Chapter 1

M-Estimators

1.1 Motivation

Let the loss function be $\rho_i(\theta) = \rho(z_i, \theta)$, where z_i is a data vector. The sample criterion is an average of $\rho_i(\theta)$:

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho_i(\theta)$$

The m-estimator minimizes the sample criterion function:

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} S_n(\theta).$$

The m-estimator includes many examples as special cases. For example, OLS, MLE, NLS, and quantile regressions are all m-estimators.

For simplicity, in this lecture we work with iid data. Let

$$S(\theta) = E[S_n(\theta)] = E[\rho_i(\theta)]$$

be the population criterion function.

Definition 1.1. We say θ is identified if $\theta_0 = \arg\min_{\theta\in\Theta} S(\theta)$ is unique. In other words, for any $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon)$ such that $\inf_{\theta\in\Theta\setminus N_\delta(\theta_0)} S(\theta) - S(\theta_0) > \varepsilon$.

1.2 Consistency

Theorem 1.1. *If (i) ULLN:* $\sup_{\theta\in\Theta}|S_n(\theta) - S(\theta)| \xrightarrow{p} 0$*; (ii)* θ_0 *is identified, then* $\hat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$.

Proof. We start from the condition of identification.

$$\begin{aligned}
\Pr\left(|\hat{\theta} - \theta| > \delta\right) &\le \Pr\left(S(\hat{\theta}) - S(\theta_0) > \varepsilon\right) \\
&= \Pr\left(S(\hat{\theta}) - S_n(\hat{\theta}) + S_n(\hat{\theta}) - S_n(\theta_0) + S_n(\theta_0) - S(\theta_0) > \varepsilon\right) \\
&\le \Pr\left(S(\hat{\theta}) - S_n(\hat{\theta}) + S_n(\theta_0) - S(\theta_0) > \varepsilon\right) \\
&\le \Pr\left(\left|S_n(\hat{\theta}) - S(\hat{\theta})\right| + |S(\theta_0) - S_n(\theta_0)| > \varepsilon\right) \\
&\le \Pr\left(\sup_{\theta\in\Theta}|S_n(\theta) - S(\theta)| \ge \frac{\varepsilon}{2}\right) \to 0
\end{aligned}$$

where the second inequality follows from the definition of the m-estimator that $S_n(\hat{\theta}) \leq S_n(\theta_0)$ $\square$

1.3 Asymptotic Normality

We go with a heuristic argument. Define $\bar{\psi}(\theta) = \frac{\partial}{\partial\theta} S_n(\theta)$. Taylor expansion of $\bar{\psi}(\hat{\theta})$ around θ_0 gives

$$0 = \bar{\psi}(\hat{\theta}) = \bar{\psi}(\theta_0) + \frac{\partial^2}{\partial\theta\partial\theta'} S_n\left(\dot{\theta}\right)\left(\hat{\theta} - \theta_0\right)$$

where $\dot{\theta}$ lies in between $\widehat{\theta}$ and θ_0. Rearrange the above inequality,

$$\sqrt{n}\left(\hat{\theta} - \theta\right) = -\left[\frac{\partial^2}{\partial\theta\partial\theta'} S_n\left(\dot{\theta}\right)\right]^{-1} \sqrt{n}\bar{\psi}(\theta_0)$$

Since $\hat{\theta} \xrightarrow{p} \theta_0$, we also have $\dot{\theta} \xrightarrow{p} \theta_0$. By the continuous mapping theorem:

$$\frac{\partial^2}{\partial\theta\partial\theta'} S\left(\dot{\theta}\right) \xrightarrow{p} \frac{\partial^2}{\partial\theta\partial\theta'} S(\theta_0) = Q$$

if $\frac{\partial^2}{\partial\theta\partial\theta'} S(\cdot)$ is continuous. In the population, $E\left[\bar{\psi}(\theta_0)\right] = E\left[\psi(\theta_0)\right] = 0$, and

$$\sqrt{n}\bar{\psi}(\theta_0) \xrightarrow{d} N(0, \Omega)$$

where $\Omega = E\left[\psi_i(\theta_0)\psi_i'(\theta_0)\right]$. As a result,

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, Q^{-1}\Omega Q^{-1}\right)$$

where the asymptotic variance follows a sandwich form.

`Zhentao Shi. Feb 7, 2023. Transcribed by Shu Shen.`