Theoretical Statistics, STOR 655

Consistency of Maximum Likelihood Estimator

Andrew Nobel

February 2023

The Maximum Likelihood Estimator

Maximum Likelihood Estimation: Setting

- Statistical model: Family $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ of densities on measurable space $(\mathcal{X}, \mathcal{A})$ with reference measure ν

- Assume parameter set $\Theta \subseteq \mathbb{R}^p$

- Let $P_\theta =$ probability measure on $(\mathcal{X}, \mathcal{A})$ with density $f(x|\theta)$, so

$$P_\theta(A) \; = \; \int_A f(x|\theta) \, d\nu(x)$$

- Observe $X_1, X_2, \ldots \in \mathcal{X}$ iid where $X \sim P_{\theta_0}$ with $\theta_0 \in \Theta$ unknown

- Goal: Point estimate of parameter θ_0 based on $X_1, \ldots, X_n$

Maximum Likelihood Estimator

Definition: Given $x_1, \dots, x_n \in \mathcal{X}$

- Likelihood function $L_n(\theta) = L_n(\theta | x_1^n) = \prod_{i=1}^n f(x_i|\theta)$
- Log likelihood function $\ell_n(\theta) = \ell_n(\theta | x_1^n) = \sum_{i=1}^n \log f(x_i|\theta)$
- Maximum likelihood estimator (MLE) is any function $\hat{\theta}_n : \mathcal{X}^n \to \Theta$ s.t.

$$\hat{\theta}_n(x_1^n) \; = \; \operatorname*{argmax}_{\theta \in \Theta} L_n(\theta) \; = \; \operatorname*{argmax}_{\theta \in \Theta} \ell_n(\theta)$$

Consistency: Does $\hat{\theta}_n(X_1^n) \to \theta_0$ wp1 as $n \to \infty$?

Uniform LLN for an Indexed Family of Functions

Semicontinuous Functions

Definition: Suppose $\mathcal{X} \subseteq \mathbb{R}^p$. A function $f : \mathcal{X} \to \mathbb{R}$ is upper semicontinous (usc) at a point $x_0 \in \mathcal{X}$ if

$$\limsup_{x \to x_0} f(x) = \lim_{\delta \to 0} \left\{ \sup_{x \in B(x_0, \delta)} f(x) \right\} = f(x_0)$$

- Equivalently, for every $\epsilon > 0$ there exist $\delta > 0$ such that $||x - x_0|| < \delta$ implies $f(x) \leq f(x_0) + \epsilon$
- Equivalently, $\limsup_n f(x_n) \leq f(x_0)$ for every sequence $x_n \to x_0$

The function f is usc if it is usc at each point $x_0 \in \mathcal{X}$

Definition: A function $f : \mathcal{X} \to \mathbb{R}$ is *lower semicontinous* (lsc) if $-f$ is usc

Semicontinuous Functions, cont.

Fact: Let $f : \mathcal{X} \to \mathbb{R}$ where $\mathcal{X} \subseteq \mathbb{R}^p$

1. f is usc iff upper level sets $\{x : f(x) \geq \alpha\}$ are closed for each $\alpha \in \mathbb{R}$
2. If f is usc then it achieves its maximum value on any compact set
3. If $f_1, f_2, \ldots$ are usc, then so is $f = \inf_n f_n$
4. f is continuous iff it is usc and lsc

Examples of upper semi-continuous functions

- $f(x) = \mathbb{I}(x \in C)$ where C is closed
- $f(x) = \sin(1/x)\mathbb{I}(x \neq 0) + \mathbb{I}(x = 0)$
- f continuous

Indexed Family of Functions

Definition: Let $\Gamma \subseteq \mathbb{R}^p$ be an index set, and let $U : \mathbb{R}^d \times \Gamma \to \mathbb{R}$ be s.t.

1. $U(x, \gamma)$ is measurable in x for every $\gamma \in \Gamma$
2. $U(x, \gamma)$ is usc in γ for every $x \in \mathbb{R}^d$

Let $\mathcal{U} = \{U(\cdot, \gamma) : \gamma \in \Gamma\}$, family of functions on $\mathbb{R}^d$ indexed by $\gamma \in \Gamma$

Definition: A function $K : \mathbb{R}^d \to [0, \infty)$ is an envelope for $\mathcal{U}$ if for every x

$$\sup_{\gamma \in \Gamma} |U(x, \gamma)| \ \leq \ K(x)$$

Proposition: If (1) and (2) hold and $\mathcal{U}$ has envelope K such that $\mathbb{E}K(X) < \infty$ then $\mu(\gamma) := \mathbb{E}[U(X, \gamma)]$ is usc on Γ

One-Sided Uniform Law of Large Numbers (LLN)

Thm: Let $X_1, X_2, \ldots, X \in \mathbb{R}^d$ be iid. Suppose that (1) and (2) hold and

(3) Γ is compact

(4) $\mathcal{U}$ has envelope K such that $\mathbb{E}K(X) < \infty$

(5) $\forall \gamma \in \Gamma\ \exists \delta > 0$ such that $\sup_{\gamma' \in B(\gamma,\delta)} U(x, \gamma')$ is measurable in x

Then we have a one-sided LLN

$$\limsup_{n\to\infty} \sup_{\gamma\in\Gamma} \frac{1}{n} \sum_{i=1}^{n} U(X_i, \gamma) \ \leq \ \sup_{\gamma\in\Gamma} \mathbb{E}U(X, \gamma)$$

Cor: If $U(x, \gamma)$ is continuous in γ for every x, then

$$\sup_{\gamma\in\Gamma} \left| \frac{1}{n} \sum_{i=1}^{n} U(X_i, \gamma) - \mathbb{E}U(X, \gamma) \right| \ \to\ 0 \ \text{ wp1 as } n \to \infty$$

Examples

1. Sinusoids. Index set $\Gamma \subseteq \mathbb{R}^d$ compact, function $U : \mathbb{R}^d \times \Theta \to \mathbb{R}$ defined by

$$U(x,\gamma) \;=\; \prod_{i=1}^{d} \sin(2\pi x_i \gamma_i)$$

2. Neural Network. Index set $\Gamma \subseteq \mathbb{R}^{d+1}$ compact, function $U : \mathbb{R}^d \times \Gamma \to \mathbb{R}$ defined by

$$U(x,\gamma) \;=\; \tanh\left(\sum_{i=1}^{d} x_i \gamma_i + \gamma_0\right)$$

Consistency of MLE

Consistency of MLE: Basic Assumptions

Given: Family $\mathcal{P} = \{f(x|\theta) : \theta \in \Theta\}$ of densities on $(\mathcal{X}, \mathcal{A})$ with reference measure ν. Let $P_\theta =$ distribution with density $f(x|\theta)$. Assume

(1) $\Theta \subseteq \mathbb{R}^p$ is compact

(2) $f(x|\theta)$ is usc in θ for every $x \in \mathcal{X}$

(3) There is a measurable envelope $K : \mathcal{X} \to [0, \infty)$ such that $|\log(f(x|\theta)/f(x|\theta_0))| \leq K(x)$ for every x and $\mathbb{E}_{\theta_0} K(X) < \infty$

(4) Identifiability: If $\theta \neq \theta_0$ then $P_\theta \neq P_{\theta_0}$

(5) $\forall \theta \in \Theta \; \exists \delta > 0$ such that $\sup_{\theta' \in B(\theta,\delta)} f(x|\theta')$ is measurable in x

Consistency of MLE

Theorem: Let $X_1, X_2, \ldots \in \mathcal{X}$ be iid with $X_i \sim f(x|\theta_0) \in \mathcal{P}$. If assumptions (1)-(5) hold then

1. MLE $\hat{\theta}_n = \text{argmax}_{\theta \in \Theta} \ell_n(\theta)$ exists
2. For any sequence $\hat{\theta}_n$ of MLE we have $\hat{\theta}_n(X_1^n) \to \theta_0$ wp1

Remarks

- MLE need not be measurable, but can be chosen to be so
- Envelope condition (3) cannot be dropped (see text)
- If $f(x|\theta)$ is continuous in θ for every x, then (2) and (5) are satisfied