obvious) analog of the van der Waerden conjecture fails in higher dimension

It is an easy consequence of the marriage theorem that if in a $0/1$ matrix $A$, all row sums and all column sums equal $k \geq 1$, then $\text{per } A > 0$. The analogous statement is no longer true in higher dimensions.
The (obvious) analog of the van der Waerden conjecture fails in higher dimension

It is an easy consequence of the marriage theorem that if in a 0/1 matrix $A$, all row sums and all column sums equal $k \geq 1$, then $\text{per } A > 0$. The analogous statement is no longer true in higher dimensions. Here is an example of a $4 \times 4 \times 4$ array with two zeros and two ones in every line which contains no 2-permutation.
**An example**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>∗</td>
<td>0</td>
<td>∗</td>
<td>0</td>
<td></td>
<td>0</td>
<td>∗</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>∗</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>∗</td>
<td>∗</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>∗</td>
<td>∗</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
In a recent breakthrough P. Keevash solved a 160-years old problem and showed the existence of combinatorial designs.
In a recent breakthrough P. Keevash solved a 160-years old problem and showed the existence of combinatorial designs. His work yields as well the tight lower bound on $|\mathcal{L}_n|$. 
In a recent breakthrough P. Keevash solved a 160-years old problem and showed the existence of combinatorial designs. His work yields as well the tight lower bound on $|\mathcal{L}_n|$. It is conceivable that an appropriate adaptation of his method will prove the tight lower bound in all dimensions.
The general scheme: We consider a Latin square (= a 2-dimensional permutation) $A$, layer by layer.
The general scheme: We consider a Latin square (a 2-dimensional permutation) $A$, layer by layer. Namely, $A$ is an $n \times n \times n$ array of 0/1 where every line has a single 1 entry.
The general scheme: We consider a Latin square (= a 2-dimensional permutation) \( A \), layer by layer. Namely, \( A \) is an \( n \times n \times n \) array of 0/1 where every line has a single 1 entry. Note that every layer in \( A \) is a permutation matrix.
The general scheme: We consider a Latin square (= a 2-dimensional permutation) $A$, layer by layer. Namely, $A$ is an $n \times n \times n$ array of 0/1 where every line has a single 1 entry. Note that every layer in $A$ is a permutation matrix. Given several layers in $A$, how many permutation matrices can play the role of the next layer?
How many choices for the next layer?

Let $B$ be a 0/1 matrix where $b_{ij} = 1$ iff in all previous layers the $ij$ entry is zero.
Let $B$ be a 0/1 matrix where $b_{ij} = 1$ iff in all previous layers the $ij$ entry is zero. The set of all possible next layers coincides with the collection of generalized diagonals in $B$. Therefore, there are exactly $\text{per}_B$ possibilities for the next layer.
How many choices for the next layer?

Let $B$ be a 0/1 matrix where $b_{ij} = 1$ iff in all previous layers the $ij$ entry is zero. The set of all possible next layers coincides with the collection of generalized diagonals in $B$. Therefore, there are exactly $\text{per}B$ possibilities for the next layer.
How many choices for the next layer?

To estimate the number of Latin squares we bound at each step the number of possibilities for the next layer (=\(\text{per}B\)) from above and from below using Minc-Brégman and van der Waerden, respectively.
Back to basics - Reproving Brégman’s theorem

One of the insights gained about Brégman’s theorem is that it is useful to interpret it using the notion of entropy.
One of the insights gained about Brégman’s theorem is that it is useful to interpret it using the notion of entropy.
So let us review the basics of this method.
A quick recap of entropy

If $X$ is a discrete random variable, taking the $i$-th value in its domain with probability $p_i$; then its entropy is defined as

$$H(X) := - \sum p_i \log p_i$$
A quick recap of entropy

If $X$ is a discrete random variable, taking the $i$-th value in its domain with probability $p_i$ then its entropy is defined as

$$H(X) := - \sum p_i \log p_i$$

In particular if the range of $X$ has cardinality $N$, then $H(X) \leq \log N$ with equality iff $X$ is distributed uniformly.
A quick recap of entropy

If $X$ is a discrete random variable, taking the $i$-th value in its domain with probability $p_i$ then its entropy is defined as

$$H(X) := - \sum p_i \log p_i$$

In particular if the range of $X$ has cardinality $N$, then $H(X) \leq \log N$ with equality iff $X$ is distributed uniformly.

All logarithms here are to base $e$. This is not the convention when it comes to entropy, but it will make things more convenient for us.
A quick recap of entropy (contd.)

If $X$ and $Y$ are two discrete random variables, then the conditional entropy

$$H(X|Y) := \sum_y Pr(Y = y)H(X|Y = y).$$
If $X$ and $Y$ are two discrete random variables, then the **conditional entropy**

$$H(X|Y) := \sum_y Pr(Y = y)H(X|Y = y).$$

The **chain rule** is one of the fundamental properties of entropy:
If $X$ and $Y$ are two discrete random variables, then the conditional entropy

$$H(X|Y) := \sum_y Pr(Y = y)H(X|Y = y).$$

The chain rule is one of the fundamental properties of entropy: If $X_1, \ldots, X_n$ are discrete random variables defined on the same probability space, then
A quick recap of entropy (contd.)

If $X$ and $Y$ are two discrete random variables, then the conditional entropy

$$H(X|Y) := \sum_y \Pr(Y = y) H(X|Y = y).$$

The chain rule is one of the fundamental properties of entropy: If $X_1, \ldots, X_n$ are discrete random variables defined on the same probability space, then

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots$$
Let us fix an $n \times n$ 0/1 matrix $A$ in which there are exactly $r_i$ 1-entries in the $i$-th row for every $i$. The number of generalized diagonals contained in $A$ is exactly $\text{per}_A$. Let $X$ be a random variable that takes values which are these generalized diagonals, with uniform distribution. Clearly, $H(X) = \log(\text{per}_A)$. Therefore, an upper bound on $H(X)$ yields an upper bound on $\text{per}_A$, which is what we want.
Let us fix an $n \times n$ 0/1 matrix $A$ in which there are exactly $r_i$ 1-entries in the $i$-th row for every $i$. The number of generalized diagonals contained in $A$ is exactly $\text{per } A$. 

Let $X$ be a random variable that takes values which are these generalized diagonals, with uniform distribution. Clearly, $H(X) = \log(\text{per } A)$. Therefore, an upper bound on $H(X)$ yields an upper bound on $\text{per } A$, which is what we want.
Let us fix an $n \times n$ 0/1 matrix $A$ in which there are exactly $r_i$ 1-entries in the $i$-th row for every $i$. The number of generalized diagonals contained in $A$ is exactly $\text{per } A$. Let $X$ be a random variable that takes values which are these generalized diagonals, with uniform distribution. Clearly,
Proving Brégman’s theorem using entropy

Let us fix an $n \times n$ 0/1 matrix $A$ in which there are exactly $r_i$ 1-entries in the $i$-th row for every $i$. The number of generalized diagonals contained in $A$ is exactly $\text{per } A$. Let $X$ be a random variable that takes values which are these generalized diagonals, with uniform distribution. Clearly,

$$H(X) = \log(\text{per } A).$$
Let us fix an $n \times n$ 0/1 matrix $A$ in which there are exactly $r_i$ 1-entries in the $i$-th row for every $i$. The number of generalized diagonals contained in $A$ is exactly $\text{per } A$. Let $X$ be a random variable that takes values which are these generalized diagonals, with uniform distribution. Clearly,

$$H(X) = \log(\text{per } A).$$

Therefore, an upper bound on $H(X)$ yields an upper bound on $\text{per } A$, which is what we want.
We next express $X = (X_1, \ldots, X_n)$, where $X_i$ is the index of the single 1-entry that is selected by the generalized diagonal $X$ at the $i$-th row.
We next express \( X = (X_1, \ldots, X_n) \), where \( X_i \) is the index of the single 1-entry that is selected by the generalized diagonal \( X \) at the \( i \)-th row. How should we interpret the relation

\[
H(X) = H(X_1) + H(X_2 | X_1) + H(X_3 | X_1, X_2) + \ldots
\]
We next express $X = (X_1, \ldots, X_n)$, where $X_i$ is the index of the single 1-entry that is selected by the generalized diagonal $X$ at the $i$-th row. How should we interpret the relation

$$H(X) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \ldots$$

In particular, what can we say about $H(X_i|X_1, \ldots, X_{i-1})$?
Let us fix a specific value for $X$, i.e., some generalized diagonal in $A$. 
Let us fix a specific value for $X$, i.e., some generalized diagonal in $A$. There are $r_i$ 1’s in the $i$-th row.
Let us fix a specific value for $X$, i.e., some generalized diagonal in $A$. There are $r_i$ 1’s in the $i$-th row. Since $X$ takes on a generalized diagonal, we know that there are $r_i$ indices $j$ for which the index $X_j$ coincides with one of these $r_i$ positions.
Let us fix a specific value for $X$, i.e., some generalized diagonal in $A$. There are $r_i$ 1’s in the $i$-th row. Since $X$ takes on a generalized diagonal, we know that there are $r_i$ indices $j$ for which the index $X_j$ coincides with one of these $r_i$ positions. If such a $j$ is smaller than $i$, we say that $j$ is shading the $i$-th row.
Clearly $X_i$ can take on only unshaded values.
Clearly $X_i$ can take on only unshaded values. If $N_i$ is the (random) number of 1-entries in the $i$-th row that remain unshaded when the $i$-th row is reached, then clearly

$$H(X_i|X_1, \ldots, X_{i-1}) \leq \log N_i$$
Clearly $X_i$ can take on only unshaded values. If $N_i$ is the (random) number of 1-entries in the $i$-th row that remain unshaded when the $i$-th row is reached, then clearly

$$H(X_i|X_1, \ldots, X_{i-1}) \leq \log N_i$$

since the conditioned random variable $(X_i|X_1, \ldots, X_{i-1})$ can take at most $N_i$ values.
Clearly $X_i$ can take on only unshaded values. If $N_i$ is the (random) number of 1-entries in the $i$-th row that remain unshaded when the $i$-th row is reached, then clearly

$$H(X_i|X_1, \ldots, X_{i-1}) \leq \log N_i$$

since the conditioned random variable $(X_i|X_1, \ldots, X_{i-1})$ can take at most $N_i$ values.

Very nice. The trouble is that we know very little about the random variable $N_i$. 
The way around this difficulty is not to sum the terms $H(X_i|X_1,\ldots,X_{i-1})$ in the normal order, but rather introduce $\sigma$, a random ordering of the rows.
A good trick

The way around this difficulty is not to sum the terms $H(X_i|X_1,\ldots,X_{i-1})$ in the normal order, but rather introduce $\sigma$, a random ordering of the rows. We add the corresponding terms in the order $\sigma$ and finally we average over the random choice of $\sigma$. What can we say about the expectation of $\log N_{\sigma_i}$?
A good trick

The way around this difficulty is not to sum the terms $H(X_i|X_1,\ldots,X_{i-1})$ in the normal order, but rather introduce $\sigma$, a random ordering of the rows. We add the corresponding terms in the order $\sigma$ and finally we average over the random choice of $\sigma$.

What can we say about the expectation of $\log N_i^\sigma$?
This can be restated as follows: You are expecting $r$ visitors who arrive at random, independently chosen times.
This can be restated as follows: You are expecting $r$ visitors who arrive at random, independently chosen times. One of the visitors is your guest of honor and you are interested in this guest’s (random) arrival rank among the $r$ visitors.
Clearly this rank $N$ is uniformly distributed over $1, \ldots, r$. 
Clearly this rank \( N \) is uniformly distributed over \( 1, \ldots, r \). In particular, the expectation

\[
\mathbb{E}(\log N) = \frac{1}{r} \sum_{j=1,\ldots,r} \log j = \frac{\log r!}{r} = \log(r!)^{1/r}.
\]
Clearly this rank $N$ is uniformly distributed over $1, \ldots, r$. In particular, the expectation

$$
\mathbb{E}(\log N) = \frac{1}{r} \sum_{j=1, \ldots, r} \log j = \frac{\log r!}{r} = \log(r!)^{1/r}.
$$

By summing over all rows, the Brégman bound is established

$$
H(X) = \log(\text{per } A) \leq \sum_i \log(r_i!)^{1/r_i}.
$$
If $A$ is an $n^{d+1} = n \times n \times \ldots \times n$ array of 0/1 we define $\text{per}_d(A)$ to be the number of $d$-permutations that are included in $A$. Let’s consider all lines in $A$ in the same direction, say lines of the form $l_i = (i_1, \ldots, i_d, \ast)$. Let $r_i$ be the number of 1’s in the line $l_i$. 
Theorem

Let $A$ be an $[n]^{d+1}$ array of 0/1, with $r_i$ 1’s in the line $l_i$. Then

$$per_d(A) \leq \prod_i \exp(f(d, r_i)).$$
Theorem
Let $A$ be an $[n]^{d+1}$ array of 0/1, with $r_i$ 1's in the line $l_i$. Then

$$per_d(A) \leq \prod_i \exp(f(d, r_i)).$$

$f(d, r)$ is defined via $f(0, r) = \log r$, and

$$f(d, r) = \frac{1}{r} \sum_{k=1,\ldots,r} f(d - 1, k).$$
Theorem

Let $A$ be an $[n]^{d+1}$ array of 0/1, with $r_i$ 1's in the line $l_i$. Then

$$per_d(A) \leq \prod_i \exp(f(d, r_i)).$$

$f(d, r)$ is defined via $f(0, r) = \log r$, and

$$f(d, r) = \frac{1}{r} \sum_{k=1,\ldots,r} f(d - 1, k).$$

It can be shown that

$$f(d, r) = \log r - d + O_d(\frac{\log^d r}{r})$$
Recall that a real nonnegative matrix $A$ is called **doubly stochastic** if all row and column sums in $A$ equal 1.
Recall that a real nonnegative matrix $A$ is called **doubly stochastic** if all row and column sums in $A$ equal 1. The set $\Omega_n$ of $n \times n$ doubly-stochastic matrices is clearly a polytope. The vertex set of $\Omega_n$ has a very pleasing description:
Recall that a real nonnegative matrix $A$ is called doubly stochastic if all row and column sums in $A$ equal 1. The set $\Omega_n$ of $n \times n$ doubly-stochastic matrices is clearly a polytope. The vertex set of $\Omega_n$ has a very pleasing description:

**Theorem (Birkhoff von-Neumann ’46)**

The vertex set of $\Omega_n$ is the set of all $n \times n$ permutation matrices.
An $n \times n \times n$ array $A$ of nonnegative reals is called a tri-stochastic array if the sum of entries in every line in $A$ equals 1.
An $n \times n \times n$ array $A$ of nonnegative reals is called a tri-stochastic array if the sum of entries in every line in $A$ equals 1. It is not hard to see that...
An $n \times n \times n$ array $A$ of nonnegative reals is called a tri-stochastic array if the sum of entries in every line in $A$ equals $1$. It is not hard to see that

- The set of all order-$n$ tri-stochastic arrays is a polytope $\Pi_n$. 
Higher-dimensional Birkhoff-von Neumann?

An $n \times n \times n$ array $A$ of nonnegative reals is called a\[\text{tri-stochastic array}\] if the sum of entries in every line in $A$ equals 1. It is not hard to see that

- The set of all order-$n$ tri-stochastic arrays is a polytope $\Pi_n$.
- Every Latin square of order $n$ is a vertex of $\Pi_n$. 
Higher-dimensional Birkhoff-von Neumann?

An $n \times n \times n$ array $A$ of nonnegative reals is called a tri-stochastic array if the sum of entries in every line in $A$ equals 1. It is not hard to see that

- The set of all order-$n$ tri-stochastic arrays is a polytope $\Pi_n$.
- Every Latin square of order $n$ is a vertex of $\Pi_n$.

Is the converse also true?
Theorem (NL and Zur Luria ’14)

The polytope $\Pi_n$ has at least $|\mathcal{L}_n|^{3/2}$ vertices.
No two-dimensional Birkhoff-von Neumann

Theorem (NL and Zur Luria ’14)

The polytope $\Pi_n$ has at least $|\mathcal{L}_n|^{3/2}$ vertices.
(So Latin squares are just a negligible minority of all vertices).
It is not hard to see that $\Pi_n$ has at most $|\mathcal{L}_n|^3$ vertices.
No two-dimensional Birkhoff-von Neumann - some comments

- It is not hard to see that $\Pi_n$ has at most $|\mathcal{L}_n|^3$ vertices.
- Presumably something similar holds in higher dimensions as well, but we do not know how to prove it.
It is not hard to see that $\Pi_n$ has at most $|L_n|^3$ vertices.

Presumably something similar holds in higher dimensions as well, but we do not know how to prove it.

The basic idea of the proof starts at the counterexample to high-dimensional van der Waerden that we saw before.
This is a vertex for tri-stochastic arrays

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>0</td>
<td>*</td>
<td>0</td>
<td>0</td>
<td>*</td>
<td>0</td>
<td>*</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>*</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
As every(?) undergraduate knows:

Theorem (Erdős-Szekeres '35)

Every permutation in $S_n$ has a monotone subsequence of length $\geq \sqrt{n}$. The bound is tight.

Several years later Ulam asked:

Problem (Ulam)

Let $L_n(\pi)$ be the length of the longest increasing subsequence in the permutation $\pi \in S_n$.

How is $L_n$ distributed when $\pi$ is drawn uniformly from $S_n$?
As every (?) undergraduate knows:

**Theorem (Erdős-Szekeres ’35)**

Every permutation in $S_n$ has a monotone subsequence of length $\geq \sqrt{n}$. The bound is tight.
As every(?) undergraduate knows:

**Theorem (Erdős-Szekeres ’35)**

Every permutation in $S_n$ has a monotone subsequence of length $\geq \sqrt{n}$. The bound is tight.

Several years later Ulam asked:

**Problem (Ulam)**

Let $L_n(\pi)$ be the length of the longest increasing subsequence in the permutation $\pi \in S_n$. 
As every(?) undergraduate knows:

**Theorem (Erdős-Szekeres ’35)**

*Every permutation in $S_n$ has a monotone subsequence of length $\geq \sqrt{n}$. The bound is tight.*

Several years later Ulam asked:

**Problem (Ulam)**

*Let $L_n(\pi)$ be the length of the longest increasing subsequence in the permutation $\pi \in S_n$. How is $L_n$ distributed when $\pi$ is drawn uniformly from $S_n$?*
It is not hard to see that with high probability

\[ c_1 \sqrt{n} > L_n(\pi) > c_2 \sqrt{n} \]

for some absolute \( c_1 > c_2 > 0 \).
It is not hard to see that with high probability

\[ c_1 \sqrt{n} > L_n(\pi) > c_2 \sqrt{n} \]

for some absolute \( c_1 > c_2 > 0 \).

A lot of beautiful work was done over the years. Here are two highlights.
It is not hard to see that with high probability
\[ c_1 \sqrt{n} > L_n(\pi) > c_2 \sqrt{n} \]
for some absolute \( c_1 > c_2 > 0 \).
A lot of beautiful work was done over the years. Here are two highlights.

**Theorem (Logan and Shepp ’77, Vershik and Kerov ’77)**

\[ \lim_{n \to \infty} \frac{\mathbb{E} L_n}{\sqrt{n}} = 2. \]
Theorem (Baik, Deift, Johansson '99)

\[ \frac{L_n - 2\sqrt{n}}{n^{1/6}} \overset{d}{\to} \text{Tracy-Widom distribution.} \]
Theorem (NL and Michael Simkin)

- Every d-dimensional length-n permutation has a monotone subsequence of length $\Omega_d(\sqrt{n})$. The bound is tight (but we still do not know the $d$-dependent factor).
Theorem (NL and Michael Simkin)

- Every $d$-dimensional length-$n$ permutation has a monotone subsequence of length $\Omega_d(\sqrt{n})$. The bound is tight (but we still do not know the $d$-dependent factor).
- The longest monotone subsequence in almost every permutation has length $\Theta(n^{\frac{d}{d+1}})$. 
Theorem (NL and Michael Simkin)

- Every $d$-dimensional length-$n$ permutation has a monotone subsequence of length $\Omega_d(\sqrt{n})$. The bound is tight (but we still do not know the $d$-dependent factor).

- The longest monotone subsequence in almost every permutation has length $\Theta(n^{\frac{d}{d+1}})$. We have no further details about this distribution.
Here are two examples of this important concept
Discrepancy in geometry

Here are two examples of this important concept

Theorem (van Aardenne-Ehrenfest ’45, Schmidt ’75)

▶ There is a set of $N$ points $X \subset [0, 1]^2$, s.t.
$|X \cap R| - N \cdot \text{area}(R)| \leq O(\log N)$ for every axis-parallel rectangle $R \subset [0, 1]^2$. 

On the other hand, for every set of $N$ points $X \subset [0, 1]^2$ there is an axis-parallel rectangle $R$ for which $|X \cap R| - N \cdot \text{area}(R)| \geq \Omega(\log N)$. 

Nati Linial
High-dimensional permutations
Discrepancy in geometry

Here are two examples of this important concept

Theorem (van Aardenne-Ehrenfest ’45, Schmidt ’75)

▶ There is a set of $N$ points $X \subset [0, 1]^2$, s.t.
\[| |X \cap R| - N \cdot \text{area}(R)| \leq O(\log N) \text{ for every axis-parallel rectangle } R \subseteq [0, 1]^2.\]

▶ On the other hand, for every set of $N$ points $X \subset [0, 1]^2$ there is an axis-parallel rectangle $R$ for which $| |X \cap R| - N \cdot \text{area}(R)| \geq \Omega(\log N)$. 

Nati Linial
High-dimensional permutations
Discrepancy in graph theory

Theorem (Alon Chung, ’88 ”The expander mixing lemma”)

Let $G = (V, E)$ be an $n$-vertex $d$-regular graph, and let $\lambda$ be the largest absolute value of a nontrivial eigenvalue of $G$’s adjacency matrix. Then for every $A, B \subset V$,

$$\left| \left| e(A, B) - dn|A||B| \right| \right| \leq \lambda \sqrt{|A||B|}.$$
Theorem (Alon Chung, ’88 ”The expander mixing lemma”)

Let $G = (V, E)$ be an $n$-vertex $d$-regular graph, and let $\lambda$ be the largest absolute value of a nontrivial eigenvalue of $G$’s adjacency matrix. Then for every $A, B \subset V$,

$$\left| e(A, B) - \frac{d}{n} |A||B| \right| \leq \lambda \sqrt{|A||B|}.$$
Discrepancy in high-dimensional permutations

Conjecture (NL and Zur Luria ’15)
Discrepancy in high-dimensional permutations

**Conjecture (NL and Zur Luria ’15)**

There exist order-$N$ Latin squares such that for every $A, B, C \subseteq [N]$ there holds

$$\left| L \cap (A \times B \times C) \right| - \frac{|A||B||C|}{N} \leq O(\sqrt{|A||B||C|}).$$

Moreover, this holds for almost every Latin square.
Conjecture (NL and Zur Luria ’15)

There exist order-$N$ Latin squares such that for every $A, B, C \subseteq [N]$ there holds

$$\left| |L \cap (A \times B \times C)| - \frac{|A||B||C|}{N} \right| \leq O(\sqrt{|A||B||C|}).$$

Moreover, this holds for almost every Latin square.
Discrepancy in high-dimensional permutations

It is interesting to restrict this conjecture to the case of empty boxes, i.e., deal with the case where \( L \cap (A \times B \times C) = \emptyset \). The conjecture reads

**Conjecture**

There exist order-N Latin squares in which every empty box has volume \( O(N^2) \).

Moreover, this holds for almost every Latin square.

Note: Every Latin square has an empty box of volume \( \Omega(N^2) \).
Discrepancy in high-dimensional permutations

It is interesting to restrict this conjecture to the case of empty boxes, i.e., deal with the case where \( L \cap (A \times B \times C) = \emptyset \). The conjecture reads

**Conjecture**

*There exist order-\( N \) Latin squares in which every empty box has volume \( O(N^2) \).*
Discrepancy in high-dimensional permutations

It is interesting to restrict this conjecture to the case of empty boxes, i.e., deal with the case where \( L \cap (A \times B \times C) = \emptyset \). The conjecture reads

**Conjecture**

There exist order-\( N \) Latin squares in which every empty box has volume \( O(N^2) \). Moreover, this holds for almost every Latin square.
It is interesting to restrict this conjecture to the case of empty boxes, i.e., deal with the case where $L \cap (A \times B \times C) = \emptyset$. The conjecture reads

**Conjecture**

*There exist order-$N$ Latin squares in which every empty box has volume $O(N^2)$.* Moreover, this holds for *almost every* Latin square.

**Note**

*Every* Latin square has an empty box of volume $\Omega(N^2)$. 
Theorem (NL and Zur Luria)

- There exist order-N Latin squares in which every empty box has volume $O(N^2)$. 
Discrepancy in high-dimensional permutations

Theorem (NL and Zur Luria)

- There exist order-$N$ Latin squares in which every empty box has volume $O(N^2)$.
- In almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$. 

Nati Linial
High-dimensional permutations
Discrepancy in high-dimensional permutations

Can we explicitly construct such Latin squares?
Can we explicitly construct such Latin squares? The multiplication table of a finite group is a Latin square.
Can we explicitly construct such Latin squares? The multiplication table of a finite group is a Latin square. However,

**Theorem (Kedlaya ’95)**

*The Latin square of every order-$N$ group contains an empty box of volume $\geq \Omega(N^{2.357\ldots})$ (this exponent is $\frac{33}{14}$).*

Gowers has examples of groups where all empty boxes have volume $N^{8/3}$. 
Discrepancy in high-dimensional permutations

Theorem (NL and Zur Luria)

- There exist order-$N$ Latin squares in which every empty box has volume $O(N^2)$.
- In almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$. 
A word about the proof

We construct Latin squares with no large empty boxes using Keevash’s construction of Steiner systems.

To every Steiner triple system $X$ we associate a Latin square $L$ where $\{i, j, k\} \in X$ implies $L(i, j, k) = \cdots = L(k, j, i) = 1$ (six terms).

Also, for all $i$, let $L(i, i, i) = 1$.

Keevash’s method starts with a random greedy choice of triples. His main argument shows that whp this triple system can be completed to an STS. We show that this initial phase suffices to hit all boxes of volume above $Cn^{2/3}$.
A word about the proof

We construct Latin squares with no large empty boxes using Keevash’s construction of Steiner systems.

- To every Steiner triple system $X$ we associate a Latin square $L$ where $\{i, j, k\} \in X$ implies $L(i, j, k) = \ldots = L(k, j, i) = 1$ (six terms). Also, for all $i$, let $L(i, i, i) = 1$. 

Keevash’s method starts with a random greedy choice of triples. His main argument shows that whp this triple system can be completed to an STS. We show that this initial phase suffices to hit all boxes of volume above $Cn^2$. 

Nati Linial
High-dimensional permutations
A word about the proof

We construct Latin squares with no large empty boxes using Keevash’s construction of Steiner systems.

- To every Steiner triple system $X$ we associate a Latin square $L$ where $\{i, j, k\} \in X$ implies $L(i, j, k) = \ldots = L(k, j, i) = 1$ (six terms). Also, for all $i$, let $L(i, i, i) = 1$.

- Keevash’s method starts with a random greedy choice of triples. His main argument shows that whp this triple system can be completed to an STS. We show that this initial phase suffices to hit all boxes of volume above $Cn^2$. 

Nati Linial  High-dimensional permutations
A word about the proof

Why is it that in almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$?
A word about the proof

Why is it that in almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$? Fix a box $A \times B \times C$ and note that the probability that it is empty in a random Latin square is:
A word about the proof

Why is it that in almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$? Fix a box $A \times B \times C$ and note that the probability that it is empty in a random Latin square is:

\[
\per_d X
\begin{array}{c}
\frac{}{L_n}
\end{array}
\]

where $X$ is the $n \times n \times n$ array whose entries are zero in $A \times B \times C$ and one otherwise.
A word about the proof

Why is it that in almost every order-$N$ Latin squares all empty boxes have volume $O(N^2 \log^2 N)$? Fix a box $A \times B \times C$ and note that the probability that it is empty in a random Latin square is:

$$\frac{\text{per}_d X}{|L_n|}$$

where $X$ is the $n \times n \times n$ array whose entries are zero in $A \times B \times C$ and one otherwise. We apply our Brégman-type upper bound on $\text{per}_d X$ and derive the conclusion fairly straightforwardly.
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
- Do Birkhoff von-Neumann in higher dimensions.

What we just did was a poor man's substitute to randomness in this domain.

Nati Linial

High-dimensional permutations
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
- Do Birkhoff von-Neumann in higher dimensions.
- Settle the discrepancy problem.
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
- Do Birkhoff von-Neumann in higher dimensions.
- Settle the discrepancy problem.
- Find explicit constructions of low-discrepancy high-dimensional permutations.
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
- Do Birkhoff von-Neumann in higher dimensions.
- Settle the discrepancy problem.
- Find explicit constructions of low-discrepancy high-dimensional permutations.
- Find how to sample high-dimensional permutations and determine their typical behavior.
Many questions remain open....

- Find the asymptotic number of high-dimensional permutations.
- Do Birkhoff von-Neumann in higher dimensions.
- Settle the discrepancy problem.
- Find explicit constructions of low-discrepancy high-dimensional permutations.
- Find how to sample high-dimensional permutations and determine their typical behavior. What we just did was a poor man’s substitute to randomness in this domain.
Many happy returns,
Noga