On Paramtetric and Implicit Algebraic Descriptions of Maximum Entropy models

Ambedkar Dukkipati

Model of independence for Binary random variables

Independence model in "implicit" form

$X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.
Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

Model of independence for Binary random variables

Independence model in “implicit” form

$X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.
Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

- ▷ X_1 is independent of X_2 translates into

$$p_{11}p_{22} - p_{12}p_{21} = 0$$

  a polynomial equations in $k[p_{11}, p_{12}, p_{21}, p_{22}]$.

  (when all probabilities are positive then this equivalent to the “odds ratio condition” for independence in 2×2 contingency tables.)

Model of independence for Binary random variables

Independence model in "implicit" form

$X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.
Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

- ▷ X_1 is independent of X_2 translates into

  $$p_{11}p_{22} - p_{12}p_{21} = 0$$

  a polynomial equations in $k[p_{11}, p_{12}, p_{21}, p_{22}]$.

  (when all probabilities are positive then this equivalent to the "odds ratio condition" for independence in 2×2 contingency tables.)

- ▷ We say that this model is in the *algebraic variety*

  $$\mathcal{V}(p_{11}p_{22} - p_{12}p_{21})$$

  - ○ On normalization
  - ○ Complex solutions

Model of independence for Binary random variables (Contd...)

Independence model in "parametric" form

$X = (X_1, X_2)$ be a random vector, where X_1 takes values from $[m_1] = \{1, \ldots, m_1\}$ and X_2 takes values from $[m_2] = \{1, \ldots, m_2\}$. $X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.

▷ Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

Model of independence for Binary random variables (Contd...)

Independence model in "parametric" form

$X = (X_1, X_2)$ be a random vector, where X_1 takes values from $[m_1] = \{1, \ldots, m_1\}$ and X_2 takes values from $[m_2] = \{1, \ldots, m_2\}$. $X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.

▷ Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

▷ Let $\Theta = \{(\theta_1, \theta_2 \,:\, 0 \leq \theta_1, \theta_2 \leq 1\}$ and consider a polynomial map

$$\begin{array}{rrcl} f : & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^{2\times 2} \\ & (\theta_1, \theta_2) & \mapsto & (\underbrace{\theta_1\theta_2}_{p_{11}}, \underbrace{\theta_1(1-\theta_2)}_{p_{12}}, \underbrace{(1-\theta_1)\theta_2}_{p_{21}}, \underbrace{(1-\theta_1)(1-\theta_2)}_{p_{22}}) \end{array}$$

Model of independence for Binary random variables (Contd...)

Independence model in "parametric" form

$X = (X_1, X_2)$ be a random vector, where X_1 takes values from $[m_1] = \{1, \ldots, m_1\}$ and X_2 takes values from $[m_2] = \{1, \ldots, m_2\}$. $X = (X_1, X_2)$ be a random vector, where X_1 and X_2 taking values from $\{0, 1\}$.

▷ Probability simplex in this case is

$$\Delta = \{(p_{11}, p_{12}, p_{21}, p_{22}) \in \mathbb{R}^4_{\geq 0} : p_{11} + p_{12} + p_{21} + p_{22} = 1\}$$

with the understanding that $P(X_1 = i, X_2 = j) = p_{ij}$, $i, j \in \{1, 2\}$.

▷ Let $\Theta = \{(\theta_1, \theta_2 \,:\, 0 \leq \theta_1, \theta_2 \leq 1\}$ and consider a polynomial map

$$\begin{aligned} f : \quad \mathbb{R}^2 &\longrightarrow \mathbb{R}^{2\times 2} \\ (\theta_1, \theta_2) &\mapsto (\underbrace{\theta_1\theta_2}_{p_{11}}, \underbrace{\theta_1(1-\theta_2)}_{p_{12}}, \underbrace{(1-\theta_1)\theta_2}_{p_{21}}, \underbrace{(1-\theta_1)(1-\theta_2)}_{p_{22}}) \end{aligned}$$

▷ $f(\Theta)$ represents the independence model. Note that f is a polynomial function.

Maximum Likelihood Estimation

Model: Say our statistical model is given by the mapping

$$f \quad \Theta \to \mathbb{R}^m \tag{1}$$

assigning probabilities as $p_j = f_j(\theta)$, $j = 1, \ldots, m$, where $f_j, : \; j = 1, \ldots, m$ are polynomial (could be rational) functions.

Maximum Likelihood Estimation

Model: Say our statistical model is given by the mapping

$$f \quad \Theta \to \mathbb{R}^m \tag{2}$$

assigning probabilities as $p_j = f_j(\theta)$, $j = 1, \ldots, m$, where $f_j, : \; j = 1, \ldots, m$ are polynomial (could be rational) functions.

Log-likelihood: Let $(u_1, \ldots, u_m) \in \mathbb{Z}^m_{\geq 0}$ sufficient statistic corresponding to sequence of an iid observations. We have *log-likelihood function*

$$l(\theta) = \sum_{j=1}^{m} u_j \log f_j(\theta)$$

Maximum Likelihood Estimation

Model: Say our statistical model is given by the mapping

$$f \quad \Theta \to \mathbb{R}^m \tag{3}$$

assigning probabilities as $p_j = f_j(\theta)$, $j = 1, \ldots, m$, where $f_j, : \ j = 1, \ldots, m$ are polynomial (could be rational) functions.

Log-likelihood: Let $(u_1, \ldots, u_m) \in \mathbb{Z}^m_{\geq 0}$ sufficient statistic corresponding to sequence of an iid observations. We have *log-likelihood function*

$$l(\theta) = \sum_{j=1}^{m} u_j \log f_j(\theta)$$

Likelihood Equations:

$$\frac{\partial l}{\partial \theta_j} = \frac{u_1}{f_1(\theta)} \frac{\partial f_1(\theta)}{\partial \theta_j} + \ldots + \frac{u_m}{f_m(\theta)} \frac{\partial f_m(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, d.$$

Maximum Likelihood Estimation (Contd. . .)

Estimation by algebraic methods: The ideal which represents the solutions is

$$\mathfrak{a} = \widehat{\mathfrak{a}} \cap \mathbb{R}[\theta_1, \ldots, \theta_d] \ ,$$

where

$$\widehat{\mathfrak{a}} = \left\langle \underbrace{y_1 f_1(\theta) - 1, \ldots, y_m f_m(\theta) - 1}, \underbrace{\sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_1}, \ldots, \sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_d}} \right\rangle$$

Solving by Gröbner bases method *Hoșten, Khetan & Sturmfels, Solving the Likelihood equations*, 2005.

What Commutative Algebra & Algebraic Geometry could offer to Statistics

General role

- ▷ Estimation
- ▷ Elimination
- ▷ Enumeration

What Commutative Algebra & Algebraic Geometry could offer to Statistics

General role

▷ Estimation

▷ Elimination

▷ Enumeration

▷ Algebraic Geometric insights???

- (Inspired by *Geometric Ideas in Minimum Cross-Entropy* by *Campbell, 2003*)
- for which we need to embed models in algebraic varieties

What Commutative Algebra & Algebraic Geometry could offer to Statistics

General role

- ▷ Estimation
- ▷ Elimination
- ▷ Enumeration
- ▷ Algebraic Geometric insights???
  - ◦ (Inspired by *Geometric Ideas in Minimum Cross-Entropy* by *Campbell, 2003*)
  - ◦ for which we need to embed models in algebraic varieties

Broad Approaches

- ▷ Reverse engineering of gene regulatory networks (*Laubenbacher* et al. *, 2000* onwards)
- ▷ Experimental design (*Pistone* et al. *, 1998* onwards)
- ▷ Analysis of graphical models and sampling (*Diaconis, Sturmfels, Pactor* et al. *, 1998* onwards)

Outline of the Talk

- ▷ Estimation of ME models as polynomial system solving
- ▷ Gröbner Bases (*Buchberger, 1965*) Fundamentals
- ▷ Elimination Theorem and Its Application to Estimation
- ▷ Embedding ME models in algebraic varieties
- ▷ Concluding Remarks and Discussion

Estimation of ME models as polynomial system solving

Kullback's minimum discrimination theorem

(*Kullback 1959*)
Given a probability space, $(X, \mathfrak{M}, R)$ define a probability measure P as

$$P(A) = Z^{-1} \int_A \exp(T)\,\mathrm{d}R \ , \quad \forall A \in \mathfrak{M}$$

where T a real valued function on X such that $Z = \mathrm{E}_{[R]} \exp(T) < \infty$. Suppose T is P integrable Then

$$I(P' \| R) \geq I(P \| R) = \mathrm{E}_{[P]} T - \ln Z$$

ME model

Set Up

- ▷ X is discrete random variable taking values from the set $[m] = \{1, 2, \dots, m\}$.
- ▷ The available information is in the form of expected values of some functions $t_i : [m] \to \mathbb{R}$, $i = 1, \dots, d$ (*feature functions*). That is

$$\sum_{j=1}^{m} t_i(j) p_j = T_i \ \ , i = 1, \dots d,$$

where T_i, $i = 1, \dots, d$, are assumed to be known.

ME model

Set Up

- ▷ X is discrete random variable taking values from the set $[m] = \{1, 2, \ldots, m\}$.
- ▷ The available information is in the form of expected values of some functions $t_i : [m] \to \mathbb{R}$, $i = 1, \ldots, d$ (*feature functions*). That is

$$\sum_{j=1}^{m} t_i(j) p_j = T_i \ , i = 1, \ldots d,$$

where T_i, $i = 1, \ldots, d$, are assumed to be known.

Principle (Information theoretic approach to statistics)

- ▷ choose the pmf that maximize the Shannon entropy functional

$$S(p) = -\sum_{j=1}^{m} p_j \ln p_j$$

(*equivalent to minimizing Bayes loss in decision theory framework*)

ME model (Contd...)

Lagrangian

$$\Xi(p,\xi) \equiv S(p) - \xi_0 \left(\sum_{j=1}^{m} p_j - 1 \right) - \sum_{i=1}^{d} \xi_d \left(\sum_{j=1}^{m} t_i(j) p_j - T_i \right)$$

model

- ▷ Holding $\xi = (\xi_1, \ldots, \xi_d)$ fixed, the unconstrained maximum of Lagrangian $\Xi(p,\xi)$ over all $p \in \Delta_{m-1}$ is given by an exponential family

$$p_j(\xi) = Z(\xi)^{-1} \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ , j = 1, \ldots, m,$$

where $Z(\xi)$ is normalizing constant given by

$$Z(\xi) = \sum_{j=1}^{m} \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ .$$

(*For various values of $\xi \in \mathbb{R}$, this is known as "maximum entropy model"*.)

ME model (Contd...)

Lagrangian

$$\Xi(p,\xi) \equiv S(p) - \xi_0 \left(\sum_{j=1}^{m} p_j - 1 \right) - \sum_{i=1}^{d} \xi_d \left(\sum_{j=1}^{m} t_i(j) p_j - T_i \right)$$

model

▷ Holding $\xi = (\xi_1, \ldots, \xi_d)$ fixed, the unconstrained maximum of Lagrangian $\Xi(p,\xi)$ over all $p \in \Delta_{m-1}$ is given by an exponential family

$$p_j(\xi) = Z(\xi)^{-1} \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ , j = 1, \ldots, m,$$

where $Z(\xi)$ is normalizing constant given by

$$Z(\xi) = \sum_{j=1}^{m} \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ .$$

(*For various values of $\xi \in \mathbb{R}$, this is known as "maximum entropy model"*.)

Various Formulations

(i) Primal **(ii)** Dual **(iii)** Kullback-Csiszár Iteration

Towards Algebraic methods

Proposition: *The estimation of maximum entropy model amounts to solving a set of polynomial equations provided that the feature functions* t_i, $i = 1, \ldots, d$ *are positive and integer valued.*

Proof: By setting $\xi_i = -\ln \theta_i$, $i = 1, \ldots, d$, we obtain maximum entropy model as

$$p_j = Z(\theta)^{-1} \prod_{i=1}^{d} \theta_i^{t_i(j)} \ ,$$

where

$$Z(\theta) = \sum_{j=1}^{m} \prod_{i=1}^{d} \theta_i^{t_i(j)} \ .$$

Given the information in the form of expected values of feature functions t_i, $i = 1, \ldots, d$, θ_i, $i = 1, \ldots, d$ should satisfy following set of polynomial equations

$$\sum_{j=1}^{m} (t_i(j) - T_i) \prod_{i=1}^{d} \theta_i^{t_i(j)} = 0.$$

■

Towards Algebraic methods (Contd...)

proposition: *Given the "sample mean" hypothesis the problem of estimating the ME-model in duel-method amounts to solving set of Laurent polynomial equations.*

Proof: To retain the integer valued exponents in our final solution we consider the constrains of the form

$$N \sum_{j=1}^{m} t_i(j) p_j = \widetilde{S}_i \ , \quad i = 1, \ldots d \ , \tag{4}$$

where $\widetilde{S}_i = \sum_{l=1}^{N} t_i(O_l)$ denotes the sample sum. In this case Lagrangian is

$$\widetilde{\Xi}(p, \xi) \equiv S(p) - \xi_0 \left(\sum_{j=1}^{m} p_j - 1 \right) - \sum_{i=1}^{d} \widetilde{\xi}_d N \left(\sum_{j=1}^{m} p_j t_i(j) - \widetilde{S}_i \right) \ . \tag{5}$$

$$p_j(\xi) = \widetilde{Z}(\xi)^{-1} \exp \left(-N \sum_{i=1}^{d} \widetilde{\xi}_i t_i(j) \right) \ , \quad j = 1, \ldots, m.$$

where $\widetilde{Z}(\xi)$ is a normalizing constant.

Towards Algebraic methods (Contd...)

To calculate the parameters we maximize the dual $\widetilde{\Psi}(\widetilde{\xi})$ of $\widetilde{\Xi}(p,\xi)$. That is we maximize the functional

$$\widetilde{\Psi}(\widetilde{\xi}) = \ln \widetilde{Z} + \sum_{i=1}^{d} \widetilde{\xi}_i \widetilde{S}_i \quad . \tag{6}$$

It is equivalent to optimizing the functional

$$\widetilde{\Psi}'(\widetilde{\xi}) = \sum_{j=1}^{m} \exp\left(\sum_{i=1}^{d} \widetilde{\xi}_i \widetilde{S}_i - N \sum_{i=1}^{d} \widetilde{\xi}_i t_i(j) \right)$$

By setting $\ln \widetilde{\theta}_i = \widetilde{\xi}$ we have

$$\tilde{\Psi}'(\widetilde{\theta}) = \sum_{j=1}^{m} \prod_{i=1}^{d} \widetilde{\theta}_i^{(\widetilde{S}_i - N t_i(j))} \tag{7}$$

The solution is given by solving the following set of equations

$$\frac{\partial \widetilde{\Psi}'}{\partial \widetilde{\theta}_j} = 0 \quad , j = 1, \ldots d. \tag{8}$$

We have

$$\frac{\partial \widetilde{\Psi}'}{\partial \widetilde{\theta}_j} \in k[\widetilde{\theta}_1^{\pm}, \ldots, \widetilde{\theta}_d^{\pm}] \quad , i = 1, \ldots, d. \tag{9}$$

Estimation by Minimum I-Divergence Principle

▷ Given a prior estimate $r \in \Delta_m$ one would choose the pmf $p \in \Delta_m$ that minimizes the Kullback-Leibler divergence

$$I(p \| r) = \sum_{j=1}^{m} p_j \ln \frac{p_j}{r_j} \tag{10}$$

with respect to the given constraints.

▷ The corresponding minimum entropy distributions are in the form of

$$p_j(\xi) = Z(\xi)^{-1} r_j \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ , j = 1, \ldots, m, \tag{11}$$

where $Z(\xi)$ is normalizing constant given by

$$Z(\xi) = \sum_{j=1}^{m} r_j \exp\left(- \sum_{i=1}^{d} \xi_i t_i(j) \right) \ . \tag{12}$$

Estimation by Minimum I-Divergence Principle

▷ Given a prior estimate $r \in \Delta_m$ one would choose the pmf $p \in \Delta_m$ that minimizes the Kullback-Leibler divergence

$$I(p\|r) = \sum_{j=1}^{m} p_j \ln \frac{p_j}{r_j} \tag{13}$$

with respect to the given constraints.

▷ The corresponding minimum entropy distributions are in the form of

$$p_j(\xi) = Z(\xi)^{-1} r_j \exp\left(-\sum_{i=1}^{d} \xi_i t_i(j)\right) \ , j = 1, \ldots, m, \tag{14}$$

where $Z(\xi)$ is normalizing constant given by

$$Z(\xi) = \sum_{j=1}^{m} r_j \exp\left(-\sum_{i=1}^{d} \xi_i t_i(j)\right) \ . \tag{15}$$

▷ Estimation in this case can be translated to solving polynomial equations, when the feature functions are integer valued. Polynomial system one would solve in this case is

$$\sum_{j=1}^{m} r_j (t_i(j) - T_i) \prod_{i=1}^{d} \theta_i^{t_i(j)} = 0. \tag{16}$$

Estimation by Kullback-Csiszár Iteration

Algorithm: The algorithm computes the distribution $p^{(N)}$ which minimizes $I(p^{(N)} \| p^{(N-1)})$ with respect the i^{th} constraint, $1 \le i \le d$ if $N = ad + i$, for any positive integer a.

Estimation by Kullback-Csiszár Iteration

Algorithm: The algorithm computes the distribution $p^{(N)}$ which minimizes $I(p^{(N)} \| p^{(N-1)})$ with respect the i^{th} constraint, $1 \leq i \leq d$ if $N = ad + i$, for any positive integer a.

at 0^{th} **iteration:** $p^{(0)} = r$

Estimation by Kullback-Csiszár Iteration

Algorithm: The algorithm computes the distribution $p^{(N)}$ which minimizes $I(p^{(N)} \| p^{(N-1)})$ with respect the i^{th} constraint, $1 \leq i \leq d$ if $N = ad + i$, for any positive integer a.

at 0^{th} **iteration:** $p^{(0)} = r$

at 1^{st} **iteration:** $p^{(1)}$ is given by

$$p_j^{(1)} = r_j \left(Z^{(1)} \right)^{-1} \zeta_1^{t_1(j)},$$

where $\left(Z^{(1)} \right)^{-1} = \sum_{j=1}^{m} r_j \zeta_1^{t_1(j)}$. Considering the first constraint it can be estimated by solving polynomial equation

$$\sum_{j=1}^{m} r_j (t_1(j) - T_1) \zeta_1^{t_1(j)} = 0 \ , \tag{17}$$

with indeterminate ζ_1.

Estimation by Kullback-Csiszár Iteration

Algorithm: The algorithm computes the distribution $p^{(N)}$ which minimizes $I(p^{(N)} \| p^{(N-1)})$ with respect the i^{th} constraint, $1 \leq i \leq d$ if $N = ad + i$, for any positive integer a.

at 0^{th} **iteration:** $p^{(0)} = r$

at 1^{st} **iteration:** $p^{(1)}$ is given by

$$p_j^{(1)} = r_j \left(Z^{(1)} \right)^{-1} \zeta_1^{t_1(j)},$$

where $\left(Z^{(1)} \right)^{-1} = \sum_{j=1}^{m} r_j \zeta_1^{t_1(j)}$. Considering the first constraint it can be estimated by solving polynomial equation

$$\sum_{j=1}^{m} r_j (t_1(j) - T_1) \zeta_1^{t_1(j)} = 0 \ , \tag{18}$$

with indeterminate ζ_1.

at 2^{st} **iteration:** Similarly we have

$$p_j^{(2)} = r_j \left(Z^{(1)} \right)^{-1} \left(Z^{(2)} \right)^{-1} \zeta_1^{t_1(j)} \zeta_2^{t_2(j)} \ ,$$

where $\left(Z^{(2)} \right)^{-1} = \sum_{j=1}^{m} \zeta_2^{t_1(j)}$.

Estimation by Kullback-Csiszár Iteration (Contd...)

Considering the first two constrains in ME distribution can be estimated by solving

$$\sum_{j=1}^{m} r_j(t_2(j) - T_2)\zeta_1^{t_1(j)}\zeta_2^{t_2(j)} = 0 \ , \tag{19}$$

along with the previous equations.

Estimation by Kullback-Csiszár Iteration (Contd...)

Considering the first two constrains in ME distribution can be estimated by solving

$$\sum_{j=1}^{m} r_j(t_2(j) - T_2)\zeta_1^{t_1(j)}\zeta_2^{t_2(j)} = 0 \ , \tag{20}$$

along with the previous equations.

at N^{th} iteration: In general, when $N = ad + i$ for some positive integer a, $p_j^{(N)}$, for $N = 1, 2 \ldots$ is given by

$$p_j^{(N)} = r_j\left(Z^{(1)}\right)^{-1} \ldots \left(Z^{(N)}\right)^{-1} \zeta_1^{t_1(j)} \ldots \zeta_N^{t_N(j)}$$

and is determined by the following system of polynomial equations

$$\left.\begin{array}{ll}
\sum_{j=1}^{m} r_j(t_1(j) - T_1)\zeta_1^{t_1(j)} & = 0 \ , \\
\sum_{j=1}^{m} r_j(t_2(j) - T_2)\zeta_1^{t_1(j)}\zeta_2^{t_2(j)} & = 0 \ , \\
\vdots & \\
\sum_{j=1}^{m} r_j(t_i(j) - T_i)\zeta_1^{t_1(j)}\zeta_2^{t_2(j)} \ldots \zeta_N^{t_i(j)} & = 0 \ .
\end{array}\right\} \tag{21}$$

Gröbner Bases (*Buchberger, 1965*) Fundamentals

Dictionary of Algebra & Geometry

Basic problem of algebraic geometry is to understand the set of points $a = (a_1, \ldots, a_n) \in k^n$ satisfying a system of polynomial equations $f_1(x_1, \ldots, x_n) = 0, \ldots, f_s(x_1, \ldots, x_n) = 0$ where $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$.

Dictionary of Algebra & Geometry

Basic problem of algebraic geometry is to understand the set of points $a = (a_1, \dots, a_n) \in k^n$ satisfying a system of polynomial equations $f_1(x_1, \dots, x_n) = 0, \dots, f_s(x_1, \dots, x_n) = 0$ where $f_1, \dots, f_s \in k[x_1, \dots, x_n]$.

Algebraic Variety: *A set $V \subset k^n$ is said to be algebraic variety if there exists $f_1, \dots, f_s \in k[x_1, \dots, x_n]$ such that*

$$\mathcal{V}(f_1, \dots, f_s) = V = \{(c_1, \dots c_n) \in k^n : f_i(c_1, \dots c_n) = 0, 1 \leq i \leq s\} \ .$$

▷ $\mathcal{V}(f_1, \dots, f_s)$ is uniquely determined by the ideal generated by $f_1, \dots, f_s$.

Dictionary of Algebra & Geometry

Basic problem of algebraic geometry is to understand the set of points $a = (a_1, \ldots, a_n) \in k^n$ satisfying a system of polynomial equations $f_1(x_1, \ldots, x_n) = 0, \ldots, f_s(x_1, \ldots, x_n) = 0$ where $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$.

Algebraic Variety: *A set $V \subset k^n$ is said to be algebraic variety if there exists $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$ such that*

$$\mathcal{V}(f_1, \ldots, f_s) = V = \{(c_1, \ldots c_n) \in k^n : f_i(c_1, \ldots c_n) = 0, 1 \leq i \leq s\} \ .$$

▷ $\mathcal{V}(f_1, \ldots, f_s)$ is uniquely determined by the ideal generated by $f_1, \ldots, f_s$.

Ideal: *A subset $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ is said to be ideal if it satisfies:*

(i) $0 \in \mathfrak{a}$

(ii) $f, g \in \mathfrak{a}$, then $f + g \in \mathfrak{a}$

(iii) $f \in \mathfrak{a}$ and $h \in k[x_1, \ldots, x_n]$ and then $hf \in \mathfrak{a}$.

▷ Ideal generated by $f_1, \ldots, f_s \in k[x_1, \ldots, x_n]$:

$$\langle f_1, \ldots, f_s \rangle = \left\{ f \in k[x_1, \ldots, x_n] : \sum_{i=1}^{s} h_s f_s, h_s \in k[x_1, \ldots, x_s] \right\}$$

Dictionary of Algebra & Geometry (Contd...)

Vanishing Ideal of a variety: *Let $E \subset k^n$ be an variety. Then vanishing ideal of a variety is defined as*

$$\mathcal{I}(E) = \{f \in k[x_1, \ldots, x_n] : f(a) = 0, \forall a \in E\}$$

(Can be extended to any arbitrary subset of k^n)

Dictionary of Algebra & Geometry (Contd...)

Vanishing Ideal of a variety: *Let $E \subset k^n$ be an variety. Then vanishing ideal of a variety is defined as*

$$\mathcal{I}(E) = \{f \in k[x_1, \ldots, x_n] : f(a) = 0, \forall a \in E\}$$

(Can be extended to any arbitrary subset of k^n)

Ideal Variety relations

- $\triangleright$ $\mathcal{V}(f_1, \ldots, f_s) = \mathcal{V}(\langle f_1, \ldots, f_s \rangle)$
- $\triangleright$ $\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle \Longrightarrow \mathcal{V}(f_1, \ldots, f_s) = \mathcal{V}(g_1, \ldots, g_t)$
- $\triangleright$ $\langle f_1, \ldots, f_s \rangle \subseteq \mathcal{I}(\mathcal{V}(f_1, \ldots, f_s))$
- $\triangleright$ $V = W \Longleftrightarrow \mathcal{I}(V) = \mathcal{I}(W)$

Dictionary of Algebra & Geometry (Contd...)

Vanishing Ideal of a variety: *Let $E \subset k^n$ be an variety. Then vanishing ideal of a variety is defined as*

$$\mathcal{I}(E) = \{f \in k[x_1, \ldots, x_n] : f(a) = 0, \forall a \in E\}$$

(Can be extended to any arbitrary subset of k^n)

Ideal Variety relations

- $\triangleright$ $\mathcal{V}(f_1, \ldots, f_s) = \mathcal{V}(\langle f_1, \ldots, f_s \rangle)$
- $\triangleright$ $\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle \Longrightarrow \mathcal{V}(f_1, \ldots, f_s) = \mathcal{V}(g_1, \ldots, g_t)$
- $\triangleright$ $\langle f_1, \ldots, f_s \rangle \subseteq \mathcal{I}(\mathcal{V}(f_1, \ldots, f_s))$
- $\triangleright$ $V = W \Longleftrightarrow \mathcal{I}(V) = \mathcal{I}(W)$

(**Hilbert Nullstellensatz:** Given a variety, we can recover the ideal up to its radical only in the case of algebraically closed fields.)

Multivariate Division Algorithm and Hilbert Bases Theorem

Division algorithm in $\boldsymbol{k}[\boldsymbol{x}]$: Let $g \in k[x]$. Then for any $f \in k[x]$, $\exists$ unique $q, r \in k[x]$ and deg $r <$ deg g or $r = 0$ such that $f = qg + r$.

$\triangleright$ $k[x]$ is a principle ideal domain (PID).

Multivariate Division Algorithm and Hilbert Bases Theorem

Division algorithm in $k[x]$: Let $g \in k[x]$. Then for any $f \in k[x]$, $\exists$ unique $q, r \in k[x]$ and deg $r <$ deg g or $r = 0$ such that $f = qg + r$.

▷ $k[x]$ is a principle ideal domain (PID).

Division algorithm in $k[x_1, \ldots, x_n]$: Consider a monomial order $\prec$ on $\mathbb{Z}^n_{\geq 0}$, and let $F = (f_1, \ldots, f_s)$ be an ordered s-tuple of polynomials in $k[x_1, \ldots, x_n]$. Then every $f \in k[x_1, \ldots, x_n]$ can be written as

$$f = \sum_{i=1}^{s} q_i f_i + r$$

where $q_i \in k[x_1, \ldots, x_n]$, $i = 1, \ldots, s$ and $r \in k[x_1, \ldots, x_n]$ such that $r = 0$ or none of the terms in r are divisible by any of $\mathrm{LT}_\prec(f_1), \ldots, \mathrm{LT}_\prec(f_s)$

Multivariate Division Algorithm and Hilbert Bases Theorem

Division algorithm in $\boldsymbol{k[x]}$: Let $g \in k[x]$. Then for any $f \in k[x]$, $\exists$ unique $q, r \in k[x]$ and deg $r <$ deg g or $r = 0$ such that $f = qg + r$.

$\triangleright$ $k[x]$ is a principle ideal domain (PID).

Division algorithm in $\boldsymbol{k[x_1, \ldots, x_n]}$: Consider a monomial order $\prec$ on $\mathbb{Z}^n_{\geq 0}$, and let $F = (f_1, \ldots, f_s)$ be an ordered s-tuple of polynomials in $k[x_1, \ldots, x_n]$. Then every $f \in k[x_1, \ldots, x_n]$ can be written as

$$f = \sum_{i=1}^{s} q_i f_i + r$$

where $q_i \in k[x_1, \ldots, x_n]$, $i = 1, \ldots, s$ and $r \in k[x_1, \ldots, x_n]$ such that $r = 0$ or none of the terms in r are divisible by any of $\mathrm{LT}_\prec(f_1), \ldots, \mathrm{LT}_\prec(f_s)$

Hilbert Basis theorem: *Every ideal in $k[x_1, \ldots, x_n]$ has a finite generating set.*

Multivariate Division Algorithm and Hilbert Bases Theorem

Division algorithm in $\boldsymbol{k[x]}$: Let $g \in k[x]$. Then for any $f \in k[x]$, $\exists$ unique $q, r \in k[x]$ and deg $r <$ deg g or $r = 0$ such that $f = qg + r$.

$\triangleright$ $k[x]$ is a principle ideal domain (PID).

Division algorithm in $\boldsymbol{k[x_1, \ldots, x_n]}$: Consider a monomial order $\prec$ on $\mathbb{Z}^n_{\geq 0}$, and let $F = (f_1, \ldots, f_s)$ be an ordered s-tuple of polynomials in $k[x_1, \ldots, x_n]$. Then every $f \in k[x_1, \ldots, x_n]$ can be written as

$$f = \sum_{i=1}^{s} q_i f_i + r$$

where $q_i \in k[x_1, \ldots, x_n]$, $i = 1, \ldots, s$ and $r \in k[x_1, \ldots, x_n]$ such that $r = 0$ or none of the terms in r are divisible by any of $\mathsf{LT}_\prec(f_1), \ldots, \mathsf{LT}_\prec(f_s)$

Hilbert Basis theorem: *Every ideal in $k[x_1, \ldots, x_n]$ has a finite generating set.*

The ascending chain condition: *Let $\mathfrak{a}_n \subset k[x_1, \ldots, x_n]$ be a sequence of ideals such that*

$$\mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \ldots \mathfrak{a}_n \subset \ldots$$

Then $\exists N \geq 1$ such that $\mathfrak{a}_N = \mathfrak{a}_{N+i}$ for all $i \geq 1$.

Monomial Ideals & Dickson's Lemma

Monomial Order or term order on $k[x_1, \ldots, x_n]$ is a relation $\prec$ (we use $\succ$ for corresponding 'greater than' on $\mathbb{Z}^n_{\geq 0}$ which satisfies following conditions

(i) $\prec$ is a total (or linear) ordering on $\mathbb{Z}^n_{\geq 0}$,

(ii) if $\alpha \succ \beta$, for $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$ and for any $\gamma \in \mathbb{Z}^n_{\geq 0}$ we have $\alpha + \gamma \succ \beta + \gamma$, and

(iii) $\prec$ is a well-ordering on $\mathbb{Z}^n_{\geq 0}$.

Monomial Ideals & Dickson's Lemma

Monomial Order or term order on $k[x_1, \ldots, x_n]$ is a relation $\prec$ (we use $\succ$ for corresponding 'greater than' on $\mathbb{Z}^n_{\geq 0}$ which satisfies following conditions

(i) $\prec$ is a total (or linear) ordering on $\mathbb{Z}^n_{\geq 0}$,

(ii) if $\alpha \succ \beta$, for $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$ and for any $\gamma \in \mathbb{Z}^n_{\geq 0}$ we have $\alpha + \gamma \succ \beta + \gamma$, and

(iii) $\prec$ is a well-ordering on $\mathbb{Z}^n_{\geq 0}$.

Monomial Ideal

An ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ is monomial ideal if there is a set $A \subset \mathbb{Z}^n_{\geq 0}$ (possibly infinite) such that $\mathfrak{a} = \langle x^\alpha : \alpha \in A \rangle$.

Monomial Ideals & Dickson's Lemma

Monomial Order or term order on $k[x_1, \ldots, x_n]$ is a relation $\prec$ (we use $\succ$ for corresponding 'greater than' on $\mathbb{Z}^n_{\geq 0}$ which satisfies following conditions

(i) $\prec$ is a total (or linear) ordering on $\mathbb{Z}^n_{\geq 0}$,

(ii) if $\alpha \succ \beta$, for $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$ and for any $\gamma \in \mathbb{Z}^n_{\geq 0}$ we have $\alpha + \gamma \succ \beta + \gamma$, and

(iii) $\prec$ is a well-ordering on $\mathbb{Z}^n_{\geq 0}$.

Monomial Ideal

An ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ is monomial ideal if there is a set $A \subset \mathbb{Z}^n_{\geq 0}$ (possibly infinite) such that $\mathfrak{a} = \langle x^\alpha : \alpha \in A \rangle$.

Dickson's Lemma: Any monomial ideal is finitely generated.

Proof of Hilbert Basis theorem

if $\mathfrak{a} = 0$ then we are done

Proof of Hilbert Basis theorem

if $\mathfrak{a} = 0$ then we are done
from Dickson's lemma $\exists f_1, \ldots, f_s \in \mathfrak{a}$ such that

$$\mathsf{LT}_{\prec}(\mathfrak{a}) = \langle \mathsf{LT}_{\prec}(\mathfrak{a}) \rangle = \langle \mathsf{LT}(f_1), \ldots, \mathsf{LT}(f_s) \rangle$$

Proof of Hilbert Basis theorem

if $\mathfrak{a} = 0$ then we are done
from Dickson's lemma $\exists f_1, \ldots, f_s \in \mathfrak{a}$ such that

$$\mathsf{LT}_{\prec}(\mathfrak{a}) = \langle \mathsf{LT}_{\prec}(\mathfrak{a}) \rangle = \langle \mathsf{LT}(f_1), \ldots, \mathsf{LT}(f_s) \rangle$$

claim: $\mathfrak{a} = \langle f_1, \ldots, f_s \rangle$

Enough to show that $\mathfrak{a} \subset \langle f_1, \ldots, f_s \rangle$

Pick $f \in \mathfrak{a}$ arbitrary.

Apply division algorithm:

$$f = \sum_{i=1}^{s} q_i f_i + r$$

we have: no term of r is divisible by $LT(f_i)$ for $i = 1, \ldots s$.

We have $r \in \mathfrak{a}$ and hence $LT(r) \in \langle \mathsf{LT}(\mathfrak{a}) \rangle$.

Hence $\mathsf{LT}(r)$ is divided by one of $\mathsf{LT}(g_i)$, $\Longrightarrow\Longleftarrow$

Hence $r = 0$ and hence $f \in \langle f_1, \ldots, f_s \rangle$

Gröbner Bases

definition

Consider $k[x_1, \ldots, x_n]$ and fix a monomial order. Given any ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$, a finite subset $G = \{g_1, \ldots, g_s\} \subset \mathfrak{a}$ is said to be Gröbner basis if

$$\langle \mathsf{LT}(\mathfrak{a}) \rangle = \langle \mathsf{LT}(f_1), \ldots, \mathsf{LT}(f_s) \rangle$$

Gröbner Bases

definition

Consider $k[x_1, \ldots, x_n]$ and fix a monomial order. Given any ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$, a finite subset $G = \{g_1, \ldots, g_s\} \subset \mathfrak{a}$ is said to be Gröbner basis if

$$\langle \mathsf{LT}(\mathfrak{a}) \rangle = \langle \mathsf{LT}(f_1), \ldots, \mathsf{LT}(f_s) \rangle$$

Other characterizations of Gröbner Bases

- ▷ $G \subset k[x_1, \ldots, x_n]$ is a Gröbner bases if and only if for any $f \in \mathfrak{a}$ there exists $g \in G$ such that $\mathsf{LM}(g)|\mathsf{LM}(f)$
- ▷ $G \subset k[x_1, \ldots, x_n]$ is a Gröbner bases if and only if for any $f \in k[x_1, \ldots, x_n]$ we have $f \in \mathfrak{a} \Leftrightarrow \eta_G(f) = 0$

Gröbner Bases

definition

Consider $k[x_1, \ldots, x_n]$ and fix a monomial order. Given any ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$, a finite subset $G = \{g_1, \ldots, g_s\} \subset \mathfrak{a}$ is said to be Gröbner basis if

$$\langle \mathsf{LT}(\mathfrak{a}) \rangle = \langle \mathsf{LT}(f_1), \ldots, \mathsf{LT}(f_s) \rangle$$

Other characterizations of Gröbner Bases

- ▷ $G \subset k[x_1, \ldots, x_n]$ is a Gröbner bases if and only if for any $f \in \mathfrak{a}$ there exists $g \in G$ such that $\mathsf{LM}(g) | \mathsf{LM}(f)$
- ▷ $G \subset k[x_1, \ldots, x_n]$ is a Gröbner bases if and only if for any $f \in k[x_1, \ldots, x_n]$ we have $f \in \mathfrak{a} \Leftrightarrow \eta_G(f) = 0$

Buchberger's algorithm ....

Some Applications

- ▷ Solving system of polynomial equations
- ▷ Intersection Ideals
- ▷ Kernel of ring homeomorphism
- ▷ Quotient Ideals
- ▷ Basis for k vector space $k[x_1, \ldots, x_n]/\mathfrak{a}$

$$\mathcal{B} = \{x^\alpha : \alpha \in \mathbb{Z}^n_{\geq 0}, \quad \mathsf{LM}(g_i) \nmid x^\alpha, \quad i = 1, 2, \ldots, s\}$$

  where $G = \{g_1, \ldots, g_s\}$ is a Gröbner basis.
- ▷ Elimination methods

Elimination Theorem and Its Application to Estimation

Elimination Theorem

(*Buchberger 1987*)

Elimination order: Consider $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ a polynomial ring in indeterminate $x_1, \ldots, x_n$, $y_1, \ldots, y_m$. We refer to $\{x_1, \ldots, x_n\}$ as x-variables and $\{y_1, \ldots, y_m\}$ as y-variables. Let $\prec_x$ and $\prec_y$ be monomial orderings on x and y variables respectively. Define an ordering relation $\prec$ on $\mathbb{Z}_{\geq 0}^{n+m}$ (i.e set of all monomials in indeterminate $x_1, \ldots, x_n, y_1, \ldots, y_m$) as follows:

$$x^{\alpha^{(1)}} y^{\beta^{(1)}} \prec_{[x \succ y]} x^{\alpha^{(2)}} y^{\beta^{(2)}} \iff \begin{cases} \alpha^{(1)} \prec_x \alpha^{(2)} \\ \text{or} \\ \alpha^{(1)} = \alpha^{(2)} \text{ and } \beta^{(1)} \prec_{\mathrm{y}} \beta^{(2)} \end{cases},$$

where $\alpha^{(1)}, \alpha^{(2)} \in \mathbb{Z}_{\geq 0}^{n}$ and $\beta^{(1)}, \beta^{(2)} \in \mathbb{Z}_{\geq 0}^{m}$. The term order $\prec_{[x \succ y]}$ is called elimination order with the x variables larger than the y variables (which is indeed a term order).

Elimination Theorem

(*Buchberger 1987*)

Elimination order: Consider $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ a polynomial ring in indeterminate $x_1, \ldots, x_n$, $y_1, \ldots, y_m$. We refer to $\{x_1, \ldots, x_n\}$ as x-variables and $\{y_1, \ldots, y_m\}$ as y-variables. Let $\prec_x$ and $\prec_y$ be monomial orderings on x and y variables respectively. Define an ordering relation $\prec$ on $\mathbb{Z}^{n+1}_{\geq 0}$ (i.e set of all monomials in indeterminate $x_1, \ldots, x_n, y_1, \ldots, y_m$) as follows:

$$x^{\alpha^{(1)}} y^{\beta^{(1)}} \prec_{[x \succ y]} x^{\alpha^{(2)}} y^{\beta^{(2)}} \Longleftrightarrow \begin{cases} \alpha^{(1)} \prec_x \alpha^{(2)} \\ \text{or} \\ \alpha^{(1)} = \alpha^{(2)} \text{ and } \beta^{(1)} \prec_{\text{y}} \beta^{(2)} \end{cases},$$

where $\alpha^{(1)}, \alpha^{(2)} \in \mathbb{Z}^n_{\geq 0}$ and $\beta^{(1)}, \beta^{(2)} \in \mathbb{Z}^m_{\geq 0}$. The term order $\prec_{[x \succ y]}$ is called elimination order with the x variables larger than the y variables (which is indeed a term order).

Elimination Ideal: l^{th}-elimination ideal of an ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ with respect to an elimination order $x_1 \succ x_2 \ldots \succ x_n$ is defined as $\mathfrak{a}_l = \mathfrak{a} \cap k[x_{l+1}, \ldots, x_n]$.

Elimination Theorem

(*Buchberger 1987*)

Elimination order: Consider $k[x_1, \ldots, x_n, y_1, \ldots, y_m]$ a polynomial ring in indeterminate $x_1, \ldots, x_n$, $y_1, \ldots, y_m$. We refer to $\{x_1, \ldots, x_n\}$ as x-variables and $\{y_1, \ldots, y_m\}$ as y-variables. Let $\prec_x$ and $\prec_y$ be monomial orderings on x and y variables respectively. Define an ordering relation $\prec$ on $\mathbb{Z}_{\geq 0}^{n+1}$ (i.e set of all monomials in indeterminate $x_1, \ldots, x_n, y_1, \ldots, y_m$) as follows:

$$x^{\alpha^{(1)}} y^{\beta^{(1)}} \prec_{[x \succ y]} x^{\alpha^{(2)}} y^{\beta^{(2)}} \Longleftrightarrow \begin{cases} \alpha^{(1)} \prec_x \alpha^{(2)} \\ \text{or} \\ \alpha^{(1)} = \alpha^{(2)} \text{ and } \beta^{(1)} \prec_{\text{y}} \beta^{(2)} \end{cases},$$

where $\alpha^{(1)}, \alpha^{(2)} \in \mathbb{Z}_{\geq 0}^{n}$ and $\beta^{(1)}, \beta^{(2)} \in \mathbb{Z}_{\geq 0}^{m}$. The term order $\prec_{[x \succ y]}$ is called elimination order with the x variables larger than the y variables (which is indeed a term order).

Elimination Ideal: l^{th}-elimination ideal of an ideal $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ with respect to an elimination order $x_1 \succ x_2 \ldots \succ x_n$ is defined as $\mathfrak{a}_l = \mathfrak{a} \cap k[x_{l+1}, \ldots, x_n]$.

Elimination Theorem: *Let $\mathfrak{a} \subset k[x_1, \ldots, x_n]$ be an ideal and let $G \subset k[x_1, \ldots, x_n]$ be a Gröbner basis of $\mathfrak{a}$ with respect to term order $x_1 \succ x_2 \succ \ldots \succ x_n$. Then for every $0 \leq l \leq n$ the set $G_l = G \cap k[x_{l+1}, \ldots, x_n]$ is a Gröbner basis of lth elimination ideal $\mathfrak{a}_l$.*

Application to Maximum Likelihood Estimation

Recall...

The ideal which represents the solutions to likelihood equations

$$\frac{\partial l}{\partial \theta_j} = \frac{u_1}{f_1(\theta)}\frac{\partial f_1(\theta)}{\partial \theta_j} + \ldots + \frac{u_m}{f_m(\theta)}\frac{\partial f_m(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, d.$$

is

$$\mathfrak{a} = \widehat{\mathfrak{a}} \cap \mathbb{R}[\theta_1, \ldots, \theta_d] \ ,$$

where

$$\widehat{\mathfrak{a}} = \left\langle \underbrace{y_1 f_1(\theta) - 1, \ldots, y_m f_m(\theta) - 1}, \underbrace{\sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_1}, \ldots, \sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_d}} \right\rangle$$

Application to Maximum Likelihood Estimation

Recall...

The ideal which represents the solutions to likelihood equations

$$\frac{\partial l}{\partial \theta_j} = \frac{u_1}{f_1(\theta)}\frac{\partial f_1(\theta)}{\partial \theta_j} + \ldots + \frac{u_m}{f_m(\theta)}\frac{\partial f_m(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, d.$$

is

$$\mathfrak{a} = \widehat{\mathfrak{a}} \cap \mathbb{R}[\theta_1, \ldots, \theta_d] \ ,$$

where

$$\widehat{\mathfrak{a}} = \left\langle \underbrace{y_1 f_1(\theta) - 1, \ldots, y_m f_m(\theta) - 1}, \underbrace{\sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_1}, \ldots, \sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_d}} \right\rangle$$

Estimation by Gröbner bases

step 1: Calculate a Gröbner basis $\widehat{G}$ for the ideal $\widehat{\mathfrak{a}}$.

Application to Maximum Likelihood Estimation

Recall...
The ideal which represents the solutions to likelihood equations

$$\frac{\partial l}{\partial \theta_j} = \frac{u_1}{f_1(\theta)} \frac{\partial f_1(\theta)}{\partial \theta_j} + \ldots + \frac{u_m}{f_m(\theta)} \frac{\partial f_m(\theta)}{\partial \theta_j} = 0, \quad j = 1, \ldots, d.$$

is

$$\mathfrak{a} = \widehat{\mathfrak{a}} \cap \mathbb{R}[\theta_1, \ldots, \theta_d] \ ,$$

where

$$\widehat{\mathfrak{a}} = \left\langle \underbrace{y_1 f_1(\theta) - 1, \ldots, y_m f_m(\theta) - 1}, \underbrace{\sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_1}, \ldots, \sum_{j=1}^{m} u_j y_j \frac{\partial f_j}{\partial \theta_d}} \right\rangle$$

Estimation by Gröbner bases

step 1: Calculate a Gröbner basis $\widehat{G}$ for the ideal $\widehat{\mathfrak{a}}$.

step 2: Polynomials in $\widehat{G}$ which doesn't involve $y_1, \ldots, y_m$ forms a Gröbner basis whose variety gives the solution to the ML estimation.

Application to Minimax estimation and Feature selection

Recall... Ideal which represents the maximum entropy solution

$$\mathfrak{a} = \left\langle \sum_{j=1}^{m} (t_i(j) - T_i) \prod_{i=1}^{d} \theta_i^{t_i(j)} \;:\; i = 1, \ldots, d \right\rangle$$

Application to Minimax estimation and Feature selection

Recall... Ideal which represents the maximum entropy solution

$$\mathfrak{a} = \left\langle \sum_{j=1}^{m}(t_i(j) - T_i)\prod_{i=1}^{d}\theta_i^{t_i(j)} \;:\; i = 1, \ldots, d \right\rangle$$

Zhu-Mumford method of estimation and feature selection

- ▷ Let X be a rv, let l be the true distribution. Let p^* be the ME distribution with respect to the feature functions t_i, $i = 1, \ldots, d$.
- ▷ KL-distance from p^* to l is

$$I(l \| p^*) = S(p*) - S(l)$$

- ▷ Feature selection by minimizing $I(l \| p^*)$ with respect to the feature subsets.
- ▷ Zhu-Mumford algorithm requires estimation of ME-distribution with respect to the various feature subsets.

Application to Minimax estimation and Feature selection

Recall... Ideal which represents the maximum entropy solution

$$\mathfrak{a} = \left\langle \sum_{j=1}^{m} (t_i(j) - T_i) \prod_{i=1}^{d} \theta_i^{t_i(j)} \ : \ i = 1, \ldots, d \right\rangle$$

Zhu-Mumford method of estimation and feature selection

▷ Let X be a rv, let l be the true distribution. Let p^* be the ME distribution with respect to the feature functions t_i, $i = 1, \ldots, d$.

▷ KL-distance from p^* to l is

$$I(l \| p^*) = S(p*) - S(l)$$

▷ Feature selection by minimizing $I(l \| p^*)$ with respect to the feature subsets.

▷ Zhu-Mumford algorithm requires estimation of ME-distribution with respect to the various feature subsets.

Application of elimination theorem to Zhu-Mumford algorithms...

Embedding ME models in Algebraic Varieties

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Statistical model as image of a rational function:
Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}$ be a semi-algebraic set. Let $\kappa : \mathbb{R} \to \mathbb{R}^{\geqslant}$ be a rational function such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a *parametric algebraic statistical model*.

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Statistical model as image of a rational function:
Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}$ be a semi-algebraic set. Let $\kappa : \mathbb{R} \to \mathbb{R}^{\geqslant}$ be a rational function such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a *parametric algebraic statistical model*.

How Commutative algebra and Algebraic geometry plays a role:
Parametric representation of statistical models $p_j = f_j(\theta_1, \ldots, \theta_d), \ \ j = 1, \ldots, m$ can be viewed in three different ways

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Statistical model as image of a rational function:
Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}$ be a semi-algebraic set. Let $\kappa : \mathbb{R} \to \mathbb{R}^{>}$ be a rational function such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a *parametric algebraic statistical model*.

How Commutative algebra and Algebraic geometry plays a role:
Parametric representation of statistical models $p_j = f_j(\theta_1, \ldots, \theta_d), \ \ j = 1, \ldots, m$ can be viewed in three different ways

- ▷ parametric description of a curve in m dimentional space

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Statistical model as image of a rational function:
Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}$ be a semi-algebraic set. Let $\kappa : \mathbb{R} \rightarrow \mathbb{R}^{>}$ be a rational function such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a *parametric algebraic statistical model*.

How Commutative algebra and Algebraic geometry plays a role:
Parametric representation of statistical models $p_j = f_j(\theta_1, \ldots, \theta_d), \ \ j = 1, \ldots, m$ can be viewed in three different ways

- ▷ parametric description of a curve in m dimentional space
- ▷ homeomorphism of the ring $k[p_1, \ldots, p_m]$ to $k[\theta_1, \ldots, \theta_d]$ identified by $p_j \mapsto f_j(\theta_1, \ldots, \theta_d)$

Algebraic Formulations

Semi-algebraic set: A set $\Theta \subseteq \mathbb{R}$ is called *semi-algebraic set*, if there are two finite collection of polynomials $F \subset k[x_1, \ldots, x_d]$ and $G \subset k[x_1, \ldots, x_d]$ such that

$$\Theta = \{\theta \in \mathbb{R}^d : f(\theta) = 0, \forall f \in F \text{ and } g(\theta) \geq 0, \forall g \in G\} \ .$$

Statistical model as image of a rational function:
Let Δ_{m-1} be a probability simplex and $\Theta \subset \mathbb{R}$ be a semi-algebraic set. Let $\kappa : \mathbb{R} \to \mathbb{R}^{>}$ be a rational function such that $\kappa(\Theta) \subseteq \Delta_{m-1}$. Then the image $\mathcal{M} = \kappa(\Theta)$ is a *parametric algebraic statistical model*.

How Commutative algebra and Algebraic geometry plays a role:
Parametric representation of statistical models $p_j = f_j(\theta_1, \ldots, \theta_d), \ \ j = 1, \ldots, m$ can be viewed in three different ways

- ▷ parametric description of a curve in m dimentional space
- ▷ homeomorphism of the ring $k[p_1, \ldots, p_m]$ to $k[\theta_1, \ldots, \theta_d]$ identified by $p_j \mapsto f_j(\theta_1, \ldots, \theta_d)$
- ▷ solution of polynomial equations (then projected to k^m) $p_j - f_j(\theta_1, \ldots, \theta_d) = 0$

Implicit ME-model: Main Theorem

Given positive integer valued functions $t_i, i = 1, \dots, d$ we have maximum entropy model as image of

$$\begin{aligned} f : k^d &\rightarrow k^m - W \\ (\theta_1, \dots, \theta_d) &\mapsto \left(\frac{\prod_{i=1}^d \theta_i^{t_i(1)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(j)}}, \dots, \frac{\prod_{i=1}^d \theta_i^{t_i(m)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)}} \right) \quad . \end{aligned} \tag{22}$$

where $W = \mathcal{V}(\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)})$

Implicit ME-model: Main Theorem

Given positive integer valued functions $t_i, i = 1, \ldots, d$ we have maximum entropy model as image of

$$\begin{aligned} f : k^d &\to k^m - W \\ (\theta_1, \ldots, \theta_d) &\mapsto \left(\frac{\prod_{i=1}^d \theta_i^{t_i(1)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(j)}}, \ldots, \frac{\prod_{i=1}^d \theta_i^{t_i(m)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)}} \right) \quad . \end{aligned} \tag{23}$$

where $W = \mathcal{V}(\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)})$

theorem Let f be a polynomial functions which parameterize maximum entropy model with respect to sufficient statistic $t_i : \mathbb{R} \to \mathbb{Z}_{\geq 0}$ according to (28). Then

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(\ker(\tilde{\mathrm{f}}^*))} \cap \underbrace{\mathcal{V}(\sum_{j=1}^m p_j - 1)} \tag{24}$$

Implicit ME-model: Main Theorem

Given positive integer valued functions $t_i, i = 1, \ldots, d$ we have maximum entropy model as image of

$$\begin{aligned} f : k^d &\rightarrow k^m - W \\ (\theta_1, \ldots, \theta_d) &\mapsto \left(\frac{\prod_{i=1}^d \theta_i^{t_i(1)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(j)}}, \ldots, \frac{\prod_{i=1}^d \theta_i^{t_i(m)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)}} \right) \quad . \end{aligned} \tag{25}$$

where $W = \mathcal{V}(\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)})$

theorem Let f be a polynomial functions which parameterize maximum entropy model with respect to sufficient statistic $t_i : \mathbb{R} \rightarrow \mathbb{Z}_{\geq 0}$ according to (28). Then

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(\ker(\tilde{\mathrm{f}}^*))} \cap \underbrace{\mathcal{V}(\sum_{j=1}^m p_j - 1)} \tag{26}$$

where $\tilde{f}^*$ is a k-algebra homeomorphism

$$\begin{aligned} \tilde{f}^* : k[p_1, \ldots, p_m] &\rightarrow k[\theta_0, \ldots, \theta_d] \\ p_j &\mapsto \theta_0 \prod_{i=1}^d \theta_i^{t_i(j)} \quad . \end{aligned} \tag{27}$$

Implicit ME-model: Main Theorem

Given positive integer valued functions $t_i, i = 1, \ldots, d$ we have maximum entropy model as image of

$$\begin{aligned} f : k^d &\to k^m - W \\ (\theta_1, \ldots, \theta_d) &\mapsto \left(\frac{\prod_{i=1}^d \theta_i^{t_i(1)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(j)}}, \ldots, \frac{\prod_{i=1}^d \theta_i^{t_i(m)}}{\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)}} \right) \quad . \end{aligned} \tag{28}$$

where $W = \mathcal{V}(\sum_{j=1}^m \prod_{i=1}^d \theta_i^{t_i(m)})$

theorem Let f be a polynomial functions which parameterize maximum entropy model with respect to sufficient statistic $t_i : \mathbb{R} \to \mathbb{Z}_{\geq 0}$ according to (28). Then

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(\ker(\tilde{\mathrm{f}}^*))} \cap \underbrace{\mathcal{V}(\sum_{j=1}^m p_j - 1)} \tag{29}$$

where $\tilde{f}^*$ is a k-algebra homeomorphism

$$\begin{aligned} \tilde{f}^* : k[p_1, \ldots, p_m] &\to k[\theta_0, \ldots, \theta_d] \\ p_j &\mapsto \theta_0 \prod_{i=1}^d \theta_i^{t_i(j)} \quad . \end{aligned} \tag{30}$$

***The inclusion should be understand from the point of “closure” or “Zariski Closure”

Corollary:

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(\mathfrak{a} \cap k[p_1, \ldots, p_m])} \cap \underbrace{\mathcal{V}(\sum_{j=1}^{m} p_j - 1)} \quad , \tag{31}$$

where

$$\mathfrak{a} = \left\langle p_j - \theta_0 \prod_{i=1}^{d} \theta_i^{t_1(j)} : j = 1, \ldots, m \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d]$$

Ideal and Gröbner bases representation

Corollary:

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(\mathfrak{a} \cap k[p_1, \ldots, p_m])} \cap \underbrace{\mathcal{V}(\sum_{j=1}^{m} p_j - 1)} \quad , \tag{32}$$

where

$$\mathfrak{a} = \left\langle p_j - \theta_0 \prod_{i=1}^{d} \theta_i^{t_1(j)} : j = 1, \ldots, m \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d]$$

Corollary: (By Elimination theorem)

$$\mathrm{im(f)} \subseteq \underbrace{\mathcal{V}(G \cap k[p_1, \ldots, p_m])} \cap \underbrace{\mathcal{V}(\sum_{j=1}^{m} p_j - 1)} \quad , \tag{33}$$

where G is the Gröbner basis of

$$\mathfrak{a} = \left\langle p_j - \theta_0 \prod_{i=1}^{d} \theta_i^{t_1(j)} : j = 1, \ldots, m \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d]$$

Ideal and Gröbner bases representation

Corollary:

$$\mathrm{im}(\mathrm{f}) \subseteq \underbrace{\mathcal{V}(\mathfrak{a} \cap k[p_1, \ldots, p_m])} \cap \underbrace{\mathcal{V}(\sum_{j=1}^{m} p_j - 1)} \ , \tag{34}$$

where

$$\mathfrak{a} = \left\langle p_j - \theta_0 \prod_{i=1}^{d} \theta_i^{t_1(j)} : j = 1, \ldots, m \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d]$$

Corollary: (By Elimination theorem)

$$\mathrm{im}(\mathrm{f}) \subseteq \underbrace{\mathcal{V}(G \cap k[p_1, \ldots, p_m])} \cap \underbrace{\mathcal{V}(\sum_{j=1}^{m} p_j - 1)} \ , \tag{35}$$

where G is the Gröbner basis of

$$\mathfrak{a} = \left\langle p_j - \theta_0 \prod_{i=1}^{d} \theta_i^{t_1(j)} : j = 1, \ldots, m \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d]$$

with a term order $\{p_1, \ldots, p_m\} \prec \{\theta_1, \ldots, \theta_d\}$

ME model as a rational parameterization

Theorem

$$\mathrm{im(f)} \subseteq \mathcal{V}(\mathfrak{b} \cap \mathrm{k}[\mathrm{p}_1, \ldots, \mathrm{p}_\mathrm{m}]) \ , \tag{36}$$

where

$$\mathfrak{b} = \left\langle Zp_j - \prod_{i=1}^{d} \theta_i^{t_i(j)} : j = 1, \ldots, m, 1 - Z^m y \right\rangle \subseteq k[p_1, \ldots, p_m, \theta_1, \ldots, \theta_d, y]$$

(Z is the partition function)
**** probably difficult to compute

Example

$$\begin{aligned}
p_1 &= \theta_0 {\theta_1}^2 \theta_2 \\
p_2 &= \theta_0 \theta_1 {\theta_2}^2 \\
p_3 &= \theta_0 {\theta_1}^3 \theta_2 \\
p_4 &= \theta_0 \theta_1 {\theta_2}^4 \\
p_5 &= \theta_0 {\theta_1}^5 {\theta_2}^3 \\
p_6 &= \theta_0 {\theta_1}^2 {\theta_2}^3 \\
p_7 &= \theta_0 \theta_1 \theta_2
\end{aligned}$$

Example (Contd...)

```
g + x y, a y - c, a z - b y, f z - d y, a g - c x, a b - f g, b f - d a, b - d g, b y - c z, g z y - a, b z y - f,
f g y - b c, b c y - e x, g z^2 - b, c z^2 - f y, b z^2 - d, e x z - f c, c x z - b g y, g^2 z - a x, e g z - f c y,
c g z - a^2, b g z - f x, c^2 z - e g, d c z - f^2, b c z - f a, f g^2 - b c x, b c g - e x^2, a^3 - e g x, c a^2 - e g^2,
f a^2 - b e x, d a^2 - f^2 g, f c a - b e g, f^2 a - d e x, d c^2 - f e x, b c^2 - a e x, f^2 c - d e g, f y^3 - e, d g y^2 - f a,
f c y^2 - a e, d c y^2 - b e, f^2 y^2 - b e z, e g^2 y - a c^2, d g^2 y - f a x, b g^2 y - a^2 x, b e g y - f c^2, d c g y - b e x,
b d g y - f^2 x, f c^2 y - a^2 e, d^2 c y - f^3, f^3 y - b d e, e g^3 - c^2 a x, d g^3 - f a x^2, b g^3 - a^2 x^2, b e g^2 - f c^2 x,
d c g^2 - b e x^2, b d g^2 - f^2 x^2, f c^2 g - a^2 e x, d^2 c g - f^3 x, f^3 g - b d e x, f c^3 - e^2 g x, d y^4 - e z, d^3 e g y - f^5,
d^3 e g^2 - f^5 x, f^6 - b d^3 c e]
```

Change of symbols

$p_1 \longrightarrow a$, $p_2 \longrightarrow b$, $p_3 \longrightarrow c$, $p_4 \longrightarrow d$, $p_5 \longrightarrow e$, $p_6 \longrightarrow f$, $p_7 \longrightarrow g$

and

$\theta_0 \longrightarrow x$, $\theta_1 \longrightarrow y$, $\theta_2 \longrightarrow z$

Example (Contd...)

Maximum entropy model is contained in

$$
\begin{aligned}
V_{model} = \mathcal{V}(p_1p_2 - p_6p_7, \\
p_2p_6 - p_4p_1, \\
p_2 - p_4p_7, \\
p_3p_1^2 - p_5p_7^2, \\
p_4p_1^2 - p_6^2p_7, \\
p_6p_3p_1 - p_2p_5p_7, \\
p_6^2p_3 - p_4p_5p_7, \\
p_6^6 - p_2p_4^3p_3p_5 \\
\sum_{i=1}^{7} p_i - 1)
\end{aligned}
$$

Example (Contd...)

Maximum entropy model is contained in

$$
\begin{aligned}
V_{model} = \mathcal{V}(p_1p_2 - p_6p_7,\\
p_2p_6 - p_4p_1,\\
p_2 - p_4p_7,\\
p_3p_1^2 - p_5p_7^2,\\
p_4p_1^2 - p_6^2p_7,\\
p_6p_3p_1 - p_2p_5p_7,\\
p_6^2p_3 - p_4p_5p_7,\\
p_6^6 - p_2p_4^3p_3p_5\\
\sum_{i=1}^{7} p_i - 1)
\end{aligned}
$$

Maximum entropy distribution is contained in $V_{model} \cap V_{data}$.

Summary

Parametric

▷ Estimation of ME models can be transformed to solving system of polynomial equations

- Primal – System of polynomial equations
- Dual – System of Laurent polynomial equations
- Kullback-Csiszár – A triangular system (A decreasing sequence of dimension of quotient vector spaces modulo ideals)

Summary

Parametric

- ▷ Estimation of ME models can be transformed to solving system of polynomial equations
  - ◦ Primal – System of polynomial equations
  - ◦ Dual – System of Laurent polynomial equations
  - ◦ Kullback-Csiszár – A triangular system (A decreasing sequence of dimension of quotient vector spaces modulo ideals)
- ▷ Gröbner bases can be used in the estimation of ME-models

Summary

Parametric

▷ Estimation of ME models can be transformed to solving system of polynomial equations

- ∘ Primal – System of polynomial equations
- ∘ Dual – System of Laurent polynomial equations
- ∘ Kullback-Csiszár – A triangular system (A decreasing sequence of dimension of quotient vector spaces modulo ideals)

▷ Gröbner bases can be used in the estimation of ME-models

▷ Elimination theorem comes in handy (for both ML and ME)

Summary

Parametric

- ▷ Estimation of ME models can be transformed to solving system of polynomial equations
  - ◦ Primal – System of polynomial equations
  - ◦ Dual – System of Laurent polynomial equations
  - ◦ Kullback-Csiszár – A triangular system (A decreasing sequence of dimension of quotient vector spaces modulo ideals)
- ▷ Gröbner bases can be used in the estimation of ME-models
- ▷ Elimination theorem comes in handy (for both ML and ME)

Implicit

- ▷ ME-models can be treated with Toric ideals (One can relax positivity of feature functions)

Summary

Parametric

- ▷ Estimation of ME models can be transformed to solving system of polynomial equations
  - ◦ Primal – System of polynomial equations
  - ◦ Dual – System of Laurent polynomial equations
  - ◦ Kullback-Csiszár – A triangular system (A decreasing sequence of dimension of quotient vector spaces modulo ideals)
- ▷ Gröbner bases can be used in the estimation of ME-models
- ▷ Elimination theorem comes in handy (for both ML and ME)

Implicit

- ▷ ME-models can be treated with Toric ideals (One can relax positivity of feature functions)
- ▷ ME-models can be embedded in Toric varieties (elegant but can we characterize the margin)

Concluding Remarks

Potential Problems

- ▷ Elimination in “implicit models”

Concluding Remarks

Potential Problems

- Elimination in "implicit models"
- "Zero dimensional ideals" and "k-vector dimension of quotient ring modulo ideal" describing ME models

Concluding Remarks

Potential Problems

▷ Elimination in "implicit models"

▷ "Zero dimensional ideals" and "k-vector dimension of quotient ring modulo ideal" describing ME models

▷ Universal Gröbner bases, term order

Concluding Remarks

Potential Problems

- ▷ Elimination in "implicit models"
- ▷ "Zero dimensional ideals" and "k-vector dimension of quotient ring modulo ideal" describing ME models
- ▷ Universal Gröbner bases, term order

Finally...

`In the case of ME, both model and data can be represented by algebraic varieties ''implicitly''`

Concluding Remarks

Potential Problems

- ▷ Elimination in "implicit models"
- ▷ "Zero dimensional ideals" and "k-vector dimension of quotient ring modulo ideal" describing ME models
- ▷ Universal Gröbner bases, term order

Finally...

`In the case of ME, both model and data can be represented by algebraic varieties ‘‘implicitly’’--- probably this result paves a way to algebraic geometry of information theoretic statistics`