Eric Slud Spring 2009, corrected 2014

Handout on Wald, Score and LR Test Asymptotics

As described in class, we consider parametric statistical problems with *iid* data $X_1, X_2, \ldots, X_n$ for large n, where $X_i \sim f(x, \vartheta)$ and $\vartheta \in int(\Theta)$, $\Theta \subset \mathbf{R}^p$, $p > 1$, and we consider estimates and hypothesis tests related to the question of whether the governing parameter vector ϑ actually lies in the lower-dimensional parameter subset $\Theta_0 \equiv \Theta \cap (\mathbf{R}^q \times \mathbf{0}_{p-q})$, i.e., whether the final $p-q$ coordinates of ϑ are 0, where $q < p$.

Denote by $\hat{\vartheta}$ the usual ML estimator, and by $\hat{\vartheta}_{(r)}$ the so-called *restricted Maximum Likelihood Estimator*,

$$\hat{\vartheta}_{(r)} \;\equiv\; \text{argmax}_{\tau \in \Theta_0} \sum_{i=1}^{n} \log f(X_i, \tau)$$

We assume that $\hat{\vartheta}$ is consistent for the true parameter vector ϑ (whether $\vartheta \in \Theta_0$ or not) and that whenever $\vartheta \in \Theta_0$ the restricted MLE $\hat{\vartheta}_{(r)}$ is consistent for ϑ.

We further assume that all of the large-sample regularity conditions used in MLE theory are valid for the densities $f(x, \vartheta)$, both in the p-dimensional parameter space Θ and its q-dimensional subspace Θ_0. Among these conditions (or consequences of them), the information matrix

$$I(\vartheta) \;=\; -E_\vartheta(\nabla^{\otimes 2} \log f(X_1, \vartheta)) \;=\; E_\vartheta((\nabla \log f(X_1, \vartheta))^{\otimes 2}) \;\equiv\; \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

is positive definite and continuous as a function of ϑ. In the final term of the last display, we have block-decomposed the symmetric $p \times p$ matrix $I(\vartheta)$ into an upper-left $q \times q$ block I_{11} which is the information matrix for the parameter subvector $\vartheta_{sub} = (\vartheta_1, \ldots, \vartheta_q)$ when $\vartheta \in \Theta_0$, along with $q \times (p-q)$ block I_{12}, and $I_{21} = I_{21}^{tr}$, and $(p-q) \times (p-q)$ lower block I_{22}.

Define the further notations

$$U(\vartheta) \;\equiv\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \log f(X_i, \vartheta) \;\equiv\; \begin{pmatrix} U^{(1)}(\vartheta) \\ U^{(2)}(\vartheta) \end{pmatrix}$$

where $U^{(1)}(\vartheta)$ is the vector of the first q derivatives of log-likelihood and $U^{(2)}(\vartheta)$ the vector of the last $p-q$ derivatives. Moreover, let $\hat{\vartheta}_{sub}$ be the vector of the first q coordinates of $\hat{\vartheta}_{(r)}$ (i.e., the nonzero ones), and ϑ_{tst} with corresponding estimator $\hat{\vartheta}_{tst}$ be the sub-vector of the last $p-q$ coordinates respectively of ϑ and $\hat{\vartheta}$.

Then the consequences of asymptotic MLE theory that we use are the following, holding under P_ϑ as $n \to \infty$:

$$\sqrt{n}\,(\hat{\vartheta} - \vartheta) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(\mathbf{0},\,(I(\vartheta))^{-1}) \qquad \forall \quad \vartheta \in \Theta \tag{1}$$

$$U(\vartheta) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(\mathbf{0},\,I(\vartheta)) \qquad \forall \quad \vartheta \in \Theta \tag{2}$$

$$\sqrt{n}\,(\hat{\vartheta}_{tst} - \vartheta_{tst}) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(\mathbf{0},\,((I(\vartheta))^{-1})_{22})\,) \qquad \forall \quad \vartheta \in \Theta \tag{3}$$

$$U^{(2)}(\vartheta) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(\mathbf{0},\,I_{22}) \qquad \forall \quad \vartheta \in \Theta_0 \tag{4}$$

$$\sqrt{n}\,(\hat{\vartheta} - \vartheta) \;=\; (I(\vartheta))^{-1}\,U(\vartheta) \,+\, o_P(1) \qquad \forall \quad \vartheta \in \Theta \tag{5}$$

$$\sqrt{n}\,(\hat{\vartheta}_{sub} - \vartheta_{sub}) \;=\; (I_{11})^{-1}\,U^{(1)}(\vartheta) \,+\, o_P(1) \qquad \forall \quad \vartheta \in \Theta_0 \tag{6}$$

The other ingredient we need for our manipulations is the linear-algebraic identity

$$\left.\begin{matrix}(2,2)\\(2,1)\end{matrix}\right\} \text{ blocks of } \quad (I(\vartheta))^{-1} \quad = \left\{\begin{matrix}(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}\\ -\,(I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}\,I_{21}\,I_{11}^{-1}\end{matrix}\right.$$

As an immediate consequence of this identity and (5), we have under $\vartheta \in \Theta$,

$$\sqrt{n}\,(\hat{\vartheta}_{tst} - \vartheta_{tst}) \;=\; (I_{22} - I_{21}I_{11}^{-1}I_{12})^{-1}\left(U^{(2)}(\vartheta) - I_{21}\,I_{11}^{-1}\,U^{(1)}(\vartheta)\right) + o_P(1) \tag{7}$$

Next, we check by Taylor-expanding $U^{(2)}(\hat{\vartheta}_{(r)})$ around $\vartheta \in \Theta_0$ using

$$n^{-1/2}\,\nabla_\vartheta^{tr}\,U(\vartheta_n) \;=\; n^{-1}\,\nabla^{\otimes 2}\sum_{i=1}^{n}\log f(X_i, \vartheta_n) \;=\; -I(\vartheta) \,+\, o_P(1)$$

when $\vartheta_n \to \vartheta$ as $n \to \infty$, that

$$U^{(2)}(\hat{\vartheta}_{(r)}) \;=\; U^{(2)}(\vartheta) \,-\, \left(I_{21}\,\middle|\,I_{22}\right) n^{1/2}\begin{pmatrix}\hat{\vartheta}_{sub} - \vartheta_{sub}\\ \mathbf{0}\end{pmatrix} \,+\, o_P(1)$$

$$= U^{(2)}(\vartheta) \;-\; I_{21}\, I_{11}^{-1} U^{(1)}(\vartheta) \;+\; o_P(1)$$

where we have used (6) in the last line. Finally, the right hand side of the last equation is obviously multivariate normal with mean $\mathbf{0}$ and variance

$$I_{22} - I_{21}\, I_{11}^{-1}\, I_{12} \;-\; I_{21}\,(I_{21}\, I_{11}^{-1})^{tr} \;+\; I_{21}\, I_{11}^{-1}\, I_{11}\, I_{11}^{-1}\, I_{12}$$

$$= \; I_{22} \;-\; I_{21}\, I_{11}^{-1}\, I_{12} \;=\; \Big(((I(\vartheta))^{-1})_{22}\Big)^{-1}$$

Note that this argument together with (7) shows that

$$(I_{22}{-}I_{21}\, I_{11}^{-1}\, I_{12})^{1/2}\, \sqrt{n}\, (\hat{\vartheta}_{tst}{-}\vartheta_{tst}) \;=\; (I_{22}{-}I_{21}\, I_{11}^{-1}\, I_{12})^{-1/2}\, U^{(2)}(\hat{\vartheta}_{(r)}) + o_P(\frac{1}{\sqrt{n}}) \tag{8}$$

Our objective has been to demonstrate the equivalence of the three hypothesis tests (one-sided, versus alternatives $H_A : \vartheta_p > 0$) with approximate large-sample significance level α of $H_0 : \vartheta_p = 0$: Wald, Score, and (Generalized) Likelihood Ratio, in this setting, when $p = q + 1$.

$$\text{Wald Test rejects when} \quad \sqrt{n}\, (\hat{\vartheta}_p{-}\vartheta_p) \;>\; z_\alpha \cdot \sqrt{(I(\vartheta)^{-1})_{22}} \;=\; \frac{z_\alpha}{\sqrt{I_{22} - I_{21} I_{11}^{-1} I_{12}}} \tag{9}$$

$$\text{Score Test rejects when} \qquad U^{(2)}(\hat{\vartheta}_r) \;>\; z_\alpha \;\cdot\; \sqrt{I_{22} - I_{21} I_{11}^{-1} I_{12}} \tag{10}$$

$$\text{LR Test rejects when} \qquad \sum_{i=1}^{n} \log\left(\frac{f(X_i, \hat{\vartheta})}{f(X_i, \hat{\vartheta}_r)}\right)\, I_{[\hat{\vartheta}_p > 0]} \;\geq\; \frac{1}{2}\, z_\alpha^2 \tag{11}$$

When $p \;=\; q + 1$, the one-sided test (9) is simply the test based on the standardized estimator of ϑ_p. On the other hand, (8) shows that the standardized Score Statistic is asymptotically equivalent (up to $o_P(n^{-1/2})$ remainders) to the Wald statistic. It remains only to complete the verification of asymptotic equivalence for the Likelihood Ratio Statistic, which we do in class in the course of proving the Wilks Theorem on asymptotic chi-square distribution of LR tests.