Proof of Wilk's phenomenon and Wald/Rao test

November 1, 2021

In this note, we prove Theorem 7.3.1 in p.314,315 of the textbook.

(**Proof of (a)&(b) of Theorem 7.3.1**) For notational simplicity, let $\hat{\theta} \equiv \hat{\theta}_n^{\Omega}$. Since $\dot{\ell}(\hat{\theta}) = 0$, by Taylor's exapnsion,

$$\begin{aligned}
2(\ell(\hat{\theta}) - \ell(\theta_0)) &= 2\ell(\hat{\theta}) - 2[\ell(\hat{\theta}) + \dot{\ell}(\hat{\theta})(\theta_0 - \hat{\theta}) + \frac{1}{2}(\theta_0 - \hat{\theta})^t \ddot{\ell}(\theta^*)(\theta_0 - \hat{\theta})] \\
&= -(\theta_0 - \hat{\theta})^t \ddot{\ell}(\theta^*)(\theta_0 - \hat{\theta}) \\
&= \sqrt{n}(\theta_0 - \hat{\theta})^t \left[-\frac{1}{n}\ddot{\ell}(\theta^*)\right] \sqrt{n}(\theta_0 - \hat{\theta}),
\end{aligned}$$

for some θ^* such that $|\theta^* - \theta_0| \le |\hat{\theta} - \theta_0|$. Then, by Taylor's expansion

$$\frac{1}{n}\ddot{\ell}(\theta^*) = \frac{1}{n}\ddot{\ell}(\hat{\theta}) + \frac{1}{n}\ell^{(3)}(\theta^{**})(\theta^* - \theta_0), \tag{1}$$

for some θ^{**} such that $|\theta^{**} - \theta_0| \le |\theta^* - \theta_0|$. Let $R_n = \frac{1}{n}\ell^{(3)}(\theta^{**})(\theta^* - \theta_0)$. Fix $\epsilon > 0$ and choose $k > 0$. Then,

$$\begin{aligned}
\mathbb{P}(|\theta_0 - \theta^*| > \frac{\epsilon}{k}) &\le \mathbb{P}(|\hat{\theta} - \theta_0| > \frac{\epsilon}{k}) \xrightarrow{P} 0, \qquad (2) \\
\mathbb{P}(|\frac{1}{n}\ell^{(3)}(\theta^{**})| > k) &\le \frac{1}{k}\mathbb{E}_{\theta_0}|\frac{1}{n}\ell^{(3)}(\theta^{**})| \\
&\le \frac{1}{k}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}M(X_1) \\
&= \frac{1}{k}\mathbb{E}_{\theta_0}M(X_1) \to 0 \text{ as } k \to \infty. \qquad (3)
\end{aligned}$$

For the inequality in (2), we used (R6), which implies the consistency of $\hat{\theta}$. Also, for the second inequality in (3), we used (R7), and the limit in the last equation holds because $\mathbb{E}_{\theta_0}M(X_1) < \infty$ by (R7). Since

$$\mathbb{P}(|R_n| > \epsilon) \le \mathbb{P}(|\frac{1}{n}\ell^{(3)}(\theta^{**})| > k) + \mathbb{P}(|\theta_0 - \theta^*| > \epsilon/k),$$

(2) and (3) imply $R_n = o_p(1)$ by choosing sufficiently large k. Recall that $\frac{1}{n}\ddot{\ell}(\theta_0) \xrightarrow{P} -I(\theta_0)$ by weak law of large numbers (WLLN), and so $\frac{1}{n}\ddot{\ell}(\theta_0) = -I(\theta_0) + o_p(1)$. Thus, this together with $|R_n| = o_p(1)$ implies that

$$-\frac{1}{n}\ddot{\ell}(\theta^*) = I(\theta_0) + o_p(1) \tag{4}$$

by (1). Therefore,

$$\begin{aligned}
2(\ell(\hat{\theta}) - \ell(\theta_0)) &= \sqrt{n}(\theta_0 - \hat{\theta})^t \left[-\frac{1}{n}\ddot{\ell}(\theta^*)\right] \sqrt{n}(\theta_0 - \hat{\theta}) \\
&= \sqrt{n}(\theta_0 - \hat{\theta})^t \left[I(\theta_0) + o_p(1)\right] \sqrt{n}(\theta_0 - \hat{\theta}) \\
&= \sqrt{n}(\theta_0 - \hat{\theta})^t I(\theta_0)\sqrt{n}(\theta_0 - \hat{\theta}) + \sqrt{n}(\theta_0 - \hat{\theta})o_p(1)\sqrt{n}(\theta_0 - \hat{\theta}).
\end{aligned}$$

Recall that $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1})$ by (R0)-(R7), and so $\sqrt{n}(\hat{\theta}-\theta_0) = O_p(1)$. Hence,

$$\sqrt{n}(\theta_0-\hat{\theta})o_p(1)\sqrt{n}(\theta_0-\hat{\theta}) = o_p(1). \tag{5}$$

Also,

$$\sqrt{n}(\theta_0-\hat{\theta})^t I(\theta_0)\sqrt{n}(\theta_0-\hat{\theta}) = [I(\theta_0)^{\frac{1}{2}}\sqrt{n}(\theta_0-\hat{\theta})]^t I(\theta_0)^{\frac{1}{2}}\sqrt{n}(\theta_0-\hat{\theta})$$

and

$$I(\theta_0)^{\frac{1}{2}}\sqrt{n}(\theta_0-\hat{\theta}) \xrightarrow{d} N(0, I_k),$$

by $\Delta-$method. Since the function $g : \mathbb{R}^k \to \mathbb{R}$ defined by $g(\mathbf{u}) = \mathbf{u}^t\mathbf{u}$ is twice continuously differentiable, by continuous mappng theorem,

$$g(I(\theta_0)^{\frac{1}{2}}\sqrt{n}(\theta_0-\hat{\theta})) = \sqrt{n}(\theta_0-\hat{\theta})^t I(\theta_0)\sqrt{n}(\theta_0-\hat{\theta}) \xrightarrow{d} \chi^2(k), \tag{6}$$

which proves Wald test (**(b) of Theorem 7.3.1**). Furthermore, (5) and (6) imply

$$2(\ell(\hat{\theta})-\ell(\theta_0)) \xrightarrow{d} \chi^2(k)$$

by Slutsky's theorem, which proves Wilk's phenomenon (**(a) of Theorem 7.3.1**). Hence, if one approximates LRT by Wilk's phenonmenon, the rejection rule is given by

$$\text{Reject } H_0 \text{ if } 2(\ell(\hat{\theta})-\ell(\theta_0)) > \chi^2_\alpha(k).$$

Similarly, if one conducts Wald test, the rejection rule is given by

$$\text{Reject } H_0 \text{ if } \sqrt{n}(\theta_0-\hat{\theta})^t I(\theta_0)\sqrt{n}(\theta_0-\hat{\theta}) > \chi^2_\alpha(k).$$

(**Proof of (c) of Theorem 7.3.1**) By Taylor's expansion,

$$\begin{aligned}\dot{\ell}(\theta_0) &= \dot{\ell}(\hat{\theta}) + \ddot{\ell}(\hat{\theta})(\theta_0-\hat{\theta}) + \frac{1}{2}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)\\ &= \ddot{\ell}(\hat{\theta})(\theta_0-\hat{\theta}) + \frac{1}{2}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0) \end{aligned} \tag{7}$$

for some θ^* such that $|\theta^*-\theta_0| \le |\hat{\theta}-\theta_0|$. The second equality holds because $\dot{\ell}(\hat{\theta}) = 0$ by (5). By (7),

$$\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) = \left(-\frac{1}{n}\ddot{\ell}(\hat{\theta})\right)\sqrt{n}(\hat{\theta}-\theta_0) + \frac{1}{2\sqrt{n}}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)$$

and so

$$\sqrt{n}(\hat{\theta}-\theta_0) = \left(-\frac{1}{n}\ddot{\ell}(\hat{\theta})\right)^{-1}\left(\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) - \frac{1}{2\sqrt{n}}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)\right). \tag{8}$$

Let $Q_n = \frac{1}{2\sqrt{n}}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)$. Fix $\epsilon > 0$ and take $k > 0$. Then,

$$\begin{aligned}\mathbb{P}(|Q_n| > \epsilon) &\le \mathbb{P}(|\frac{1}{2\sqrt{n}}(\hat{\theta}-\theta_0)^t\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)| > \epsilon)\\ &\le \mathbb{P}(|\sqrt{n}(\hat{\theta}-\theta_0)| > k) + \mathbb{P}(|\frac{1}{2n}\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)| > \epsilon/k)\\ &= \mathbb{P}(|\sqrt{n}(\hat{\theta}-\theta_0)| > k) + \mathbb{P}(|\frac{1}{n}\ell^{(3)}(\theta^*)(\hat{\theta}-\theta_0)| > 2\epsilon/k)\\ &= \mathbb{P}(|\sqrt{n}(\hat{\theta}-\theta_0)| > k) + \mathbb{P}(|\frac{1}{n}\ell^{(3)}(\theta^*)| > k) + \mathbb{P}(|\hat{\theta}-\theta_0| > 2\epsilon/k^2). \end{aligned} \tag{9}$$

As in the proof of $R_n = o_p(1)$ in the proof of (a) and (b) of Theorem 7.3.1, by taking sufficiently large k, $\mathbb{P}(|\frac{1}{n}\ell^{(3)}(\theta^*)| > k)$ can be made suffuciently small. Also, by (R5), $\mathbb{P}(|\hat{\theta}-\theta_0| > 2\epsilon/k^2)$ tends to 0 as $n \to \infty$, as $\hat{\theta}$ is a consistent estimator of θ_0. Since $\sqrt{n}(\hat{\theta} - \theta_0)$ converges to $Z \sim N(0, I(\theta_0)^{-1})$ in distribution, $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$. Hence, $\mathbb{P}(|\sqrt{n}(\hat{\theta} - \theta_0)| > k)$ can be made sufficiently small if k is chosen to be sufficiently large. Therefore, by (9), $Q_n = o_p(1)$.

Also, as

$$-\frac{1}{n}\ddot{\ell}(\hat{\theta}) = -\frac{1}{n}\ddot{\ell}(\theta_0) - \frac{1}{n}\ell^{(3)}(\theta^{**})(\hat{\theta} - \theta_0),$$

$$\frac{1}{n}\ell^{(3)}(\theta^{**}) = o_p(1),$$

using the similar argument in the proof of $R_n = o_p(1)$ in the proof of (a) and (b) of Theorem 7.3.1. Since $\hat{\theta}$ is a consistent estimator of θ_0 by (R5), $\hat{\theta} - \theta_0 = o_p(1)$. Also, as $-\frac{1}{n}\ddot{\ell}(\theta_0)$ converges in probability to $I(\theta_0)$ by WLLN, one can see that

$$-\frac{1}{n}\ddot{\ell}(\hat{\theta}) = I(\theta_0) + o_p(1). \tag{10}$$

By (8), (9) and (10),

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) &= (I(\theta_0) + o_p(1))^{-1}(\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1)) \\ &= (I(\theta_0)^{-1} + o_p(1))(\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1)) \\ &= I(\theta_0)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1).\end{aligned}$$

Recall that under (R0)-(R7),

$$\sqrt{n}(\hat{\theta} - \theta_0)^t I(\theta_0)\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

Substituting $\sqrt{n}(\hat{\theta} - \theta_0) = I(\theta_0)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1)$ gives

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0)^t I(\theta_0)\sqrt{n}(\hat{\theta} - \theta_0) &= \left(I(\theta_0)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1)\right)^t I(\theta_0)\left(I(\theta_0)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\theta_0) + o_p(1)\right) \\ &= \frac{1}{n}\dot{\ell}(\theta_0)^t I(\theta_0)^{-1}\dot{\ell}(\theta_0) + o_p(1).\end{aligned}$$

Again, by Slutsky's theorem,

$$\frac{1}{n}\dot{\ell}(\theta_0)^t I(\theta_0)^{-1}\dot{\ell}(\theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

So if one conducts Rao test, the rejection rule is given by

$$\text{Reject } H_0 \text{ if } \frac{1}{n}\dot{\ell}(\theta_0)^t I(\theta_0)^{-1}\dot{\ell}(\theta_0) > \chi^2_\alpha(k).$$