M and Z Estimators

M2RI - Toulouse University

Clément Lalanne

M Estimators

Setup $\Theta \subset \mathbb{R}^p$, $\{\mathcal{L}_\theta ; \theta \in \Theta\}$ is a statistical model.

$(M_n)_{n\in\mathbb{N}}$ a sequence of random functions from Θ to $\mathbb{R}^d$: $\forall n, \forall \theta, M_n(\theta)$ is a random vector in $\mathbb{R}^d$ and $\{M_n(\theta) ; \theta \in \Theta\}$ are defined on the same space.

Definition A M-Estimator is a sequence of random $(\hat{\theta}_n)$ taking values in Θ such that

$$\forall n, \text{ almost surely, } \hat{\theta}_n \in \underset{\theta \in \Theta}{\operatorname{argmax}}\, M_n(\theta)$$

M = "maximizer".

Exercice : Given random vectors $X_1, \dots, X_n \in \mathbb{R}^d$, express the empirical mean $\bar{X}_n = \frac{1}{n} \sum_i X_i$ as a M-estimator.

Maximum Likelihood

<u>Assumption:</u> $\forall \theta$, $\mathcal{L}_\theta$ has a density f_θ w.r.t. a reference measure μ.

$$L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) \quad \text{(Likelihood)}.$$

$$\ell_n(\theta) = \log(L_n(\theta)) = \sum_{i=1}^{n} \log(f_\theta(X_i)) \quad \text{(Log-Likelihood)}.$$

<u>Log-Likelihood estimator:</u>

$$\hat{\theta}_n \in \underset{\theta \in \Theta}{\arg\max}\, M_n(\theta) \quad \text{with } M_n(\theta) = \ell_n(\theta).$$

<u>Idea:</u> Choose the model that maximizes the "probability" of observing the samples.

Excercice : What is the maximum likelihood estimator in the Gaussian model ?

Consistency of M Estimators

Theorem: Consider a sequence $(\hat{M}_n)$ of random functions from $\Theta \subset \mathbb{R}^d$ to $\mathbb{R}$. Consider a deterministic function $M : \Theta \to \mathbb{R}$.

If $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0$

and $\exists \theta_0 \in \Theta$ s.t. $\forall \varepsilon > 0$, $\sup_{\substack{\theta \in \Theta \\ \|\theta - \theta_0\| \geq \varepsilon}} M(\theta) < M(\theta_0)$,

and $(\hat{\theta}_n)$ is a sequence s.t. $M_n(\hat{\theta}_n) \geq \left(\sup_{\theta \in \Theta} M_n(\theta)\right) + o_{\mathbb{P}}(1)$

Then $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

Z Estimators

<u>Setup</u> $\Theta \subset \mathbb{R}^p$, $\{\mathcal{L}_\theta ; \theta \in \Theta\}$ is a statistical model.

$(Z_n)_{n\in\mathbb{N}}$ a sequence of random functions from Θ to $\mathbb{R}^d$: $\forall n$, $\forall \theta$, $M_n(\theta)$ is a random vector in $\mathbb{R}^d$ and $\{M_n(\theta) ; \theta \in \Theta\}$ are defined on the same space.

Definition A M-Estimator is a sequence of random $(\hat{\theta}_n)$ takin values in Θ such that

$$\forall n, \text{ a.s.}, \quad Z_n(\hat{\theta}_n) = 0 .$$

<u>Remark</u> Often, M estimators are defined by $\nabla M_n(\hat{\theta}_n) = 0$, in which case they are also Z-estimators.

<u>Remark</u> The method of moments is a Z-estimator.

Consistency of Z Estimators

<u>Theorem</u> (Z_n) sequence of random functions from $\Theta \subset \mathbb{R}^p$ to $\mathbb{R}^q$ and $Z : \Theta \to \mathbb{R}^q$ a deterministic function.

If $\sup_{\theta \in \Theta} \| Z_n(\theta) - Z(\theta) \| \xrightarrow{\mathbb{P}} 0$

and $\inf_{\substack{\theta \in \Theta \\ \|\theta - \theta_0\| \geq \varepsilon}} \| Z(\theta) \| > 0 = \| Z(\theta_0) \|$

and $(\hat{\theta}_n)$ is a sequence such that $Z_n(\hat{\theta}_n) = o_{\mathbb{P}}(1)$

Then $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$.

Easy Case : Monotone in 1D

Theorem If $\Theta = \mathbb{R}$, (Z_n) is a sequence of random functions from Θ to $\mathbb{R}$
and $Z : \Theta \to \mathbb{R}$ is a deterministic function such that

- $\forall \theta$ fixed $\quad Z_n(\theta) \xrightarrow{\mathbb{P}} Z(\theta)$ (not uniform)
- $\forall n$, Z_n is non-decreasing a.s.
- $\exists \theta_0$ s.t. $\forall \varepsilon > 0$, $\quad Z(\theta_0 - \varepsilon) < 0 < Z(\theta_0 + \varepsilon)$.

Then, if $(\hat{\theta}_n)$ is such that $Z_n(\hat{\theta}_n) = o_{\mathbb{P}}(1)$,

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta.$$

Exercice $X_1, \dots \overset{iid}{\sim} X$. By considering the empirical median $\hat{\theta}_n$ defined as

$$\sum_{i=1}^{n} \operatorname{sign}(\hat{\theta}_n - X_i) = 0,$$

show its consistency. (X has a continuous density w.r.t. Lebesgue bounded >0 by below).