- First consider the standard case in which the F.O.C is smooth and allow to do a mean value expansion around the true value θ_0.

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta}$$

mean value expansion

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \underbrace{\frac{\partial Q_n(\theta_0)}{\partial \theta}}_{k \times 1} + \underbrace{\frac{\partial^2 Q_n(\theta^*)}{\partial \theta \partial \theta'}}_{k \times k} \cdot \underbrace{(\hat{\theta_n} - \theta_0)}_{k \times 1}$$

multiply $\sqrt{n}$

$$o_p(1) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + \underbrace{\frac{\partial^2 Q_n(\theta_n^*)}{\partial \theta \partial \theta'}}\ \sqrt{n}\,(\hat{\theta_n} - \theta_0)$$

This is an interesting object since we have θ_n^* and Q_n both random. (*) We will need a uniform LLN rather than the usual lemma we use.

$$o_p(1) = \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + \left(B(\theta_0) + o_p(1) \right) \sqrt{n}\,(\hat{\theta_n} - \theta_0)$$

Rewrite

$$\sqrt{n}\,(\hat{\theta_n} - \theta_0) = \left(B(\theta_0) + o_p(1) \right)^{-1} \sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1)$$

$$\sqrt{n}\,(\hat{\theta_n} - \theta_0) = B(\theta_0)^{-1} \underbrace{\sqrt{n}\,\frac{\partial Q_n(\theta_0)}{\partial \theta}}_{\xrightarrow{d} N(0, \Omega_0)} + o_p(1)$$

$$\xrightarrow{d} N(0,\ B(\theta_0)^{-1}\ \Omega_0\ B(\theta_0)^{-1}).$$

To review how we deal with (*) recall the following assumption

$$\sup_{\theta \in \Theta_0} \left\| \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} - B(\theta) \right\| \xrightarrow{p} 0$$

for some non stochastic $B(\theta)$ that is continuous at θ_0 and $B_0 := B(\theta_0)$ is non singular.

First, notice that $\left\| B(\theta_n^*) - \frac{\partial^2 Q_n(\theta_n^*)}{\partial\theta\,\partial\theta'} \right\| = o_p(1)$

<u>proof:</u>

$$\left\| B(\theta_n^*) - \frac{\partial^2 Q_n(\theta_n^*)}{\partial\theta\,\partial\theta'} \right\| \le \sup_{\theta \in \Theta_0} \left\| B(\theta) - \frac{\partial^2 Q_n(\theta)}{\partial\theta\,\partial\theta'} \right\| = o_p(1)$$

Therefore $\lim_{n\to\infty} P\left(\left\| B(\theta_n^*) - \frac{\partial^2 Q_n(\theta_n^*)}{\partial\theta\,\partial\theta'} \right\| > \varepsilon \right) \le \lim_{n\to\infty} P(o_p(1) > \varepsilon) = 0.$ ∎

Secondly, notice that $\| B(\theta_0) - B(\theta_n^*) \| = o_p(1)$ provided $\theta_n^* \xrightarrow{p} \theta_0$.

<u>proof:</u>

$$\begin{aligned}
P\left(\| B(\theta_0) - B(\theta_n^*) \| > \varepsilon \right) &= P\left(\| B(\theta_0) - B(\theta_n^*) \| > \varepsilon, \|\theta_n^* - \theta_0\| \le \delta \right) \\
&\quad + P\left(\| B(\theta_0) - B(\theta_n^*) \| > \varepsilon, \|\theta_n^* - \theta_0\| > \delta \right) \\
&\le \underbrace{P\left(\| B(\theta_0) - B(\theta_n^*) \| > \varepsilon, \|\theta_n^* - \theta_0\| \le \delta \right)}_{=0 \text{ by continuity of } B(\theta) \text{ at } \theta_0.\text{ We can choose such } \delta \text{ that allows this.}} \\
&\quad + \underbrace{P\left(\|\theta_n^* - \theta_0\| > \delta \right)}_{o(1)} \\
&= 0 + o(1)
\end{aligned}$$

Then

$$\lim_{n\to\infty} P\left(\| B(\theta_0) - B(\theta_n^*) \| > \varepsilon \right) = \lim_{n\to\infty} o(1) = 0.$$ ∎

Now, notice that our final result goes as follows

$$\begin{aligned}
\frac{\partial^2 Q_n(\theta_n^*)}{\partial\theta\,\partial\theta'} = \frac{\partial^2 Q_n(\theta_0)}{\partial\theta\,\partial\theta'} + o_p(1) &= B(\theta_0) + o_p(1) + o_p(1) \\
&= B(\theta_0) + o_p(1).
\end{aligned}$$

- A second case arises when the F.O.C is not differentiable. We can still recover asymptotic normality under some conditions. In particular, consider the case of quantile regression:

$$o_p(1/\sqrt{n}) = \frac{\partial Q_n(\hat{\beta}_{\tau,n})}{\partial \beta} = \underbrace{\frac{1}{n}\sum^{n}(\tau - \mathbb{1}\{Y_i < x_i'\hat{\beta}_{\tau n}\})X_i}_{\text{Problem: non differentiable, our solution is to make it smooth using expectation for a fixed value } \hat{\beta}_{\tau n}.}$$

Define

$$m(b) = E\left[\frac{1}{n}\sum^{n}(\tau - \mathbb{1}\{Y_i < x_i'b\})X_i\right]$$

$$\underset{\text{LIE}}{=} E[(\tau - F(X_i'b \mid X_i))X_i] \quad \text{(notice that if evaluated at } \beta_\tau\text{) this is zero, i.e. } m(\beta_\tau) = 0$$

After adding and substracting

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum^{n}\left\{(\tau - \mathbb{1}\{Y_i < x_i'\hat{\beta}_{\tau,n}\})X_i - m(\hat{\beta}_{\tau,n})\right\} + m(\hat{\beta}_{\tau,n})$$

Multiplying $\sqrt{n}$

$$o_p(1) = \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{(\tau - \mathbb{1}\{Y_i < x_i'\hat{\beta}_{\tau,n}\})X_i - m(\hat{\beta}_{\tau,n})\right\}}_{V(\hat{\beta}_{\tau,n}) \text{ and we assume this process is SE}} + \underbrace{\sqrt{n}\, m(\hat{\beta}_{\tau,n})}_{\text{now this is smooth!}}$$

$$o_p(1) = V(\beta_\tau) + o_p(1) + \sqrt{n}\, m(\hat{\beta}_{\tau,n})$$

Mean value expansion

$$o_p(1) = V(\beta_\tau) + o_p(1) + \sqrt{n}\,\underbrace{m(\beta_\tau)}_{=0} + \sqrt{n}\,\underbrace{\frac{\partial m(\beta^*_{\tau n})}{\partial \beta_\tau'}}_{\text{we can use WLLN, the function is non random, only the argument is.}}(\hat{\beta}_{\tau,n} - \beta_\tau)$$

Rewrite

$$\sqrt{n}(\hat{\beta}_{\tau,n} - \beta_\tau) = \frac{\partial m(\beta_\tau)}{\partial \beta_\tau'}^{-1}\underbrace{V(\beta_\tau)}_{\xrightarrow{d} N(0, \Omega_0)} + o_p(1)$$

$$\xrightarrow{d} N\left(0, \frac{\partial m(\beta_\tau)}{\partial \beta_\tau'}^{-1}\Omega_0 \frac{\partial m(\beta_\tau)}{\partial \beta_\tau}^{-1}\right).$$

- A final application mentioned are two step estimators. Consider $\hat{T_n} \xrightarrow{p} T_0$ as the first step estimator. Define $m(\theta, T) = E\, g(w_i, \theta, T)$ and $m(\theta_0, T_0) = 0$ is the moment condition.

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^{n} g(w_i, \hat{\theta}_n, \hat{T_n})$$

mean value expansion around θ_0

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^{n} g(w_i, \theta_0, \hat{T_n}) + \underbrace{\frac{1}{n}\sum \frac{\partial g(w_i, \theta_n^*, \hat{T_n})}{\partial \theta'}}_{\xrightarrow{p} B_0} \cdot (\hat{\theta_n} - \theta_0)$$

multiply $\sqrt{n}$

$$o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(w_i, \theta_0, \hat{T_n}) + (B_0 + o_p(1))\sqrt{n}(\hat{\theta_n} - \theta_0)$$

Rewrite

$$\sqrt{n}(\hat{\theta_n} - \theta_0) = -B_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(w_i, \theta_0, \hat{T_n}) + o_p(1)$$

$$\pm B_0^{-1}\sqrt{n}\, E g(w_i, \theta_0, \hat{T_n})$$

$$\sqrt{n}(\hat{\theta_n} - \theta_0) = -B_0^{-1}\Big(\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[g(w_i, \theta_0, \hat{T_n}) - m(\theta_0, \hat{T_n})]}_{V(\hat{T_n})}\Big) - B_0^{-1}\sqrt{n}\, m(\theta_0, \hat{T_n}) + o_p(1)$$

and assume it's SE

$$\sqrt{n}(\hat{\theta_n} - \theta_0) = -B_0^{-1} V(T_0) - B_0^{-1}\sqrt{n}\, m(\theta_0, \hat{T_n}) + o_p(1)$$

Mean value exp

$$\sqrt{n}(\hat{\theta_n} - \theta_0) = -B_0^{-1}\Big\{\frac{1}{\sqrt{n}}\sum g(w_i, \theta_0, T_0) + \underbrace{\sqrt{n}\, m(\theta_0, T_0)}_{=0 \text{ by moment condition}} + \sqrt{n}\underbrace{\frac{\partial m(\theta_0, T_n^*)}{\partial T}}_{\xrightarrow{p} \Delta_0}(\hat{T_n} - T_0)\Big\}$$

$$\xrightarrow{d} -B_0^{-1} N\left(0, [I_K \vdots \Delta_0]\begin{bmatrix} V_{10} & V_{20} \\ V_{20} & V_{30}\end{bmatrix}[I_K \vdots \Delta_0]'\right)$$

$$= -B_0^{-1} N\left(0, [V_{10} + \Delta_0 V_{20} \quad V_{20} + \Delta_0 V_{30}]\begin{bmatrix} I_K \\ \hdashline \Delta_0' \end{bmatrix}\right)$$

$$= -B_0^{-1} N(0, \underbrace{V_{10} + \Delta_0 V_{20} + V_{20}\Delta_0' + \Delta_0 V_{30}\Delta_0'}_{\Omega_0})$$

$$= N(0, B_0^{-1}\Omega_0 B_0^{-1})$$

provided

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\sum g(w_i, \theta_0, T_0) \\ \sqrt{n}(\hat{T_n} - T_0)\end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_{10} & V_{20} \\ V_{20} & V_{30}\end{bmatrix}\right).$$

What if we didn't take the SE approach?

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^{n} g(W_i, \hat{\theta}_n, \hat{T}_n)$$

mean value expansion around θ_0

$$o_p\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^{n} g(W_i, \theta_0, \hat{T}_n) + \underbrace{\frac{1}{n}\sum \frac{\partial g(W_i, \theta_n^*, \hat{T}_n)}{\partial \theta'}}_{\xrightarrow{p} B_0} \cdot (\hat{\theta}_n - \theta_0)$$

multiply $\sqrt{n}$

$$o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(W_i, \theta_0, \hat{T}_n) + (B_0 + o_p(1))\sqrt{n}(\hat{\theta}_n - \theta_0)$$

mean value expansion around T_0

$$o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(W_i, \theta_0, T_0) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{\partial g(W_i, \theta_0, T_n^*)}{\partial T}(\hat{T}_n - T_0) + (B_0 + o_p(1))\sqrt{n}(\hat{\theta}_n - \theta_0)$$

Rewrite

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -B_0^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(W_i, \theta_0, T_0) + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i, \theta_0, T_n^*)}{\partial T}}_{\xrightarrow{p} \Delta_0}\sqrt{n}(\hat{T}_n - T_0)\right\} + o_p(1)$$

under different conditions!

and we still get the same asymptotic distribution.

where $\Delta_0 = \dfrac{\partial E g(W_i, \theta_0, T_0)}{\partial T}$. To get this result now we require

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i, \theta_0, T_n^*)}{\partial T} = \Delta_0 + o_p(1).$$

This can be achieved by assuming, for example,

(*) I believe Vadim has some Lemmas with sufficient conditions for this:

$$\sup_{T \in \mathcal{T}} \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i, \theta_0, T)}{\partial T} - \underbrace{\frac{\partial E g(W_i, \theta_0, T)}{\partial T}}_{\Delta(T)} \right\| = o_p(1)$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i, \theta_0, T_n^*)}{\partial T} = \Delta(T_n^*) + o_p(1) = \Delta(T_0) + o_p(1)$$

provided $T_n^* \xrightarrow{p} T_0$ and $\Delta(T_0)$ is continuous at T_0.

- Efficient IVs as a special case + Bracketing :

$$E(Y_i - X_i'\beta_0 \mid Z_i) = 0$$

Ⓡ Notice that GLS is the case of $Z_i = X_i$.

$$g^*(Z_i) = \frac{E(X_i \mid Z_i)}{E(u_i^2 \mid Z_i)}$$

and suppose we have a consistent estimator

$$\hat{g}_n(Z_i) \xrightarrow{p} g^*(Z_i)$$

which is our first step estimator.

$$\sqrt{n}(\hat{\beta_n} - \beta_0) = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} \hat{g}_n(Z_i) X_i'\right)^{-1}}_{\xrightarrow{p} (E g^*(Z_i) X_i')^{-1}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_n(Z_i) u_i$$

$$= (E g^*(Z_i) X_i')^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{g}_n(Z_i) u_i + o_p(1)$$

$$\pm \; (E g^*(Z_i) X_i')^{-1} \sqrt{n}\, E\, g^*(Z_i) u_i$$

$$= (E g^*(Z_i) X_i')^{-1} \left\{ \underbrace{\frac{1}{\sqrt{n}} \sum (\hat{g}_n(Z_i) u_i - E\hat{g}(Z_i) u_i)}_{\mathbb{G}(\hat{g}_n) \text{ and assume it's SE}} + \sqrt{n}\, E\hat{g}(Z_i) u_i \right\} + o_p(1)$$

$$= (E g^*(Z_i) X_i')^{-1} \left\{ \frac{1}{\sqrt{n}} \sum (g^*(Z_i) u_i - \underbrace{E g^*(Z_i) u_i}_{=0}) + \sqrt{n}\, E\, \underbrace{\hat{g}_n(Z_i) u_i}_{} \right\} + o_p(1)$$

$=0$ by moment condition, recall that $\hat{g}_n(.)$ is some measurable function.

$$= (E g^*(Z_i) X_i')^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g^*(Z_i) u_i \right\} + o_p(1)$$

$$\xrightarrow{d} (E g^*(Z_i) X_i')^{-1} \; N(0, \; E\, u_i^2\, g^*(Z_i) g^*(Z_i)')$$

Recall now that $\mathbb{G}_n$ is a functional, so the definition we use is

$$\limsup_{n\to\infty} P\Big(\sup_{g_1 \in \mathcal{G}} \; \sup_{g_2 : \|g_1 - g_2\| < \delta} |\mathbb{G}_n(g_1 - g_2)| > \varepsilon \Big) < \varepsilon$$

↳ L_2 norm

Given two functions l, u of W_i such that $l(W_i) \le u(W_i)$. We can define a bracket $[l, u] := \{ g \in \mathcal{G} : l \le g \le u \}$. An ε-bracket satisfies $\|u - l\|_2 \le \varepsilon$.

We say a collection of brackets $\{[l,u]_\alpha : \alpha \in A\}$ covers $\mathcal{G}$ if $\mathcal{G} \subset \bigcup_{\alpha \in A} [l,u]_\alpha$. The bracketing number $N_{[\,]}(\varepsilon, \mathcal{G}, L_p)$ is the smallest number of ε-brackets $[l, u]$ needed to cover $\mathcal{G}$.

Ⓡ Roughly $\varepsilon \to \infty \Rightarrow N_{[\,]} \to 1$
$\varepsilon \to 0 \Rightarrow N_{[\,]} \to \infty$ or some large number.

The entropy with respect to bracketing is $\log N_{[\,]}(\varepsilon, \mathcal{G}, L_p)$, which measures the size/complexity of the family of functions $\mathcal{G}$.

Allow me to rewrite

$$\limsup_{n\to\infty} P\Big(\sup_{g_1 \in \mathcal{G}} \sup_{\substack{g_2 : \|g_1 - g_2\| < \delta \\ \hookrightarrow L_2 \text{ norm}}} |\mathbb{G}_n(g_1 - g_2)| > \varepsilon \Big) < \varepsilon$$

$$=$$

$$\limsup_{n\to\infty} P\Big(\sup_{\substack{g : \|g\| < \delta, \\ g \in \mathcal{G}}} |\mathbb{G}_n(g)| > \varepsilon \Big) < \varepsilon$$

Use Markov Ineq

$$\limsup_{n\to\infty} P\Big(\sup_{\substack{g : \|g\| < \delta, \\ g \in \mathcal{G}}} |\mathbb{G}_n(g)| > \varepsilon \Big) \leq \limsup_{n\to\infty} \Big\{ E \sup_{\substack{g : \|g\| < \delta, \\ g \in \mathcal{G}}} |\mathbb{G}_n(g)| \cdot \frac{1}{\varepsilon} \Big\}$$

OK... that doesn't seem helpful, but there's a proposition we could use.

<u>Proposition.</u>- Suppose that (i) $\|g\|_2 < \delta$ for all $g \in \mathcal{G}$

(ii) There's an envelope function $\bar{G}$ such that $|g| \leq \bar{G}$ for all $g \in \mathcal{G}$.

Then, for some constant $c > 0$,

$$E \sup_{g \in \mathcal{G}} |\mathbb{G}_n g| \leq c \cdot \Big\{ \int_0^\delta \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon + \sqrt{n}\, E\Big[\bar{G}\, \mathbb{1}\Big(\bar{G} > \delta \sqrt{\frac{n}{\log N_{[\,]}(\delta, \mathcal{G}, L_2)}} \Big) \Big] \Big\}$$

Using this proposition gives

$$\limsup_{n\to\infty} P\Big(\sup_{\substack{g : \|g\| < \delta, \\ g \in \mathcal{G}}} |\mathbb{G}_n(g)| > \varepsilon \Big) \leq \overline{\lim_{n\to\infty}} \frac{c}{\varepsilon} \Big\{ \underbrace{\int_0^\delta \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon}_{A} + \underbrace{\sqrt{n}\, E\Big[\bar{G}\, \mathbb{1}\Big(\bar{G} > \delta \sqrt{\frac{n}{\log N_{[\,]}(\delta, \mathcal{G}, L_2)}} \Big) \Big]}_{B} \Big\}$$

A $\quad \overline{\lim_{n\to\infty}} \lim_{\delta\to 0} \int_0^\delta \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon = 0$ using dominated convergence theorem.

We require a dominating function, like $K \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon$ for some $K > 1$. Then we also need that it's integrable

$$\int_0^\infty K \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon = O(1)$$, so a sufficient condition is

$$\int_0^\infty \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon = O(1).$$

B $\quad \overline{\lim_{n\to\infty}} \lim_{\delta\to 0} \sqrt{n}\, E\left[\bar{G}\, \mathbb{1}\left(\bar{G} > \sqrt{n} \underbrace{\frac{\delta}{\sqrt{\log N_{[\,]}(\delta, \mathcal{G}, L_2)}}}_{a(\delta)} \right)\right]$

This says that $\frac{\bar{G}}{\sqrt{n}} \cdot \frac{1}{a(\delta)} > 1$, so if we multiply it we increase this number.

$$\leq \overline{\lim_{n\to\infty}} \lim_{\delta\to 0} \sqrt{n}\, E\left[\bar{G}\, \frac{\bar{G}}{\sqrt{n}} \cdot \frac{1}{a(\delta)}\, \mathbb{1}\left(\bar{G} > \sqrt{n}\, a(\delta)\right)\right]$$

goes outside, it's non random

$$= \overline{\lim_{n\to\infty}} \lim_{\delta\to 0} \frac{1}{a(\delta)} E\left[\bar{G}^2\, \mathbb{1}\left(\bar{G} > \sqrt{n}\, a(\delta)\right)\right]$$

we want this to enter inside

By the dominated convergence theorem with $\bar{G}^2$ as dominating function and given that $E\,\bar{G}^2 < \infty$, which can be shown if we choose, for instance, $\bar{G} = |e| + |L|$.

$$= \lim_{\delta\to 0} \frac{1}{a(\delta)} E\left[\bar{G}^2 \underbrace{\overline{\lim_{n\to\infty}} \mathbb{1}\left(\bar{G} > \sqrt{n}\, a(\delta)\right)}_{=0}\right] = 0.$$

Putting it all together implies

$$\limsup_{n\to\infty} P\left(\sup_{\substack{g:\, \|g\| < \delta, \\ g \in \mathcal{G}}} |\mathbb{G}_n(g)| > \varepsilon \right) \leq \overline{\lim_{n\to\infty}} \frac{C}{\varepsilon} \left\{ \underbrace{\int_0^\delta \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{G}, L_2)}\, d\varepsilon}_{A} + \underbrace{\sqrt{n}\, E\left[\bar{G}\, \mathbb{1}\left(\bar{G} > \delta \sqrt{\frac{n}{\log N_{[\,]}(\delta, \mathcal{G}, L_2)}} \right)\right]}_{B} \right\}$$

$$= 0 \quad . \quad \blacksquare$$