MATH 557 – ASYMPTOTIC NORMALITY OF THE MLE

We consider a Taylor expansion of the function $\ell(x;\theta) = \log f_X(x;\theta)$ with respect to θ around θ_0. We have any value of θ

$$\ell(x;\theta) = \ell(x;\theta_0) + \dot{\ell}(x;\theta_0)(\theta - \theta_0) + \frac{1}{2}(\theta-\theta_0)^\top \ddot{\ell}(x;\theta_0)(\theta-\theta_0) + \mathcal{R}_3(x;\theta^*) \tag{1}$$

where $\mathcal{R}_3(x;\theta^*)$ is a remainder term, for some θ^* such that $\|\theta_0 - \theta^*\| \leqslant \|\theta_0 - \theta\|$. Evaluating (1) for each of $x_1,\ldots,x_n$ and summing the result, we have

$$\ell_n(\theta) = \ell_n(\theta_0) + \dot{\ell}_n(\theta_0)^\top(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^\top \ddot{\ell}_n(\theta_0)(\theta-\theta_0) + \mathcal{R}_3(x_{1:n};\theta^*). \tag{2}$$

Evaluating this expression at $\theta = \widehat{\theta}_n$ and rearranging we have

$$\ell_n(\widehat{\theta}_n) - \ell_n(\theta_0) = \dot{\ell}_n(\theta_0)^\top(\widehat{\theta}_n - \theta_0) + \frac{1}{2}(\widehat{\theta}_n-\theta_0)^\top \ddot{\ell}_n(\theta_0)(\widehat{\theta}_n-\theta_0) + \mathcal{R}_3(x_{1:n};\theta^*) \tag{3}$$

where $\|\theta_0 - \theta^*\| \leqslant \|\theta_0 - \widehat{\theta}_n\|$. The left hand side of (3) converges to zero by previously established results. Consider now the right hand side of (3). The first term is

$$\dot{\ell}_n(\theta_0)^\top(\widehat{\theta}_n - \theta_0) \quad = \quad \left\{\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0)\right\}^\top \left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\} = \left\{\sqrt{n}\left(\frac{1}{n}\dot{\ell}_n(\theta_0)\right)\right\}^\top \left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\}.$$

Consider now a Taylor expansion of $\dot{\ell}_n(\theta)$ around θ_0 evaluated at $\widehat{\theta}_n$:

$$\mathbf{0}_k = \dot{\ell}_n(\widehat{\theta}_n) = \dot{\ell}_n(\theta_0) + \ddot{\ell}_n(\theta_0)(\widehat{\theta}_n - \theta_0) + \frac{1}{2}(\widehat{\theta}_n-\theta_0)^\top \dddot{\ell}_n(\theta^\dagger)(\widehat{\theta}_n-\theta_0)$$

where $\|\theta_0 - \theta^\dagger\| \leqslant \|\theta_0 - \widehat{\theta}_n\|$. On rearrangement, we obtain that

$$\dot{\ell}_n(\theta_0) = -\ddot{\ell}_n(\theta_0)(\widehat{\theta}_n - \theta_0) - \frac{1}{2}(\widehat{\theta}_n-\theta_0)^\top \dddot{\ell}_n(\theta^\dagger)(\widehat{\theta}_n-\theta_0)$$

and hence, dividing through by $\sqrt{n}$ we have

$$\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) \quad = \quad -\frac{1}{\sqrt{n}}\ddot{\ell}_n(\theta_0)(\widehat{\theta}_n - \theta_0) - \frac{1}{\sqrt{n}}\frac{1}{2}(\widehat{\theta}_n-\theta_0)^\top \dddot{\ell}_n(\theta^\dagger)(\widehat{\theta}_n-\theta_0). \tag{4}$$

Note that the right hand side of (4) can be rewritten

$$\left[-\left\{\frac{1}{n}\ddot{\ell}_n(\theta_0)\right\} - \frac{1}{2}(\widehat{\theta}_n-\theta_0)^\top\left\{\frac{1}{n}\dddot{\ell}_n(\theta^\dagger)\right\}\right]\left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\}. \tag{5}$$

- In its random variable form, the left hand side of (4) is

$$\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{\ell}(X_i;\theta_0) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n U(X_i;\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U(X_i;\theta_0)\right)$$

  that is, a sample average quantity scaled by $\sqrt{n}$. But by definition of θ_0,

$$\mathbb{E}_{f_0}[U(X_i;\theta_0)] = \int \dot{\ell}(y;\theta_0) f_0(y)\, dy = \mathbf{0}_k$$

as, by definition θ_0 minimizes $KL(f_0, f_X(X;\theta))$, and therefore must be a solution of this equation. Therefore, by the Central Limit Theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U(X_i;\theta_0) \xrightarrow{d} \text{Normal}_k(\mathbf{0}_k, \mathcal{J}_{f_0}(\theta_0)) \tag{6}$$

where

$$\mathcal{J}_{f_0}(\theta_0) = \mathbb{E}_{f_0}[U(X;\theta_0)U(X;\theta_0)^\top] \equiv \text{Var}_{f_0}[U(X;\theta_0)] \qquad (k \times k).$$

Thus the left hand side of (4) converges in distribution to a Normal random variable given by (6).

- For the right hand side of (4), consider the terms in (5). Specifically, suppose that the third-derivative term $\dddot{\ell}(X;\theta)$ is bounded in expectation, that is, for all θ

  $$\mathbb{E}_{f_0}[\dddot{\ell}(X;\theta)] < \mathbf{M}(\theta) \qquad (k \times k \times k).$$

  Then we have by the strong law of large numbers that

  $$\frac{1}{n}\dddot{\ell}_n(\theta^\dagger) \xrightarrow{a.s.} \mathbb{E}_{f_0}[\dddot{\ell}(X;\theta^\dagger)]$$

  with $\mathbb{E}_{f_0}[\dddot{\ell}(X;\theta^\dagger)]$ a **finite** array. Hence, as by earlier results $\widehat{\theta}_n \xrightarrow{p} \theta_0$, we have by Slutsky's theorem that

  $$\frac{1}{2}(\widehat{\theta}_n - \theta_0)^\top \left\{\frac{1}{n}\dddot{\ell}_n(\theta^\dagger)\right\} \xrightarrow{p} \mathbf{0}_{k\times k}$$

  and hence we may write that

  $$\left[\frac{1}{2}(\widehat{\theta}_n - \theta_0)^\top \left\{\frac{1}{n}\dddot{\ell}_n(\theta^\dagger)\right\}\right]\left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\} \xrightarrow{p} \mathbf{0}_k$$

  which may be alternately denoted

  $$\left[\frac{1}{2}(\widehat{\theta}_n - \theta_0)^\top \left\{\frac{1}{n}\dddot{\ell}_n(\theta^\dagger)\right\}\right]\left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\} = \text{o}_p(1).$$

  Therefore we write from (4) that

  $$\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) = \left\{-\frac{1}{n}\ddot{\ell}_n(\theta_0)\right\}\left\{\sqrt{n}(\widehat{\theta}_n - \theta_0)\right\} + \text{o}_p(1)$$

  where the distribution of the left hand size is given by (6). Under regularity conditions, we have that

  $$-\frac{1}{n}\ddot{\ell}_n(\theta_0) \xrightarrow{p} \mathbb{E}_{f_0}[\ddot{\ell}(Y;\theta_0)] = \mathcal{I}_{f_0}(\theta_0) \qquad (k \times k)$$

  where we presume that $\mathcal{I}_{f_0}(\theta_0)$ is non-singular. By Slutsky's theorem, we therefore have that

  $$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} \text{Normal}_k(\mathbf{0}_k, \{\mathcal{I}_{f_0}(\theta_0)\}^{-1}\mathcal{J}_{f_0}(\theta_0)\{\mathcal{I}_{f_0}(\theta_0)\}^{-1}).$$

- The remainder term in (3), when considered as a random quantity $\mathcal{R}_3(X_{1:n};\theta^*)$, can be shown to have the property

  $$\frac{1}{\sqrt{n}}\mathcal{R}_3(X_{1:n};\theta^*) = \text{o}_p(1)$$

  as $\mathcal{R}_3(x_{1:n};\theta^*)$ depends on the third derivative $\dddot{\ell}$, which is presumed above to be bounded in expectation, and also the term is $\text{O}(||\widehat{\theta}_n - \theta_0||^3)$, and we established that $\widehat{\theta}_n \xrightarrow{p} \theta_0$.