Assignment 1

Weihao Li

2020/8/29

1

(a)

$X_T = \overline{Y}_T - \mu_0$ is of order $T^{-\frac{1}{2}}$ in probability. $T^{-\frac{1}{2}}$ is a deterministic sequence of T that converges to 0 as $T \to \infty$.

Proof:

By definition of $O_p(.)$,

if for every $\varepsilon > 0$, there exists a constant M_ε, such that:

$$\sup_T Pr(|\frac{X_T}{R_T}| > M_\varepsilon) < \varepsilon$$

then we can write:

$$X_T = O_p(R_T)$$

We know

$$\sqrt{T}(\overline{Y}_T - \mu_0) = O_p(1)$$

$$\Longrightarrow$$

$$\sup_T Pr(|\frac{\sqrt{T}(\overline{Y}_T - \mu_0)}{1}| > M_\varepsilon) < \varepsilon$$

$$\sup_T Pr(|\frac{(\overline{Y}_T - \mu_0)}{\frac{1}{\sqrt{T}}}| > M_\varepsilon) < \varepsilon$$

$$\sup_T Pr(|\frac{(\overline{Y}_T - \mu_0)}{T^{-\frac{1}{2}}}| > M_\varepsilon) < \varepsilon$$

$$(\overline{Y}_T - \mu_0) = X_T = O_p(T^{-\frac{1}{2}})$$

(b)

$X_T = \overline{Y}_T - \mu_0$ is of smaller order in probability than $T^{-\frac{1}{4}}$

Proof:

By definition of $o_p(.)$,

if for every $\varepsilon > 0$,

$$\lim_{T \to \infty} Pr(|\frac{X_T}{R_T} - 0| > \varepsilon) = 0$$

then we can write:

$$X_T = o_p(R_T)$$

$$\frac{\overline{Y}_T - \mu_0}{T^{-\frac{1}{4}}} = T^{\frac{1}{4}}(\overline{Y}_T - \mu_0)$$
$$= T^{-\frac{1}{4}}\sqrt{T}(\overline{Y}_T - \mu_0)$$

Given $\sqrt{T}(\overline{Y}_T - \mu_0) = O_p(1)$ and $T^{-\frac{1}{4}} = o_p(1)$. Then using the rules of engagement $O_p(1)o_p(1) = o_p(1)$,
$T^{-\frac{1}{4}}\sqrt{T}(\overline{Y}_T - \mu_0) = o_p(1)$

for every $\varepsilon > 0$,

$$\lim_{T \to \infty} Pr(|\frac{T^{-\frac{1}{4}}\sqrt{T}(\overline{Y}_T - \mu_0)}{1} - 0| > \varepsilon) = 0$$
$$\lim_{T \to \infty} Pr(|\frac{(\overline{Y}_T - \mu_0)}{T^{-\frac{1}{4}}} - 0| > \varepsilon) = 0$$

$$X_T = \overline{Y}_T - \mu_0 = o_p(T^{-\frac{1}{4}})$$

$X_T = \overline{Y}_T - \mu_0$ is of smaller order in probability than $T^{-\frac{1}{4}}$

(c)

$X_T = \overline{Y}_T - \mu_0$ is of smaller order in probability than $T^{-\frac{1}{4}}$

informal explanation:

Using $\overline{Y}_T - \mu_0 \overset{approx}{\sim} N(0, \frac{\sigma_0^2}{T})$

$$\frac{\overline{Y}_T - \mu_0}{T^{-\frac{1}{4}}} \approx T^{\frac{1}{4}} \times N(0, \frac{\sigma_0^2}{T})$$
$$= N(0, T^{-\frac{1}{2}}\sigma_0^2)$$

since $T^{-\frac{1}{2}} \to 0$ as $T \to \infty$, $N(0, T^{-\frac{1}{2}}\sigma_0^2)$ will become a degenerate distribution concentrate onto 0. In other words, $\frac{\overline{Y}_T - \mu_0}{T^{-\frac{1}{4}}} = \frac{X_T}{T^{-\frac{1}{4}}} = o_p(1)$ and X_T is of smaller order in probability than $T^{-\frac{1}{4}}$.

2

(d)

Given the conditions provided in the question 1 and by the central limit theorem,

$$\frac{\sqrt{T}(\overline{Y}_T - \mu_0)}{\sigma_0} \xrightarrow{d} N(0,1)$$

Now $\hat{\sigma}$ is any consistent estimator of $\hat{\sigma}_0$, which means

$$\hat{\sigma}_0 \xrightarrow{p} \hat{\sigma}$$

By the continuous mapping theorem, if $X_T \xrightarrow{d} X$ and $C_T \xrightarrow{p} C$, then $X_T/C_T \xrightarrow{d} X/C$. Thus,

$$\frac{\sqrt{T}(\overline{Y}_T - \mu_0)}{\hat{\sigma}} \xrightarrow{d} N(0,1)$$

2

(a)

$$lnL_T(\boldsymbol{\theta}) = \frac{1}{T} ln f(\mathbf{Y}|\boldsymbol{\theta})$$

independence

$$= \frac{1}{T} ln \prod_{t=1}^{T} f(y_t|\boldsymbol{\theta})$$

identically distributed

$$= \frac{1}{T} ln \prod_{t=1}^{T} (2\pi)^{-1/2}(\sigma^2)^{-1/2} exp\{-\frac{1}{2\sigma^2}(y_t - \mu)^2\}$$

$$= \frac{1}{T} ln[(2\pi)^{-T/2}(\sigma^2)^{-T/2} \prod_{t=1}^{T} exp\{-\frac{1}{2\sigma^2}(y_t - \mu)^2\}]$$

$$= \frac{1}{T} ln[(2\pi)^{-T/2}(\sigma^2)^{-T/2} exp\{\sum_{t=1}^{T} -\frac{1}{2\sigma^2}(y_t - \mu)^2\}]$$

$$= \frac{1}{T} ln[(2\pi)^{-T/2}(\sigma^2)^{-T/2} exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu)^2\}]$$

$$= \frac{1}{T}[ln((2\pi)^{-T/2}) + ln((\sigma^2)^{-T/2}) + ln(exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu)^2\})]$$

$$= \frac{1}{T}[ln((2\pi)^{-T/2}) + ln((\sigma^2)^{-T/2}) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu)^2]$$

$$= \frac{1}{T}[\frac{-T}{2}ln(2\pi) + \frac{-T}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - \mu)^2]$$

$$= -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu)^2$$

(b)

$$G_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\boldsymbol{\theta})$$

$$= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial ln f(y_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

identically distributed

$$= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial ln((2\pi)^{-1/2}(\sigma^2)^{-1/2} exp\{-\frac{1}{2\sigma^2}(y_t - \mu)^2\})}{\partial \boldsymbol{\theta}}$$

$$= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial ln((2\pi)^{-1/2}) + ln((\sigma^2)^{-1/2}) - \frac{1}{2\sigma^2}(y_t - \mu)^2}{\partial \boldsymbol{\theta}}$$

$$= \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{\partial ln((2\pi)^{-1/2}) + ln((\sigma^2)^{-1/2}) - \frac{1}{2\sigma^2}(y_t - \mu)^2}{\partial \mu} \\ \frac{\partial ln((2\pi)^{-1/2}) + ln((\sigma^2)^{-1/2}) - \frac{1}{2\sigma^2}(y_t - \mu)^2}{\partial \sigma} \end{bmatrix}$$

$$= \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{1}{\sigma^2}(y_t - \mu) \\ -\frac{1}{\sigma} + (y_t - \mu)^2\sigma^{-3} \end{bmatrix}$$

(c)

$$H_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} h_t(\boldsymbol{\theta})$$

$$= \frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 ln f(y_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$$

$$= \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} \frac{\partial^2 ln f(y_t|\boldsymbol{\theta})}{\partial \mu^2} & \frac{\partial^2 ln f(y_t|\boldsymbol{\theta})}{\partial \mu \partial \sigma} \\ \frac{\partial^2 ln f(y_t|\boldsymbol{\theta})}{\partial \sigma \partial \mu} & \frac{\partial^2 ln f(y_t|\boldsymbol{\theta})}{\partial \sigma^2} \end{bmatrix}$$

using the result from part b

$$= \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} -\frac{1}{\sigma^2} & -2(y_t - \mu)\sigma^{-3} \\ -2(y_t - \mu)\sigma^{-3} & \frac{1}{\sigma^2} - 3(y_t - \mu)^2\sigma^{-4} \end{bmatrix}$$

(d)

To derive the MLE, we set the first order derivative of $ln L_T(\boldsymbol{\theta})$ to $\mathbf{0}$

4

$$G_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \frac{1}{\sigma^2}(y_t - \mu) \\ -\frac{1}{\sigma} + (y_t - \mu)^2 \sigma^{-3} \end{bmatrix} = \mathbf{0}$$

$$\begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sigma^2}(y_t - \mu) \\ \frac{1}{T} \sum_{t=1}^{T} -\frac{1}{\sigma} + (y_t - \mu)^2 \sigma^{-3} \end{bmatrix} = \mathbf{0}$$

$$\hat{\mu}_{MLE} = \overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$$

$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \overline{y})^2}$$

$$\hat{\boldsymbol{\theta}}_{MLE} = (\hat{\mu}_{MLE}, \hat{\sigma}_{MLE})'$$

We define a non-zero column vector $\boldsymbol{z}$ with real entries a and b

$$\boldsymbol{z}' H_T(\boldsymbol{\theta}) \boldsymbol{z} = \begin{bmatrix} a & b \end{bmatrix} (\frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} -\frac{1}{\sigma^2} & -2(y_t - \mu)\sigma^{-3} \\ -2(y_t - \mu)\sigma^{-3} & \frac{1}{\sigma^2} - 3(y_t - \mu)^2 \sigma^{-4} \end{bmatrix}) \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= \begin{bmatrix} a\frac{1}{T}\sum_{t=1}^{T} -\frac{1}{\sigma^2} + b\frac{1}{T}\sum_{t=1}^{T} -2(y_t - \mu)\sigma^{-3} & a\frac{1}{T}\sum_{t=1}^{T} -2(y_t - \mu)\sigma^{-3} + b\frac{1}{T}\sum_{t=1}^{T} \frac{1}{\sigma^2} - 3(y_t - \mu)^2 \sigma^{-4} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= a^2 \frac{1}{T} \sum_{t=1}^{T} -\frac{1}{\sigma^2} + ab\frac{1}{T} \sum_{t=1}^{T} -2(y_t - \mu)\sigma^{-3} + ab\frac{1}{T} \sum_{t=1}^{T} -2(y_t - \mu)\sigma^{-3} + b^2 \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sigma^2} - 3(y_t - \mu)^2 \sigma^{-4}$$

$$= a^2 \frac{1}{T} \sum_{t=1}^{T} -\frac{1}{\sigma^2} + 2ab\frac{1}{T} \sum_{t=1}^{T} -2(y_t - \mu)\sigma^{-3} + b^2 \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\sigma^2} - 3(y_t - \mu)^2 \sigma^{-4}$$

$$= -\frac{a^2}{\sigma^2} + \frac{b^2}{\sigma^2} - \frac{b^2}{T} \sum_{t=1}^{T} 3(y_t - \mu)^2 \sigma^{-4} - \frac{2ab}{T} \sum_{t=1}^{T} 2(y_t - \mu)\sigma^{-3}$$

$$= \frac{b^2 - a^2}{\sigma^2} - 3\frac{b^2}{T} \sigma^{-4} \sum_{t=1}^{T} (y_t - \mu)^2 - \frac{4ab}{T} \sigma^{-3} \sum_{t=1}^{T} (y_t - \mu)$$

$$\boldsymbol{z}' H_T(\hat{\boldsymbol{\theta}}_{MLE}) \boldsymbol{z} = \frac{b^2 - a^2}{\hat{\sigma}^2_{MLE}} - 3\frac{b^2}{T} \hat{\sigma}^{-4}_{MLE} \sum_{t=1}^{T} (y_t - \hat{\mu}_{MLE})^2 - \frac{4ab}{T} \hat{\sigma}^{-3}_{MLE} \sum_{t=1}^{T} (y_t - \hat{\mu}_{MLE})$$

$$= \frac{b^2 - a^2}{\hat{\sigma}^2_{MLE}} - 3\frac{b^2}{T} \hat{\sigma}^{-4}_{MLE} T \hat{\sigma}^2_{MLE} - 0$$

$$= \frac{-2b^2 - a^2}{\hat{\sigma}^2_{MLE}}$$

Since $-2b^2 - a^2 < 0$ and $\hat{\sigma}^2_{MLE} > 0$,

$$\frac{-2b^2 - a^2}{\hat{\sigma}^2_{MLE}} < 0$$

$H_T(\hat{\boldsymbol{\theta}}_{MLE})$ is a negative definite matrix. The second-order condition for a maximum holds.

(e)

Using the equation in part a, and replace σ^2 with $\frac{1}{\sigma^{-2}}$

$$lnL_T(\boldsymbol{\eta}) = -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\frac{1}{\sigma^{-2}}) - \frac{1}{2\frac{1}{\sigma^{-2}}}\frac{1}{T}\sum_{t=1}^{T}(y_t - \mu)^2$$

$$G_T(\boldsymbol{\eta}) = \frac{1}{T}\sum_{t=1}^{T}g_t(\boldsymbol{\eta})$$

$$= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial lnf(y_t|\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}$$

identically distributed

$$= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial ln((2\pi)^{-1/2}(\frac{1}{\sigma^{-2}})^{-1/2}exp\{-\frac{1}{2\frac{1}{\sigma^{-2}}}(y_t - \mu)^2\})}{\partial \boldsymbol{\eta}}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial ln((2\pi)^{-1/2}) + ln((\frac{1}{\sigma^{-2}})^{-1/2}) - \frac{1}{2\frac{1}{\sigma^{-2}}}(y_t - \mu)^2}{\partial \boldsymbol{\eta}}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{\partial ln((2\pi)^{-1/2})+ln((\frac{1}{\sigma^{-2}})^{-1/2})-\frac{1}{2\frac{1}{\sigma^{-2}}}(y_t-\mu)^2}{\partial \mu} \\ \frac{\partial ln((2\pi)^{-1/2})+ln((\frac{1}{\sigma^{-2}})^{-1/2})-\frac{1}{2\frac{1}{\sigma^{-2}}}(y_t-\mu)^2}{\partial \sigma^{-2}} \end{bmatrix}$$

$$= \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} \frac{1}{\sigma^2}(y_t - \mu) \\ \frac{1}{2\sigma^{-2}} - \frac{1}{2}(y_t - \mu)^2 \end{bmatrix}$$

$$H_T(\boldsymbol{\eta}) = \begin{bmatrix} -\frac{1}{\sigma^2} & y_t - \mu \\ y_t - \mu & -\frac{1}{2}(\sigma^{-2})^{-2} \end{bmatrix}$$

Set $G_T(\boldsymbol{\eta}) = \mathbf{0}$,

$$\hat{\mu}_{MLE} = \bar{y} = \frac{1}{T}\sum_{t=1}^{T}y_t$$

$$\widehat{\sigma^{-2}}_{MLE} = [\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2]^{-1}$$

$$\hat{\boldsymbol{\eta}}_{MLE} = (\hat{\mu}_{MLE}, \widehat{\sigma^{-2}}_{MLE})'$$

We define a non-zero column vector $\boldsymbol{z}$ with real entries a and b

$$z'H_T(\boldsymbol{\theta})z = \begin{bmatrix} a & b \end{bmatrix} \left(\frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} -\frac{1}{\sigma^2} & y_t - \mu \\ y_t - \mu & -\frac{1}{2}(\sigma^{-2})^{-2} \end{bmatrix}\right) \begin{bmatrix} a \\ b \end{bmatrix}$$

$$= \frac{a^2}{T}\sum_{t=1}^{T} -\frac{1}{\sigma^2} + 2\frac{ab}{T}\sum_{t=1}^{T}(y_t - \mu) + \frac{b^2}{T}\sum_{t=1}^{T} -\frac{1}{2}(\sigma^{-2})^2$$

$$= -\frac{a^2}{\sigma^2} + 2\frac{ab}{T}\sum_{t=1}^{T}(y_t - \mu) - \frac{b^2}{2}(\sigma^{-2})^2$$

$$z'H_T(\boldsymbol{\theta}_{MLE})z = -a^2\widehat{\sigma^{-2}}_{MLE} + 2\frac{ab}{T}\sum_{t=1}^{T}(y_t - \hat{\mu}_{MLE}) - \frac{b^2}{2}(\widehat{\sigma^{-2}}_{MLE})^2$$

$$= -a^2\widehat{\sigma^{-2}}_{MLE} - \frac{b^2}{2}(\widehat{\sigma^{-2}}_{MLE})^2$$

Since $\widehat{\sigma^{-2}}_{MLE} > 0$, $(\widehat{\sigma^{-2}}_{MLE})^2 > 0$ $a^2 > 0$ and $\frac{b^2}{2} > 0$, so

$$-a^2\widehat{\sigma^{-2}}_{MLE} - \frac{b^2}{2}(\widehat{\sigma^{-2}}_{MLE})^2 < 0$$

$H_T(\hat{\boldsymbol{\eta}}_{MLE})$ is a negative definite matrix. The second-order condition for a maximum holds. The second-order condition for a maximum holds.

Meanwhile,

$$\widehat{\sigma^{-2}}_{MLE} = \left(\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})^2}\right)^{-2} = (\hat{\sigma}_{MLE})^{-2} = C(\hat{\sigma}_{MLE})$$

and

$$\hat{\mu}_{MLE} = \hat{\mu}_{MLE} = C(\hat{\mu}_{MLE})$$

.

We also know

$C(\sigma_0) = \sigma_0^{-2}$ and $C(\mu_0) = \mu_0$

$\implies$ The functions apply to the true parameters are the same as the functions we obtain from the MLE. Thus,

$$\hat{\boldsymbol{\eta}}_{MLE} = C(\hat{\boldsymbol{\theta}}_{MLE})$$

if $\boldsymbol{\eta}_0 = C(\boldsymbol{\theta}_0)$ and $\hat{\boldsymbol{\theta}}_{MLE}$ is the MLE of $\boldsymbol{\theta}_0$.

3

(a)

In the derivation of the limiting distribution of MLE, we first using the Taylor's theorem to expand the average of derivatives $G_T(\theta)$.

$$\frac{1}{T}\sum_{t=1}^{T} g_t(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0) + (\theta - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0)] + \frac{1}{2!}(\theta - \theta_0)^2[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)]$$

The reason we use this tool is because we want the term $\theta - \theta_0$, which can then be evaluate at $\hat{\theta}_{MLE}$. Then we rearrange the terms in the equation

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) = \sqrt{T}[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)]/\{-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0) - \frac{1}{2!}(\hat{\theta}_{MLE} - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)]\}$$

From this equation, we can see that if we can prove that $\frac{1}{2!}(\hat{\theta}_{MLE} - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)]$ is negligible, then we will have a nice distribution form for $\sqrt{T}(\hat{\theta}_{MLE} - \theta_0)$, given $\sqrt{T}[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)]/\{-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0)\}$ will become a normal distribution asymptotically.

To prove $\frac{1}{2!}(\hat{\theta}_{MLE} - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)]$ is negligible we need to use the **preliminary result of consistency**.

By the result of consistency,

$$(\hat{\theta}_{MLE} - \theta_0) = o_p(1)$$

By the regularity condition 3, $q_t(\theta^*) < |q_t(\theta^*)|$ will be bounded by some values B_t. And by the WLLN, the average of B_t converges to $E[B]$, which means the average of $q_t(\theta^*)$ will be always smaller than a value $E[B]$. Therefore, by the definition of $O_p(.)$, $\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*) = O_p(1)$.

Finally, by the rules of engagement, $kO_p(1)o_p(1) = o_p(1)$

$$\frac{1}{2!}(\hat{\theta}_{MLE} - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)] = o_p(1)$$

Now,

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) = \sqrt{T}[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)]/\{-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0) - o_p(1)\}$$

Given in the i.i.d case, the information equality holds, which means

$I(\theta_0) = -H(\theta_0) = J(\theta_0)$

By the WLLN, $-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0)$ is a consistent estimator of $-H(\theta_0)$.

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) = \sqrt{T}[\frac{\frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)}{\sqrt{I(\theta_0)}}]/\{\frac{-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0) - o_p(1)}{\sqrt{I(\theta_0)}}\}$$

$$\{\frac{-\frac{1}{T}\sum_{t=1}^{T} h_t(\theta_0) - o_p(1)}{\sqrt{I(\theta_0)}}\} \xrightarrow{P} \sqrt{I(\theta)}$$

By the CLT and Lemma 2 that the expectation of the gradient at the true value is equal to 0

$$\frac{\frac{1}{T}\sum_{t=1}^{T} g_t(\theta_0)}{\sqrt{I(\theta_0)}} \xrightarrow{d} \mathbb{Z}$$

By the CMT,

$$\sqrt{T}(\hat{\theta}_{MLE} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$$

(b)

There are three regularity conditions that are not required in proof of concsistency.

R1: The true θ_0 is at some interior point (i.e. not on the boundary) of Θ

This one is required because if the θ_0 is on the boundary of the parameter space, it is possible that the $E[g(\theta_0)]$ is not equal to 0. In this situation, our derivation and proof will become much more complicated.

R3:

- (a) $L(\theta|\mathbf{y})$ is a thrice-differentiable continuous function of θ
- (b) all derivatives are bounded, when evaluated at any of θ close to θ_0

We use (a) to define the first, second and third-order derivative of the log likelihood function. The third-order derivative is only used in the Taylor's theorem application in the proof of limiting normality.

We use (b) to bound the value of the third-order derivative to prove $\frac{1}{2!}(\hat{\theta}_{MLE} - \theta_0)[\frac{1}{T}\sum_{t=1}^{T} q_t(\theta^*)]$ is negligible which is a step of the proof of limiting normality.

R5: The support of $Y_1, Y_2, ..., Y_T$ is independent of θ.

We use this condition to put the differentiation inside the integral. More particular, we use it in the proof of Lemma 2 and Lemma 3.

4

(a)

With reference to Q2, which is a i.i.d case, where

$$I(\theta_0) = -H(\theta_0) = J(\theta_0)$$

holds

$$I^{-1}(\boldsymbol{\theta}_0) = -H(\boldsymbol{\theta}_0) = -\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} -\frac{1}{\sigma_0^2} & -2(y_t - \mu_0)\sigma_0^{-3} \\ -2(y_t - \mu_0)\sigma_0^{-3} & \frac{1}{\sigma_0^2} - 3(y_t - \mu_0)^2\sigma_0^{-4} \end{bmatrix}$$

(b)