t-Scrambling Permutations

Anant Godbole   Samantha Pinella   Yan Zhuang

East Tennessee State University

University of Edinburgh

Brandeis University

Permutation Patterns in Paris, July 4 2013

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.
- Other titles we have used include:

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.
- Other titles we have used include:
- Vapnik-Červonenkis (VC) dimension of random permutations (Statistics, learning theory);

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.
- Other titles we have used include:
- Vapnik-Červonenkis (VC) dimension of random permutations (Statistics, learning theory);
- Shattering n permutations in any t positions (also Statistics and learning theory);

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.
- Other titles we have used include:
- Vapnik-Červonenkis (VC) dimension of random permutations (Statistics, learning theory);
- Shattering n permutations in any t positions (also Statistics and learning theory);
- t-covering arrays (combinatorial design theory);

Collaborators and Titles

- Three recent collaborators who have proved some new results along these lines are Stephanie deGraaf, Zoe Koch, and Kathleen Lan.
- Other titles we have used include:
- Vapnik-Červonenkis (VC) dimension of random permutations (Statistics, learning theory);
- Shattering n permutations in any t positions (also Statistics and learning theory);
- t-covering arrays (combinatorial design theory);
- The phrase "scrambling" has been used before in the permutations context, as pointed out by a referee of our submitted paper.

VC Dimension

DEFINITION: A class $\mathcal{F}$ of subsets of a set X is said to **shatter** a subset $A = \{a_1, \ldots, a_t\} \subseteq X$ if

$$\forall S \subseteq A, \exists F \in \mathcal{F} \text{ such that } A \cap F = S,$$

or equivalently, if

$$|\{A \cap F\} : F \in \mathcal{F}| = 2^t.$$

DEFINITION: The **VC dimension** of $\mathcal{F}$, $\mathrm{VC}(\mathcal{F})$, is the cardinality of the smallest subset not shattered by $\mathcal{F}$. If all subsets of finite size are shattered by $\mathcal{F}$, then the $\mathrm{VC}(\mathcal{F}) = \infty$.

Examples

- $X = \mathbb{R}, \mathcal{F} = \{(-\infty, t]; t \in \mathbb{R}\}$; VC($\mathcal{F}$)=2;

Examples

- $X = \mathbb{R}, \mathcal{F} = \{(-\infty, t]; t \in \mathbb{R}\}$; VC($\mathcal{F}$)=2;
- $X = \mathbb{R}^2, \mathcal{F} = \{\text{all convex sets}\}$; VC($\mathcal{F}$)=3.

Examples

- $X = \mathbb{R}, \mathcal{F} = \{(-\infty, t]; t \in \mathbb{R}\}$; VC($\mathcal{F}$)=2;
- $X = \mathbb{R}^2, \mathcal{F} = \{\text{all convex sets}\}$; VC($\mathcal{F}$)=3.
- Many authors define the VC dimension to be the size of the largest shattered set, in this case our values would be 1 and ∞ respectively.

Examples

- $X = \mathbb{R}, \mathcal{F} = \{(-\infty, t]; t \in \mathbb{R}\}$; VC($\mathcal{F}$)=2;
- $X = \mathbb{R}^2, \mathcal{F} = \{\text{all convex sets}\}$; VC($\mathcal{F}$)=3.
- Many authors define the VC dimension to be the size of the largest shattered set, in this case our values would be 1 and ∞ respectively.
- Often, the two numbers are off by one, as in Example 1.

Covering Arrays

- A $k \times n$ array with entries from the alphabet $\{0, 1, \ldots, q-1\}$ is said to be a (t, q, n, k)-covering array, or briefly a t-covering array, if for each of the $\binom{n}{t}$ choices of t columns, each of the q^t q-ary words of length t can be found at least once among the rows of the selected columns.

Covering Arrays

- A $k \times n$ array with entries from the alphabet $\{0, 1, \ldots, q-1\}$ is said to be a (t, q, n, k)-covering array, or briefly a t-covering array, if for each of the $\binom{n}{t}$ choices of t columns, each of the q^t q-ary words of length t can be found at least once among the rows of the selected columns.
- If $q = 2$, we can interpret any row as the characteristic vector of a subset of $[n]$ – by making a correspondence between the positions where the row has ones, and the set of those positions.

Covering Arrays

- A $k \times n$ array with entries from the alphabet $\{0, 1, \ldots, q-1\}$ is said to be a (t, q, n, k)-covering array, or briefly a t-covering array, if for each of the $\binom{n}{t}$ choices of t columns, each of the q^t q-ary words of length t can be found at least once among the rows of the selected columns.
- If $q = 2$, we can interpret any row as the characteristic vector of a subset of $[n]$ – by making a correspondence between the positions where the row has ones, and the set of those positions.
- We thus have the following alternative formulation of covering arrays: A family $\mathcal{F}$ of subsets of $[n]$ is a t-covering array if for each $\{a_1, \ldots, a_t\} \subset [n]$,

$$\left|\{\{a_1, \ldots, a_t\} \cap F\} : F \in \mathcal{F}\right| = 2^t.$$

Connections, More Terminology, Scrambling

- Thus if an array is binary t-covering, then its smallest unshattered set must be of size $\geq t+1$ and thus $\mathrm{VC}(\mathcal{F}) \geq t+1$.

Connections, More Terminology, Scrambling

- Thus if an array is binary t-covering, then its smallest unshattered set must be of size $\geq t+1$ and thus VC($\mathcal{F}$) $\geq t+1$.
- If $q \geq 3$ we do not have an exact analogy with shattering sets and dimension, but we can make a parallel with shattering multisets.

Connections, More Terminology, Scrambling

- Thus if an array is binary t-covering, then its smallest unshattered set must be of size $\geq t+1$ and thus $\mathrm{VC}(\mathcal{F}) \geq t+1$.
- If $q \geq 3$ we do not have an exact analogy with shattering sets and dimension, but we can make a parallel with shattering multisets.
- What about permutations, which we focus on in this talk?

Connections, More Terminology, Scrambling

- Thus if an array is binary t-covering, then its smallest unshattered set must be of size $\geq t+1$ and thus $\mathrm{VC}(\mathcal{F}) \geq t+1$.
- If $q \geq 3$ we do not have an exact analogy with shattering sets and dimension, but we can make a parallel with shattering multisets.
- What about permutations, which we focus on in this talk?
- We say that a $k \times n$ rectangular array of k permutations on $[n]$ is t-scrambling if for each set of t columns, each of the $t!$ permutations on $[t]$ may be found in an order isomorphic fashion at least once among the rows of the selected columns.

Connections, More Terminology, Scrambling

- Thus if an array is binary t-covering, then its smallest unshattered set must be of size $\geq t+1$ and thus $\mathrm{VC}(\mathcal{F}) \geq t+1$.
- If $q \geq 3$ we do not have an exact analogy with shattering sets and dimension, but we can make a parallel with shattering multisets.
- What about permutations, which we focus on in this talk?
- We say that a $k \times n$ rectangular array of k permutations on $[n]$ is t-scrambling if for each set of t columns, each of the $t!$ permutations on $[t]$ may be found in an order isomorphic fashion at least once among the rows of the selected columns.
- There are clear parallels with VC, shattering, t-covering arrays etc, but the original terminology is "scrambling".

Dartmouth PP Conference

- At PP2010 in Dartmouth, Cibulka and Kynčl investigated VC dimension of permutation arrays also;

Dartmouth PP Conference

- At PP2010 in Dartmouth, Cibulka and Kynčl investigated VC dimension of permutation arrays also;
- They used the second, i.e., "largest shattered set" definition to assemble arrays of permutations with the following property:

Dartmouth PP Conference

- At PP2010 in Dartmouth, Cibulka and Kynčl investigated VC dimension of permutation arrays also;
- They used the second, i.e., "largest shattered set" definition to assemble arrays of permutations with the following property:
- If we write each of the $\sim 4^n$ 123-avoiders in an array, there would be no 123 in any set of 3 columns, no matter what row we choose;

Dartmouth PP Conference

- At PP2010 in Dartmouth, Cibulka and Kynčl investigated VC dimension of permutation arrays also;
- They used the second, i.e., "largest shattered set" definition to assemble arrays of permutations with the following property:
- If we write each of the $\sim 4^n$ 123-avoiders in an array, there would be no 123 in any set of 3 columns, no matter what row we choose;
- How much larger can such an array be?

Dartmouth PP Conference

- At PP2010 in Dartmouth, Cibulka and Kynčl investigated VC dimension of permutation arrays also;
- They used the second, i.e., "largest shattered set" definition to assemble arrays of permutations with the following property:
- If we write each of the $\sim 4^n$ 123-avoiders in an array, there would be no 123 in any set of 3 columns, no matter what row we choose;
- How much larger can such an array be?
- They provided superexponential bounds on the size of the extremal such array.

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then
- $m(n, t) \leq \frac{t \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$ (Spencer, 1972).

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then
- $m(n, t) \leq \frac{t \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$ (Spencer, 1972).
- For $t = 3$, however, $m(n, 3) \leq 2 \lg n$ (Tarui, 2008).

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then
- $m(n, t) \leq \frac{t \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$ (Spencer, 1972).
- For $t = 3$, however, $m(n, 3) \leq 2 \lg n$ (Tarui, 2008).
- Lower bounds of Füredi (1996) were improved by Radhakrishnan (2003).

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then
- $m(n, t) \leq \frac{t \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$ (Spencer, 1972).
- For $t = 3$, however, $m(n, 3) \leq 2 \lg n$ (Tarui, 2008).
- Lower bounds of Füredi (1996) were improved by Radhakrishnan (2003).
- We were able to improve the Spencer 1972 bound using the Lovász Local Lemma:

Known Results

- Let $k = m(t, n)$ be the size (i.e. number of rows) of the *smallest* array that is t-scrambling, i.e., all $t!$ perms are present in any set of t columns, i.e. arrays with VC dimension $\geq t + 1$. Then
- $m(n, t) \leq \frac{t \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$ (Spencer, 1972).
- For $t = 3$, however, $m(n, 3) \leq 2 \lg n$ (Tarui, 2008).
- Lower bounds of Füredi (1996) were improved by Radhakrishnan (2003).
- We were able to improve the Spencer 1972 bound using the Lovász Local Lemma:
- (deGraaf, G, Koch, Lan, 2013+): $m(n, t) \leq \frac{(t-1) \lg n}{\lg(t!/(t!-1))}$; $t \geq 4$. Furthermore, a $\log\log$ result holds:

Improved Upper Bounds (deGraaf, G, Koch, Lan, 2013+)

- If $m(n, t, \lambda)$ is the smallest number of rows so that for any choice of t columns, each of the $t!$ permutations are present at least λ times among the rows of the selected columns, then....

Improved Upper Bounds (deGraaf, G, Koch, Lan, 2013+)

- If $m(n, t, \lambda)$ is the smallest number of rows so that for any choice of t columns, each of the $t!$ permutations are present at least λ times among the rows of the selected columns, then....
- Reverting to Spencer's bound for a start, we have for $t \geq 3$,

Improved Upper Bounds (deGraaf, G, Koch, Lan, 2013+)

- If $m(n, t, \lambda)$ is the smallest number of rows so that for any choice of t columns, each of the $t!$ permutations are present at least λ times among the rows of the selected columns, then....
- Reverting to Spencer's bound for a start, we have for $t \geq 3$,
- (deGraaf, G, Koch, Lan, 2013+): $m(n, t, \lambda) \leq \frac{(t \lg n + (\lambda - 1) \lg\lg n)}{\lg(t!/(t! - 1))}$; $t \geq 3$.

Improved Upper Bounds (deGraaf, G, Koch, Lan, 2013+)

- If $m(n, t, \lambda)$ is the smallest number of rows so that for any choice of t columns, each of the $t!$ permutations are present at least λ times among the rows of the selected columns, then....
- Reverting to Spencer's bound for a start, we have for $t \geq 3$,
- (deGraaf, G, Koch, Lan, 2013+): $m(n, t, \lambda) \leq \frac{(t \lg n + (\lambda - 1) \lg \lg n)}{\lg(t!/(t!-1))}$; $t \geq 3$.
- We conjecture that the t can be replaced by $t - 1$.

Outline of Proof

- We will prove the $t \to t-1$ improvement using the Lovász lemma:

Outline of Proof

- We will prove the $t \to t-1$ improvement using the Lovász lemma:
- If E_i are events in some probability space with $\mathbb{P}(E_i) \le p \forall i$, and if

Outline of Proof

- We will prove the $t \to t-1$ improvement using the Lovász lemma:
- If E_i are events in some probability space with $\mathbb{P}(E_i) \leq p \forall i$, and if
- E_i is dependent on at most d other E_j, with

Outline of Proof

- We will prove the $t \to t-1$ improvement using the Lovász lemma:
- If E_i are events in some probability space with $\mathbb{P}(E_i) \leq p \forall i$, and if
- E_i is dependent on at most d other E_j, with
- $ep(d+1) < 1$, then

Outline of Proof

- We will prove the $t \to t-1$ improvement using the Lovász lemma:
- If E_i are events in some probability space with $\mathbb{P}(E_i) \leq p \forall i$, and if
- E_i is dependent on at most d other E_j, with
- $ep(d+1) < 1$, then
- $\mathbb{P}(\text{none of the } E_i \text{ occur}) = \mathbb{P}(\cap E_i^C) > 0$

Implementing the Lovász Local Lemma

- E_i is the event that the ith set of t columns is missing at least one permutation.

Implementing the Lovász Local Lemma

- E_i is the event that the ith set of t columns is missing at least one permutation.
- $\mathbb{P}(E_i) \leq t!(t! - 1/t!)^k := p.$

Implementing the Lovász Local Lemma

- E_i is the event that the ith set of t columns is missing at least one permutation.
- $\mathbb{P}(E_i) \leq t!(t! - 1/t!)^k := p.$
- $d = O(n^{t-1})$ (why?)

Implementing the Lovász Local Lemma

- E_i is the event that the ith set of t columns is missing at least one permutation.
- $\mathbb{P}(E_i) \leq t!(t! - 1/t!)^k := p$.
- $d = O(n^{t-1})$ (why?)
- This, on simplification, yields the required bound; $\mathbb{P}(\cap E_i^C) > 0$ means that a construction exists, which yields an upper bound for the minimum size of a scrambling array.

Key Results of this talk

Theorem

Let $t = 3$. Then, for $\phi(n)$ growing to infinity arbitrarily slowly we have

$$k \le (3\lg n - \phi(n))/lg(6/5) \Rightarrow \mathbb{P}(\text{array is } 3-\text{scrambling}) \to 0; n \to \infty$$

and

$$k \ge (3\lg n + \phi(n))/lg(6/5) \Rightarrow \mathbb{P}(\text{array is } 3-\text{scrambling}) \to 1; n \to \infty.$$

Key Results

- In fact, we have recently proved with deGraaf, Koch, Lan, that the above result actually holds for all t at the threshold given by the "Spencer bound." The correlation analysis is much harder than for the $t = 3$ case, however.

Key Results

- In fact, we have recently proved with deGraaf, Koch, Lan, that the above result actually holds for all t at the threshold given by the "Spencer bound." The correlation analysis is much harder than for the $t = 3$ case, however.
- IDEA OF PROOF:

Key Results

- In fact, we have recently proved with deGraaf, Koch, Lan, that the above result actually holds for all t at the threshold given by the "Spencer bound." The correlation analysis is much harder than for the $t = 3$ case, however.
- IDEA OF PROOF:
- Let X denote the number of sets of defective columns, i.e. those that do not contain at least one 3-permutation. Then,

Key Results

- In fact, we have recently proved with deGraaf, Koch, Lan, that the above result actually holds for all t at the threshold given by the "Spencer bound." The correlation analysis is much harder than for the $t = 3$ case, however.
- IDEA OF PROOF:
- Let X denote the number of sets of defective columns, i.e. those that do not contain at least one 3-permutation. Then,
- $\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \leq \binom{n}{3} \cdot 6 \cdot (5/6)^k \to 0$ if $k \geq (3\lg n + \phi(n))/lg(6/5)$.

Key Results

- In fact, we have recently proved with deGraaf, Koch, Lan, that the above result actually holds for all t at the threshold given by the "Spencer bound." The correlation analysis is much harder than for the $t = 3$ case, however.
- IDEA OF PROOF:
- Let X denote the number of sets of defective columns, i.e. those that do not contain at least one 3-permutation. Then,
- $\mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \leq \binom{n}{3} \cdot 6 \cdot (5/6)^k \to 0$ if $k \geq (3 \lg n + \phi(n))/lg(6/5)$.
- Thus the chance that the array is 3-scrambling tends to one.

Outline of Proof, continued

- For the lower bound, we proceed as follows:

Outline of Proof, continued

- For the lower bound, we proceed as follows:
- $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}^2(X)}$,

Outline of Proof, continued

- For the lower bound, we proceed as follows:
- $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}^2(X)}$,
- by Chebychev's inequality, and an intricate correlation analysis reveals that this quantity tends to zero with n if $k \leq (3\lg n - \phi(n))/lg(6/5)$.

Outline of Proof, continued

- For the lower bound, we proceed as follows:
- $\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}^2(X)}$,
- by Chebychev's inequality, and an intricate correlation analysis reveals that this quantity tends to zero with n if $k \leq (3\lg n - \phi(n))/lg(6/5)$.
- Thus, the chance that the array is scrambling tends to zero!

Current Work

- With deGraaf, Koch, and Lan, we are currently working on ways to reduce the upper bound even more;

Current Work

- With deGraaf, Koch, and Lan, we are currently working on ways to reduce the upper bound even more;
- Several techniques are being explored, including packing permutations, various statistics for, e.g., 312-avoiding permutations, etc.

Current Work

- With deGraaf, Koch, and Lan, we are currently working on ways to reduce the upper bound even more;
- Several techniques are being explored, including packing permutations, various statistics for, e.g., 312-avoiding permutations, etc.
- With Yuan and Koch, we are investigating similar questions for t-covering arrays, in which we are to shatter sets ($q = 2$) and words/multisets, $q \geq 3$.