1. Introduction

Basic notions
Statistics: probability density
Algebra: ring, field.

Outline
Discrete and Gaussian statistical models
Maximum likelihood estimation
Algebraic geometry basics
Algebraic models

Statistical models

Definition. Let Ω be a measurable space ('sample space'), $\mathrm{Prob}(\Omega)$ the set of all probability densities on Ω. A *statistical model* is a subset $\mathcal{M} \subseteq \mathrm{Prob}\,\Omega$. A *parametric* statistical model is a statistical model $\mathcal{M}$ together with a set $\Theta \subseteq \mathbb{R}^d$ and a surjection $\Theta \to \mathcal{M}$.

Maximum likelihood estimation problem: given $e = (e_1, \ldots, e_N) \in \Omega^N$ independent and identically distributed (i.i.d.) samples, find $x \in M$ such that its density $f_x \in \mathrm{Prob}\,\Omega$ maximises (since i.i.d.)

$$f_x^N(e) = \prod_{i=1}^{n} f_x(e_i)$$

This is equivalent to the problem $(\star)$

$$\max_{x \in \mathcal{M}} \sum_{i=1}^{n} \log f_x(e_i)$$

Example 1
$\Omega = (0, 1, 2, 3)$
$\Delta_3 = \{(p_0, p_1, p_2, p_3) \in \mathbb{R}^4 \mid p_j \geq 0 \text{ for all } j, \sum_j p_j = 1\}$
Independence model (see Fig. 1):
Θ = $\{(\theta, \eta) \in (0,1) \times (0,1)\}$
$p : \Theta \to \Delta_3, \quad (\theta, \eta) \mapsto (\theta\eta, (1-\theta)\eta, (1-\theta)\eta, (1-\theta)(1-\eta))$
$\mathcal{M} = \mathrm{im}\, p = \{p \in \Delta_3 \mid p_0 p_3 - p_1 p_3 = 0\}$.
If $e = (0, 0, 1, 3, 3)$ then

$$(\star) \Leftrightarrow \max_{p \in \mathcal{M}} 2 \log p_0 + \log p_1 + 2 \log p_3.$$

Data vector: $e \to u = (2,1,0,2) \in \mathbb{N}^4$.

Definition. A *discrete statistical model on* $n+1$ *outcomes* is a subset $\mathcal{M} \subseteq \Delta_n$, where

$$\Delta_n = \{p \in \mathbb{R}^{n+1} \mid p_j \geq 0 \text{ for all } j, \sum_j p_j = 1\}.$$

A *data vector* is $u \in \mathbb{N}^{n+1}$.
An *empirical distribution* is $q = u/|u|$ where $|u| = \sum_j u_j$ or just any $q \in \Delta_n$.
The *log-likelihood function* given $q \in \Delta_n$ is

$$\ell : \mathcal{M} \times \Delta_n \to \mathbb{R}, \quad \ell(p,q) = \sum_{j=0}^{n} q_j \log p_j$$

The *ML estimation problem* given q is

$$\max_{p \in \mathcal{M}} \sum_{j=0}^{n} q_j \log p_j.$$

Example 2
Let $\Omega = \mathbb{R}^n$. The Gaussian probability densities on Ω with mean zero correspond 1:1 with the positive definite $n \times n$ symmetric matrices. These are interpreted either as covariance matrices Σ or concentration matrices $K = \Sigma^{-1}$.

$$\Sigma = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) & \mathrm{Cov}(X,Z) \\ \dots & \mathrm{Var}(Y) & \mathrm{Cov}(Y,Z) \\ \dots & \dots & \mathrm{Var}(Z) \end{pmatrix}$$

Σ covariance matrix $\rightsquigarrow$

$$f_\Sigma(x) = (2\pi)^{-n/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right).$$

$n = 3 \rightsquigarrow$ Gaussian random variables X, Y, Z.
$\mathrm{PD}_3 = \{M \in \mathrm{Sym}_2(n \times n, \mathbb{R})\}$
$X \perp\!\!\!\perp Y \Leftrightarrow \mathrm{Cov}(X,Y) = 0$ (since Gaussian)
Independence covariance model: $\mathcal{M} = \{\Sigma = (\sigma_i j) \in \mathrm{PD}_3 \mid \sigma_{12} = 0\}$
Independence concentration model:
$\mathcal{M} = \{K \in \mathrm{PD}_3 \mid K^{-1} \in \mathcal{M}\} = \{K = (k_{ij}) \in \mathrm{PD}_3 \mid k_{12}k_{23} - k_{13}k_{33} = 0\}$.
If $e = (e_1, e_2, e_3) \in (\mathbb{R}^3)^3$, then by [Sullivant] Prop. 5.3.7:

$$(\star) \Leftrightarrow \max_{\Sigma \in \mathcal{M}} -\log\det(\Sigma) - \mathrm{tr}(S\Sigma^{-1}),$$

with $S = (1/3)\sum_{j=1}^{3} e_j e_j^T$ ("sample covariance").

Definition. A *Gaussian statistical model* on n random variables is a subset $\mathcal{M} \subseteq \mathrm{PD}_n$, where

$$\mathrm{PD}_n = \{\Sigma \in \mathrm{Mat}(n \times n, \mathbb{R}) \mid \Sigma \text{ symmetric pos. def.}\}$$

A *data vector* is $e \in (R^n)^N$.
An *empirical distribution* is $S = (1/n)\sum_{j=1}^n e_j e_j^T$ or just any $S \in \mathrm{PD}_n$.
The *negated log-likelihood function* is

$$\ell : \mathcal{M} \times \mathrm{PD}_n \to \mathbb{R}$$

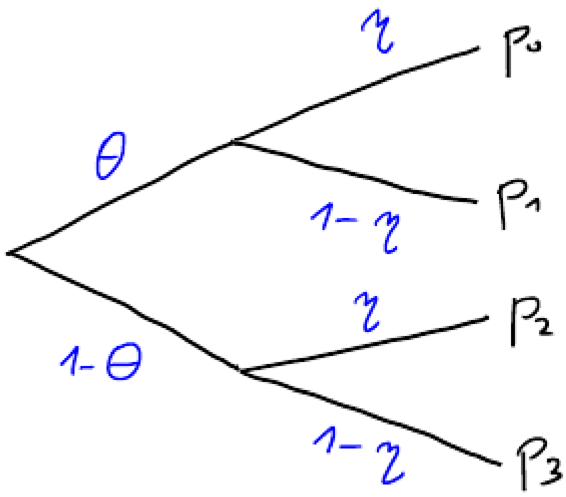
given by

$$\ell_{\mathrm{cov}}(\Sigma, S) = \log\det(\Sigma) + \mathrm{tr}(S\Sigma^{-1}) \qquad \text{(covariance version)}$$
$$\ell_{\mathrm{con}}(K, S) = \log\det(K) - \mathrm{tr}(SK) \qquad \text{(concentration version).}$$

The *ML estimation problem* given S is then

$$\min_{x \in \mathcal{M}_{\mathrm{cov}}} \ell_{\mathrm{cov}}(\Sigma, S) \qquad \text{(covariance version)}$$
$$\max_{x \in \mathcal{M}_{\mathrm{con}}} \ell_{\mathrm{con}}(\Sigma, S) \qquad \text{(concentration version).}$$

Fig. 1



Algebraic geometry basics

Definitions.
K a field $\rightsquigarrow$

$K[x_1, \ldots, x_n] = \{\text{polynomials in the variables } x_1, \ldots, x_n \text{ with coefficients in } K\}$ *polynomial*

$$K^n = \{(x_1, \ldots, x_n) \mid x_i \in K\} \textit{ affine space}$$

$$K(x_1, \ldots, x_n) = \{f/g \mid f, g \in K[x_1, \ldots, x_n], g \neq 0\} \textit{ field of rational functions}$$

$f_1, \ldots, f_k$ polynomials $\rightsquigarrow$

$$I = \langle f_1, \dots, f_k \rangle = \{\sum_{j=1}^{k} g_j f_j \mid g_j \in K[x_1, \dots, x_n]\} \text{ the } \textit{ideal} \text{ generated by } f_1, \dots, f_k$$

$$V(I) = \{x \in K^n \mid f_1(x) = \dots = f_k(x) = 0\} = \{x \mid g(x) = 0 \text{ for all } g \in I\} \text{ the } \textit{algebraic vari}$$

An *ideal*: any such $I = \langle f_1, \dots, f_k$.
A *variety*: any such $V = V(I)$.
$S \subseteq K^n$ a subset $\rightsquigarrow$

$$\overline{S} = \bigcap_{V \supseteq S \text{ variety}} V \quad \text{the } \textit{closure} \text{ of } S$$

$$I(S) = \{f \in k[x_1, \dots, x_n] \mid f(x) = 0 \text{ for all } x \in S\} \text{ the } \textit{ideal} \text{ of } S$$

$I \subseteq K[x_1, \dots, x_n]$ an ideal $\rightsquigarrow$

$$\sqrt{I} \text{ or } \operatorname{rad}(I) = \{g \in K[x_1, \dots, x_n] \mid \exists k \in \mathbb{N} : g^k \in I\} \text{ the } \textit{radical} \text{ of } I$$

Hilbert's Nullstellensatz: if $K = \mathbb{C}$ then for all ideals I and sets $S \subseteq K^n$:

$$I(V(I)) = \operatorname{rad}(I),$$

$$V(I(S)) = \overline{S}.$$

Examples.

1. Let $I = \langle (x^2) \rangle \subseteq K[x]$. Then $\operatorname{rad}(I) = \langle x \rangle$ and $V(I) = V(\langle x \rangle) = \{0\}$.
2. From the Nullstellensatz it follows that $V(I_1) = V(I_2) \Leftrightarrow \operatorname{rad}(I_1) = \operatorname{rad}(I_2)$.
3. The Nullstellensatz does not work over $\mathbb{R}$: take $I_1 = \langle x^2 + 1 \rangle \subset \mathbb{R}[x]$. Then $V(I_1) = \emptyset = V(\mathbb{R}[x])$ but $\operatorname{rad}(I_1) \neq \mathbb{R}[x]$.

Definition. Let $K = \mathbb{R}$ and $f_1, \dots, f_k, g_1, \dots, g_\ell \in \mathbb{R}[x_1, \dots, x_n]$. Then

$$\operatorname{semialg}(f_1, \dots, f_k \mid g_1, \dots, g_\ell) = \{x \in \mathbb{R}^n \mid f_i(x) = 0, g_j(x) > 0 \text{ for all } i, j\}$$

is the *semialgebraic set* defined by the f_i, g_j.
A *semialgebraic set*: any such $\mathcal{M} = \operatorname{semialg}(f_i \mid g_j)$.
Its *complexification* is $\mathcal{M}_\mathbb{C} = V(f_1^\mathbb{C}, \dots, f_k^\mathbb{C}) \subseteq \mathbb{C}^n$, where $f_i^\mathbb{C}$ is the image of f_i in $\mathbb{C}[x_1, \dots, x_n]$.

Algebraic models

Definition. An *algebraic model* $\mathcal{M} \subseteq \mathcal{N}$ is a semialgebraic subset $\mathcal{M}$ of a semi algebraic set $\mathcal{N}$ of probability densities on some space Ω.
$\mathcal{N}$ is the space of *empirical distributions*.
Its *complexification* is $\mathcal{M}_\mathbb{C} \subseteq \mathcal{N}_\mathbb{C}$

Examples: Example 1, Example 2.