1 Maximum Likelihood Method

Definition 1. A **(statistical) model** is a parametric family $\mathcal{M}_\Theta = \{P_\theta : \theta \in \Theta\}$ of probability measures of some fixed measurable space $\mathcal{X}$, where Θ is some nonempty set.

In most examples we consider, $\mathcal{X}$ is a subset of $\mathbb{R}^m$ for some $m \in \mathbb{N}$ with Lebegsue measure or counting measure. We will assume hereafter every $P_\theta \in M_\Theta$ has a density function $p_\theta : \mathcal{X} \to [0,1]$ with respect to the chosen measure μ on $\mathcal{X}$:

$$P_\Theta\{A\} = \int_A p_\theta(x)\, d\mu(x)$$

for all measurable sets $A \subseteq \mathcal{X}$.

Definition 2. A **sample** for a model $\mathcal{M}_\Theta$ is a family of random variables $X = (X^{(i)})_{i \le n}$ that have the same distribution $P_\theta \in \mathcal{M}_\Theta$ for all $i \le n$.

Problem. Fix a model $\mathcal{M}_\Theta$. Given a sample X with distribution $P_\theta \in \mathcal{M}_\Theta$, find θ.

Definition 3. Let $\mathcal{M}_\Theta$ be a model. The **likelihood function** of a sample X of size n is the mapping $L_X : \Theta \to [0,1]$, given by

$$L_X(\theta) = \prod_{i=1}^{n} p_\theta\left(X^{(i)}\right), \quad \theta \in \Theta.$$

The **log-likelohood function** of X is $\ell_X : \Theta \to (-\infty, 0]$, $\ell_X(\theta) = \ln L_x(\theta)$ for $\theta \in \Theta$.

Definition 4. Given a model $\mathcal{M}_\Theta$ and a sample X, the **maximum likelihood estimator** is the random variable

$$\hat{\theta} := \arg\max_{\theta \in \Theta} \ell_X(\theta) = \arg\max_{\theta \in \Theta} L_X(\theta).$$

For any realization $(x_1, \ldots, x_n) \in \mathcal{X}^n$ of the sample X, the corresponding realization θ_{MLE} of $\hat{\theta}$ is called the **maximum likelihood estimate (MLE)** for the data $(x_1, \ldots, x_n)$.

Example 1. Let $\mathcal{X} = \mathbb{R}^m$ and consider a rational curve C in $\mathcal{X}$, given by a polynomial parametrization $g = (g_1, \ldots, g_m) : \mathbb{R} \to \mathbb{R}^m$. Define a model

$$M_\Theta = \{\mathcal{N}(\mu, I_m) : \mu = g(t)\}_{t \in \mathbb{R}},$$

where $\mathcal{N}(\mu, \Sigma)$ denotes the multivariate normal distribution with parameters $\mu \in \mathbb{R}^m$, $\Sigma \in \mathbb{R}^{m \times m}$, which can be defined by its density function

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x} - \mu)^T \Sigma^{-1} (\vec{x} - \mu)\right), \quad \vec{x} \in \mathbb{R}^m.$$

Then for some constant C,

$$\ell_n(t) = -\frac{n}{2} \left\| g(t) - \frac{1}{n} \sum_{i=1}^{n} X^{(i)} \right\|_2 + C.$$

Observe that $\ell_n(t)$ is a polynomial in $\mathbb{R}[t]$, so we may use algebraic methods to describe the MLE. Let $(\cdot)'$ denote the (formal) derivative of a polynomial, then

$$t_{\mathrm{MLE}} \in \mathbf{V}(\langle \ell_n' \rangle).$$

The example file `gaussian-models.m2` demonstrates a realization this method.

2 Implicit models

In this section we will consider the distributions on a finite set $\mathcal{X} = \{a_1, \dots, a_k\}$. We assume that each value has a nonzero probability.

The probability measures P on $\mathcal{X}$ are in one-to-one correspondence with vectors $\vec{p} \in (0,1)^k$ of probabilities: $\vec{p} = (p_1, \dots, p_k) \subseteq [0,1]^k$, $\sum p_i = 1$. We will write $P = P_{\vec{p}}$ if $P(a_i) = p_i$ for all $i \le k$.

We may describe any probability vector $\vec{p}$ algebraically using polynomials in the ring $\mathbb{R}[x_1, \dots, x_k]$. First observe that

$$\vec{p} \in \Delta_{k-1}^{\circ} := (0,\infty)^k \cap \mathbf{V}\left(\sum_{i=1}^{k} x_i - 1\right) \subseteq \mathbb{R}^k.$$

Let us identify Δ_{k-1}° with the subset of the complex projective space $\mathbb{P}^{k-1}(\mathbb{C})$:

$$\Delta_{k-1}^{\circ} = \{(p_1 : \ldots : p_k) : \operatorname{Re} p_i > 0, \operatorname{Im} p_i = 0 \ \forall i \le k\} \subseteq \mathbb{P}^{k-1}(\mathbb{C}).$$

Then any projective variety in $\mathbb{P}^{k-1}(\mathbb{C})$ defines a set of probability vectors. We will consider the varieties of the form $\mathbf{V}(I)$ where $I \subseteq \mathbb{R}[x_1, \dots, x_k]$ is a homogenous prime ideal that satisfies the reqularity condition

$$\mathbf{V}(I) = \overline{\mathbf{V}_{\Delta_{k-1}^{\circ}}(I)} = \mathbf{V}(\mathbf{I}(\mathbf{V}(I) \cap \Delta_{k-1}^{\circ})). \tag{1}$$

Definition 5. Let $I \subseteq \mathbb{R}[x_1, \dots, x_k]$ be a homogenous prime ideal satisfying (1). We define the **implicit model** $\mathcal{M}(I)$ as the statistical model with the parameter $\vec{p}$:

$$\mathcal{M}(I) := \{P_{\vec{p}} \mid \vec{p} \in \mathbf{V}(I) \cap \Delta_{k-1}^{\circ}\}.$$

Definition 6. Let $X = \left(X^{(i)}\right)_{i \le n}$ be a sample distributed over $\mathcal{X} = \{a_1, \dots, a_k\}$. The **vector of counts** is the random vector $\vec{u} \in \mathbb{Z}_{\ge 0}^k$, where

$$u_j = \#\left\{i \le n : X^{(i)} = a_j\right\}, \quad j \le k.$$

A **data vector** is a realization of $\vec{u}$ for some realization $(x_1, \dots, x_n)$ of X.

Given a vector of counts $\vec{u}$, we may calculate the likelihood function of X :

$$L_X(\vec{p}) = \prod_{i=1}^{n} P_{\vec{p}}\left(X^{(i)}\right) = \prod_{j=1}^{k} \prod_{X^{(i)}=a_i} P_{\vec{p}}\left(X^{(i)}\right) = \prod_{j=1}^{k} \prod_{X^{(i)}=a_i} p_k = p_1^{u_1} \cdots p_k^{u_k}.$$

For any $\vec{p} \in \Delta^\circ_{k-1}$, $L_X(\vec{p})$ coincides with the rational function $f_{\vec{u}} : \mathbb{P}^{k-1}(\mathbb{C}) \dashrightarrow \mathbb{R}$, given by

$$f_{\vec{u}}(p_1 : \ldots : p_k) = \frac{p_1^{u_1} \cdots p_k^{u_k}}{(p_1 + \ldots + p_k)^n}.$$

It follows that the MLE for a given data vector $\vec{u}$ is a critical point of $f_{\vec{u}}$ inside $\mathbf{V}(I) \cap \Delta^\circ_{k-1}$.

Definition 7. Given a homogenous prime ideal $I \subseteq \mathbb{R}[x_1, \ldots, x_k]$ satisfying (1), define

$$\mathcal{U}(I) = \mathbf{V}_{\text{reg}}(I) \setminus \mathbf{V}(p_1 \cdots p_k \cdot (p_1 + \ldots + p_k)).$$

The **likelihood locus** $Z_{\vec{u}}(I)$ for a data vector $\vec{u}$ is the set of all vectors $\vec{p} \in \mathcal{U}(I)$ such that the gradient $f'_{\vec{u}}(\vec{p})$ lies in the tangent space of $\mathbf{V}(I)$ at $\vec{p}$.

Definition 8. If $I = \langle g_1, \ldots, g_s \rangle$, define the **augmented Jacobian matrix** $J(p)$ of I by

$$J(p) = \begin{pmatrix} p_1 & p_2 & \cdots & p_k \\ p_1 \frac{\partial g_1}{\partial p_1} & p_2 \frac{\partial g_1}{\partial p_2} & \cdots & p_k \frac{\partial g_1}{\partial p_k} \\ p_1 \frac{\partial g_2}{\partial p_1} & p_2 \frac{\partial g_2}{\partial p_2} & \cdots & p_k \frac{\partial g_2}{\partial p_k} \\ \vdots & \vdots & \ddots & \vdots \\ p_1 \frac{\partial g_s}{\partial p_1} & p_2 \frac{\partial g_s}{\partial p_2} & \cdots & p_k \frac{\partial g_s}{\partial p_k} \end{pmatrix}$$

Proposition 1. *A vector $\vec{p} \in \mathcal{U}(I)$ is in the likelihood locus $Z_{\vec{u}}$ if and only if the data vector $\vec{u}$ lies in the row span of the augmented Jacobian matrix $J(p)$.*

Recall that the kernel of a matrix is the orthogonal complement of its row span. Then Z_u lies in the variety of the ideal$\langle \vec{u}, \vec{\varphi}(p) : \vec{\varphi}(p) \in \ker J(p) \rangle$.

The example file `likelihood-ideal.m2` shows an application of 1: we calculate a basis for the ideal $I_{\vec{u}}$ of the likelihood locus $Z_{\vec{u}}$ and use numeric methods to find the variety of $I_{\vec{u}}$. The we may check each of the resulting points using the Hessian test for critical points.