$(\mathcal{X}, \Sigma, P)$, $X_1, \cdots, X_n \overset{iid}{\sim} P$, $X_i \in \mathcal{X}$.

$\{m_\theta ; \theta \in \Theta\}$ is a class of measurable function: $\mathcal{X} \mapsto \mathbb{R}$

Define $M(\theta) = P m_\theta$, $M_n(\theta) = P_n m_\theta$, if

$\theta_0 = \arg\max_{\theta \in \Theta} M(\theta)$

we called $\hat{\theta}_n = \arg\max_{\theta \in \Theta} M_n(\theta)$ is a M-estimator of θ_0.

Example MLE

Suppose that $X_1, \cdots, X_n \sim p_{\theta_0}(x)$, than

$\theta_0 = \arg\min_{\theta \in \Theta} \underbrace{KL(p_{\theta_0} \| p_\theta)}_{} \propto -\int \log p_\theta(x) \, dP_{\theta_0}(x)$

$\Rightarrow$ $\theta_0 = \arg\max_{\theta \in \Theta} P \log p_\theta$, then the MLE

$\hat{\theta}_n = \arg\max_{\theta \in \Theta} P_n \log p_\theta = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} p_\theta(X_i)$

is M-estimator.

Theorem (Consistency) $\mathcal{F} = \{m_\theta, \theta \in \Theta\}$, if

① $\|P_n - P\|_{\mathcal{F}} \to_p 0$ ($\mathcal{F}$ is P-GC class)

② $\sup_{\{\theta: d(\theta, \theta_0) \geq \varepsilon\}} M(\theta) < M(\theta_0)$ (Well seperation)

Then $\forall \hat{\theta}_n$ s.t. $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$ has

$$\hat{\theta}_n \to_p \theta_0$$

Pf: $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t.

$\{d(\theta_0, \hat{\theta}_n) \geq \varepsilon\} \subset \underbrace{\{M(\hat{\theta}_n) < M(\theta_0) - \delta\}}$

$\subset \underline{\{M(\theta_0) - M_n(\theta_0) > \frac{\delta}{3}\}} \cup \underline{\{M_n(\theta_0) - M_n(\hat{\theta}_n) > \frac{\delta}{3}\}}$

$\cup \underline{\{M_n(\hat{\theta}_n) - M(\hat{\theta}_n) > \frac{\delta}{3}\}}$

"—", "—", "—" all are $o_p(1)$. $\square$

Remark: △1 If M is continuous and has unique Maximum then ② holds.

Pf: Since $\{\theta;\ d(\theta,\theta_0)\geq\varepsilon\}$ is closed, $\exists\ \theta_\varepsilon$ s.t.

$$M(\theta_\varepsilon) = \sup_{\{\theta;\ d(\theta,\theta_0)\geq\varepsilon\}} M(\theta) < M(\theta_0) \text{ (Uniqueness)}$$

$\square$

△2 If (Θ,ρ) is compact and $\theta\mapsto m_\theta(x)$ is continuous $\forall x\in\mathcal{X}$. More, the envolope function of $\mathcal{F}$: F s.t. $PF<\infty$, then ① holds.

Pf: Claim that $(\Theta,\rho)\mapsto(\mathcal{F}, d^{(1)})$ is continuus since $\forall\varepsilon>0$,

$$\mathbb{1}\{|m_{\theta_1}(X)-m_{\theta_2}(X)|>\varepsilon\} \to 0,\ \text{if } \theta_1\to\theta_2,\ \forall\omega\in\Omega$$

$$\Rightarrow\quad \mathbb{P}(|m_{\theta_1}(X)-m_{\theta_2}(X)|>\varepsilon)\to 0$$

$\Rightarrow\quad P|m_{\theta_1}-m_{\theta_2}|\to 0$. Claim holds.

$\Rightarrow\ (\Theta,\rho)$ is compact $\Rightarrow (\mathcal{F}, d^{(1)})$ is compact

$\square$

△3 If Θ is convex subset of $\mathbb{R}^d$ and $\theta\mapsto m_\theta(x)$ is continuous and concave for $\forall x\in\mathcal{X}$. And $PF<\infty$. $\hat\theta_n \to_p \theta_0$.

Pf: $\forall M>0$, $\underbrace{\{\theta;\ d(\theta,\theta_0)\leq M\}}_{\Theta_M}$ is compact, then

$$(\underbrace{\{m_\theta;\ \theta\in\Theta_M\}}_{P\text{-}GC}, d^{(1)}) \text{ is compact}$$

① holds.

More, we assume ② holds.

For $\forall \hat{\theta}_n \in \Theta$, exist $\alpha_n \in [0,1]$

s.t. $\tilde{\theta}_n = \alpha_n \hat{\theta}_n + (1-\alpha_n)\theta_0 \in \Theta_M$

$$M_n(\tilde{\theta}_n) \geq \alpha_n M_n(\hat{\theta}_n) + (1-\alpha_n) M_n(\theta_0)$$
$$\geq M_n(\theta_0) - o_p(1)$$

$\Rightarrow$ $\tilde{\theta}_n \to_p \theta_0$, and

$$\hat{\theta}_n - \theta_0 = \frac{\tilde{\theta}_n - (1-\alpha_n)\theta_0}{\alpha_n} - \theta_0$$
$$= \frac{\tilde{\theta}_n - \theta_0}{\alpha_n} = o_p(1) \qquad \square$$

Wald's Consistency Proof

Let $\theta \mapsto m_\theta(x)$ is upper-semicontinuous $\forall x \in \mathcal{X}$.

Lemma $\{U_n\}$ is a sequence of open ball s.t. $\theta \in U_n$, $U_n \downarrow \theta$. Let

$$m_U(x) := \sup_{\theta \in U} m_\theta(x)$$

And $P|m_{U_1}| < \infty$, then

$$P m_{U_n} \downarrow P m_\theta$$

Pf: $\forall n$, $\exists\, \theta_n \in U_n$ s.t.

$$m_{U_n}(x) \leq m_{\theta_n}(x) + \frac{1}{n}$$

$\Rightarrow$ $\limsup_n m_{U_n}(x) \leq m_\theta(x)$, since $\theta_n \to \theta$

$\Rightarrow$ $\limsup_n P m_{U_n} \leq P \limsup_n m_{U_n}$
$$\leq P m_\theta \qquad \square$$

Theorem $\forall\ \hat{\theta}_n$ s.t. $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_p(1)$ and compact $K \subset \Theta$, then

$$\mathbb{P}(\hat{\theta}_n \in B_\varepsilon) \to 0, \quad B_\varepsilon = \{\theta \in K; d(\theta, \Theta_0) \geq \varepsilon\}$$

where $\Theta_0 = \{\theta_1 \in \Theta; M(\theta_1) = \operatorname{argmax}_{\theta \in \Theta} M(\theta)\}$

Pf: Notice that B_ε is compact,

$\forall\ \theta \in B_\varepsilon$, $M(\theta) < M(\theta_0)$, $\theta_0 \in \Theta_0$

Let $U_n \downarrow \theta$, then $\exists\ U_\theta$ s.t.

$$P m_{U_\theta} < M(\theta_0)$$

Due to the compactness, $\exists$ finite $\{U_{\theta_i}\}$ cover B_ε.

$$\Rightarrow \mathbb{P}(\hat{\theta}_n \in B_\varepsilon) \leq \sum_i \mathbb{P}(\hat{\theta}_n \in U_{\theta_i})$$

If $\hat{\theta}_n \in U_{\theta_i}$,

$$M_n(\hat{\theta}_n) \leq P_n m_{U_{\theta_i}} \xrightarrow{a.s.} M(\theta_i) < M(\theta_0)$$

$$\Rightarrow \{\hat{\theta}_n \in U_{\theta_i}\} \subset \{M_n(\hat{\theta}_n) < M(\theta_0) - \delta + o_p(1)\}$$

$(\exists\ \delta > 0)$

And $\mathbb{P}(\underline{M_n(\hat{\theta}_n) < M(\theta_0) - \delta + o_p(1)})$

$$= \mathbb{P}(\text{—} \cap \underline{\{M_n(\hat{\theta}_n) \geq M(\theta_0) - o_p(1)\}})$$

$$= \mathbb{P}(\{o_p(1) > \delta\} \cap \text{—})$$

$$= \mathbb{P}(o_p(1) > \delta) \to 0 \qquad \square$$

Asymptotic Normal

We assume that Θ is convex subset of $\mathbb{R}^d$, $\theta \mapsto m_\theta(x)$ is continuous and convex $\forall x \in \mathcal{X}$.

The envelope function of $\{m_\theta\}$: F s.t. $(PF^2)^{\frac{1}{2}} < \infty$.

The condition above will ensure that $\hat{\theta}_n \to \theta_0$

More, we assume that Θ is open or $\theta_0 \in ri(\Theta)$, $\nabla M(\theta_0) \triangleq \Psi(\theta_0)$ exists, then $\Psi(\theta_0) = 0$.

If $\nabla \Psi(\theta_0)$ exist, $\nabla \Psi(\theta_0) \succeq 0$.
We assume that $\nabla \Psi(\theta_0) \succ 0$.

Def Z-estimator

Assume that $\nabla_\theta m_\theta(x)$ exists, $\forall x \in \mathcal{X}$, let $\psi_\theta(x) = \nabla_\theta m_\theta(x)$ and $\Psi_n(\theta) = P_n \psi_\theta$

If $\hat{\theta}_n$ s.t. $\Psi_n(\hat{\theta}_n) = 0$, $\hat{\theta}_n$ is Z-estimator of θ_0.

Now we consider the conditions on $\psi_\theta(x)$ s.t.

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d Z$$

where Z is a normal vector.

Theorem (classical condition)

If $\nabla^2_\theta \psi_\theta(x)|_{\theta=\theta_0}$ exists and continuous at U_{θ_0}, for $\forall x \in \mathcal{X}$. Moreover, $\exists\, L$ s.t. $\forall \theta \in U_{\theta_0}$, $\|\nabla^2_\theta \psi_\theta(X)\| \le L(X)$, $P|L| < \infty$

then $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d Z$, if $\Psi_n(\hat{\theta}_n) = o_p(\frac{1}{\sqrt{n}})$.

Pf:

$$\Psi_n(\hat{\theta}_n) = \Psi_n(\theta_0) + \Psi_n'(\theta_0)(\hat{\theta}_n - \theta_0) + \frac{1}{2}\left(\Psi_n''(\theta_0)(\hat{\theta}_n - \theta_0)\right)(\hat{\theta}_n - \theta_0)$$

$\hat{\theta}_n \in \Theta$.

Since $\|\Psi_n''(\tilde{\theta}_n)\| \le \frac{1}{n}\sum_{i=1}^{n} L(X_i) \to_{a.s.} PL$

$\Rightarrow \|\Psi_n''(\tilde{\theta}_n)\| = O_p(\frac{1}{\sqrt{n}})$

$\Rightarrow o_p(\frac{1}{\sqrt{n}}) = \Psi_n(\theta_0) + \nabla\Psi_n(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(\frac{1}{\sqrt{n}})$

$\Rightarrow \nabla\Psi_n(\theta_0)(\sqrt{n}(\hat{\theta}_n - \theta_0)) = -\sqrt{n}\Psi_n(\theta_0) + o_p(1)$

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0)$ is tight.

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) = -\sqrt{n}(\nabla\Psi(\theta_0))^{-1}\Psi_n(\theta_0) + o_p(1)$ $\square$

Theorem

If $\exists\, L$ s.t. $PL^2 < \infty$,

$\| \psi_{\theta_1}(x) - \psi_{\theta_2}(x) \| \le L(x) \|\theta_1 - \theta_2\|$

Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is tight and asymptotic normal.

Pf: Let $\mathcal{F} = \{\psi_\theta,\ \theta \in \Theta\}$, then $\mathcal{F}$ is P-D class, $G_n \to_d G_p$. Since

$\int \| \psi_{\hat{\theta}_n}(x) - \psi_{\theta_0}(x) \|^2 dF(x) \le \|\hat{\theta}_n - \theta_0\|^2 PL^2 \to_p 0$

$\Rightarrow G_n \psi_{\hat{\theta}_n} = G_p \psi_{\theta_0} + o_p(1)$

$\Rightarrow -\sqrt{n}(\Phi(\hat{\theta}_n) - \Phi(\theta_0)) = G_p \psi_{\theta_0} + o_p(1)$

$\Rightarrow -\sqrt{n}(\nabla\Phi(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(\|\hat{\theta}_n - \theta_0\|)) = G_p \psi_{\theta_0} + o_p(1)$

$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) = -(\nabla\Phi(\theta_0))^{-1} G_p \psi_{\theta_0} + o_p(\|\hat{\theta}_n - \theta_0\|) + o_p(1)$ $\square$

Remark: $\hat{\theta}_n - \theta_0 = o_p(\frac{1}{\sqrt{n}}) + \Phi_n(\theta_0) + o_p(\|\hat{\theta}_n - \theta\|)$

Note that

$\mathbb{P}(|\Phi_n(\theta_0)| > \varepsilon)$

$= \mathbb{P}(|P_n\psi_{\theta_0} - P\psi_{\theta_0}| > \varepsilon)$

$\le \mathbb{P}(\|P_n - P\|_{\mathcal{F}} > \varepsilon) \le \frac{1}{\varepsilon}\sqrt{\frac{d}{n}}$

$\Rightarrow \Phi_n(\theta_0) = O_p(\sqrt{\frac{d}{n}})$

$\Rightarrow \hat{\theta}_n - \theta_0 = O_p(\sqrt{\frac{d}{n}})$