7.8.2 Asymptotic Normality of M-Estimators

Next we give an asymptotic normality theorem that is the direct generalization of Theorem 6.7 (p. 286).

Theorem 7.2. *Let $Y_1, \ldots, Y_n$ be iid with distribution function $F(y)$. Assume that*

1. *$\boldsymbol{\psi}(y, \boldsymbol{\theta})$ and its first two partial derivatives with respect to $\boldsymbol{\theta}$ exist for all y in the support of F and for all $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, where $\boldsymbol{G}_{\mathrm{F}}(\boldsymbol{\theta}_0) = 0$.*
2. *For each $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, there exists a function $g(y)$ (possibly depending on $\boldsymbol{\theta}_0$) such that for all j, k and $l \in \{1, \ldots, b\}$,*

$$\left| \frac{\partial^2}{\partial\theta_j \partial\theta_k} \psi_l(y, \boldsymbol{\theta}) \right| \le g(y)$$

for all y and where $\int g(y)\, dF(y) < \infty$.

3. *$A(\boldsymbol{\theta}_0) = E\left\{-\boldsymbol{\psi}'(Y_1, \boldsymbol{\theta}_0)\right\}$ exists and is nonsingular.*
4. *$B(\boldsymbol{\theta}_0) = E\left\{\boldsymbol{\psi}(Y_1, \boldsymbol{\theta}_0)\boldsymbol{\psi}(Y_1, \boldsymbol{\theta}_0)^T\right\}$ exists and is finite.*

If $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{o}_p(n^{-1/2})$ and $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, then

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left[\boldsymbol{0}, \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\theta}_0)\left\{\boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\right\}^T\right] \quad \text{as } n \to \infty.$$

Proof. The proof uses a component-wise expansion of $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})$ similar to that in (6.21, p. 289) used in the proof of Theorem 6.10 (p. 288). By assumption $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) = \boldsymbol{o}_p(n^{-1/2})$ and thus a Taylor series expansion of the jth component of $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})$ results in

$$\begin{aligned} o_p(n^{-1/2}) &= \boldsymbol{G}_{n,j}(\widehat{\boldsymbol{\theta}}) \\ &= \boldsymbol{G}_{n,j}(\boldsymbol{\theta}_0) + \boldsymbol{G}'_{n,j}(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{G}''_{n,j}(\widetilde{\boldsymbol{\theta}}_j^*)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \boldsymbol{G}_{n,j}(\boldsymbol{\theta}_0) + \left\{\boldsymbol{G}'_{n,j}(\boldsymbol{\theta}_0) + \frac{1}{2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{G}''_{n,j}(\widetilde{\boldsymbol{\theta}}_j^*)\right\}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \end{aligned}$$

where $\widetilde{\boldsymbol{\theta}}_j^*$ is on the line segment joining $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, $j = 1, \ldots, b$. Writing these b equations in matrix notation we have

$$\boldsymbol{o}_p(n^{-1/2}) = \boldsymbol{G}_n(\boldsymbol{\theta}_0) + \left\{\boldsymbol{G}'_n(\boldsymbol{\theta}_0) + \frac{1}{2}\widetilde{Q}^*\right\}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\widetilde{Q}^*$ is the $b \times b$ matrix with jth row given by $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T \boldsymbol{G}''_{n,j}(\widetilde{\boldsymbol{\theta}}_j^*)$. Note that under Condition 2, each entry in $\widetilde{Q}^*$ is bounded by $||\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|| n^{-1}\sum g(Y_i) = o_p(1)$,

and thus $\widetilde{Q}^* = \boldsymbol{o}_p(1)$. By the WLLN $\boldsymbol{G}'_n(\boldsymbol{\theta}_0) \xrightarrow{p} -\boldsymbol{A}(\boldsymbol{\theta}_0)$ which is nonsingular under Condition 3. Thus for n sufficiently large, the matrix in brackets above is nonsingular with probability approaching 1. On the set where the matrix in brackets is nonsingular (call that set S_{N}) we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left\{\boldsymbol{G}'_n(\boldsymbol{\theta}_0) + \frac{1}{2}\widetilde{Q}^*\right\}^{-1}\left\{\sqrt{n}\boldsymbol{G}_n(\boldsymbol{\theta}_0) + \boldsymbol{o}_p(1)\right\}.$$

Slutsky's Theorem and the CLT then give the result when we note that $P(S_{\mathrm{N}}) \to 1$. As in Problem 6.6 (p. 293), we could also add and subtract terms to give an approximation-by-averages representation, where $h_{\mathrm{F}}(Y_i, \boldsymbol{\theta}_0) = \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\psi}(Y_i, \boldsymbol{\theta}_0)$. ∎

7.8.3 Weak Law of Large Numbers for Averages with Estimated Parameters

One of the most useful aspects of the M-estimator approach is the availability of the empirical sandwich estimator (7.12, p. 302). Thus, it is important that the pieces of this estimator, $\boldsymbol{A}_n(\boldsymbol{Y}, \widehat{\boldsymbol{\theta}})$ and $\boldsymbol{B}_n(\boldsymbol{Y}, \widehat{\boldsymbol{\theta}})$, converge in probability to $\boldsymbol{A}(\boldsymbol{\theta}_0)$ and $\boldsymbol{B}(\boldsymbol{\theta}_0)$, respectively. But note that this convergence would follow immediately from the WLLN except for the presence of $\widehat{\boldsymbol{\theta}}$. Thus, the next two theorems give conditions for the WLLN to hold for averages whose summands are a function of $\widehat{\boldsymbol{\theta}}$ (and thus dependent). The first theorem assumes differentiability and a bounding function similar to Theorem 5.28 (p. 249). The second uses monotonicity.

Theorem 7.3. *Suppose that $Y_1, \ldots, Y_n$ are iid with distribution function F and assume that the real-valued function $q(Y_i, \boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$, $E_{\mathrm{F}}|q'(Y_1, \boldsymbol{\theta}_0)| < \infty$, and there exists a function $M(y)$ such that for all $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$ and all $j \in \{1, \ldots, b\}$,*

$$\left| \frac{\partial}{\partial\theta_j} q(y, \boldsymbol{\theta}) \right| \le M(y),$$

where $E_{\mathrm{F}}\{M(Y_1)\} < \infty$. If $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$, then $n^{-1}\sum_{i=1}^n q(Y_i, \widehat{\boldsymbol{\theta}}) \xrightarrow{p} E_{\mathrm{F}}q(Y_1, \boldsymbol{\theta}_0)$ as $n \to \infty$.

Proof.

$$\begin{aligned} \left| \frac{1}{n}\sum_{i=1}^n q(Y_i, \widehat{\boldsymbol{\theta}}) - E_{\mathrm{F}}q(Y_1, \boldsymbol{\theta}_0) \right| &\le \left| \frac{1}{n}\sum_{i=1}^n q(Y_i, \widehat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{i=1}^n q(Y_i, \boldsymbol{\theta}_0) \right| \\ &\quad + \left| \frac{1}{n}\sum_{i=1}^n q(Y_i, \boldsymbol{\theta}_0) - E_{\mathrm{F}}q(Y_1, \boldsymbol{\theta}_0) \right| \end{aligned}$$