STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Feb 8

Lecturer: Xianyang Zhang

1 Maximum likelihood estimation

1.1 Basic setup

We have a family $\{P_\theta\}_{\theta\in\Theta}$ of distributions on χ, where $\Theta \subseteq \mathbb{R}^d$.

Assumption: Suppose P_θ has a density w.r.t a base measure μ on χ, that is $p_\theta = \frac{\partial P_\theta}{\partial \mu}$.

Definiation: The log likelihood $l_\theta(x) = \log p_\theta(x)$ with

$$\nabla l_\theta(x) = \left[\frac{\partial}{\partial\theta_1} l_\theta(x), \dots, \frac{\partial}{\partial\theta_d} l_\theta(x)\right]^\top,$$
$$\nabla^2 l_\theta(x) = \left[\frac{\partial^2 l_\theta(x)}{\partial\theta_i\partial\theta_j}\right]_{i,j=1}^d.$$

Observe that $\{X_1,\dots,X_n\} \stackrel{i.i.d}{\sim} P_{\theta_0}$ with $\theta_0 \in \Theta$. We aim to estimate θ_0 based on $\{X_1,\dots,X_n\}$. A standard estimator for θ_0 is the maximum likelihood estimator (MLE) given by

$$\hat{\theta}_n \in \text{argmax}_{\theta\in\Theta} P_n l_\theta \quad \text{where} \quad P_n l_\theta = \frac{1}{n}\sum_{i=1}^n l_\theta(X_i).$$

1.2 Main questions

1. Consistency: Whether the MLE converges to the true parameter, that is $\hat{\theta}_n \xrightarrow{p} \theta_0$? It comprises of two components!

- Identifiability of the paremeter;
- Convergence of $\hat{\theta}_n$.

2. Does there exist a r_n such that

- $r_n(\hat{\theta}_n - \theta_0) = O_p(1)$,
- $r_n(\hat{\theta}_n - \theta_0) \xrightarrow{d}$? (some distribution)

3. Optimality : Is the MLE better than the MOME?

1.3 Identifiability Condition

A family of models $\{P_\theta\}_{\theta\in\Theta}$ is identifiable if $P_{\theta_1} \neq P_{\theta_2}$ for all $\theta_1 \neq \theta_2$ and $\theta_1, \theta_2 \in \Theta$. If $P_{\theta_1} \neq P_{\theta_2}$, then we have

- There exists $A \subseteq \chi$ such that $P_{\theta_1}(A) \neq P_{\theta_2}(A)$,
- $D_{KL}(P_{\theta_1}||P_{\theta_2}) > 0$.

1.4 Proposition

Suppose $\{P_\theta\}_{\theta\in\Theta}$ is identifiable and cardinlity$(\Theta) < \infty$. Then, if $\hat{\theta}_n \in \text{argmax}_{\theta\in\Theta} P_n l_\theta$, we can say that,

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

1.4.1 Proof

Since $\{X_1, \ldots, X_n\} \overset{i.i.d}{\sim} P_{\theta_0}$, by the strong law of large numbers, we have

$$P_n l_\theta \xrightarrow{a.s} P_{\theta_0} l_\theta, \quad \forall \theta \in \Theta.$$

Note that

$$P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_\theta = E_{\theta_0}\left[\log \frac{p_{\theta_0}(X)}{p_\theta(X)}\right] = D_{KL}(P_{\theta_0}||P_\theta), \quad X \sim P_{\theta_0}$$

which implies that $P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_\theta > 0$ if $\theta \neq \theta_0$. As $P_n l_\theta \xrightarrow{a.s} P_{\theta_0} l_\theta$ and cardinlity$(\Theta) < \infty$, there exists A such that $P(A) = 1$ and for any $\omega \in A$

$$P_n l_\theta(\omega) \to P_{\theta_0} l_\theta \text{ uniformly over } \Theta.$$

Note that this is possible as cardinlity$(\Theta) < \infty$ (a more general result requires empirical process theory). There exists $N(\omega)$ such that when $n \geq N(\omega)$

$$\begin{aligned} P_n l_{\theta_0}(\omega) - P_n l_\theta(\omega) &= [P_n l_{\theta_0}(\omega) - P_{\theta_0} l_{\theta_0}] - [P_n l_\theta(\omega) - P_{\theta_0} l_\theta] \\ &+ [P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_\theta] > 0, \end{aligned}$$

for $\theta \neq \theta_0$. Then, from the definition of $\hat{\theta}_n(\omega)$, we have $\hat{\theta}_n(\omega) \to \theta_0$, which implies that $\hat{\theta}_n \to^{a.s} \theta_0$.

1.5 Proposition

Assume that

- $\sup_{\theta\in\Theta} |P_n l_\theta - P_{\theta_0} l_\theta| \xrightarrow{p} 0$,
- $P_{\theta_0} l_{\theta_0} > \sup_{\theta:||\theta-\theta_0||>\epsilon} P_{\theta_0} l_\theta$ for any $\epsilon > 0$.

Then we have $\hat{\theta} \xrightarrow{p} \theta_0$.

1.5.1 Proof

For every $\epsilon > 0$, there exists $\eta > 0$ such that

$$P_{\theta_0} l_\theta < P_{\theta_0} l_{\theta_0} - \eta$$

whenever $\|\theta_0 - \theta\| > \epsilon$. Notice that

$$P(||\hat{\theta}_n - \theta_0|| > \epsilon) \leq P(P_{\theta_0} l_{\hat{\theta}_n} < P_{\theta_0} l_{\theta_0} - \eta) = P(\eta < P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_{\hat{\theta}_n}).$$

To complete the proof, we only need to show $P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_{\hat{\theta}_n} \leq o_p(1)$. To this end, we note that

$$\begin{aligned} P_n l_{\hat{\theta}_n} &\geq P_n l_{\theta_0}, \\ P_n l_{\theta_0} &\xrightarrow{p} P_{\theta_0} l_{\theta_0}, \end{aligned}$$

where the second result follows from Condition 1. Thus

$$\begin{aligned} P_n l_{\hat{\theta}_n} &\geq P_n l_{\theta_0} - P_{\theta_0} l_{\theta_0} + P_{\theta_0} l_{\theta_0} \\ &\geq P_{\theta_0} l_{\theta_0} - |P_n l_{\theta_0} - P_{\theta_0} l_{\theta_0}| \\ &= P_{\theta_0} l_{\theta_0} - o_p(1). \end{aligned}$$

We then have

$$\begin{aligned} P_{\theta_0} l_{\theta_0} - P_{\theta_0} l_{\hat{\theta}_n} &\leq \quad P_n l_{\hat{\theta}_n} - P_{\theta_0} l_{\hat{\theta}_n} + o_p(1) \\ &\leq \sup_{\theta \in \Theta} |P_n l_\theta - P_{\theta_0} l_\theta| + o_p(1) \\ &= o_p(1). \end{aligned}$$