On Entropy and Extensions of Posets

Tom Leighton
ftl@math.mit.edu
MIT Math Dept, CSAIL
and Akamai Technologies, Inc
Cambridge, MA 02139

Ankur Moitra *
moitra@ias.edu
School of Mathematics
Institute for Advanced Study
Princeton, NJ 08540

September 12, 2011

Abstract

A vast body of literature in combinatorics and computer science aims at understanding the structural properties of a poset P implied by placing certain marginal constraints on the uniform distribution over linear extensions of P. These questions are typically concerned with whether or not P must be a total order (or have small width). Here, we instead consider whether or not these marginal constraints imply a non-trivial bound on the entropy of the uniform distribution (over the set of linear extensions of P). We prove such a result and establish that local bounds do indeed yield global bounds.

*This research was supported in part by an NSF Computing and Innovation Fellowship.

1 Introduction

A vast body of literature in combinatorics (more specifically, order theory) and computer science has focused on understanding the properties of linear extensions of posets. Given a poset P and a pair of elements x and y, throughout this paper we will use the notation $Pr[x \succ y|P]$ to denote the fraction of linear extensions[1] of P for which x is larger than y. Additionally, we will use $\delta(P)$ to denote the largest real number δ so that there is a pair x and y with

$$\delta \leq Pr[x \succ y|P] \leq 1 - \delta.$$

The famous 1/3-2/3 Conjecture asserts that if P is not a total order, $\delta(P) \geq \frac{1}{3}$. This question was posed independently by Kislitsyn (1968) [14], Fredman (1976) [7] and Linial (1984) [18]. Kahn and Saks [12] gave a computer science motivation for this question (which actually was not contained in Fredman's original paper, although Fredman's paper did focus on sorting): If the 1/3-2/3 Conjecture is true, it guarantees that no matter what comparisons have already been made, there is always some comparison that can be made whose outcome will reduce the number of possible sorted orders by at least a 2/3-factor. This question has connections to convex geometry, where a nice analogy can be drawn with Grunbaum's Theorem (see [11]). The body of literature on this subject is too vast to cover comprehensively here, but we refer the interested reader to a survey on this question [3]. The record bounds on $\delta(P)$ are due to Brightwell, Felsner and Trotter (1995) [4] who proved that $\delta(P) \geq \frac{1}{2} - \frac{\sqrt{5}}{10}$ by making use of the Alexandrov-Fenchel Inequality. Additionally, Kahn and Kim [9] gave a polynomial time algorithm based on graph entropy for sorting starting from a poset, and making $O(\log E)$ comparisons – where E is the number of linear extensions of P. Despite this vast line of work, the conjecture remains unsolved.

Additionally, there are many generalizations and related questions but which have not had as much progress. An interesting, related conjecture is due to Kahn and Saks [12]: If the 1/3-2/3 Conjecture is true, it is clearly tight since a poset P with three elements x, y and z with $x \succ y$ and z incomparable to both x and y has $\delta(P) = \frac{1}{3}$. This poset also has width two – i.e. the size of the maximum anti-chain is two (see Definition 2.4). What if the width of a poset P is larger than two? Is $\delta(P)$ any larger? Kahn and Saks conjecture that as the width w tends to infinity, $\delta(P)$ tends to $\frac{1}{2}$. The smallest known value of $\delta(P)$ is $\frac{14}{39}$ [25]. Another direction of generalization is that even though $\delta(P)$ can be $\frac{1}{3}$ in the worst-case, asymptotically the number of comparisons needed to sort a poset P is believed to have logarithm whose base is ϕ, the golden ratio. This conjecture is known as the Golden Partition Conjecture [21].

We can reformulate many of these results and conjectures, in a way that will lead to the main question considered in this paper.

Question 1.1 *Suppose we are given constraints on the values of $Pr[x \succ y|P]$. What can we determine about P?*

The 1/3-2/3 Conjecture asserts that if each value $Pr[x \succ y|P]$ is outside the range $[1/3, 2/3]$ then P must be a total order. The work of Brightwell, Felsner and Trotter [4] proves that if each value is outside the range $[1/2-\sqrt{5}/10, 1/2+\sqrt{5}/10]$ then P is indeed a total order. Similarly, the conjecture of Kahn and Saks [12] asserts that if $Pr[x \succ y|P]$ is outside the range $[1/2 - f(w), 1/2 + f(w)]$, then the width of P is at least w and moreover $f(w)$ tends to zero as w tends to infinity.

In fact, we ask a more general question. Let Σ be a distribution on total orderings. We can define $Pr[x \succ y|\Sigma]$ as the probability that x is larger than y in a total ordering sampled from Σ.

[1]For convenience of the reader, we give the standard definitions (including e.g. the definition of a linear extension) related to posets in Section 2

Note that if Σ_P is the uniform distribution on linear extensions of P, then $Pr[x \succ y|\Sigma_P]$ is exactly $Pr[x \succ y|P]$ as defined earlier. Also, let $H_2(\Sigma)$ denote the (binary) entropy of Σ.

Question 1.2 *Suppose we are given constraints on the values of $Pr[x \succ y|\Sigma]$. What can we determine about $H_2(\Sigma)$?*

In this general form, what we are asking for is a local to global bound for the entropy of permutations. A constraints on $Pr[x \succ y|\Sigma] \geq p$ is a local constraint in the sense that the comparison of x and y is a random variable whose outcome has entropy at most $H_2(p)$. Can these local constraints on entropy be patched together to lead to a (non-trivial) global constraint on entropy?

Interestingly, the maximum entropy distribution Σ^* subject to marginal constraints of the form $Pr[x \succ y|\Sigma] \geq p_{x,y}$ has a particularly nice structure. For a total ordering π, let $\succ_\pi$ denote the comparison relation for π. For some (non-negative) choice of the values $\Lambda_{x,y}$, the probability of a particular total ordering π in Σ^* is proportional to the product of $\Lambda_{x,y}$ over all pairs x and y for which $x \succ_\pi y$ - i.e.

$$Pr_{\Sigma^*}[\pi] \propto \prod_{x \succ_\pi y} \Lambda_{x,y}$$

This fact has been observed in a more general form in a number of contexts, such as statistical mechanics [22], population genetics [23], [24], equilibrium selection [19] and combinatorics [9], [15], [16]. Often times, this observation is used to give a local update rule whose only fixed points are the maximum entropy distribution. Our questions differ in that we are interested in understanding what the maximum entropy of this distribution is, and current techniques seem only to apply to understanding local issues in convergence.

1.1 Our Results

We would like to prove that for any distribution on total orderings Σ of n elements, if all pairs of elements x and y have $Pr[x \succ y|\Sigma]$ is outside the interval $[1/2 - \epsilon, 1/2 + \epsilon]$ then the entropy of Σ is at most $(1 - f(\epsilon))n \log_2 n$. However, we need an additional technical restriction in order to establish this theorem.

Definition 1.3 *A distribution Σ on total orderings of n elements is ϵ-stable if there is some ordering of the elements $x_1, x_2, ...x_n$ so that for all $i < j$*

$$Pr[x_i \succ x_j|\Sigma] \geq \frac{1}{2} + \epsilon.$$

Hence, in addition to assuming that for all pairs of elements the quantity $Pr[x \succ y|\Sigma]$ is outside the interval $[1/2 - \epsilon, 1/2 + \epsilon]$ we also need to assume that these comparisons are "consistent" in the sense that there is some ordering $x_1, x_2, ...x_n$ and when comparing x_i and x_j (for $i < j$) not only is the probability outside the interval $[1/2 - \epsilon, 1/2 + \epsilon]$ but in fact $x_i \succ x_j$ is more probable.

Our main theorem is:

Theorem 1.4 *For any ϵ-stable distribution Σ on total orderings of n elements,*

$$H_2(\Sigma^*) \leq (1 - C\epsilon^4)n \log_2 n$$

It is easy to construct a distribution Σ on total orderings that is ϵ-stable and has entropy at least $(1-\epsilon)n\log_2 n - O(n)$ (See Claim 4.14). In Section 3 we give a detailed outline of our proof.

This result fits into a more general framework in which the goal is to maximize the entropy of a distribution subject to marginal constraints. As noted earlier, previous work has studied this general version of the question (in the context of hypergraphs, distributions on hyper-edges and marginal constraints on nodes [9], [15], [16]). However, our goal is not to characterize the maximum entropy distribution but rather to given bounds on the entropy of this distribution. We can write a general entropy maximization problem as:

$$
\text{(M):} \qquad \begin{array}{rl} \max & H(\vec{p}) = \sum_i -p_i \log p_i \\ \text{s.t.} & A\vec{p} \geq \vec{b} \\ & \vec{p} \geq \vec{0} \\ & \sum_i p_i = 1. \end{array}
$$

Indeed, the constraint that a distribution on total orderings be ϵ-stable can be encoded into this form for an appropriate choice of A and $\vec{b}$. And in this context, we prove our main theorem by interpreting the rows of $A\vec{p} \leq \vec{b}$ as *tests*. We carefully group these linear inequalities in a structured way, to derive entropy bounds for p.

Our main theorem implies that, for any poset P, if the uniform distribution on linear extensions of P is ϵ-stable, P must be non-trivially close to a total order in the sense that the number of linear extensions of P is at most $(n!)^{1-C\epsilon^4}$. However, we stress that our aim in this paper is to introduce a method of proving global entropy bounds from only local constraints. Previous techniques shed no light on this quantity apart from describing the functional form of the maximum entropy distribution. In principle, the outcome of any comparison (for a distribution Σ) across different pairs of elements can be arbitrarily correlated – and this is precisely the source of the technical challenge. Our hope is that the techniques that we introduce here may lead to a more general understanding of constrained entropy maximization problems.

As we indicated, a more appealing bound in the context of distributions on total orderings would be:

Conjecture 1.5 *Let Σ be a distribution on total orderings on n elements, so that for all pairs of elements x and y, $Pr[x \succ y|\Sigma]$ is outside the interval $[1/2-\epsilon, 1/2+\epsilon]$. Then $H_2(\Sigma) \leq (1-f(\epsilon))n\log_2 n$.*

Intuitively, making these comparisons disagree (i.e. Σ is not ϵ-stable) should have a destructive interference and make it *harder* to have large entropy. This points to another interesting question:

Question 1.6 *Is there a monotonicity principle for constraint entropy maximization problems?*

To put this another way, if we have a global entropy bound for some system $A\vec{p} \leq \vec{b}$, are there other systems $A'\vec{p} \leq \vec{b'}$ whose entropy must also be bounded (by, say the same bound as for the system $A\vec{p} \leq \vec{b}$)? Perhaps this question has connections to correlation inequalities – indeed correlation inequalities have found uses in attacks on the 1/3-2/3 Conjecture [4]. We also hope that the techniques we introduce here can be useful in making progress on some of the well-known questions in order theory, especially the conjecture of Kahn and Saks [12].

2 Definitions

Here we give standard definitions for terminology related to posets. We only use this notation in the introduction, when describing literature in order theory but we include these definitions for completeness.

Definition 2.1 *A poset P is a binary relation $\succeq_P$ over a set of elements that satisfies:*

- *(reflexivity) $a \succeq_P a$*
- *(anti-symmetry) $a \succeq_P b$ and $b \succeq_P a \Rightarrow a = b$*
- *(transitivity) $a \succeq_P b$ and $b \succeq_P c \Rightarrow a \succeq_P c$*

We will use $\succeq$ instead of $\succeq_P$ when the underlying poset is clear. Note that the binary relation $\succeq_P$ need not be defined for all pairs of elements. A relation that is defined for all pairs is a total ordering.

Definition 2.2 *A linear extension E of P is a total ordering that is consistent with P - i.e. $x \succeq_P y \Rightarrow x \succeq_E y$. Let $E(P)$ denote the set of linear extensions of P.*

Definition 2.3 *A chain $x_1, x_2, ...x_r$ in P is an ordered list that satisfies $x_1 \succeq_P x_2 \succeq_P .... \succeq_P x_r$. A chain of r elements is called a length r chain. Similarly an anti-chain is a list of incomparable (according to P) elements and a r anti-chain is a set of r elements that form an anti-chain.*

Definition 2.4 *The height of a poset P is the maximum length of a chain, and the width is the maximum size of an anti-chain.*

The height and width of a poset are typically used as a measure of the complexity of a poset, and for small values of either the height or the width the 1/3-2/3 Conjecture, and even some other related conjectures, have been established. Of course, throughout this paper our measure of choice of the complexity of a poset is $\log_2 |E(P)|$ – which is the entropy of the uniform distribution on the set of linear extensions of P.

3 Outline

Throughout the rest of the paper, we will describe Σ as a distribution on permutations (rather than a distribution on total orderings). The reason we choose this convention is because we will be interested in ϵ-stable distributions Σ, and if $x_1, x_2, ...x_n$ is the ordering so that (for all $i < j$)

$$Pr[x_i \succ x_j | \Sigma] \geq \frac{1}{2} + \epsilon,$$

we may as well regard a total ordering π on $x_1, x_2, ...x_n$ as a permutation in which $\pi(1)$ is the number of elements x_j for which x_j is at least x_i (i.e. the rank of x_i). We will continue to use the symbols $x_1, x_2, ...x_n$ to attempt to avoid confusion. We could have renamed these symbols $1, 2, ..n$ but we choose x_i instead to emphasize that it is an element of the *domain* of π and we will regard the range of π as $1, 2, ...n$.

So we can regard an ϵ-stable distribution Σ on total orderings as a distribution on permutations that satisfies certain marginal constraints, and we will abuse notation and call this an ϵ-stable distribution on permutations.

Definition 3.1 *Let Σ be a distribution on permutations of n elements. Σ is ϵ-stable if for all $i < j$,*

$$Pr_{\pi \leftarrow \Sigma}[\pi(x_i) < \pi(x_j)] \geq \frac{1}{2} + \epsilon$$

Here we use the notation that $\pi \leftarrow \Sigma$ denotes sampling a permutation π from Σ.

Definition 3.2 *Let H_ϵ be the maximum entropy (measured base 2) of a distribution on permutations that is ϵ-stable.*

Next, we describe our general approach – which is based on elementary methods.To simplify the discussion, we will specialize the description to our problem of interest - namely proving entropy bounds for an ϵ-stable distribution on permutations.

How is the constraint that a distribution on permutations be ϵ-stable encoded in the notation of Problem (M)? The vector $\vec{p}$ is length $n!$. Each index into $\vec{p}$ represents a permutation. Moreover, each row in A corresponds to a pair of elements x_i and x_j $(i < j)$and the entry in this row, corresponding to some permutation π is one if and only if $\pi(i) < \pi(j)$. Hence setting the corresponding entry of $\vec{b}$ to be $\frac{1}{2} + \epsilon$ enforces that for a $\frac{1}{2} + \epsilon$ fraction of the permutations (weighted according to Σ), the comparison of x_i and x_j favors x_i.

1. **Choosing Tests:** We use the combinatorial structure of the problem to group linear inequalities $a_k^T \vec{p} \leq b_k$ into sets. For each set we sum the linear inequalities in the set and get an aggregate linear inequality $t_l^T \vec{p} \leq s_l$. Each such linear inequality defines a test that we will perform on a permutation π.

2. **Sampling from Σ:** We sample a permutation π from the distribution Σ and we consider the characteristic vector $\vec{p}_\pi \in \{0,1\}^{n!}$ for this permutation. We apply each aggregate linear inequality to this characteristic vector. If $t_l^T \vec{p}_\pi \leq s_l$ holds then the test is PASSED, and otherwise the test is FAILED. We concatenate the test outcomes and get a test outcome vector $\vec{r}(\pi)$ that contains information about which tests have been passed.

3. **The Implications of Passing a Test:** For any test outcome vector $\vec{r}$, we consider the set of permutations $\Pi_{\vec{r}}$ that result in this test outcome vector. We define a potential function ϕ on test outcome vectors. For any test outcome vector $\vec{r}$, the potential function $\phi(\vec{r})$ gives a rough estimate of $|\Pi_{\vec{r}}|$.

4. **"Good" and "Bad" Permutations:** We define a permutation to be "good" iff $\phi(\vec{r}(\pi)) \geq T$ i.e. the permutation π is assigned a potential larger than some target threshold T. We argue via convexity arguments that on average a permutation sampled from Σ is "good", and we use the potential function ϕ to count the space of "good" permutations.

The intuition for this approach is that while a permutation π sampled from Σ can on average fail many linear inequalities $a_k^T \vec{p}_\pi \leq b_k$, we can group linear inequalities into sets. For an appropriate grouping (which we choose based on the combinatorial structure of the space of permutations), when a test is passed we can characterize *how* - in a combinatorial sense - such a test restricts the space of permutations. This notion of grouping linear inequalities based on the combinatorial structure of the space of permutations is inspired by de la Vega's proof that there are graphical tournaments T for which $fit(T) \leq O(n^{3/2})$ [6].[2]

[2] $fit(T)$ is defined as the maximum over all orderings of the vertices of a di-graph, of the number of forward edges minus the number of backward edges.

4 A Local to Global Entropy Bound

4.1 Choosing Tests

For notational convenience (because often we will be working in either the domain or range of the permutations) let us consider a permutation π to be a mapping from the set $x_1, x_2, ...x_n$ to $y_1, y_2, ...y_n$. But x_i, y_i are really just placeholders for the value i.

Given a permutation π sampled from Σ we will apply local tests of the following form: For any two disjoint sets $S, T \subset \{x_1, x_2, ...x_n\}$ in the *domain*, we will measure inversions across these two sets:

Definition 4.1 $inv(\pi, S, T) = \{(x_i, x_j) | x_i \in S, x_j \in T$ s.t. EITHER $x_i < x_j$ and $\pi(x_i) > \pi(x_j)$ OR $x_i > x_j$ and $\pi(x_i) < \pi(x_j)\}$

If Σ is an ϵ-stable distribution on permutations, the expected number of inversions is at most $(\frac{1}{2} - \epsilon)|S||T|$. We relax this linear inequality and define a test $t(\pi, S, T)$:

Definition 4.2

$$t(\pi, S, T) = \; PASS \; if \; |inv(\pi, S, T)| \leq (\frac{1-\epsilon}{2})|S||T| \; and$$

$$t(\pi, S, T) = \; FAIL \; if \; |inv(\pi, S, T)| > (\frac{1-\epsilon}{2})|S||T|$$

Notice that the threshold for a test to PASS is higher than the expected number of inversions. We relax the inequality so that each test has a constant chance to be PASSED:

Lemma 4.3 $Pr_{\pi \leftarrow \Sigma}[t(\pi, S, T) = PASS] > 1 - \frac{\frac{1}{2}-\epsilon}{\frac{1}{2}-\frac{\epsilon}{2}} = \Omega(\epsilon)$

Proof: Let $Pr_{\pi \leftarrow \Sigma}[t(\pi, S, T) = PASS] = p$. Then $E_{\pi \leftarrow \Sigma}[|inv(\pi, S, T)|] > (1-p)\frac{1-\epsilon}{2}|S||T|$. Also linearity of expectation yields that $(\frac{1}{2} - \epsilon)|S||T| \geq E_{\pi \leftarrow \Sigma}[|inv(\pi, S, T)|]$. Combining these inequalities yields:

$$\frac{\frac{1}{2} - \epsilon}{\frac{1}{2} - \frac{\epsilon}{2}} > 1 - p$$

■

We have only chosen the structure of the tests that we will apply. Passing a test constrains the distribution Σ, and we will carefully choose (offline) a set of tests so that any large enough fraction of these tests that are passed can be patched together to yield a global constraint on the entropy of Σ. The choice of tests needs to be made offline because distinct tests are not be independent, and if we allow which test we apply next to depend on the outcome of a previous test then we can no longer assume that the next test has a constant chance to be PASSED.

4.2 The Implications of Passing a Test

Here we characterize *how* a PASSED test restricts the space of permutations. The key observation is that if a test is PASSED there must be some constant fraction of the range space that is non-trivially biased under Σ towards either the set S or the set T (in a test $t(\pi, S, T)$).

Let $\epsilon = \frac{1}{2^l}$ for $l > 0$. Then partition the range $\{y_1, y_2, ...y_n\}$ into $A_1 = \{y_1, y_2, ...y_{\frac{1}{2^{l+1}}n}\}$, $A_2 = \{y_{\frac{1}{2^{l+1}}n+1}, ...y_{\frac{2}{2^{l+1}}n}\}$ .... $A_{2^{l+1}} = \{y_{\frac{2^{l+1}-1}{2^{l+1}}n+1}, ...y_n\}$.

Let $U = \{x_1, x_2, ...x_{\frac{n}{2}}\}$, $V = \{x_{\frac{n}{2}+1}, x_{\frac{n}{2}+2}, ...x_n\}$.

We consider π^{-1} to be a function that maps a point in the range $y_i \in \{y_1, y_2, ...y_n\}$ to the point in the domain $x_j \in \{x_1, x_2, ...x_n\}$ s.t. $\pi(x_j) = y_i$. For a subset S of the range, we define $\pi^{-1}(S) = \cup_{y_i \in S} \pi^{-1}(y_i)$. So π^{-1} maps subsets of the range to subsets of the domain:

Definition 4.4

$$\delta(\pi, l+1) = 2 \min_{A_i} \frac{\min(|\pi^{-1}(A_i) \cap U|, |\pi^{-1}(A_i) \cap V|)}{|A_i|}$$

The quantity $\min(|\pi^{-1}(A_i) \cap U|, |\pi^{-1}(A_i) \cap V|)/|A_i|$ measures the skew of the subset A_i i.e. are the points in A_i when mapped by π^{-1} to points in the domain significantly biased to either U or V? Suppose that π is sampled uniformly at random from all permutations. We would expect that for any subset A_i of the range, $|\pi^{-1}(A_i) \cap U| \approx |\pi^{-1}(A_i) \cap V|$. So if π is sampled uniformly at random from all permutations, we expect $\delta(\pi, l+1) \approx 1$.

Suppose π is instead sampled from Σ. Suppose we expect $\delta(\pi, l+1) < f(\epsilon)$ for some $f(\epsilon) < 1$ that is independent of n. Then there is some set A_i (which makes up a $\frac{\epsilon}{2}$-fraction of the range) which is biased to either U or V - and we can use this to get a non-trivial entropy bound. Roughly, we prove such a statement in the contrapositive - a bound on the number of inversions across the sets U, V implies a bound on the parameter $\delta(\pi, l+1)$.

We can lower bound the number of inversions across U, V with respect to π using $\delta(\pi, l+1)$:

Lemma 4.5 $|inv(\pi, U, V)| \geq \delta(\pi, l+1)^2 n^2 (\frac{1}{8} - \frac{1}{2^{l+4}})$

Proof: The sets $\pi^{-1}(A_1) \cap U, \pi^{-1}(A_2) \cap U, ...\pi^{-1}(A_n) \cap U$ and $\pi^{-1}(A_1) \cap V, \pi^{-1}(A_2) \cap V, ...\pi^{-1}(A_n) \cap V$ partition the domain $\{x_1, x_2, ...x_n\}$. So

$$|inv(\pi, U, V)| = \sum_{i,j} |inv(\pi, \pi^{-1}(A_i) \cap U, \pi^{-1}(A_j) \cap V)|$$

And for $i > j$ we obtain:

$$\begin{aligned} |inv(\pi, \pi^{-1}(A_i) \cap U, \pi^{-1}(A_j) \cap V)| &= |\pi^{-1}(A_i) \cap U| \times |\pi^{-1}(A_j) \cap V| \\ &\geq \frac{\delta(\pi, l+1)}{2} |A_i| \frac{\delta(\pi, l+1)}{2} |A_j| \end{aligned}$$

Using $|A_i| = |A_j| = \frac{n}{2^{l+2}}$ we obtain $|inv(\pi, \pi^{-1}(A_i) \cap U, \pi^{-1}(A_j) \cap V)| \geq [\frac{\delta(\pi, l+1)n}{2^{l+2}}]^2$ and this implies

$$\begin{aligned} |inv(\pi, U, V)| &\geq [\frac{\delta(\pi, l+1)n}{2^{l+2}}]^2 (1 + 2 + ...2^{l+1} - 1) \\ &\geq \frac{\delta(\pi, l+1)^2 n^2}{2^{2l+4}} \frac{1}{2} (2^{l+1})(2^{l+1} - 1) \\ &\geq \delta(\pi, l+1)^2 n^2 \left(\frac{1}{8} - \frac{1}{2^{l+4}}\right) \end{aligned}$$

■

In the above lemma, we counted only guaranteed inversions and ignored the diagonal inversions such as:

$$|inv(\pi, \pi^{-1}(A_i) \cap U, \pi^{-1}(A_i) \cap V)|$$

We ignored these diagonal inversion because apriori this quantity can be anywhere between 0 and $|\pi^{-1}(A_i) \cap U| \times |\pi^{-1}(A_i) \cap V|$. It depends on *which* elements of A_i are mapped by π^{-1} to U and which are mapped to V. So the number of diagonal inversions does not depend on $\delta(\pi, l+1)$ as do the other inversions that we counted.

Because these diagonal inversions are not counted, we need to choose the number of sets in the partition depending on ϵ - for smaller ϵ, we need a finer partition of the range $\{y_1, y_2, ...y_n\}$. If we chose a static partition size independent of ϵ then we woudl lose too much in the previous bound by not counting diagonal inversions.

Claim 4.6 *If $t(\pi, U, V) = PASS$ then $\delta(\pi, l+1) \leq \sqrt{\gamma}$ where $1 > \gamma = \frac{1-\frac{1}{2^l}}{1-\frac{1}{2^{l+1}}} = 1 - \Omega(\epsilon)$*

Proof: Suppose $t(\pi, U, V) = PASS$. Then

$$(\frac{1}{2} - \frac{1}{2^{l+1}})\frac{n^2}{4} \geq |inv(\pi, U, V)| \geq \delta(\pi, l+1)^2 n^2 (\frac{1}{8} - \frac{1}{2^{l+4}})$$

■

So whenever a test is PASSED, there is a set A_i (which makes up an $\frac{\epsilon}{2}$-fraction of the range) which is significantly biased i.e. either

$$|\pi^{-1}(A_i) \cap U| = (1 + \Omega(\epsilon))\frac{|A_i|}{2} \text{ OR } |\pi^{-1}(A_i) \cap V| = (1 + \Omega(\epsilon))\frac{|A_i|}{2}$$

Such a set yields a corresponding reduction in the entropy of Σ (from the trivial bound) that we can "charge" to A_i.

4.3 An Offline Testing Strategy

Next, we need to choose a set of tests to apply in the hope that passing any large enough fraction of these tests will imply enough constraints on Σ that we can bound the entropy.

Recall that $U = \{x_1, x_2, ...x_{\frac{n}{2}}\}$, $V = \{x_{\frac{n}{2}+1}, x_{\frac{n}{2}+2}, ...x_n\}$.

Our testing strategy will first apply the test $t(\pi, U, V)$ and then will only apply tests $t(\pi, S, T)$ (for disjoint S, T) such that either $S, T \subset U$ or $S, T \subset V$. So later tests that depend on $\pi(x_i)$ where $x_i \in U$ will depend only on the outcome of comparing $\pi(x_i)$ to any $\pi(x_j)$ for $x_j \in U$. So we can "collapse" a permutation π and forget about the value $\pi(x_i)$ and remember only the relative value of $\pi(x_i)$ when compared to all $\pi(x_j)$ for $x_j \in U$.

Let π be a permutation mapping $\{x_1, x_2, ...x_n\}$ to $\{y_1, y_2, ...y_n\}$ and consider a set $S \subset \{x_1, x_2, ...x_n\}$. Let $s = |S|$ and suppose $S = \{x_{i_1}, ...x_{i_s}\}$ where $x_{i_1} \leq x_{i_2} \leq ...x_{i_s}$:

Definition 4.7 *Define the permutation π_S - which is the permutation π collapsed on a set S - as a permutation mapping the set $\{x_1, x_2, ...x_{|S|}\}$ to the set $\{y_1, y_2, ...y_{|S|}\}$ s.t. $\forall_{x_{i_p}, x_{i_q} \in S} \pi_S(x_p) \leq \pi_S(x_q) \iff \pi(x_{i_p}) \leq \pi(x_{i_q})$.*

Using this notation, we can succinctly describe the sequence of tests that will be performed on π. For a given permutation π, the test outcome vector $\vec{r}(\pi)$ is defined recursively. Let $\circ$ denote the concatenation operator:

Definition 4.8 $\vec{r}(\pi) = t(\pi, U, V) \circ \vec{r}(\pi_U) \circ \vec{t}(\pi_V)$ and $\vec{r}(\pi) = \emptyset$ *if* π is a permutation on one element

Similarly define a potential vector $\vec{c}(n)$:

Definition 4.9 $\vec{c}(n) = n \circ \vec{c}(\frac{n}{2}) \circ \vec{c}(\frac{n}{2})$ and $\vec{c}(1) = \emptyset$

See Figure 1. The sequence of tests just tests $t(\pi, U, V)$ and then recurses on both π_U and π_V. We also define the vector $c(\vec{n})$ which will help us characterize *how* a test outcome vector $\vec{r}$ restricts the space of permutations.

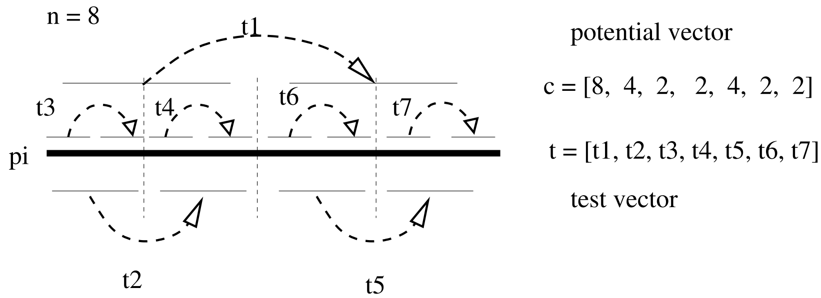

Figure 1: Our sequence of tests $\vec{r}$ applied to a permutation on $n = 8$ elements.

4.4 Choosing a Potential Function

Not all tests in our testing strategy are created equal. Passing an early test more significantly restricts the distribution Σ since passing an early test implies a large sized set A_i is biased, but passing a later test only means that a smaller set is biased. Here we define a potential function, with the goal being if any weighted-constant (by the potential function) fraction of the the tests are passed, we want to be able to non-trivially bound the entropy of Σ.

Let $\vec{r}$ be any vector of test outcomes – i.e. $\vec{r}$ is a $0, 1$-vector (in $\Theta(n)$-dimensional space) so that the value r_i in any coordinate i is 1 iff the corresponding test is passed. We can then consider the set $\Pi_{\vec{r}}$ which we will define as all permutations π for which $\vec{r}(\pi) = \vec{r}$. Here we choose a potential function ϕ that (for any test outcome vector $\vec{r}$) gives a rough estimate of $|\Pi_{\vec{r}}|$.

Theorem 4.10 $|\Pi_{\vec{r}}| \leq 2^{n \log n - d\phi + O(n)}$ where $d = \frac{1 - H_2(\frac{\sqrt{\gamma}}{2})}{2^{l+1}}$ and $\phi = \vec{c}(n)^T \vec{r}$ and γ is defined as in Claim 4.6

Proof: Every test except the first is only a function of either π_U or π_V. Assume that the bound in this theorem is true for all values $n_0 < n$. We can write the test vector $\vec{r}$ as the concatenation of the test outcome $t(\pi, U, V)$, and the test outcome vectors for π_U and π_V. Let the resulting test outcome vectors be $r_1, \vec{r_2}, \vec{r_3}$ respectively. And so $\vec{r} = r_1 \circ \vec{r_2} \circ \vec{r_3}$.

Let R_n be the number of ways of choosing which elements from $\{y_1, y_2, ..y_n\}$ are mapped by π^{-1} to U and which are mapped to V s.t. the first test $t(\pi, U, V) = r_1$.

Let $T_n(\phi) = 2^{n \log n - d\phi + O(n)}$. Let $\phi_2 = \vec{c}(\frac{n}{2})^T \vec{r_2}$ and $\phi_3 = \vec{c}(\frac{n}{2})^T \vec{r_3}$. Then the number of permutations π s.t. $\vec{r}(\pi) = \vec{r}$ is at most $|\Pi_r| \leq R_n |\Pi_{r_2}||\Pi_{r_3}| \leq R_n T_n(\phi_2) T_n(\phi_3)$

If $r_1 = 0$ then we can choose the trivial upper bound of 2^n for R_n. If $r_1 = 1$ using Claim 4.6, there is some A_i that (when mapped by π^{-1}) is significantly biased to either U or V. So if $r_1 = 1$ then

$$R_n \leq \binom{\epsilon n}{\frac{\epsilon n}{2}}^{\frac{1}{\epsilon} - 1} \binom{\epsilon n}{\frac{\sqrt{\gamma}}{2} \epsilon n}$$

Then using the well-known inequality $\log \binom{n}{k} \leq H_2(\frac{k}{n})n + O(1)$ we obtain

$$\log R_n \leq \Big(\frac{1}{2^{l+1}} H_2\Big(\frac{\sqrt{\gamma}}{2}\Big) + \frac{2^{l+1} - 1}{2^{l+1}} \Big) n + O(1)$$

and this implies the theorem. ■

Note that if no tests are passed in $\vec{r}$, then the above theorem gives a trivial bound of $2^{n \log n}$ on $|\Pi_{\vec{r}}|$ which is larger than $n!$. So we need to argue that on average a permutation π sampled from Σ is "good" and must be assigned a potential $\phi(\vec{r}(\pi))$ that is larger than some target threshold T. We can then use the potential function to count the space of "good" permutations.

4.5 Good and Bad Permutations

Here we complete the argument, by first establishing that a permutation π sampled from Σ has a constant chance of accumulating enough weight from tests that are passed (using the potential function defined in the previous section, which credits different tests with different weights). Our main theorem then follows from a union bound.

Choose $f >> 1$. Consider a permutation π sampled from Σ and the test vector $\vec{r}(\pi)$ for π.

Definition 4.11 *A permutation π is "good" if $\phi = \vec{c}(n)^T \vec{r}(\pi) \geq \frac{1}{f} n \log n$ and "bad" if $\phi = \vec{c}(n)^T \vec{r}(\pi) < \frac{1}{f} n \log n$*

Notice that there are $\Theta(n)$ tests performed, so there are at most $2^{\Theta(n)}$ possible test outcome vectors. We also note that sampling from Σ can result in a "bad" permutation. Sampling from Σ can even result in a permutation π for which *no* tests are passed! But such samples must be infrequent:

Lemma 4.12 *For large but fixed $f >> 1$, the probability that π is "good" is $\Omega(\epsilon)$*

Proof: Let Σ' be a distribution on permutations such that $Pr[\pi' \leftarrow \Sigma'] = Pr[\pi' = \pi, \pi \leftarrow \Sigma | \pi \text{ is "bad" }]$. So Σ' is a restriction of Σ to "bad" permutations.

All "bad" permutations are assigned a potential $< \frac{1}{f} n \log n$. So $E_{\pi' \leftarrow \Sigma'}[\vec{c}(n)^T \vec{r}(\pi')] < \frac{1}{f} n \log n$. Also $\sum_i \vec{c}(n)_i = \Theta(n \log n)$. so there must be a test $t(\pi, S, T)$ such that

$$Pr_{\pi' \leftarrow \Sigma'}[t(\pi', S, T) == PASS] \leq \Theta\Big(\frac{1}{f}\Big)$$

Using Lemma 4.3, we know that for any test $t(\pi, S, T)$: $Pr_{\pi \leftarrow \Sigma}[t(\pi, S, T) == PASS] \geq \Omega(\epsilon)$.

Also $Pr_{\pi \leftarrow \Sigma}[t(\pi, S, T) == PASS] \leq Pr_{\pi' \leftarrow \Sigma'}[t(\pi', S, T) == PASS] + Pr_{\pi \leftarrow \Sigma}[\pi \text{ is "good" }]$. This implies that for $f = \Theta(\frac{1}{\epsilon})$, $Pr_{\pi \leftarrow \Sigma}[\pi \text{ is "good" }] \geq \Omega(\epsilon)$. ■

Theorem 4.13 $H_\epsilon \leq (1 - C\epsilon^4) n \log n - O(n)$

Proof: Using Theorem 4.10, and applying the union bound over all possible test outcome vectors $\vec{r}$, there are at most $2^{n \log n - \Theta(d) n \log n + O(n)}$ permutations π s.t. $\vec{r}(\pi)$ is "good" and using Lemma 4.12 these permutations contain at least $\Omega(\epsilon)$ fraction of the weight in Σ. So for any ϵ-stable distribution Σ we have that $H(\Sigma) \leq n \log n - \Omega(\epsilon)\Omega(d) n \log n - O(n)$ (where d is defined in Theorem 4.10). So

$$H_\epsilon \leq n \log n - \Omega\Big(\epsilon^2 - \epsilon^2 H_2\Big(\frac{1}{2}\sqrt{1 - \frac{\epsilon}{2}}\Big)\Big) n \log n - O(n) \leq n \log n - C\epsilon^4 n \log n - O(n)$$

■

Claim 4.14 $H_\epsilon \geq (1 - \epsilon) n \log n - O(n)$

Proof: Let Σ to be a distribution that with probability $1 - 2\epsilon$ chooses a permutation π uniformly at random, and with probability 2ϵ chooses the permutation π s.t. $\pi(x_i) = y_i$. Then for each pair $i < j$, $Pr[\pi(i) < \pi(j)] = \frac{1}{2}(1 - 2\epsilon) + 2\epsilon = \frac{1}{2} + \epsilon$. And so Σ is ϵ-stable.

Also $H(\Sigma) > (1 - 2\epsilon) \log n! = (1 - 2\epsilon) n \log n - O(n)$ ■

5 Acknowledgments

We would like to thank Noga Alon, Jeff Kahn and Joel Spencer for many helpful discussions.

References

[1] M. Aigner. A note on merging. *Order*, pages 257–264, 1985.

[2] M. Ajtai, J. Komlos, and E. Szemeredi. An $\mathrm{O}(n \log n)$ sorting network. *Combinatorica*, pages 1–19, 1983.

[3] G. Brightwell. Balanced pairs in partial orders. *Discrete Mathematics*, pages 25–52, 1999.

[4] G. Brightwell, S. Felsner, and W. Trotter. Balancing pairs and the cross product conjecture. *Order*, pages 327–349, 1995.

[5] G. Brightwell and P. Winkler. Counting linear extensions is sharp P-complete. *Symposium on Theory of Computation*, pages 175–181, 1991.

[6] F. de la Vega. On the maximal cardinality of a consistent set of arcs in a random tournament. *Journal of Combinatorial Theory Series B*, pages 328–332, 1983.

[7] M. Fredman. How good is the information theory bound in sorting? *Theoretical Computer Science*, pages 355–361, 1976.

[8] J. Kahn. Asymptotics of the chromatic index for multigraphs. *Journal of Combinatorial Theory Series B*, 68:233–254, 1996.

[9] J. Kahn and P. Kayll. On the stochastic independence properties of hard-core distributions. *Combinatoria*, 17:369–391, 1997.

[10] J. Kahn and J. Kim. Entropy and sorting. *Symposium on Theory of Computation*, pages 178–187, 1992.

[11] J. Kahn and N. Linial. Balancing extensions via brunn-minkowski. *Combinatorica*, 11:363–368, 1991.

[12] J. Kahn and M. Saks. Balancing poset extensions. *Order*, 1:113–126, 1984.

[13] L. Khachiyan. Optimal algorithms in convex programming, decomposition and sorting. *Computers and Decision Problems*, pages 161–205, 1989.

[14] S. Kislitsyn. Finite partially ordered sets and their associated sets of permutations. *Matematicheskiye Zametki*, pages 511–518, 1968.

[15] C. Lee. Some recent results on convex polytopes. *Contemporary Mathematics*, 114:3–19, 1990.

[16] C. Lee. Convex polytopes, the moment map, and canonical convex combinations. *manuscript*, 1994.

[17] T. Leighton and G. Plaxton. A (fairly) simple circuit that (usually) sorts. *Symposium on Foundations of Computer Science*, pages 264–274, 1990.

[18] N. Linial. The information-theoretic bound is good for merging. *SIAM Journal on Computing*, pages 795–801, 1984.

[19] L. Ortiz, R. Shapire, and S. Kakade. Maximum entropy correlated equilibrium. *Journal of Machine Learning Research*, pages 347–354, 2007.

[20] M. Paterson. Improved sorting networks with $\mathrm{O}(\log n)$ depth. *Algorithmica*, 5:75–92, 1990.

[21] M. Peczarski. The golden partition conjecture. *Order*, pages 89–95, 2006.

[22] O. Penrose. Foundations of statistical mechanics. *Pergamon Press*, 1970.

[23] Y. Rabinovich, A. Sinclair, and A. Wigderson. Quadratic dynamical systems (preliminary version). *Symposium on Foundations of Computer Science*, pages 304–313, 1992.

[24] Y. Rabinovich and A. Wigderson. Analysis of a simple genetic algorithm. *Proceedings of the 4th International Conference on Genetic Algorithms*, pages 215–221, 1991.

[25] M. Saks. Balancing linear extensions of ordered sets, "Unsolved problems". *Order*, pages 327–330, 1985.