

Correlation and Sorting

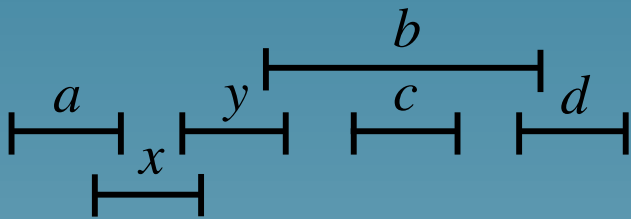
Tom Trotter

September 20, 2001

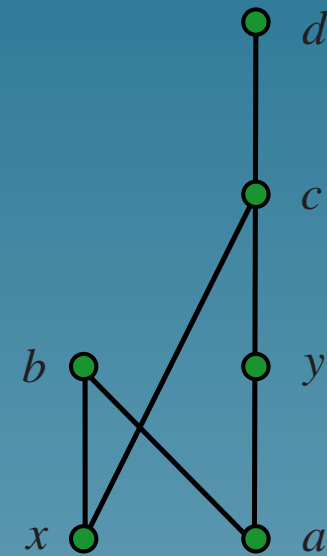
Research Interests

- Graph theory
- Discrete geometry
- On-line algorithms
- Adversarial algorithms
- Ramsey theory
- Extremal Problems
- Probabilistic methods
- Partially ordered sets

Interval Orders

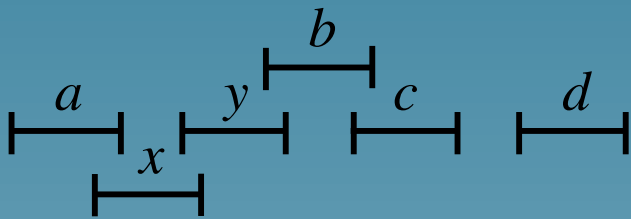


Intervals from a linear order

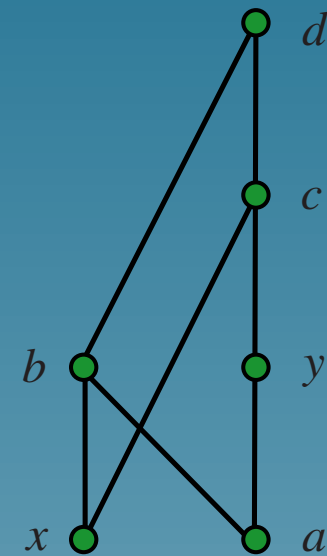


The associated interval order

Semiorders



Constant Length Intervals



The associated semiorder

Classical Sorting Problem

Problem 1. Determine an unknown linear order L of a set X of n elements by asking a series of questions of the form:

Is $x < y$ in L ?

Classical Sorting Problem

Problem 1. Determine an unknown linear order L of a set X of n elements by asking a series of questions of the form:

Is $x < y$ in L ?

Remark. There are $n!$ possible linear orders, so at least $\lg n! \sim n \log n$ questions are required. Of course, several well known algorithms sort in $O(n \log n)$ rounds.

Sorting with Partial Information

Problem 2. Determine an unknown linear extension L of a poset \mathbf{P} of n elements by asking a series of questions of the form:

Is $x < y$ in L ?

Sorting with Partial Information

Problem 2. Determine an unknown linear extension L of a poset \mathbf{P} of n elements by asking a series of questions of the form:

Is $x < y$ in L ?

Question. If \mathbf{P} has t linear extensions, can we always determine L by asking $O(\log t)$ questions?

An Elementary Probability Space

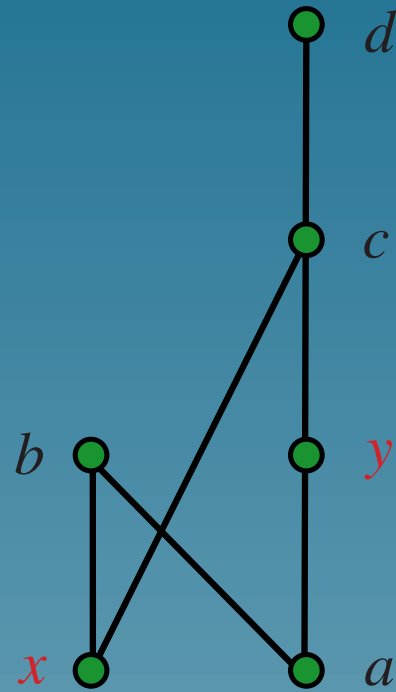
Notation. Consider $\mathcal{E}(P)$ as a probability space with each $L \in \mathcal{E}(P)$ an equally likely outcome.

Notation. For distinct elements x and y , the event $[x < y]$ is then the subset of $\mathcal{E}(P)$ consisting of those L with $x < y$ in L .

Notation.

$$\text{Prob}[x < y] = \frac{|[x < y]|}{|\mathcal{E}(P)|}$$

Example



$$\text{Prob}[x < y] = \frac{8}{11}$$

The $1/3$ – $2/3$ Conjecture

Conjecture. [Kislytsin, 1966] *If P is a finite poset and is not a chain, then there exist distinct x, y in P with*

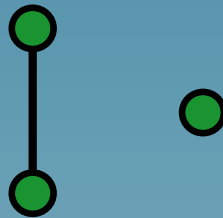
$$\frac{1}{3} \leq \text{Prob}[x < y] \leq \frac{2}{3}$$

The $1/3$ – $2/3$ Conjecture

Conjecture. [Kislytsin, 1966] *If P is a finite poset and is not a chain, then there exist distinct x, y in P with*

$$\frac{1}{3} \leq \text{Prob}[x < y] \leq \frac{2}{3}$$

Remark. *If true, the inequality is best possible.*



Observation

Remark. *It is not at all clear that there is any $\delta > 0$ so that for any P , there exists a pair x, y with*

$$\delta \leq \text{Prob}[x < y] \leq 1 - \delta$$

Observation

Remark. *It is not at all clear that there is any $\delta > 0$ so that for any P , there exists a pair x, y with*

$$\delta \leq \text{Prob}[x < y] \leq 1 - \delta$$

Remark. *If there exists such a δ , then we can determine an unknown linear extension of a poset P with t linear extensions in $O(\log t)$ rounds.*

The Kahn/Saks Theorem

Theorem. [Kahn and Saks] *If P is a finite poset and is not a chain, then there exist distinct x, y in P with*

$$\frac{3}{11} < \text{Prob}[x < y] < \frac{8}{11}$$

The Kahn/Saks Theorem

Theorem. [Kahn and Saks] *If P is a finite poset and is not a chain, then there exist distinct x, y in P with*

$$\frac{3}{11} < \text{Prob}[x < y] < \frac{8}{11}$$

Remark. *Over the next few years, several other papers on balancing pairs appeared, all with weaker results but somewhat shorter proofs.*

A Basic Pigeon-hole

Notation. Given $x \in P$ and $L \in \mathcal{E}(P)$, let $h_L(x)$ denote the *height* of x in L . Also, let $h(x)$ denote the *average height* of x .

A Basic Pigeon-hole

Notation. Given $x \in P$ and $L \in \mathcal{E}(P)$, let $h_L(x)$ denote the *height* of x in L . Also, let $h(x)$ denote the *average height* of x .

Remark. When P is not a chain, there exists a pair x, y with

$$0 \leq h(y) - h(x) < 1$$

A Basic Pigeon-hole

Notation. Given $x \in P$ and $L \in \mathcal{E}(P)$, let $h_L(x)$ denote the *height* of x in L . Also, let $h(x)$ denote the *average height* of x .

Remark. When P is not a chain, there exists a pair x, y with

$$0 \leq h(y) - h(x) < 1$$

Remark. Kahn and Saks show that for any such pair, we always have

$$\frac{3}{11} < \text{Prob}[x < y] < \frac{8}{11}$$

Linear Constraints (1)

Let x and y be incomparable elements of P with $0 \leq h(y) - h(x) \leq 1$. Define

$$a_i = \text{Prob}[h_L(y) - h_L(x) = i] \quad \text{and}$$

$$b_i = \text{Prob}[h_L(x) - h_L(y) = i]$$

Linear Constraints (1)

Let x and y be incomparable elements of P with $0 \leq h(y) - h(x) \leq 1$. Define

$$a_i = \text{Prob}[h_L(y) - h_L(x) = i] \quad \text{and}$$

$$b_i = \text{Prob}[h_L(x) - h_L(y) = i]$$

Then,

$$\text{Prob}[x < y] = \sum_{i \geq 1} a_i$$

Linear Constraints (2)

$$1 = \sum_{i \geq 1} a_i + \sum_{i \geq 1} b_i$$

$$0 \leq \sum_{i \geq 1} i a_i - \sum_{i \geq 1} i b_i \leq 1$$

$$a_1 = b_1$$

$$a_2 + b_2 \leq a_1 + b_1$$

$$a_{i+1} \leq a_i + a_{i+2} \quad \text{and} \quad b_{i+1} \leq b_i + b_{i+2}$$

Linear Constraints (3)

Remark. *These linear constraints are **not** enough, since the optimal solution is still:*

$$\text{Prob}[x < y] = 1$$

The Alexandrov/Fenchel Inequalities

Let K_0 and K_1 be convex bodies in \mathbb{R}^d , and let $K_\lambda = (1 - \lambda)K_0 + \lambda K_1$. Then there exist unique numbers a_0, a_1, \dots, a_d so that the volume of K_λ is given by:

$$\text{Vol}(K_\lambda) = \sum_{i=0}^d \binom{d}{i} a_i (1 - \lambda)^{d-i} \lambda^i$$

The Alexandrov/Fenchel Inequalities

Let K_0 and K_1 be convex bodies in \mathbb{R}^d , and let $K_\lambda = (1 - \lambda)K_0 + \lambda K_1$. Then there exist unique numbers a_0, a_1, \dots, a_d so that the volume of K_λ is given by:

$$\text{Vol}(K_\lambda) = \sum_{i=0}^d \binom{d}{i} a_i (1 - \lambda)^{d-i} \lambda^i$$

Furthermore, the sequence $\{a_i : 0 \leq i \leq d\}$ is log-concave, i.e.,

$$a_{i+1}^2 \geq a_i a_{i+2} \quad \text{for } i = 0, 1, \dots, d - 2.$$

Height Sequences are Log-Concave

Notation. Let $x \in P$ and let $h_i = \text{Prob}[h_L(x) = i]$.

Height Sequences are Log-Concave

Notation. Let $x \in P$ and let $h_i = \text{Prob}[h_L(x) = i]$.

Theorem. [Stanley] For any poset P and any $x \in P$, the height sequence $\{h_i : 1 \leq i \leq |P|\}$ is log-concave, i.e.,

$$h_{i+1}^2 \geq h_i h_{i+2} \quad \text{for } i = 0, 1, \dots, |P| - 2.$$

Height Sequences are Log-Concave

Notation. Let $x \in P$ and let $h_i = \text{Prob}[h_L(x) = i]$.

Theorem. [Stanley] For any poset P and any $x \in P$, the height sequence $\{h_i : 1 \leq i \leq |P|\}$ is log-concave, i.e.,

$$h_{i+1}^2 \geq h_i h_{i+2} \quad \text{for } i = 0, 1, \dots, |P| - 2.$$

Remark. No combinatorial proof of this result is known.

Differential Height Sequences are Log-Concave

Remark. Kahn and Saks extended Stanley's technique to show that $\{a_i : i > 0\}$ and $\{b_i : i > 0\}$ are log-concave, i.e.,

$$a_{i+1}^2 \geq a_i a_{i+2} \quad \text{and} \quad b_{i+1}^2 \geq b_i b_{i+2}$$

Differential Height Sequences are Log-Concave

Remark. Kahn and Saks extended Stanley's technique to show that $\{a_i : i > 0\}$ and $\{b_i : i > 0\}$ are log-concave, i.e.,

$$a_{i+1}^2 \geq a_i a_{i+2} \quad \text{and} \quad b_{i+1}^2 \geq b_i b_{i+2}$$

Remark. Adding these non-linear constraints to the linear ones listed previously yields the desired inequality:

$$\frac{3}{11} \leq \text{Prob}[x < y] \leq \frac{8}{11}$$

Improving the Pigeon Hole

Theorem. [Felsner and Trotter] *Let x and y be distinct points in a poset P with $|h(y) - h(x)| \leq 2/3$. Then*

$$\frac{1}{3} \leq \text{Prob}[x < y] \leq \frac{2}{3}$$

Improving the Pigeon Hole

Theorem. [Felsner and Trotter] *Let x and y be distinct points in a poset P with $|h(y) - h(x)| \leq 2/3$. Then*

$$\frac{1}{3} \leq \text{Prob}[x < y] \leq \frac{2}{3}$$

Definition. *For a finite poset P , let $\gamma(P)$ denote the minimum value of $|h(y) - h(x)|$ taken over all pairs $x, y \in P$.*

Improving the Pigeon Hole

Theorem. [Felsner and Trotter] *Let x and y be distinct points in a poset P with $|h(y) - h(x)| \leq 2/3$. Then*

$$\frac{1}{3} \leq \text{Prob}[x < y] \leq \frac{2}{3}$$

Definition. *For a finite poset P , let $\gamma(P)$ denote the minimum value of $|h(y) - h(x)|$ taken over all pairs $x, y \in P$.*

Remark. [Saks, 85] *There exists a poset P with $\gamma(P) \sim 0.8657$.*

Could $3/11-8/11$ be Tight

Perhaps the $1/3-2/3$ conjecture is false and $3/11-8/11$ is the right answer. This would require that for every $\epsilon > 0$, we can find a poset P with $\gamma(P) > 1 - \epsilon$.

Could $3/11-8/11$ be Tight

Perhaps the $1/3-2/3$ conjecture is false and $3/11-8/11$ is the right answer. This would require that for every $\epsilon > 0$, we can find a poset P with $\gamma(P) > 1 - \epsilon$.

Can this happen? And even if it does, can we use another line of reasoning to improve the Kahn/Saks bound?

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

1. *P has width 2 (Linial).*

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

1. *P has width 2 (Linial).*
2. *P is a semiorder (Brightwell).*

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

- 1. P has width 2 (Linial).*
- 2. P is a semiorder (Brightwell).*
- 3. For every $x \in P$, there are at most 5 elements incomparable with x (Brightwell and Wright).*

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

1. *P has width 2 (Linial).*
2. *P is a semiorder (Brightwell).*
3. *For every $x \in P$, there are at most 5 elements incomparable with x (Brightwell and Wright).*
4. *$\gamma(P) \leq 2/3$ (Felsner and Trotter).*

Progress on the $1/3$ – $2/3$ Conjecture

Theorem. *The $1/3$ – $2/3$ conjecture holds for any poset P which satisfies any one of the following conditions:*

1. *P has width 2 (Linial).*
2. *P is a semiorder (Brightwell).*
3. *For every $x \in P$, there are at most 5 elements incomparable with x (Brightwell and Wright).*
4. *$\gamma(P) \leq 2/3$ (Felsner and Trotter).*
5. *P has height 2 (Fishburn, Gehrlein and Trotter).*

A Conjecture on Large Width

Conjecture. [Kahn and Saks] *For every $\epsilon > 0$, there exists an integer w so that if the width of P is at least w , then there exists a distinct pair x, y in P with*

$$\frac{1}{2} - \epsilon < \text{Prob}[x < y] < \frac{1}{2} + \epsilon$$

A Conjecture on Large Width

Conjecture. [Kahn and Saks] For every $\epsilon > 0$, there exists an integer w so that if the width of P is at least w , then there exists a distinct pair x, y in P with

$$\frac{1}{2} - \epsilon < \text{Prob}[x < y] < \frac{1}{2} + \epsilon$$

Theorem. [Komlós] There exists a function $f : N \rightarrow N$ with $f(n) \rightarrow \infty$ and $f(n) = o(n)$ so that for every $\epsilon > 0$, there exists an integer n so that if P has n elements and at least $f(n)$ maximal elements, then there exists a distinct pair x, y in P with

$$\frac{1}{2} - \epsilon < \text{Prob}[x < y] < \frac{1}{2} + \epsilon$$

Just How Important is “Finite”

Just How Important is “Finite”

Theorem. [Brightwell and Trotter] *The $1/3$ – $2/3$ conjecture is **FALSE** for infinite posets. In fact, there exists a countably infinite poset P satisfying:*

1. P has width 2;
2. P is a semiorder;
3. For every $x \in P$, there are at most 2 elements incomparable with x ;
4. $\gamma(P) = 1$; and
5. For every incomparable pair x, y ,

$$\text{Prob}[x < y] \in \left\{ \frac{5 - \sqrt{5}}{10}, \frac{5 + \sqrt{5}}{10} \right\}$$

Incremental Progress

Theorem. [Felsner and Trotter, 96] *There exists a constant $\delta > 0$ so that if \mathbf{P} is any poset which is not a chain, then \mathbf{P} contains a distinct pair x, y so that*

$$\frac{3}{11} + \delta < \text{Prob}[x < y] < \frac{8}{11} - \delta$$

Incremental Progress

Theorem. [Felsner and Trotter, 96] *There exists a constant $\delta > 0$ so that if \mathbf{P} is any poset which is not a chain, then \mathbf{P} contains a distinct pair x, y so that*

$$\frac{3}{11} + \delta < \text{Prob}[x < y] < \frac{8}{11} - \delta$$

Remark. *We didn't bother to calculate δ since it was very very small!*

Incremental Progress

Theorem. [Felsner and Trotter, 96] *There exists a constant $\delta > 0$ so that if \mathbf{P} is any poset which is not a chain, then \mathbf{P} contains a distinct pair x, y so that*

$$\frac{3}{11} + \delta < \text{Prob}[x < y] < \frac{8}{11} - \delta$$

Remark. *We didn't bother to calculate δ since it was very very small!*

Remark. *This research made extensive use of computation in both the discovery and proof modes.*

The Cross Product Conjecture

Conjecture. [Felsner and Trotter] *Let x, y and z be three distinct points in a poset P . For positive integers i, j , define*

$$a(i, j) = \text{Prob}[h_L(y) - h_L(x) = i, \quad h_L(z) - h_L(y) = j]$$

The Cross Product Conjecture

Conjecture. [Felsner and Trotter] Let x, y and z be three distinct points in a poset P . For positive integers i, j , define

$$a(i, j) = \text{Prob}[h_L(y) - h_L(x) = i, \quad h_L(z) - h_L(y) = j]$$

Then

$$a(i, j)a(i + 1, j + 1) \leq a(i, j + 1)a(i + 1, j)$$

The Cross Product Conjecture

Conjecture. [Felsner and Trotter] Let x , y and z be three distinct points in a poset P . For positive integers i , j , define

$$a(i, j) = \text{Prob}[h_L(y) - h_L(x) = i, \quad h_L(z) - h_L(y) = j]$$

Then

$$a(i, j)a(i + 1, j + 1) \leq a(i, j + 1)a(i + 1, j)$$

Theorem. [Brightwell, Felsner and Trotter, 97] The Cross Product Conjecture holds when $i = j = 1$, i.e.,

$$a(1, 1)a(2, 2) \leq a(1, 2)a(2, 1).$$

Improving Kahn/Saks

Theorem. [Brightwell, Felsner and Trotter] *Let P be a countable poset which is not a chain. If there exists an integer k so that any element of P is incomparable with at most k others, then there exist distinct x, y with*

$$\frac{5 - \sqrt{5}}{10} \leq \text{Prob}[x < y] \leq \frac{5 + \sqrt{5}}{10}$$

Improving Kahn/Saks

Theorem. [Brightwell, Felsner and Trotter] *Let P be a countable poset which is not a chain. If there exists an integer k so that any element of P is incomparable with at most k others, then there exist distinct x, y with*

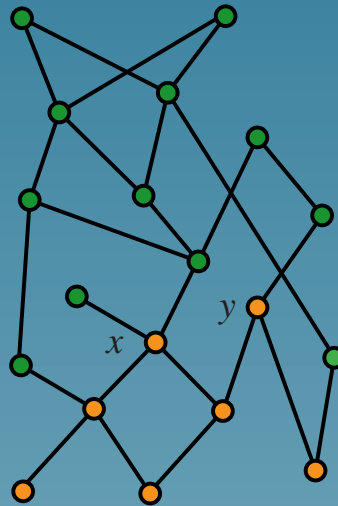
$$\frac{5 - \sqrt{5}}{10} \leq \text{Prob}[x < y] \leq \frac{5 + \sqrt{5}}{10}$$

$$\frac{3}{11} = .272727 \dots$$
$$\frac{5 - \sqrt{5}}{10} = .27639 \dots$$

Kahn's Problem

Problem. Let x and y be distinct point in a poset P and suppose that $|D[x] \cup D[y]| = n$. Is it true that

$$\max\{h(x), h(y)\} \geq n - 1 \quad ?$$



Observations

Remark. *Kahn's conjecture is true when $P = D[x] \cup D[y]$. This follows from the fact that*

$$\max\{\text{Prob}[h_L(x) = n], \text{Prob}[h_L(y) = n]\} \geq \frac{1}{2}$$

Observations

Remark. *Kahn's conjecture is true when $P = D[x] \cup D[y]$. This follows from the fact that*

$$\max\{\text{Prob}[h_L(x) = n], \text{Prob}[h_L(y) = n]\} \geq \frac{1}{2}$$

The argument depends on the log-concavity property of the height sequence.

Observations

Remark. *Kahn's conjecture is true when $P = D[x] \cup D[y]$. This follows from the fact that*

$$\max\{\text{Prob}[h_L(x) = n], \text{Prob}[h_L(y) = n]\} \geq \frac{1}{2}$$

The argument depends on the log-concavity property of the height sequence.

Remark. *When*

$$D[x] \cup D[y] \subsetneq P$$

there seems to be nothing which would prevent $\max\{h(x), h(y)\}$ from being far below n .

Error Analysis on Log-Concavity

Let $\{h_i : 1 \leq i \leq n\}$ be the height sequence of a point x in a poset P . Consider the log-concave inequality

$$h_{i+1}^2 \geq h_i h_{i+2}.$$

Error Analysis on Log-Concavity

Let $\{h_i : 1 \leq i \leq n\}$ be the height sequence of a point x in a poset P . Consider the log-concave inequality

$$h_{i+1}^2 \geq h_i h_{i+2}.$$

1. When is the inequality tight?

Error Analysis on Log-Concavity

Let $\{h_i : 1 \leq i \leq n\}$ be the height sequence of a point x in a poset P . Consider the log-concave inequality

$$h_{i+1}^2 \geq h_i h_{i+2}.$$

1. When is the inequality tight?
2. When it is not tight, what is the magnitude of the minimum error term as a function of n ?

Correlation: The XYZ Theorems

Theorem. [Shepp] *Let x, y and z be distinct points in a poset P . Then*

$$\text{Prob}[x > y] \text{Prob}[x > z] \leq \text{Prob}[x > y, x > z].$$

Correlation: The XYZ Theorems

Theorem. [Shepp] *Let x, y and z be distinct points in a poset P . Then*

$$\text{Prob}[x > y] \text{Prob}[x > z] \leq \text{Prob}[x > y, x > z].$$

Theorem. [Fishburn] *Let $\{x, y, z\}$ be a 3-element antichain in a poset P . Then*

$$\text{Prob}[x > y] \text{Prob}[x > z] < \text{Prob}[x > y, x > z].$$

Correlation: The XYZ Theorems

Theorem. [Shepp] *Let x, y and z be distinct points in a poset P . Then*

$$\text{Prob}[x > y] \text{Prob}[x > z] \leq \text{Prob}[x > y, x > z].$$

Theorem. [Fishburn] *Let $\{x, y, z\}$ be a 3-element antichain in a poset P . Then*

$$\text{Prob}[x > y] \text{Prob}[x > z] < \text{Prob}[x > y, x > z].$$

Remark. *Both proofs require the Ahlswede/Daykin four functions theorem.*

The Ahlswede/Daykin Four Functions Theorem

Theorem. [Ahlswede and Daykin] *Let L be a distributive lattice. For sets X , Y and a function f , let*

1. $f(X) = \sum_{x \in X} f(x)$.
2. $X \wedge Y = \{x \wedge y : x \in X, y \in Y\}$.
3. $X \vee Y = \{x \vee y : x \in X, y \in Y\}$.

Let α , β , γ and δ be four functions mapping L to the non-negative reals. If

$$\alpha(x)\beta(y) \leq \gamma(x \vee y)\delta(x \wedge y)$$

for all $x, y \in L$, then

$$\alpha(X)\beta(Y) \leq \gamma(X \vee Y)\delta(X \wedge Y)$$

for all $X, Y \subseteq L$.

Fishburn's Lemma

Lemma. [Fishburn] *Let A and B be down-sets in a poset P with $|A| = n$, $|B| = m$ and $|A \cap B| = k$. Then*

$$|\mathcal{E}(A)| |\mathcal{E}(B)| \binom{n+m}{n} \leq |\mathcal{E}(A \cup B)| |\mathcal{E}(A \cap B)| \binom{n+m}{n+m-k}$$

Fishburn's Lemma

Lemma. [Fishburn] *Let A and B be down-sets in a poset P with $|A| = n$, $|B| = m$ and $|A \cap B| = k$. Then*

$$|\mathcal{E}(A)| |\mathcal{E}(B)| \binom{n+m}{n} \leq |\mathcal{E}(A \cup B)| |\mathcal{E}(A \cap B)| \binom{n+m}{n+m-k}$$

Remark. *Fishburn's argument also uses the Ahlswede/Daykin theorem.*

A Combinatorial Approach to Correlation

In 1999, Brightwell and Trotter gave a combinatorial proof of Fishburn's lemma by providing an explicit injection between two sets of appropriate sizes.

A Combinatorial Approach to Correlation

In 1999, Brightwell and Trotter gave a combinatorial proof of Fishburn's lemma by providing an explicit injection between two sets of appropriate sizes.

They then extended this approach to give combinatorial proofs for the XYZ correlation results of Shepp and Fishburn.

A Combinatorial Approach to Correlation

In 1999, Brightwell and Trotter gave a combinatorial proof of Fishburn's lemma by providing an explicit injection between two sets of appropriate sizes.

They then extended this approach to give combinatorial proofs for the XYZ correlation results of Shepp and Fishburn.

These new arguments do not use the Ahlswede/Daykin theorem or any of its variant forms.

Ahlswede/Daykin and Stanley

Some modest progress to report:

Suppose x is point in a poset P so that x is incomparable with exactly two other points of P . Then there are three non-zero terms in the height sequence of x . If these terms are h_i , h_{i+1} and h_{i+2} , then it is possible to prove that

$$h_{i+1}^2 \geq h_i h_{i+2}$$

via an entirely combinatorial argument.

Ahlswede/Daykin and Stanley

Some modest progress to report:

Suppose x is point in a poset P so that x is incomparable with exactly two other points of P . Then there are three non-zero terms in the height sequence of x . If these terms are h_i , h_{i+1} and h_{i+2} , then it is possible to prove that

$$h_{i+1}^2 \geq h_i h_{i+2}$$

via an entirely combinatorial argument.

However, the proof does require the Ahlswede/Daykin four functions theorem.