Let $\xi_1 \cdots \xi_n \sim P$. iid obs.

Consider $\min_{\theta \in S} \mathbb{E} f(\theta, \xi) = \min_{\theta \in \mathbb{R}^d} \mathbb{E}[f_0(\theta, \xi) + I_S(\theta)]$, where $I_S(t) = \begin{cases} 0 & \text{if } t \in S \\ \infty & \text{if } t \notin S \end{cases}$

Denote. $\theta_0 = \operatorname{argmin}_{\theta \in \mathbb{R}^d} \mathbb{E}[f_0(\theta, \xi) + I_S(\theta)]$, $\theta_n \in \operatorname{argmin}_{\theta \in \mathbb{R}^d} n^{-1} \sum_{i=1}^n f_0(\theta, \xi_i) + I_S(\theta)$.

A1.) $\xi \mapsto f_0(\theta, \xi)$ is continuous.

$\theta \mapsto f_0(\theta, \xi)$ is Lipschitz w/ modulus $\beta(\xi)$

where $\beta(\xi)$ is $\mathbb{P}_n$-uniformly integrable for all $n \geq 1$.

A2) $\mathbb{P}_n \xrightarrow{d} P$.

<u>Lemma 4.3 [Clarke (1983)]</u>

$|h(x_1, \xi) - h(x_2, \xi)| \leq \beta(\xi) \|x_1 - x_2\|$ for open subset $U \subseteq \mathbb{R}^d$. $x_1, x_2 \in \mathcal{U}$.

and $\mathbb{E} h(\bar{x}, \xi)$ for some $\bar{x} \in \mathcal{U}$. Then.

$$\partial \mathbb{E} h(x, \xi) \subset \mathbb{E}\{\partial h(x, \xi)\}$$

Equality holds when h is subdifferentially regular a.e.

<u>Lemma 4.2 [Rockafellar (1979)]</u>

$h_1, h_2 \in \text{l.s.c}(\mathbb{R}^d)$ and. h_2 is locally Lipschitz

$\partial(h_1 + h_2)(x) \subset \partial h_1(x) + \partial h_2(x)$.

Lemma 4.5 [Dupačová and Wets (1988)]

under A1 and A2,

$$\partial \mathbb{E} f(\theta, \xi) \subset \partial \mathbb{E} f_0(\theta, \xi) + \partial I_S(\theta)$$

$$\partial\, n^{-1} \Sigma f(\theta, \xi_i) \subset \partial\, n^{-1} \Sigma f_0(\theta, \xi_i) + \partial I_S(\theta)$$

For some. $V_n \in \partial\, n^{-1}\sum (f_0 + 1_S)(\theta_0, \xi_i)$ and $V \in \partial\, \mathbb{E}(f_0 + 1_S)(\theta_n, \xi)$

B1) $\sqrt{n}\,[\, V_n(\theta_0, \xi) + V(\theta_n)\,] = o_p(1)$.

B2) $\sqrt{n}\,[\, V_S(\theta_n) - V_S(\theta_0)\,] = o_p(1)$

B3) $\sqrt{n}\, V_n(\theta_0, \xi) \xrightarrow{d} N(0, \Sigma)$

B4) $\mathbb{E} f_0 \in C^2$ and. Hessian H is invertible.

Thm 4.8

Assume. A1), A2), B1) - B4). $\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma(H^{-1})^T)$

Proof

$\theta \mapsto \mathbb{E} f_0(\theta, \xi) \in C^2$ and $\theta_n \to \theta_0$ then for n large,

$\nabla \mathbb{E} f_0(\theta_n, \xi) - \nabla \mathbb{E} f_0(\theta_0, \xi) = H(\theta_n - \theta_0) + o(\|\theta_n - \theta_0\|)$

By Lemma 4.5,

$V(\theta) \in \partial\, \mathbb{E} f_0(\theta, \xi) \Rightarrow \exists\, V_S(\theta) \in \partial 1_S(\theta)$ s.t $V(\theta) = V_0(\theta) + V_S(\theta)$

In particular $0 = V_0(\theta_0) + V_S(\theta_0)$

Similarly. $V_n(\theta) \in \partial\, n^{-1}\sum f_0(\theta, \xi_i) \Rightarrow V_n(\theta) = V_{0,n}(\theta) + V_S(\theta)$.

In particular $0 = V_{0,n}(\theta_n) + V_S(\theta_n)$

Hence. $\sqrt{n}\{\nabla \mathbb{E} f_0(\theta_n, \xi) - \nabla \mathbb{E} f_0(\theta_0, \xi)\}$

$= \sqrt{n}\{ V(\theta_n) - V_S(\theta_n) - \cancel{V(\theta_0)}^{0} + V_S(\theta_0)\}$

$= \sqrt{n}\{ V_n(\theta_0) + V(\theta_n) - V_n(\theta_0) + V_S(\theta_0) - V_S(\theta_n)\}$

$= -\sqrt{n}\, V_n(\theta_0) + \sqrt{n}\{V_n(\theta_0) + V(\theta_n)\} + \sqrt{n}\{V_S(\theta_0) - V_S(\theta_n)\}$

$\xrightarrow{d} N(0, \Sigma) + o_p(1) + o_p(1)$.

Thus.

$\sqrt{n}(\theta_n - \theta_0) \xrightarrow{d} N(0, H^{-1}\Sigma(H^{-1})^T) + o_p(1)$

Some comment on (B1)-(B3)

B1) $\sqrt{n}\left[V_n(\theta_0, \xi) + V(\theta_n) \right] = \sqrt{n}\left\{ n^{-1}\sum g(\theta_0, \xi_i) + \mathbb{E}[g(\theta_n, \xi)] \right\}$

$$= \sqrt{n}\left\{ n^{-1}\sum g(\theta_0, \xi_i) - \mathbb{E}g(\theta_0, \xi) - \left[n^{-1}\sum g(\theta_n, \xi_i) - \mathbb{E}g(\theta_0, \xi) \right] \right\}$$

$$= \mathbb{G}_n\, g(\theta_0, \cdot) - g(\theta_n \cdot)$$

$$\lesssim \sup_{\|\theta - \theta_0\| \le r_n} \left| \mathbb{G}_n g(\theta_0, \cdot) - g(\theta, \cdot) \right|$$

B3) $-\sqrt{n}\, V_n(\theta_0, \xi) = -\sqrt{n} \cdot n^{-1}\sum g(\theta_0 \xi_i)$

$$= -\sqrt{n} \cdot \left[n^{-1}\sum g(\theta_0, \xi_i) - \mathbb{E}g(\theta_0, \xi_i) \right]$$

$$\xrightarrow{d} N(0, \Sigma) \quad \text{by CLT.}$$

B2). requires constraint quantification.

Dupacova 1987.

minimize. $\mathbb{E} f_0(\theta, \xi)$ sbj to. $Ab \geq c$ $\quad A \in \mathbb{R}^{m \times d}$. $c \in \mathbb{R}^{m \times 1}$ (m. ineq. constraints)

minimize. $\frac{1}{n} \sum f_0(\theta, \xi_i)$ sbj to. $Ab \geq c$

The corresponding Lagrangian is

$$L(\theta, \lambda) = \begin{cases} \mathbb{E} f_0(\theta, \xi) - \lambda^T (A\theta - c) & \text{for } \lambda \geq 0. \\ \infty & \text{o.w.} \end{cases}$$

$$L_n(\theta, \lambda) = \begin{cases} \frac{1}{n} \sum f_0(\theta, \xi_i) - \lambda^T (A\theta - c) & \text{for } \lambda \geq 0. \\ \infty & \text{o.w.} \end{cases}$$

Let (θ_0, λ_0), (θ_n, λ_n) be saddle points of these programs.

Then $0 \in \partial_\theta L(\theta_0, \lambda_0) = \partial_\theta \mathbb{E} f_0(\theta_0) - A^T \lambda_0$.

$0 \in \partial_\lambda L(\theta_0, \lambda_0) = -(A\theta_0 - c)$

and. $0 \in \partial_\theta L_n(\theta_n, \lambda_n) = \partial_\theta \frac{1}{n} \sum f_0(\theta_n, \xi_i) - A^T \lambda_n$

$0 \in \partial_\lambda L_n(\theta_n, \lambda_n) = -(A\theta_n - c)$

Strict complementary conditions

for all $k \in \{1 \dots m\}$ $e_k^T \lambda_0 = 0 \iff \sum_{j=1}^{d} A_{kj} \theta_{0,j} > c_k$.

(In words). $e_k \lambda_0$ must be strictly positive for active constraints.

, i.e., $\sum_{j=1}^{d} A_{kj} \theta_{0,j} = c_k$. (boundary of the constraints)

Let $I \subseteq \{1, \dots, m\}$ be the indecies. of active constraints.

under. Strict. complementary conditions,

$\lambda_{n,i} = 0$ for $i \notin I$ and $\sum_{j=1}^{d} A_{k,j} \theta_{n,j} = c_k$ for $i \in I$ a.s.

Then Dupacova 1987 treats the original problem as the "unconstrained" version.

$$L_I^n(\theta, \lambda_I) = n^{-1}\sum f_0(\theta, \xi_i) - \sum_{k\in I} \lambda_k \Big[\sum_{j=1}^d A_{k,j}\theta_j - C_k\Big].$$

Hence.

B1) - B3) must be proved for

changing $V_n(\theta_0)$ to $V_{0,n}(\theta_0) - A_I^T \lambda_{0,I}$

$V_0(\theta_n)$ to $V_0(\theta_n) - A_I^T \lambda_{n,I}$.

Asymptotic normality further requires A_I to be full row rank.

Related setting is

$$S = \{\theta \mid g_i(\theta) = 0 \ \ i\in I \text{ and } g_j(\theta) \le 0 \ \ j \in J\} \text{ for "smooth" } g.$$

<u>Mangassarian-Fromovitz (1967) condition</u>

Let θ_0 be a unique minimizer $\theta \mapsto \mathbb{E} f_0(\theta, \xi) + 1_S(\theta)$.

Assume.

① $\nabla g_i(\theta_0)$, $i\in I$ are linearly independent.

② $\exists w$. s.t. $w^T \nabla g_i(\theta_0) = 0 \ \ i \in I, \quad w^T\nabla g_j(\theta_0) < 0 \ \ j \in J$

For. $L(\theta,\lambda) = \begin{cases} \mathbb{E} f_0(\theta,\xi) + \sum \lambda_i g(\theta) & \text{for } \lambda \ge 0. \\ \infty & \text{o.w.} \end{cases}$

<u>Second-order suff. cond</u>

For all non-zero w that satisfies ②, and $w^T \nabla \mathbb{E} f_0(\theta_0, \xi) \le 0$.

$$\max_\lambda w^T \nabla_\theta^2 L(\theta_0, \lambda) w > 0.$$

Shapiro (1989) claims.

MF condition + second-order sufficient condition

are not enough for asympt. normality.

He also considers Strict complementary conditions for asympt. normal.

---

example 1.

let $Z_1 \cdots Z_n \sim N(\theta_p, 1)$

Set up optimization under $\theta \geq 0$. s.t

$$L(\theta, \lambda) = \mathbb{E}(Z - \theta)^2 - \lambda\theta$$

$$= \mathbb{E}(Z - \theta_p + \theta_p - \theta)^2 - \lambda\theta$$

$$= 1 + (\theta_p - \theta)^2 - \lambda\theta$$

when $\theta_p = 0$ (boundary). and $\theta_0 = 0$.

$\frac{\partial}{\partial\theta} L(\theta_0, \lambda) = -2(\theta_p - \theta_0) - \lambda = -\lambda$. Hence $\lambda = 0$.

$\Rightarrow$ Strict complementary cond. fails.

Recall $\sqrt{n}(\theta_n - \theta_0) = \sqrt{n}(\max(\bar{Z}, 0) - \theta_0)$ is half-normal.