Covering n-permutations by $(n+1)$- (or $(n+k)$-permutations

Anant Godbole, East Tennessee State University

Permutation Patterns 2012, June 12, 2012

Collaborators

Bounds

Probabilistic Results

- This is joint work with Taylor Allison, North Carolina State University; Katie Hawley, Harvey Mudd College/University of Oregon; and Bill Kay, University of South Carolina/Emory University.

- This is joint work with Taylor Allison, North Carolina State University; Katie Hawley, Harvey Mudd College/University of Oregon; and Bill Kay, University of South Carolina/Emory University.
- One of the problems discussed today was posed by Professor Robert Brignall during the Open Problem Session at the *International Permutation Patterns Conference* held at California State Polytechnic University in June 2011.

The problem

- Let S_n be the set of all permutations on $[n] := \{1, 2, \ldots, n\}$. We denote by κ_n the smallest cardinality of a subset $\mathcal{A}$ of S_{n+1} that "covers" S_n, in the sense that each $\pi \in S_n$ may be found as an order-isomorphic subsequence of some π' in $\mathcal{A}$. [We similarly define $\kappa_{n,k}$ but will not consider these numbers today.] What are general upper bounds on κ_n? If we randomly select ν_n elements of S_{n+1}, when does the probability that they cover S_n transition from 0 to 1? Can we provide a fine-magnification analysis that provides the "probability of coverage" when ν_n is around the level given by the phase transition? In this talk we answer these questions and raise others.

Easy Results

- Small values are easy to calculate; e.g., it is easy to see that $\kappa_1 = 1, \kappa_2 = 1$, and the permutation set $\{1342, 4213\}$ reveals that $\kappa_3 = 2$ – but the situation rapidly gets out of precise control.

Easy Results

- Small values are easy to calculate; e.g., it is easy to see that $\kappa_1 = 1, \kappa_2 = 1$, and the permutation set $\{1342, 4213\}$ reveals that $\kappa_3 = 2$ – but the situation rapidly gets out of precise control.
- Pigeonhole bound: Clearly,

Easy Results

- Small values are easy to calculate; e.g., it is easy to see that $\kappa_1 = 1, \kappa_2 = 1$, and the permutation set $\{1342, 4213\}$ reveals that $\kappa_3 = 2$ – but the situation rapidly gets out of precise control.
- Pigeonhole bound: Clearly,
- $$(\kappa_n)\binom{n+1}{n} \geq n!, \text{ so that}$$

Easy Results

- Small values are easy to calculate; e.g., it is easy to see that $\kappa_1 = 1, \kappa_2 = 1$, and the permutation set $\{1342, 4213\}$ reveals that $\kappa_3 = 2$ – but the situation rapidly gets out of precise control.
- Pigeonhole bound: Clearly,
- $$(\kappa_n)\binom{n+1}{n} \geq n!, \text{ so that}$$
- $$\kappa_n \geq \frac{(n+1)!}{n^2}(1 + o(1)).$$

A Lemma

Lemma

Let $c(n, \pi)$ denote the number of permutations in S_{n+1} that cover a fixed $\pi \in S_n$. Then $c(n, \pi) = c(n, \pi') = n^2 + 1$ for each $\pi, \pi' \in S_n$.

Proof.

It is clear that any permutation pattern $\pi \in S_n$ may be realized in $\binom{n+1}{n} = n + 1$ ways, one for any choice of n numbers from $[n + 1]$. Arrange these ways lexicographically (for example if $n = 3$, we can realize the pattern 132 as 132, 142, 243, and 143, or, lexicographically, as 132, 142, 143, 243). Note that the rth and $r + 1$st lex-orderings of π differ in a single bit.

Proof Continued

Now, given any realization of π, the $n+1$st letter may clearly be inserted in $(n+1)$ ways to create an $(n+1)$- covering permutation; however, for any $1 \le r \le n-1$, the list of covering $(n+1)$-permutations for the rth and $r+1$st lex-orderings have an overlap of magnitude 2, corresponding to whether the $(n+1)$st letter is inserted before or after the non-matching bit. Thus $c(n,\pi) = c(n,\pi') = (n+1)^2 - 2n = n^2 + 1$, as asserted.

A Theorem

Theorem

$$\kappa_n \leq \frac{\log n}{n^2}(n+1)!(1+o(1)).$$

Proof.

We use the "method of alterations" as follows: Choose a random number Y of $(n+1)$-permutations by "without replacement" sampling. The expected number $\mathbb{E}(X)$ of uncovered n-permutations can easily be calculated and estimated as

$$\mathbb{E}(X) = n!\frac{\binom{(n+1)!-n^2-1}{Y}}{\binom{(n+1)!}{Y}} \leq n! \exp\{-Y(n^2+1)/(n+1)!.$$

Proof, continued

We choose a realization with $X = X_Y \leq \mathbb{E}(X)$ and cover these with at most $\mathbb{E}(X)$ additional $(n+1)$-permutations, yielding, for any initial size Y, a covering with at most

$$Y + n! \exp\{-Y(n^2+1)/(n+1)!\}$$

members. Minimizing over Y yields an initial choice of size

$$\frac{(n+1)!}{(n^2+1)} \log\left(\frac{n^2+1}{n+1}\right),$$

and an upper bound of

$$\kappa_n \leq \frac{(n+1)!}{(n^2+1)}\left(1 + \log\left(\frac{n^2+1}{n+1}\right)\right) = \frac{\log n}{n^2}(n+1)!(1+o(1)),$$

as claimed.

Lower linear cost of additional coverings

- Theorem

  Let $\kappa_{n,\lambda}$ denote the minimum number of $(n+1)$-permutations needed to cover each n-permutation $\lambda \geq 2$ times. Then,

$$\kappa_{n,\lambda} \leq \frac{(n+1)!}{n^2}\left(\log n + (\lambda - 1)\log\log n + O(1)\right).$$

Lower linear cost of additional coverings

- Theorem

  Let $\kappa_{n,\lambda}$ denote the minimum number of $(n+1)$-permutations needed to cover each n-permutation $\lambda \geq 2$ times. Then,

$$\kappa_{n,\lambda} \leq \frac{(n+1)!}{n^2} \left(\log n + (\lambda - 1) \log\log n + O(1)\right).$$

  - There are parallels to this, e.g. Covering Designs; Coupon Collection

Lower linear cost of additional coverings

- Theorem

  Let $\kappa_{n,\lambda}$ denote the minimum number of $(n+1)$-permutations needed to cover each n-permutation $\lambda \geq 2$ times. Then,

$$\kappa_{n,\lambda} \leq \frac{(n+1)!}{n^2}\left(\log n + (\lambda - 1)\log\log n + O(1)\right).$$

  - There are parallels to this, e.g. Covering Designs; Coupon Collection
  - A discussion of a parallel set of results on asymptotic coverings....

Covering Designs Analogy

- A collection $\mathcal{A}$ of sets of size k of $[n]$ is said to form a t-covering design if each t-set is contained in at least one k-set in $\mathcal{A}$.

Covering Designs Analogy

- A collection $\mathcal{A}$ of sets of size k of $[n]$ is said to form a t-covering design if each t-set is contained in at least one k-set in $\mathcal{A}$.
- If $m(n,k,t)$ denotes the smallest size of a t-covering design $\mathcal{A}$ then it is clear that $m(n,k,t) \geq \binom{n}{t} / \binom{k}{t}$;

Covering Designs Analogy

- A collection $\mathcal{A}$ of sets of size k of $[n]$ is said to form a t-covering design if each t-set is contained in at least one k-set in $\mathcal{A}$.
- If $m(n,k,t)$ denotes the smallest size of a t-covering design $\mathcal{A}$ then it is clear that $m(n,k,t) \geq \binom{n}{t}/\binom{k}{t}$;
- Erdős and Spencer proved in 1968 that $\forall n,k,t$,

$$m(n,k,t) \leq \frac{\binom{n}{t}}{\binom{k}{t}}\left(1+\log\binom{k}{t}\right);$$

Covering Designs Analogy

- A collection $\mathcal{A}$ of sets of size k of $[n]$ is said to form a t-covering design if each t-set is contained in at least one k-set in $\mathcal{A}$.
- If $m(n,k,t)$ denotes the smallest size of a t-covering design $\mathcal{A}$ then it is clear that $m(n,k,t) \geq \binom{n}{t}/\binom{k}{t}$;
- Erdős and Spencer proved in 1968 that $\forall n,k,t$,

$$m(n,k,t) \leq \frac{\binom{n}{t}}{\binom{k}{t}}\left(1+\log\binom{k}{t}\right);$$

- It was shown furthermore by G, Thompson, Vigoda, that the minimum number $m(n,k,t,\lambda)$ of k-sets needed to cover each t-set λ times satisfied

$$m(n,k,t,\lambda) \leq \frac{\binom{n}{t}}{\binom{k}{t}}\left(1+\log\binom{k}{t}+(\lambda-1)\log\log\binom{k}{t}+O(1)\right),$$

Analogy, contd...

- This was the $\log\log$ result.

Analogy, contd...

- This was the $\log\log$ result.
- Also, the Erdős-Hanani conjecture, namely that for fixed k, t,

$$\lim_{n\to\infty} \frac{m(n,k,t)}{\binom{n}{t}} = \frac{1}{\binom{k}{t}}$$

was proved by V. Rödl and, later, by J. Spencer.

Hypergraph Formulation

We now describe the hypergraph formulation of Pippenger/Spencer that was used in Spencer to prove the Erdős-Hanani conjecture using a method that involved branching processes, dynamical algorithms, hypergraph theory, and differential equations. In this formulation the vertices of the hypergraph consisted of the ensemble of t-sets; for us they would be the class of permutations in S_n. The edges in Spencer's paper were the collections of t-subsets of the k-sets, so that the hypergraph was $\binom{k}{t}$ uniform. If analogously, we let edges be the set of n-permutations covered by an $(n+1)$-permutation, then the hypergraph is no longer uniform. It is not too hard to prove, however, that each $(n+1)$-permutation π covers $n+1-s_\pi$ n-permutations, where s_π is the number of *successions* in π, where a succession is defined as an episode $\pi(i+1)=\pi(i)\pm 1$.

Hypergraphs, continued

Moreover, we know (G&Sissokho) that the number of successions in a random permutation is approximately Poisson with parameter ~ 2, so that it is reasonable to assert that most hypergraph edges consist of $n - O(1)$ vertices. This is the first deviation from the Pippenger model, which we consider to be *not too serious* insofar as the lack of uniformity of the hypergraph is concerned but *rather serious* due to the fact that the uniformity level $n - O(1)$ is not finite. The lemma proven above shows that the degree of each vertex is $O(n^2)$, and we will also prove below that the codegree of two vertices π and π' is at most $O(1)$, so that the codegree is an order of magnitude smaller than the degree. This is good. The above problems with the hypergraph formulation notwithstanding, we make the following conjecture:

Conjecture

- Conjecture

  For some constant A,

$$\limsup_{n\to\infty} \frac{\kappa_n}{(n+1)!/n^2} = A,$$

  and possibly $A \leq 2$.

Conjecture

- Conjecture

  For some constant A,

$$\limsup_{n\to\infty} \frac{\kappa_n}{(n+1)!/n^2} = A,$$

  and possibly $A \leq 2$.

  - Is $\limsup = \lim$?

Conjecture

- Conjecture

  For some constant A,

$$\limsup_{n\to\infty} \frac{\kappa_n}{(n+1)!/n^2} = A,$$

  and possibly $A \leq 2$.

  - Is $\limsup = \lim$?
  - Is $A = 1$? Is $A \leq 2$?

A Key Lemma

Lemma

For any $\pi \in S_n$*, the set*

$$\mathcal{J}_\pi := \{\pi' \in S_n : \pi \text{ and } \pi' \text{ can be jointly covered by } \rho \in S_{n+1}\}$$

has cardinality at most n^3*. Moreover, for any* $\pi, \pi' \in S_{n+1}$*, the cardinality of*

$$\mathcal{C}_{\pi,\pi'} := \{\rho \in S_{n+1} : \rho \text{ covers both } \pi \text{ and } \pi'\}$$

is at most 4*.*

Methods used in Proof of 2nd part of the Lemma

- Interestingly, in the sequel, we only need the above codegree to be $o(n^2)$, but work hard in the paper (5.5 pages) to show that the answer is O(1)=4...perhaps because

Methods used in Proof of 2nd part of the Lemma

- Interestingly, in the sequel, we only need the above codegree to be $o(n^2)$, but work hard in the paper (5.5 pages) to show that the answer is O(1)=4...perhaps because
- The proof of the conjecture may need us to have the codegree being *really* small, and

Methods used in Proof of 2nd part of the Lemma

- Interestingly, in the sequel, we only need the above codegree to be $o(n^2)$, but work hard in the paper (5.5 pages) to show that the answer is O(1)=4...perhaps because
- The proof of the conjecture may need us to have the codegree being *really* small, and
- The generalizations to $\kappa_{n,k}$ may need a similar analysis.

Methods used in Proof of 2nd part of the Lemma

- Interestingly, in the sequel, we only need the above codegree to be $o(n^2)$, but work hard in the paper (5.5 pages) to show that the answer is O(1)=4...perhaps because
- The proof of the conjecture may need us to have the codegree being *really* small, and
- The generalizations to $\kappa_{n,k}$ may need a similar analysis.
- The proof used string matching ideas, longest common matches, and some "geometry of matching." It is long but not hard.

Proof of first part of the Lemma

Proof.

Fix π. For an $(n+1)$-permutation to be able to successfully cover another $\pi' \in S_n$ (in addition to π), π must contain an $(n-1)$-subpattern of π'. This subpattern may be present in $\binom{n}{n-1} = n$ possible positions of π, and can be represented, using the numbers $\{1, 2, \ldots, n\}$ in n ways. Finally, the nth letter of π' can be inserted into this subpattern in n ways. This proves the first part of the lemma.

Threshold

Theorem

Consider the probability model that chooses each $\rho \in S_{n+1}$ with probability p, independently. Then,

$$p \leq \frac{\log n}{n}(1 + o^*(1)) \Rightarrow \mathbb{P}(\mathcal{A} \text{ is a cover of } S_n) \to 0 \ \ (n \to \infty),$$

and

$$p \geq \frac{\log n}{n}(1 + o(1)) \Rightarrow \mathbb{P}(\mathcal{A} \text{ is a cover of } S_n) \to 1 \ \ (n \to \infty).$$

Poisson Distribution

In fact, if

$$p = \frac{\log n - 1 + \frac{1}{2}\frac{\log n}{n} - \frac{K}{n}}{n}, K \in \mathbb{R},$$

then $\mathbb{E}(X) = e^{-K}$ and $\mathbb{P}(X = 0) = \exp\{-e^{-K}\}$. Much more is true: The entire probability distribution $\mathcal{L}(X)$ of X can be approximated by a Poisson random variable with mean $\lambda = \mathbb{E}(X)$ in a range of ps that allows for large means.

Theorem

Consider the model in which each $\pi \in S_{n+1}$ is independently chosen with probability p, thus creating a random ensemble $\mathcal{A}$ of $(n+1)$-permutations. Then $d_{\mathrm{TV}}(\mathcal{L}(X), \mathrm{Po}(\lambda)) \to 0$ if $p \geq \frac{\log n}{n^2}(1+\epsilon)$, where $\mathrm{Po}(\lambda)$ denotes the Poisson distribution with parameter λ, $\epsilon > 0$ is arbitrary, and the total variation distance d_{TV} is as usual.