Maximum Likelihood Estimation for Toric Varieties

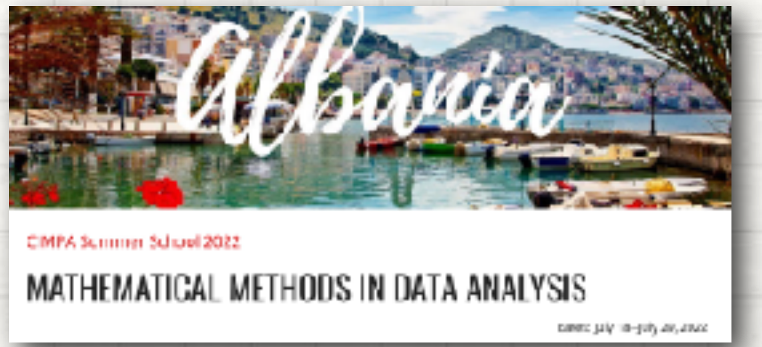
Centro de Matemática Universidade do Porto

CIMPA Summer School 2022

Mathematical Methods in Data Analysis

Eliana Duarte

CENTRO DE
MATEMÁTICA
UNIVERSIDADE DO PORTO



Lecture #1:

Toric varieties are exponential families + Examples

Lecture #2:

Fundamentals of MLE for toric varieties,
an algebraic approach

Lecture #3:

Toric varieties with rational MLE, recent trends
and open questions

$\mathcal{M}_\Theta$ = parametric statistical model $\subseteq \mathring{\Delta}_{r-1}$

$D = \{X^{(1)}, \dots, X^{(n)}\}$ i.i.d set of observations

The likelihood function is

$$L(\theta \mid D) := p_\theta(D) = \prod_{i=1}^{n} p_\theta(X^{(i)})$$

u_j = # of times j is observed in D

$$p_\theta(D) = \prod_{j=1}^{r} p_\theta(j)^{u_j}$$

$u = (u_1, \dots, u_n)$ summarizes D

The maximum likelihood estimate (MLE) $\hat{\theta}$ is the maximizer of $L(\theta \mid D)$

$$\hat{\theta} := \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta \mid D)$$

$\ell(\theta \mid D) := \log(L(\theta \mid D))$ is the log-likelihood

Example 2.1

Binomial(r) $\qquad r = \#$ of trials $\qquad D = \{X^{(1)}, \dots, X^{(n)}\}$

$u_j = \#$ of times each binomial experiment gave j successes

$$\mathcal{L}(\theta|D) = \prod_{i=1}^{n} P_\theta(X^{(i)}) = \prod_{i=0}^{r} P_\theta(i)^{u_i} = \prod_{i=0}^{r} \left[\binom{r}{i}\theta^{i}(1-\theta)^{r-i}\right]^{u_i}, \quad \theta \in (0,1)$$

$$\ell(\theta|D) = \log\left(\prod_{i=0}^{r}\left[\binom{r}{i}\theta^{i}(1-\theta)^{r-i}\right]^{u_i}\right) =$$

$$= \log\left(\prod_{i=0}^{r}\binom{r}{i}^{u_i}\theta^{\sum_{i=0}^{r} i u_i}(1-\theta)^{\sum_{i=0}^{r}(r-i)u_i}\right)$$

$$= C + \left(\sum_{i=0}^{r} i u_i\right)\log(\theta) + \left(\sum_{i=0}^{r}(r-i)u_i\right)\log(1-\theta)$$

$$\frac{\partial \ell(\theta|D)}{\partial \theta} = \frac{\sum_{i=1}^{r} i u_i}{\theta} - \frac{\sum_{i=0}(r-i)u_i}{1-\theta} = 0 \iff \hat{\theta} = \frac{\sum_{i=0}^{r} i u_i}{rn}$$

$$\hat{\theta} = \frac{\sum_{i=0}^{r} i\, u_i}{rn} \qquad 1 - \hat{\theta} = \frac{\sum_{i=0}^{r} (r-i) u_i}{rn} \qquad n = \sum_{i=0}^{r} u_i$$

$$\Phi : \mathbb{N}^r \longrightarrow \mathcal{M} \subseteq \mathring{\Delta}_r$$

Maximum likelihood Estimator

$$(u_0, \ldots, u_r) \longmapsto \Bigg(\underbrace{\binom{r}{j} \hat{\theta}^j (1-\hat{\theta})^{r-j}}_{\Phi_j} \Bigg)_{j \in \{0,1,\ldots,r\}}$$

Φ_j , j-th coordinate of Φ

Φ is a rational function of $u = (u_0, \ldots, u_r)$!

For discrete models and data, $u=(u_1,\dots,u_r)$

$$\ell(\theta|u) = \sum_{j=1}^{r} u_j \log p_j(\theta)$$

To optimize we differentiate and set $= 0$.

Score equations

$$\sum_{j=1}^{r} \frac{u_j}{p_j} \frac{\partial p_j(\theta)}{\partial \theta_i} = 0 \quad , \quad i = 1,\dots,d$$

For discrete models and data, $u=(u_1,\dots,u_r)$

$$\ell(\theta\mid u) = \sum_{j=1}^{r} u_j \log p_j(\theta)$$

To optimize we differentiate and set $= 0$.

Score equations $\left\{ \sum_{j=1}^{r} \frac{u_j}{p_j} \frac{\partial p_j(\theta)}{\partial \theta_i} = 0 \quad , \quad i = 1,\dots,d \right.$

Theorem 2.2: Let $\mathcal{M}_{\Theta} \subseteq \Delta_{r-1}$ be a statistical model. For generic data, the number of solutions to the score eqns is independent of u.

Def 2.3: The number of solutions for generic u is called the *maximum likelihood degree* of the parametric statistical model $p: \Theta \longrightarrow \mathring{\Delta}_{r-1}$

If the model $\mathcal{M}_{\Theta}$ has MLdegree one $\Rightarrow$ its MLE is a rational function of u.

$$\Phi: \mathbb{N}^r \longrightarrow \mathcal{M}_{\Theta} \quad \text{is rational}$$
$$u \longmapsto \hat{p}$$

Example 2.1

$$(u_0, \ldots, u_r) \longmapsto \Big(\underbrace{\binom{r}{j} \left(\frac{\sum i u_i}{r u_+} \right) \left(\frac{\sum (r-i) u_i}{r u_+} \right)}_{\Phi_j \text{ , } j\text{-th coordinate of } \Phi} \Big)_{j \in \{0, 1, \ldots, r\}}$$

Examples:

- The simplex Δ_n, $\Phi: u \mapsto \frac{1}{u_0 + \cdots + u_n}(u_0, \ldots, u_n)$
- The independence model

$$\Phi: (u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+} u_{+0}}{u_{++}^2}, \frac{u_{0+} u_{+1}}{u_{++}^2}, \frac{u_{1+} u_{+0}}{u_{++}^2}, \frac{u_{1+} u_{+1}}{u_{++}^2} \right)$$

$$u_{++} := \sum u_{ij}, \quad u_{i+} = u_{i0} + u_{i1}, \quad u_{+j} = u_{0j} + u_{1j}$$

- Do you know of any others?

Examples:

- The simplex Δ_n, $\Phi: u \mapsto \frac{1}{u_0 + \cdots + u_n}(u_0, \ldots, u_n)$
- The independence model

$$\Phi: (u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+}u_{+0}}{u_{++}^2}, \frac{u_{0+}u_{+1}}{u_{++}^2}, \frac{u_{1+}u_{+0}}{u_{++}^2}, \frac{u_{1+}u_{+1}}{u_{++}^2} \right)$$

$$u_{++} := \sum u_{ij}, \quad u_{i+} = u_{i0} + u_{i1}, \quad u_{+j} = u_{0j} + u_{1j}$$

- Do you know of any others?

Q: Models with rational MLE are a class of algebraic varieties, can we characterize them?

Horn Uniformization

Horn Uniformization

$H=(h_{ij})$ is a $m\times(n+1)$ integer matrix

Horn Uniformization

$H = (h_{ij})$ is a $m \times (n+1)$ integer matrix

Columns of H sum to zero $\sum_{i=1}^{m} h_{ij} = 0$, $j = 0, \dots, n$.

Horn Uniformization

$H = (h_{ij})$ is a $m \times (n+1)$ integer matrix

Columns of H sum to zero $\sum_{i=1}^{m} h_{ij} = 0$, $j = 0, \ldots, n$.

$$(Hu)^{h_j} := \prod_{i=1}^{m} (h_{i0} u_0 + h_{i1} u_1 + \cdots + h_{in} u_n)^{h_{ij}}, \quad j = 0, 1, \ldots, n$$

Horn Uniformization

$H = (h_{ij})$ is a $m \times (n+1)$ integer matrix

Columns of H sum to zero $\sum_{i=1}^{m} h_{ij} = 0$, $j = 0, \ldots, n$.

$$(Hu)^{h_j} := \prod_{i=1}^{m} (h_{i0}u_0 + h_{i1}u_1 + \cdots + h_{in}u_n)^{h_{ij}}, \quad j = 0, 1, \ldots, n$$

For $\lambda \in \mathbb{R}^{n+1}$, we say (H, λ) is a Horn pair, if the rational map $\varphi_{(H,\lambda)} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$

$$u \longmapsto \left(\lambda_0 (Hu)^{h_0}, \lambda_1 (Hu)^{h_1}, \ldots, \lambda_n (Hu)^{h_n}\right)$$

- maps positive vectors to positive vectors, and
- $\lambda_0 (Hu)^{h_0} + \lambda_1 (Hu)^{h_1} + \cdots + \lambda_n (Hu)^{h_n} = 1$

The MLE as a Horn pair

$$\Phi : (u_{00}, u_{01}, u_{10}, u_{11}) \mapsto \left(\frac{u_{0+} u_{+0}}{u_{++}^2}, \frac{u_{0+} u_{+1}}{u_{++}^2}, \frac{u_{1+} u_{+0}}{u_{++}^2}, \frac{u_{1+} u_{+1}}{u_{++}^2} \right)$$

$$\lambda = (4, 4, 4, 4)$$

$$H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -2 & -2 & -2 & -2 \end{pmatrix}, \quad Hu = \begin{pmatrix} u_{0+} \\ u_{1+} \\ u_{+0} \\ u_{+1} \\ -2u_{++} \end{pmatrix}$$

$$\varphi_{(H,\lambda)} = \Phi$$

Theorem (D., Marigliano, Sturmfels):

The following are equivalent for a statistical model $\mathcal{M}$ with MLE Φ:

(1) The model $\mathcal{M}$ has rational MLE.

(2) There exist a Horn pair (H, λ) such that $\mathcal{M}$ is the image of the Horn map $\varphi_{(H,\lambda)}$.

(3) Formulation in terms of toric geometry.

$\Phi = \varphi_{(H,\lambda)}$ on $\mathbb{R}^{n+1}_{>0}$.

Examples:

- Undirected decomposable graphical models
- Discrete Bayesian networks
- Staged tree models
- Multinomial model