Addendum OE

Rate and Limiting Distribution when θ_0 lies in the interior of Θ.

When θ_0 lies in the interior of Θ_0

- e.g. this is trivially true when $\Theta = \mathbb{R}^k$, and $\hat{\theta}_n$ is (weakly) consistent, as well as c_n is differentiable in an open neighborhood of θ_0, for almost every sample value, then $\hat{\theta}_n$ satisfies first order conditions with probability tending to one:

(*) suppose that $\frac{\partial c_n(\theta)}{\partial \theta}$ exists $\forall \theta \in B_{\theta_0}$, with B_{θ_0} an open set containing θ_0 (exists since θ_0 is interior) for almost every sample value. (In several cases $B_{\theta_0} = \mathbb{R}^k$)

* $\hat{\theta}_n \in B_{\theta_0}$ with probability converging to one, since $\hat{\theta}_n \xrightarrow{P} \theta_0$, and $\theta_0 \in B_{\theta_0}$.

* $\hat{\theta}_n \in \arg\min C_n(\theta)$ [suppose for simplicity that the optimization error is zero]

Then $\frac{\partial C_n(\hat{\theta}_n)}{\partial \theta} = 0_{q \times 1}$ with probability converging to one (A)

* If furthermore C_n is also two-times continuously differentiable on B_{θ_0}, for almost every sample value, and B_{θ_0} is convex ($\Rightarrow \forall \theta \in B_{\theta_0}$, there is a line in B_{θ_0} that connects θ with θ_0)

then due to the Mean Value Theorem,

$$\forall \theta \in B_{\theta_0}, \quad \frac{\partial C_n}{\partial \theta}(\theta) = \frac{\partial C_n}{\partial \theta}(\theta_0) + \frac{\partial^2 C_n(\theta^*)}{\partial \theta \partial \theta'}(\theta - \theta_0)$$

Can this be used in order to obtain the asymptotic approximation of θ_n?

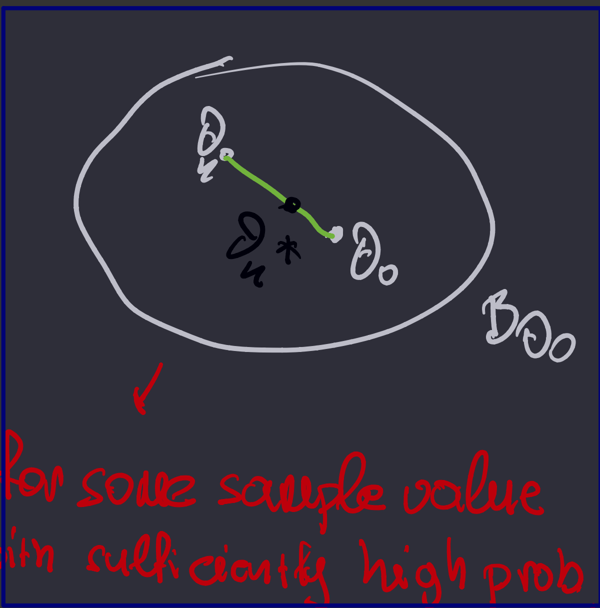
this is the Hessian of C_n evaluated at θ^* that belongs in the line that connects θ with θ_0 inside B_{θ_0}.

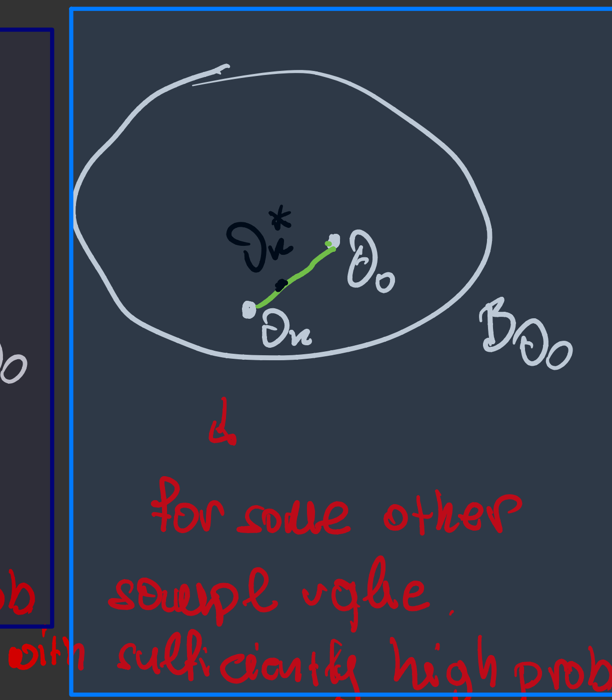
* Since $\theta_n \in B_{\theta_0}$ with probability converging to one, then the above Mean value expansion also holds with such like probability with θ_n in place of θ, i.e.

With probability converging to one:

$$\boxed{\frac{\partial C_n}{\partial \theta}(\hat{\theta}_n) = \frac{\partial C_n}{\partial \theta}(\theta_0) + \frac{\partial^2 C_n}{\partial \theta \partial \theta'}(\theta_n^*)(\hat{\theta}_n - \theta_0) \quad (B)}$$

where θ_n^* is stochastic, and lies in the line that connects $\hat{\theta}_n$ with θ_0





* Notice that $0 \leq \|\theta_n^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$ by definition. Hence, since $\|\hat{\theta}_n - \theta_0\| \xrightarrow{P} 0$, then

$\|\theta_n^* - \theta_0\| \xrightarrow{P} 0$, i.e. since $\theta_n \xrightarrow{P} \theta_0$ then by construction $\theta_n^* \xrightarrow{P} \theta_0$ (yet θ_n^* is generally latent)

* Combining (A) with (B) it is obtained that with probability converging to one

$$(\dagger)\quad \underbrace{\frac{\partial g_n}{\partial \theta}(\theta_0)}_{k\times 1} + \overbrace{\frac{\partial^2 g_n}{\partial \theta \partial \theta'}(\theta_n^*)}^{k\times k}\,\overbrace{(\theta_n - \theta_0)}^{k\times 1} = 0_{k\times 1}$$

Beware of the dimensionalities of those random elements

* In the previous we can ideneify the following compared to our general theory:

- $q_n(\theta_0) := \dfrac{\partial c_n}{\partial \theta}(\theta_0)$, and

- $g_n(\theta_n^*) = \frac{1}{2} \dfrac{\partial^2 c_n(\theta_n^*)}{\partial\theta\,\partial\theta'}$.

* If we utilize the relevant assumption parts from the general theory, i.e.

① $r_n \dfrac{\partial c_n}{\partial \theta}(\theta_0) \rightsquigarrow z_{\theta_0}$

② $\dfrac{\partial^2 c_n(\theta_n^*)}{\partial\theta\,\partial\theta} \xrightarrow{P} \breve{J}_{\theta_0}$ which is

this is then $\frac{1}{2}\breve{J}_{\theta_0}$ in the general theory

supposed ③ to be a positive definite matrix. Then:

A ② and ③ imply that

$\frac{\partial^2 c_n(\theta_n^*)}{\partial\theta\partial\theta'}$ is positive definite and

thereby invertible with probability converging to one. Combining this with ① it is obtained that:

$$\boxed{\theta_n - \theta_0 = -\left[\frac{\partial^2 c_n}{\partial\theta\partial\theta'}(\theta_n^*)\right]^{-1}\frac{\partial c_n}{\partial\theta}(\theta_0)}$$

with probability converging to one

C

C is essentially a(n) (latent)

asymptotic approximation of the form of the estimator [it is also reminiscent of the iteration form of the approximation of optimizers by some numerical optimization algorithms, like the Newton-Raphson].

Multiplying by the rate in (I) it is obtained that with probability converging to one:

$$(*)\quad r_n(\hat{\theta}_n - \theta_0) = \overbrace{-\left[\frac{\partial^2 G_n(\theta_n^*)}{\partial\theta\,\partial\theta'}\right]^{-1}}^{\xrightarrow[(2),\,(3)\ \text{CMT}]{P}\ -J_{\theta_0}^{-1}} \overbrace{r_n \frac{\partial G_n}{\partial\theta}(\theta_0)}^{\overset{(4)}{\rightsquigarrow}\ Z_{\theta_0}}$$

Using then 1, 2, 3, Slutsky's lemma and the CLT, it is obtained that:

$$\boxed{r_n(\theta_n - \theta_0) \rightsquigarrow -J_{\theta_0}^{-1} Z_{\theta_0} \quad (D)}$$

* (D) derives simultaneously the rate r_n, and the limit in distribution $-J_{\theta_0}^{-1} Z_{\theta_0}$ of $r_n(\theta_n - \theta_0)$.

* In several cases $r_n = \sqrt{n}$ and $Z_{\theta_0} \sim N(0_{k\times 1}, V_{\theta_0})$, for some - generally

latent - Covariance $k \times k$ Matrix V_{θ_0}.

Then $-J_{\theta_0}^{-1} Z_{\theta_0} \sim N\left(-J_{\theta_0}^{-1} \mathbf{0}_{k \times L}, \left(-J_{\theta_0}^{-1}\right)' V_{\theta_0} \left(-J_{\theta_0}^{-1}\right)'\right)$

$= N\left(\mathbf{0}_{k \times L}, J_{\theta_0}^{-1} V_{\theta_0} J_{\theta_0}^{-1}\right)$ (due to CLT, J_{θ_0}, and thereby $J_{\theta_0}^{-1}$, must be symmetric, as a limit of a sequence of Hessian Matrices (those are symmetric due to the continuity of the second order partial derivatives). Hence

$$\boxed{\sqrt{n}\,(\theta_n - \theta_0) \rightsquigarrow N\left(\mathbf{0}_{k \times L}, J_{\theta_0}^{-1} V_{\theta_0} J_{\theta_0}^{-1}\right)}$$

E

* there are cases where r_n is not $\sqrt{n}$ and/or Z_{θ_0} is not Gaussian.

* this is a special case of over general theory; it can be proven that since θ_0 is interior, $H = \mathbb{R}^5$.

* Given (E), the asymptotic variance of the estimator $J_{\theta_0}^{-1} V_{\theta_0} J_{\theta_0}^{-1}$ is latent, partially due to its dependence on θ_0:

if (2*) for all $\theta \to \theta_0$, $\frac{\partial^2 C_n}{\partial\theta\partial\theta'}(\theta) \xrightarrow{P} J_{\theta_0}$

then since $\theta_n \xrightarrow{P} \theta_0$, $\frac{\partial^2 C_n}{\partial\theta\partial\theta'}(\theta_n) \xrightarrow{P} J_{\theta_0}$, And due to the CLT,

$$\left(\frac{\partial^2 C_n}{\partial\theta\partial\theta'}(\hat\theta_n)\right)^{-1} \xrightarrow{P} J_{\theta_0}^{-1}.$$

If (4) Furthermore, $V_n \xrightarrow{P} V_{\theta_0}$,

i.e. the sample dependent V_n is a consistent estimator of V_{θ_0} (for lower level details, look at the examples for brevity), then due to the CLT:

$$\boxed{U_n^* := \left(\frac{\partial^2 C_n(\hat\theta_n)}{\partial\theta\partial\theta'}\right)^{-1} V_n \left(\frac{\partial^2 C_n(\hat\theta_n)}{\partial\theta\partial\theta'}\right)^{-1} \xrightarrow{P} J_{\theta_0}^{-1} V_{\theta_0} J_{\theta_0}^{-1}}$$

(F)

i.e. U_n is a consistent estimator of the asymptotic variance.

(*) The latter could be useful for the construction of Wald tests for the Hypotheses Structure:

$$\boxed{\begin{aligned} H_0 &: \theta_0 = \theta^* \\ H_1 &: \theta_0 \neq \theta^* \end{aligned}}$$

since under H_0 the Wald statistic

$$\boxed{n(\theta_n - \theta^*)' U_n^{-1} (\theta_n - \theta^*) \rightsquigarrow \chi^2_K .}$$

(given this, the testing procedure and its properties are explained in the main text).

Are there cases where r_n is not "classical" (i.e. $\sqrt{n}$) and or the limiting distribution not Gaussian

Tracing the proofs of the determination of the rate and the limiting distribution of the M-estimators shows that – given consistency – the rate and the limiting distribution of the estimator depend:

a. Smoothness of $\mathbb{G}_n$ (formulated as twice continuous differentiability), b. the rate and the limiting distribution of $\frac{\partial \mathbb{G}_n}{\partial \theta}(\theta_0)$, c. continuous convergence (at θ_0) of the Hessian of C_n to a limiting deterministic matrix, and d. the invertibility of the latter limit.

Any deviation in a. - d. could imply deviation from classical rates and/or asymptotic Gaussianity.

For example, if $r_n \frac{\partial C_n}{\partial \theta}(\theta_0) \rightsquigarrow N(0, V_{\theta_0})$ with $r_n \neq n^{1/2}$ (see for example the CLT in the Stochastic Convergence Notes where $r_n = \frac{n^{1/2}}{\ln(n)}$), yet the other conditions hold, then $r_n(\theta_n - \theta_0)$ would be asymptotically Gaussian, albeit with a non-classical rate. Different rates than $n^{1/3}$ are also compatible with non-Gaussianity in similar contexts, involving frameworks of stationarity and ergodicity, when for example $\frac{\partial C_n}{\partial \Theta}(\Theta_0) = \frac{1}{n}\sum_{i=1}^{n} q(\Theta_0, X_i)$, for some appropriate function q, but $\mathbb{E}(|q(\Theta_0, X_1)|^p) = +\infty$, for some $p \leq 2$.

Different rates for $\partial C_n / \partial \Theta(\Theta_0)$, can also appear in contexts of non-stationarity. If time permits we will examine such like behaviours for the LS case in non-stationary versions of the AR(1) model.

Furthermore, and if for example c. fails due to that the Hessian converges in distribution to a stochastic (yet non-singular) limit, this can affect both the rate and the limiting distribution of the estimator; this can happen in both linear and non linear models due to issues regarding absence of stationarity, and/or the non-existence of needed moments for the random element involved.

Finally, if d. fails, and the Hessian has an asymptotically degenerate limit, Asymptotic representations like $\circledast$ for the estimator would cease to be valid, the estimator could essentially be represented as a solution of more complicated asymptotically polynomial equations, thereby affecting the form of the rate and the limiting distribution.

The above could also affect the form of the testing procedures based on θ_n and/or the criterion, as well as their limiting distributions under the associated null-hypotheses. This makes important the identification of such-like deviations and the analogous modifications of the procedures, and/or the design of procedures that remain robust to (some of) such deviations from the classical theory.