STA3000: Note on asymptotic theory for likelihood

Assume we have a sample $Y = Y_1, \ldots, Y_n$, where the Y_i are independent, identically distributed with density $f(y_i;\theta) = f(y_i;\psi,\lambda)$, where $\theta = (\theta_1, \ldots \theta_k) = (\psi_1, \ldots, \psi_q, \lambda_1, \ldots, \lambda_{k-q})$. To fix the notation we recall:

log-likelihood function	$\ell(\theta) = \ell(\theta; y)$	$O_p(n)$
maximum likelihood estimate	$\hat\theta = \hat\theta(y) = \arg\sup_\theta L(\theta)$	$\theta + O_p(n^{-1/2})$
score function	$U(\theta) = \ell'(\theta) = \sum U_i(\theta)$	$O_p(n^{1/2})$
observed information function	$j(\theta) = \{-\ell''(\theta)\} = -\ell(\theta; Y)$	$O_p(n)$
observed (Fisher) information	$j(\hat\theta)$	
expected (Fisher) information	$i(\theta) = \frac{1}{n}E_\theta\{U(\theta)U(\theta)^T\}$	$O(1)$
profile loglikelihood	$\ell_p(\psi) = \ell(\psi, \hat\lambda_\psi)$	
profile information	$j_p(\psi) = -\ell_p''(\psi)$	
partitioned information matrix	$i(\theta) = \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix} \quad j(\theta) = \begin{pmatrix} j_{\psi\psi} & j_{\psi\lambda} \\ j_{\lambda\psi} & j_{\lambda\lambda} \end{pmatrix}$	
partitioned inverse	$i^{-1}(\theta) = \begin{pmatrix} i^{\psi\psi} & i^{\psi\lambda} \\ i^{\lambda\psi} & i^{\lambda\lambda} \end{pmatrix} \quad j(\theta) = \begin{pmatrix} j^{\psi\psi} & j^{\psi\lambda} \\ j^{\lambda\psi} & j^{\lambda\lambda} \end{pmatrix}$	

First order asymptotic theory

1. θ is a scalar

$\frac{1}{\sqrt{n}}U(\theta)/i^{1/2}(\theta) \xrightarrow{d} N(0,1)$	by the central limit theorem
standardized score statistic	$r_u = U(\theta)/j^{1/2}(\hat\theta) \xrightarrow{d} N(0,1)$
$\sqrt{n}(\hat\theta - \theta)i^{1/2}(\theta) =$	$\dfrac{1}{\sqrt{n}}\dfrac{U(\theta)}{i^{1/2}(\theta)}\{1 + o_p(1)\}$
standardized m.l.e.	$r_e = r_p = (\hat\theta - \theta)j^{1/2}(\hat\theta) \xrightarrow{d} N(0,1)$
(log) likelihood ratio statistic	$w(\theta) = 2\{\ell(\hat\theta) - \ell(\theta)\} = (\hat\theta - \theta)^2 n i(\theta)\{1 + o_p(1)\}$
	$w(\theta) \xrightarrow{d} \chi^2_1$
likelihood root	$r(\theta) = \text{sign}(\theta - \hat\theta)\{w(\theta)\}^{1/2}$
	$r(\theta) \xrightarrow{d} N(0,1)$

2. θ a vector of length k

$\frac{1}{\sqrt{n}}\{U(\theta)\} \xrightarrow{d} N_k(0, i(\theta))$	by the central limit theorem
standardized score statistic	$w_u = U(\theta)^T\{ni(\theta)\}^{-1}U(\theta)$
$\sqrt{n}(\hat\theta - \theta) =$	$\frac{1}{\sqrt{n}}U(\theta)i^{-1}(\theta)\{1 + o_p(1)\}$
standardized m.l.e.	$w_e = w_p = (\hat\theta - \theta)^T n i(\theta)(\hat\theta - \theta)$
likelihood ratio statistic	$w = 2\{\ell(\hat\theta) - \ell(\theta)\} = (\hat\theta - \theta)^T n i(\theta)(\hat\theta - \theta)\{1 + o_p(1)\}$
	$w(\theta) \xrightarrow{d} \chi^2_k$

3. $\theta = (\psi, \lambda) : \psi \in R^q$

We want to use the results in 2. to conclude the following:

$$w_u(\psi) = U(\psi, \hat{\lambda}_\psi)^T \{n^{-1} i^{\psi\psi}(\psi, \hat{\lambda}_\psi)\} U(\psi, \hat{\lambda}_\psi) \overset{d}{\to} \chi^2_q \quad (1)$$

$$w_e(\psi) = w_p(\psi) = (\hat{\psi} - \psi)\{n^{-1} i^{\psi\psi}(\psi, \hat{\lambda}_\psi)\}^{-1}(\hat{\psi} - \psi) \overset{d}{\to} \chi^2_q \quad (2)$$

$$w(\psi) = 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \overset{d}{\to} \chi^2_q \quad (3)$$

The first of these is given in (3.35) of BNC, the second in (3.33); note that their use of $_p$ for the standardized m.l.e. is to emphasize that it is not parametrization invariant. However, they also use $_P$ for the profile likelihood, so this is all a bit confusing. I prefer to use w_e for the standardized m.l.e.

From part 2., we have

$$\sqrt{n} \begin{pmatrix} \hat{\psi} - \psi \\ \hat{\lambda} - \lambda \end{pmatrix} \overset{d}{\to} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i^{\psi\psi}(\theta) & i^{\psi\lambda}(\theta) \\ i^{\lambda\psi}(\theta) & i^{\lambda\lambda}(\theta) \end{pmatrix} \right] \quad (4)$$

from which we have

$$\sqrt{n}(\hat{\psi} - \psi) \overset{d}{\to} N(0, i^{\psi\psi}(\theta)) \quad (5)$$

and hence

$$n(\hat{\psi} - \psi)^T \{i^{\psi\psi}(\psi, \lambda)\}^{-1}(\hat{\psi} - \psi) \overset{d}{\to} \chi^2_q. \quad (6)$$

Result (2) follows on verifying that $i(\psi, \hat{\lambda}_\psi) = i(\psi, \lambda)\{1 + o_p(1)\}$. We can also show the same result for $i(\hat{\psi}, \hat{\lambda})$, and for $n^{-1} j(\psi, \hat{\lambda}_\psi)$ and for $n^{-1} j(\hat{\psi}, \hat{\lambda})$. Based on some higher order asymptotic arguments, it seems likely that the best variance estimator to use is $j(\hat{\psi}, \hat{\lambda})$, i.e.

$$(\hat{\psi} - \psi)^T \{j_p(\hat{\psi})\}(\hat{\psi} - \psi) \overset{\cdot}{\sim} \chi^2_q,$$

where j_p is the information from the profile loglikelihood function.

For result (1), we start from

$$\sqrt{n} \begin{pmatrix} U_\psi(\psi, \lambda) \\ U_\lambda(\psi, \lambda) \end{pmatrix} \overset{d}{\to} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} i_{\psi\psi}(\theta) & i_{\psi\lambda}(\theta) \\ i_{\lambda\psi}(\theta) & i_{\lambda\lambda}(\theta) \end{pmatrix} \right] \quad (7)$$

from which we have

$$\frac{1}{n} U_\psi(\psi, \lambda)^T \{i_{\psi\psi}(\psi, \lambda)\}^{-1} U_\psi(\psi, \lambda) \overset{d}{\to} \chi^2_q. \quad (8)$$

The mistake I made in class was to toss in $\hat{\lambda}_\psi$ for λ in *both* U and i: this is wrong because

$$\begin{aligned} U_\psi(\psi, \hat{\lambda}_\psi) &= U_\psi(\psi, \lambda) + \ell_{\psi\lambda}(\psi, \lambda)(\hat{\lambda}_\psi - \lambda) + O_p(1) && (9) \\ &= U_\psi(\psi, \lambda) - j_{\psi\lambda}(\psi, \lambda)(\hat{\lambda}_\psi - \lambda) + O_p(1). && (10) \end{aligned}$$

Both the first two terms are $O_p(\sqrt{n})$.

Now, in the model with ψ fixed, we have again from part 2. that

$$\sqrt{n}(\hat{\lambda}_\psi - \lambda) = i^{\lambda\lambda}(\psi, \lambda) U_\lambda(\psi, \lambda)\{1 + o_p(1)\}; \tag{11}$$

using this in (1) gives

$$\begin{aligned} w_u(\psi) &= n^{-1}\{U_\psi(\theta) - i_{\psi\lambda}(\theta) i^{\lambda\lambda}(\theta) U_\lambda(\theta)\}^T i^{\psi\psi}(\theta) \\ &\qquad \{U_\psi(\theta) - i_{\psi\lambda}(\theta) i^{\lambda\lambda}(\theta) U_\lambda(\theta)\}\{1 + o_p(1)\}. \end{aligned}$$

Now we're nearly done: we need only recall that if

$$\sqrt{n}\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}\right) \tag{12}$$

that $Y - \Sigma_{YX}\Sigma_{XX}^{-1}X$ and X are uncorrelated (hence independent), and the first has covariance matrix $\Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$.