Convex rank tests

Anne Shiu
Texas A&M University

CombinaTexas
8 May 2016
OUTLINE OF TALK

- Introduction
- Main Result: \textit{convex rank tests} = \textit{semigraphoids}
- 2 counterexamples
- Application to biology

Joint work with Raymond Hemmecke, Jason Morton, Lior Pachter, Bernd Sturmfels, and Oliver Wienand.
Outline of Talk

- Introduction
- Main Result: \textit{convex rank tests} = \textit{semigraphoids}
- 2 counterexamples
- Application to biology

Joint work with Raymond Hemmecke, Jason Morton, Lior Pachter, Bernd Sturmfels, and Oliver Wienand.
Introduction
A fan in $\mathbb{R}^n$ is a finite collection $\mathcal{F}$ of polyhedral cones such that:

- if $C \in \mathcal{F}$ and $C'$ is a face of $C$, then $C' \in \mathcal{F}$, and
- if $C, C' \in \mathcal{F}$, then $C \cap C'$ is a face of $C$.

The $S_n$-arrangement (the braid arrangement) is the arrangement of hyperplanes $\{x_i = x_j\}$ in $\mathbb{R}^n$.

Example: the fan associated to the $S_3$-arrangement has 6 maximal cones.

\[ x_1 = x_3 \]
\[ x_1 = x_2 \]
What is a convex rank test?

- A rank test is a partition of $S_n$.
- A convex rank test is a partition of $S_n$ defined by a fan that coarsens the $S_n$-arrangement.
- Example: the following convex rank test partitions $S_3$ into 4 classes.

```
123  132  
213  312  (x_3 > x_1 > x_2)
231  321
```
A non-convex rank test

- This partition of $S_3$ into 4 classes is not a convex rank test.

- Remark: a convex rank test is determined by the walls removed from the $S_n$-arrangement.
Label walls by conditional-independence statements

- Two maximal cones of the $S_n$-fan, labeled by permutations $\delta$ and $\delta'$ in $S_n$, share a **wall** if $\delta$ and $\delta'$ differ by an *adjacent transposition*: there exists an index $k$ such that $\delta_k = \delta'_k + 1$, $\delta_{k+1} = \delta'_k$, and $\delta_i = \delta'_i$ for $i \neq k, k+1$.

- Label wall $\{\delta, \delta'\}$ by the **conditional-independence (CI) statement**:
  \[
  \delta_k \independent \delta_{k+1} \mid \{\delta_1, \ldots, \delta_{k-1}\}.
  \]
**Conditional independence**

Consider a collection of $n$ random variables indexed by $[n]$.

\[
\begin{align*}
[1 \perp \perp 2|\emptyset] & \quad [1 \perp \perp 3|\emptyset] & \quad [2 \perp \perp 3|\emptyset] & \quad [2 \perp \perp 3|1] & \quad [1 \perp \perp 3|2] & \quad [1 \perp \perp 2|3] \\
[1 \perp \perp 4|\emptyset] & \quad [2 \perp \perp 4|\emptyset] & \quad [3 \perp \perp 4|\emptyset] & \quad [1 \perp \perp 2|4] & \quad [1 \perp \perp 3|4] & \quad [2 \perp \perp 3|4] \\
[2 \perp \perp 4|1] & \quad [3 \perp \perp 4|1] & \quad [1 \perp \perp 4|2] & \quad [3 \perp \perp 4|2] & \quad [1 \perp \perp 4|3] & \quad [2 \perp \perp 4|3] \\
[1 \perp \perp 2|34] & \quad [1 \perp \perp 3|24] & \quad [1 \perp \perp 4|23] & \quad [2 \perp \perp 3|14] & \quad [2 \perp \perp 4|13] & \quad [3 \perp \perp 4|12] \\
[1 \perp \perp 5|\emptyset] & \quad [2 \perp \perp 5|\emptyset] & \quad \ldots & \quad [4 \perp \perp 5|123] & \quad \ldots
\end{align*}
\]

The symbol $[i \perp \perp j|K]$ represents the statement, “the random variables $i$ and $j$ are conditionally independent given the joint random variable $K$.”
Semigraphoids

(definition #1) A set $\mathcal{M}$ of CI statements on $[n]$ is a **semigraphoid** if the following axiom holds$^1$:

*(SG)* If $[i \perp \perp j \mid K \cup \ell]$ and $[i \perp \perp \ell \mid K]$ are in $\mathcal{M}$ then also $[i \perp \perp j \mid K]$ and $[i \perp \perp \ell \mid K \cup j]$ are in $\mathcal{M}$.

---

$^1$ *Probabilistic Conditional Independence Structures, Studený 2005*
Semigraphoids

(definition #1) A set $\mathcal{M}$ of CI statements on $[n]$ is a **semigraphoid** if the following axiom holds\(^1\):

(SG) If $[i \perp \perp j | K \cup \ell]$ and $[i \perp \perp \ell | K]$ are in $\mathcal{M}$ then also $[i \perp \perp j | K]$ and $[i \perp \perp \ell | K \cup j]$ are in $\mathcal{M}$.

Example:

(SG) If $[1 \perp \perp 2 | 3]$ and $[1 \perp \perp 3 | \emptyset]$ are in $\mathcal{M}$, then also $[1 \perp \perp 2 | \emptyset]$ and $[1 \perp \perp 3 | 2]$ are in $\mathcal{M}$.

So, $\mathcal{M} = \{ [1 \perp \perp 3 | \emptyset], [1 \perp \perp 2 | 3] \}$ is not a semigraphoid.

---

\(^1\) Probabilistic Conditional Independence Structures, Studený 2005
Main result
Main Theorem

- A convex rank test $\mathcal{F}$ is characterized by the collection of walls $\{\delta, \delta'\}$ that are removed from the $S_n$-arrangement. Let $\mathcal{M}_{\mathcal{F}}$ denote the CI statements that label those walls.

- Main theorem: The map $\mathcal{F} \mapsto \mathcal{M}_{\mathcal{F}}$ is a bijection between convex rank tests and semigraphoids.

- The following convex rank test corresponds to the semigraphoid $\mathcal{M} = \{ 1 \perp \perp 3|\emptyset, 1 \perp \perp 3|2 \}$. 

```
1 2 3
```

```
\begin{itemize}
  \item 1 \perp \perp 3|\emptyset
  \item 1 \perp \perp 3|\{2\}
\end{itemize}
```
Restating the Main result via the permutohedron
The Permutohedron

- The fan of the $S_n$-arrangement is the normal fan of the permutohedron $P_n$ (the convex hull of the vectors $(\rho_1, \ldots, \rho_n)$, where $\rho$ is in $S_n$).

- The edges of the permutohedron correspond to walls of the $S_n$-arrangement.
The permutohedron $P_4$

The 2-d faces of $P_n$ are squares and hexagons.
**Square and Hexagon Axioms**

**Lemma:** A set $M$ of edges of the permutohedron $P_n$ is a semigraphoid if and only if $M$ satisfies the following two axioms:

- **Square axiom:** Whenever an edge of a square is in $M$, then the opposite edge is also in $M$.
- **Hexagon axiom:** When two adjacent edges of a hexagon are in $M$, then the two opposite edges are also in $M$.

![Diagram of permutohedron with labels 123, 132, 213, 312, 231, 321]

**Main theorem, restated.**

*Coarsenings of the $S_n$-fan* are equivalent to subsets of edges of $P_n$ that satisfy the Square and Hexagon axioms.

**Generalization to other Coxeter arrangements.**

*Coarsenings = subsets of edges with the polygon property.*

(Nathan Reading 2012).
Consider $\mathbf{M} = \{1 \perp 3|\emptyset, 1 \perp 2|\{3\}\}$ (again). It is not a convex rank test, because it violates the Hexagon axiom:
Main theorem illustrated

\[ f = (16, 24, 10) \]
2 COUNTEREXAMPLES
Semigraphoids: another definition

- Each CI statement defines a linear form in $2^n$ unknowns $h_I$ for $I \subseteq [n]$:

  $$[i \perp j \mid K] \mapsto -h_{ijK} + h_{iK} + h_{jK} - h_K.$$ 

- Non-negativity of these linear forms defines the $(2^n - n - 1)$-dimensional submodular cone in $\mathbb{R}^{2^n}$.

- The linear relations among the forms are spanned by entropy equations:

  $$[i \perp j \mid K \cup \ell] + [i \perp \ell \mid K] = [i \perp j \mid K] + [i \perp \ell \mid K \cup j].$$
Semigraphoids: Another Definition

- Each CI statement defines a linear form in $2^n$ unknowns $h_I$ for $I \subseteq [n]$

\[ [i \perp j \mid K] \mapsto -h_{ijK} + h_{iK} + h_{jK} - h_K. \]

- Non-negativity of these linear forms defines the $(2^n - n - 1)$-dimensional submodular cone in $\mathbb{R}^{2^n}$.

- The linear relations among the forms are spanned by entropy equations:

\[ [i \perp j \mid K \cup \ell] + [i \perp \ell \mid K] = [i \perp j \mid K] + [i \perp \ell \mid K \cup j]. \]

- (Definition #4) A semigraphoid $\mathcal{M}$ specifies the possible zeros for a non-negative solution of the entropy equations.

- A semigraphoid $\mathcal{M}$ is submodular if it is the set of actual zeros of a point in the submodular cone.
Postnikov, Reiner and Williams (2006) asked: Is every simplicial fan which coarsens the $S_n$-fan the normal fan of convex polytope?

**Facts.** A convex rank test $\mathcal{F}$ is the normal fan of a polytope if and only if the semigraphoid $\mathcal{M}_\mathcal{F}$ is submodular. This polytope is a generalized permutohedron. It is simple iff $\mathcal{F}$ is simplicial iff the posets on $[n]$ are trees.

The answer to the PRW question is no for $n = 4$:

**Proposition.** This is simplicial, but not submodular:
**Proof: simplicial**

This simple polytope looks like a generalized permutohedron...

\[
\mathcal{M}_F = \{[2 \perp 3 | 14], [1 \perp 4 | 23], [1 \perp 2 | \emptyset], [3 \perp 4 | \emptyset]\}.
\]
**Proof: not submodular**

... but, it is not a generalized permutohedron.

\[
\begin{align*}
[1 \perp 2 | \emptyset] + [2 \perp 3 | 1] &= [1 \perp 2 | 3] + [2 \perp 3 | \emptyset] \\
[3 \perp 4 | \emptyset] + [1 \perp 4 | 3] &= [3 \perp 4 | 1] + [1 \perp 4 | \emptyset] \\
[2 \perp 3 | 14] + [3 \perp 4 | 1] &= [2 \perp 3 | 1] + [3 \perp 4 | 12] \\
[1 \perp 4 | 23] + [1 \perp 2 | 3] &= [1 \perp 4 | 3] + [1 \perp 2 | 34]
\end{align*}
\]

If \( M_\mathcal{F} \) were submodular, there would be a solution where the blue unknowns are zero and the others are positive. Adding both left- and right-hand sides yields

\[
[2 \perp 3 | \emptyset] + [1 \perp 4 | \emptyset] + [3 \perp 4 | 12] + [1 \perp 2 | 34] = 0.
\]

Contradiction!
For $n = 3$, there are 22 semigraphoids.

For $n = 4$, there are 26424 semigraphoids but only 22108 of them are submodular.

For $n \geq 5$, Studený posed many questions, including:

- Is every maximal semigraphoid submodular?

The answer is no.
NON-SUBMODULAR, BUT MAXIMAL
Everyone loves graphs

- We saw: submodular semigraphoids = generalized permutohedra.

- In statistics, the most popular semigraphoids are graphical models.

- In mathematics, the most popular polytopes are the graph associahedra (Stasheff, Bott-Taubes, ...)

- Theorem. Graphical models = graph associahedra.

- For the biological application which started all this, the corresponding graphical rank tests worked best ...
APPLICATION: BIOLOGICAL CLOCKS
Biological clocks

- Somitogenesis: process during embryonic development in vertebrates in which the somites (precursors to the segments of the backbone) are formed.

- Which genes control this molecular clock?

- Olivier Pourquié lab at the Stowers Institute, now Harvard

Search for cyclic genes

- **Microarray** experiments- a microarray chip can measure the gene expression level of tens of thousands of genes simultaneously.
- 17 experiments conducted within one cycle
- Example: the expression level of gene Axin2
  
  \((0.34204059, 0.195306068, 0.151584691, 0.215046787, -0.238626783, -0.380163626, -0.431032137, -0.41198219, -0.36420852, -0.317375356, -0.141293099, -0.191303023, 0.085202023, 0.420653258, 0.300682397, -0.002791647, 0.281696744)\in \mathbb{R}^{17}

- Its *rank vector*: \((16, 12, 11, \ldots, 14)\in S_{17}

- Convex rank test as a *statistical test*...
One convex rank test: Up-down analysis
Another test: Cyclohedron

Figure: $M_{\mathcal{F}} = \{[1 \bot 3|\emptyset], [2 \bot 4|\emptyset]\}$. 
**Cyclohedron test for gene Obox**

**Figure:** The cyclohedron test smooths the data; shown are the data vector $v$ and the height vector $h(v)$. How many permutations share a height vector?

**Result:** We identified this and other genes to be possibly part of the biological clock.
Conclusion

Summary theorem.
Convex rank tests = semigraphoids = edges of the permutohedron that satisfy the square and hexagon axioms.

Combinatorics helped us answer some questions from statistics and biology.
Thank you.
Proof of Theorem

Lemma

If $\mathcal{M}$ is a semigraphoid, then if $\delta$ and $\delta'$ lie in the same class of $\mathcal{M}$, then so do all shortest paths on $\mathbb{P}_n$ between them.

Lemma $\Rightarrow$ A semigraphoid is a pre-convex rank test.
Proof (continued)

Now, we see that a semigraphoid corresponds to a fan (convex rank test):

\[ x_i = x_j \]

Conversely, it is easy to show that a convex rank test satisfies the square and hexagon axioms.