# MS-E1621, Algebraic Statistics

Fill out http://presemo.aalto.fi/aslec1

# Goals

After the course, you can:

- list topics in algebraic statistics
- recognize problems in statistics that are answerable by algebraic methods
- assess which algebraic methods are suitable for solving a problem
- apply basic algebraic tools to solve a problem

# Material

- Main textbook: Seth Sullivant "Algebraic Statistics" (e-book available through Aalto library)
- I will upload slides/worksheets after each lecture
- Background on algebra: Cox, Little, O'Shea "Ideals, Varieties, and Algorithms"
- See MyCourses/Material for additional books
- Algebraic Statistics Seminar every two weeks

# Lectures

- Lecture 1: Algebra
- Lecture 2: Probability (reading pre-task)
- Lecture 3: Conditional independence
- Lecture 4: Statistics (reading pre-task)
- Lecture 5: Exponential families
- Lecture 6: Likelihood inference
- Fisher's exact test (reading task)
- Graphical models
- Group presentations on selected chapters (topics around lecture 4 and 5)

# Homework

- There will be six homework sets.
- Deadline for submitting homework is Fridays at 6pm (not every Friday since the course runs through two periods).
- Submissions only through MyCourses.
- It is recommended that solutions are typed.
- You can resubmit two weeks after we return the solutions.
- There will be additional reading assignments and group work.

# Exercise sessions

- Exercise session 1: Introduction to Macaulay2
- Exercise session 2, 4, 6 etc: You have the possibility to discuss in groups anything that remains unsolved and include the results of the group work in their solutions. Everyone has to write up their own solutions.
- Exercise sessions 3, 5, 7 etc: No organized activity. Possibility to ask questions.

# Grade

- This course is graded pass/fail.
- For passing the course, one has to
  - attend at least 10 lectures,
  - receive at least 70% of maximal possible points on homework sets and
  - complete all additional assignments.
- Additional reading assignments and the group project are not graded by points.
- There is no exam.

# Communication

- The communication for this course takes place in Zulip.

  https://algstat-mse1621.zulipchat.com

- There is a separate channel for each problem set and assignment.
- Please be active asking your questions!

Any questions about the organization?

# Motivating example 1: Discrete Markov chain

- Let $X_1, X_2, X_3$ be a sequence of random variables taking values in $\Sigma = \{0,1\}$
- There are 8 joint probabilities $p_{ijk} = P(X_1 = i, X_2 = j, X_3 = k)$ where $i, j, k \in \{0,1\}$
- A probability distribution associated to $X_1, X_2, X_3$ corresponds to a point in $\mathbb{R}^8$
- The sequence $X_1, X_2, X_3$ is a Markov chain if

$$P(X_3 = x_3 \,|\, X_1 = x_1, X_2 = x_2) = P(X_3 = x_3 \,|\, X_2 = x_2)$$

- When is a point in $\mathbb{R}^8$ the probability distribution associated to a Markov chain?

# Motivating example 1: Discrete Markov chain

- Conditional probabilities can be expressed in terms of the joint probabilities:

$$P(X_3 = k \mid X_1 = i, X_2 = j) = \frac{p_{ijk}}{p_{ij+}}, \text{ where } p_{ij+} = \sum_{k \in \{0,1\}} p_{ijk}$$

- Markov chain condition: $\frac{p_{ijk}}{p_{ij+}} = \frac{p_{+jk}}{p_{+j+}}$
- This gives $\frac{p_{ijk}}{p_{ij+}} = \frac{p_{i'jk}}{p_{i'j+}}$ (consider previous equality for $i$ and $i'$)
- Simplifying gives $p_{000}p_{101} - p_{001}p_{100} = 0$ and $p_{010}p_{111} - p_{011}p_{110} = 0$

# Motivating example 1: Discrete Markov chain

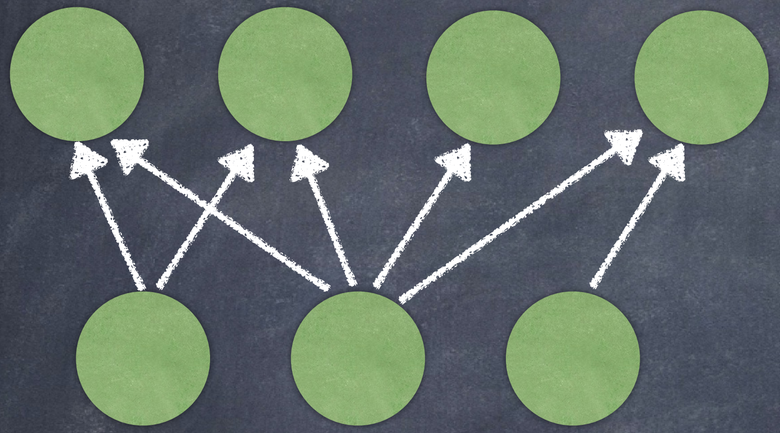
- A point $p \in \mathbb{R}^8$ is the probability distribution associated to a Markov chain if and only if
  - $p_{ijk} \geq 0$ for all $i, j, k \in \{0,1\}$,
  - $\sum_{i,j,k \in \{0,1\}} p_{ijk} = 1$
  - $p_{000}p_{101} - p_{001}p_{100} = 0$
  - $p_{010}p_{111} - p_{011}p_{110} = 0$

This Markov chain model is a semialgebraic set: It is a solution set of a system of polynomial equations and inequalities.

# Motivating example 1: Discrete Markov chain

- This is an example of a conditional independence model (Lecture 3)
- Fitting the model to data: Assuming there is a true unknown probability distribution $p$ in our model from which our data is generated. What is $p$? (Likelihood inference in Lecture 6)
- How well does the model fit the data? (Fisher's exact test in Lecture 7)

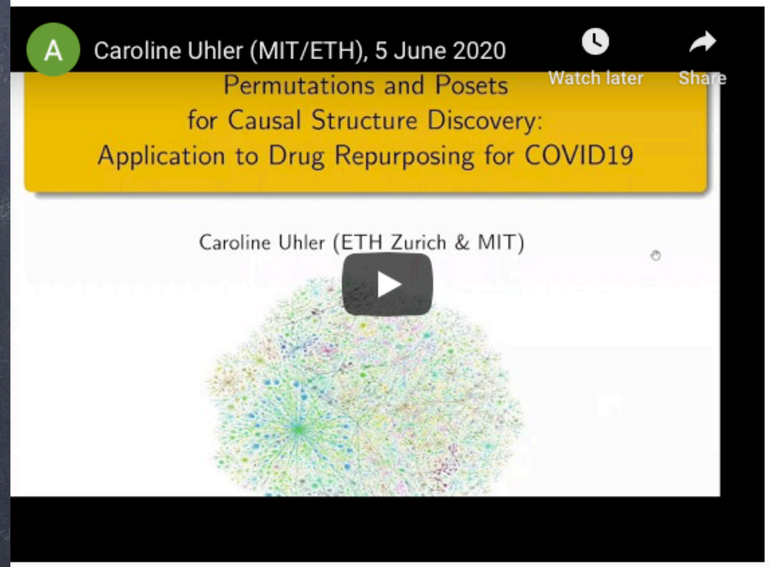
# Motivating example 2: Graphical models

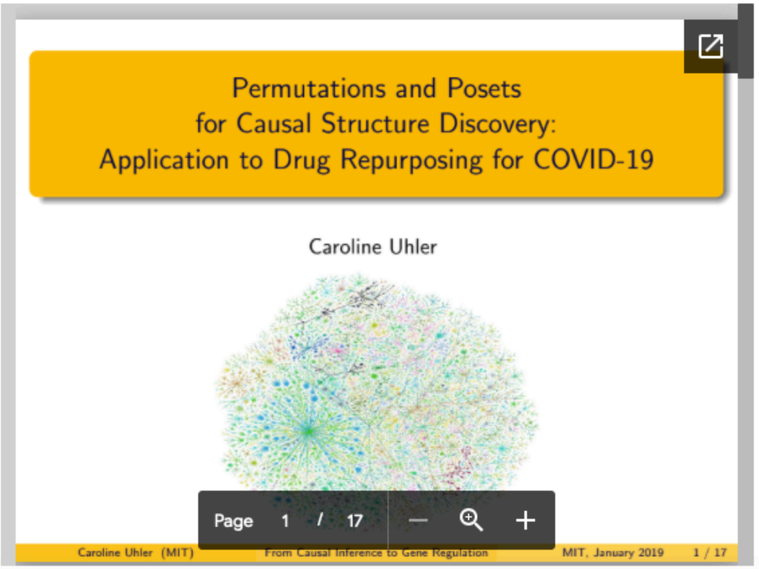


Algebraic Statistics Seminar Online -> Past Talks and Recordings

## Friday, June 5, 2020: Caroline Uhler (MIT/ETH)

- **Speaker:** Caroline Uhler (MIT/ETH)
- **Title:** Permutations and Posets for Causal Structure Discovery
- **Abstract:** Gene knockout experiments allow performing interventions in large-scale systems. This represents a unique opportunity for causal structure discovery, since it allows testing algorithms with real data and in relevant settings. We discuss the rich combinatorial, algebraic and geometric questions underlying causal structure discovery. In particular, we show that viewing causal structure discovery as an optimization problem over permutations (in the fully observed setting) or posets (in the presence of unobserved variables) can lead to algorithms with stronger consistency guarantees than previously known, which translates into better performance in terms of predicting the effect of a gene knockout experiment.



# Polynomials

- field $\mathbb{K}$ (usually $\mathbb{Q}, \mathbb{R}, \mathbb{C}$)
- polynomial variables or indeterminates $p_1, p_2, \ldots, p_r$
- monomial: $p^u := p_1^{u_1} p_2^{u_2} \cdots p_r^{u_r}$ where $u = (u_1, u_2, \ldots, u_r) \in \mathbb{N}^r$
- polynomial in $p_1, p_2, \ldots, p_r$ over $\mathbb{K}$: $f = \sum_{u \in A} c_u p^u$ where $A$ is a finite subset of $\mathbb{N}^r$ and each coefficient $c_u \in \mathbb{K}$
- polynomial ring: $\mathbb{K}[p] := \mathbb{K}[p_1, p_2, \ldots, p_r]$

# Worksheet

- You will work on worksheets in groups of 3-4 persons
- In MyCourses/Worksheets choose the worksheet according to the Breakout room number
- Edit the worksheet together with your group members in Overleaf
- You can use any tools you want (Mathematica etc)
- Ardi, Luca, Olga and myself will help you
- No worries if you cannot solve everything - last exercises will go to Homework 1

# Algebraic varieties

Def: Let $S \subset \mathbb{K}[p]$ be a set of polynomials. The variety defined by $S$ is

$$V(S) = \{a \in \mathbb{K}^r : f(a) = 0 \quad \forall f \in S\}.$$

- The variety $V(S)$ is also called the zero set of $S$.
- A variety depends on the field.
- Often the field is clear from the context. If want to emphasize the field, then write $V_{\mathbb{K}}(S)$.

# Ideals

Def: A subset $I$ of a ring $R$ is an ideal if

- $f+g \in I \quad \forall f, g \in I$
- $hf \in I \quad \forall f \in I$ and $h \in R$

Def: Let $W \subseteq \mathbb{K}^r$. The vanishing ideal of $W$ is

$$I(W) = \{f \in \mathbb{K}[p] : f(a) = 0 \quad \forall a \in W\}.$$

- $I(W)$ is an ideal

# Generating sets

Def: Let $S$ be a set of polynomials. Then we set $\langle S \rangle = \{ \sum_{i=1}^{k} h_i f_i : f_i \in S, h_i \in \mathbb{K}[p] \}$.

- The set $\langle S \rangle$ is an ideal. It is called the ideal generated by $S$.
- $\langle S \rangle \subseteq I(V(S))$
- We say that an ideal $I$ is finitely generated if there exists finite $S$ such that $I = \langle S \rangle$. Hilbert Basis Theorem says that every ideal in $\mathbb{K}[p]$ is finitely generated.

# Radical ideals

Def: An ideal $I$ is called radical if $f^k \in I$ for some polynomial $f$ and positive integer $k$ implies $f \in I$.

Def: The radical of an ideal $I$, denoted $\sqrt{I}$, is the smallest radical ideal that contains $I$:

$$\sqrt{I} = \{f \in \mathbb{K}[p] : f^k \in I \text{ for some } k \in \mathbb{N}\}.$$

Prop: Given any field $\mathbb{K}$ and set $W \subseteq \mathbb{K}^r$, the vanishing ideal $I(W)$ is radical.

Nullstellensatz: If $\mathbb{K}$ is algebraically closed, then the vanishing ideal of the variety of an ideal is the radical of the ideal, i.e. $I(V(I)) = \sqrt{I}$.

# Ideal-variety correspondence

Theorem: Let $\mathbb{K}$ be an algebraically closed field. Then the maps $V$ and $I$ are inclusion-reversing bijections between the set of radical ideals and the set of varieties.

# Univariate division algorithm

Input: A polynomial $f$ and a finite set of polynomials $\mathcal{G} = \{g_1, \ldots, g_k\}$

Output: A representation $f = \sum_{i=1}^{k} h_i g_i + r$ such that no term of $r$ is divisible by the highest degree term of any of the polynomials in $\mathcal{G}$

Algorithm:

- Set $h_i = 0$ for all $i$ and $r = f$.
- While $r$ has a term $c_a p^a$ divisible by a highest degree term $in(g_i)$ of some $g_i$, replace $h_i$ by $h_i + c_a p^a / in(g_i)$ and $r$ by $r - c_a p^a / in(g_i) \cdot g_i$

# Gröbner bases

- Main computational tool for computations with ideals

Def: Let $I \subset \mathbb{K}[p]$ be an ideal. A finite subset $\mathcal{G}$ of $I$ is called a Gröbner basis if dividing $f$ by $\mathcal{G}$ gives remainder $0$ for all $f \in I$.

- It follows that $\mathcal{G}$ is a generating set of $I$.
- See Chapter 3.3 for the definition in the multivariate case. In this course we use Gröbner bases as a black box. They will be covered in detail in Computational Algebraic Geometry (period III).

# Elimination ideal

- Let $\pi : \mathbb{K}^{r_1+r_2} \to \mathbb{K}^{r_1}$ be the coordinate projection
  $(a_1, \ldots, a_{r_1}, b_1, \ldots, b_{r_2}) \mapsto (a_1, \ldots, a_r)$ .

Prop: Let $V \subseteq \mathbb{K}^{r_1+r_2}$ be a variety and let
$I := I(V) \subseteq \mathbb{K}[p_1, \ldots, p_{r_1}, q_1, \ldots, q_{r_2}]$ be its vanishing ideal. Then
$I(\pi(V)) = I \cap \mathbb{K}[p]$.

- The ideal $I \cap \mathbb{K}[p]$ is called an elimination ideal.

# Pre-task for next time

- Read Chapters 2.1-2.4
- Write at least three questions that remained unclear in the text and submit in MyCourses before the start of the lecture
- Next time:
  - we will discuss these questions in groups in Breakout rooms
  - work on first tasks connecting probability and algebra
- Alternative pre-task if you already have strong background in probability: Read and write three questions for Chapters 3.1-3.4