STA3000: Likelihood asymptotics with nuisance parameters

Assume we have a sample $Y = (Y_1, \ldots, Y_n)$, where the Y_i are independent, identically distributed with density $f(y_i; \theta)$. Refer to an earlier handout for the definitions and orders of magnitude of the score function, maximum likelihood estimate, observed and expected Fisher information. Also there we give the first order theory for θ in the case that θ is a vector of length k, as well as the special case $k = 1$. The vector version results are repeated here:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\{U(\theta)\} &\xrightarrow{d} N_k(0, i_1(\theta)) & (1)\\
\sqrt{n}(\hat\theta - \theta) &= \frac{1}{\sqrt{n}} i_1^{-1}(\theta) U(\theta)\{1 + o_p(1)\}, & (2)\\
2\{\ell(\hat\theta) - \ell(\theta)\} &= (\hat\theta - \theta)^T i(\theta)(\hat\theta - \theta)\{1 + o_p(1)\} & (3)
\end{aligned}$$

from which we have the approximations

$$\begin{aligned}
w_u(\theta) = U(\theta)^T \{i(\theta)\}^{-1} U(\theta) &\;\dot\sim\; \chi^2_k, & (4)\\
w_e(\theta) = (\hat\theta - \theta)^T i(\theta)(\hat\theta - \theta) &\;\dot\sim\; \chi^2_k, & (5)\\
w(\theta) = 2\{\ell(\hat\theta) - \ell(\theta)\} &\;\dot\sim\; \chi^2_k. & (6)
\end{aligned}$$

Now assume that $\theta = (\theta_1, \ldots \theta_k)^T = (\psi_1, \ldots, \psi_q, \lambda_1, \ldots, \lambda_{k-q})^T$. We partition the information matrices compatibly and write

$$U(\theta) = \begin{pmatrix} U_\psi(\theta) \\ U_\lambda(\theta) \end{pmatrix},$$

$$i(\theta) = \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix} \quad j(\theta) = \begin{pmatrix} j_{\psi\psi} & j_{\psi\lambda} \\ j_{\lambda\psi} & j_{\lambda\lambda} \end{pmatrix}$$

and

$$i^{-1}(\theta) = \begin{pmatrix} i^{\psi\psi} & i^{\psi\lambda} \\ i^{\lambda\psi} & i^{\lambda\lambda} \end{pmatrix} \quad j^{-1}(\theta) = \begin{pmatrix} j^{\psi\psi} & j^{\psi\lambda} \\ j^{\lambda\psi} & j^{\lambda\lambda} \end{pmatrix}.$$

The constrained maximum likelihood estimator of λ is denoted by $\hat\lambda_\psi$, which in regular models satisfies $U_\lambda(\psi, \hat\lambda_\psi) = 0$.

Note that

$$i^{\psi\psi}(\theta) = \{i_{\psi\psi}(\theta) - i_{\psi\lambda}(\theta) i_{\lambda\lambda}^{-1}(\theta) i_{\lambda\psi}(\theta)\}^{-1}, \qquad (7)$$

using the formula for the determinant of a partitioned matrix. A similar result holds for j.

The profile log-likelihood function is $\ell_{\mathrm{P}}(\psi) = \ell(\psi, \hat\lambda_\psi)$, and the (observed) profile information function is $j_{\mathrm{P}}(\psi) = -\ell''_{\mathrm{P}}(\psi)$, a $q \times q$ matrix.

We will use results (1) – (3) to conclude the following:

$$\begin{aligned} w_u(\psi) = U_\psi(\psi,\hat\lambda_\psi)^T\{i^{\psi\psi}(\psi,\hat\lambda_\psi)\}U_\psi(\psi,\hat\lambda_\psi) &\ \dot\sim\ \chi^2_q && (8)\\ w_e(\psi) = (\hat\psi-\psi)\{i^{\psi\psi}(\hat\psi,\hat\lambda)\}^{-1}(\hat\psi-\psi) &\ \dot\sim\ \chi^2_q && (9)\\ w(\psi) = 2\{\ell(\hat\psi,\hat\lambda)-\ell(\psi,\hat\lambda_\psi)\} = 2\{\ell_{\rm P}(\hat\psi)-\ell_{\rm P}(\psi)\} &\ \dot\sim\ \chi^2_q; && (10) \end{aligned}$$

see (52), (54) and (56) in CH §9.3.

From (2) we have

$$\sqrt{n}\begin{pmatrix}\hat\psi-\psi\\ \hat\lambda-\lambda\end{pmatrix}\xrightarrow{d} N\left[\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix} i_1^{\psi\psi}(\theta) & i_1^{\psi\lambda}(\theta)\\ i_1^{\lambda\psi}(\theta) & i_1^{\lambda\lambda}(\theta)\end{pmatrix}\right] \tag{11}$$

from which we have

$$\sqrt{n}(\hat\psi-\psi)\xrightarrow{d} N(0,i_1^{\psi\psi}(\theta)) \tag{12}$$

and hence

$$(\hat\psi-\psi)^T\{i^{\psi\psi}(\psi,\lambda)\}^{-1}(\hat\psi-\psi)\xrightarrow{d}\chi^2_q. \tag{13}$$

Result (9) follows on verifying that $i(\hat\psi,\hat\lambda)=i(\psi,\lambda)\{1+o_p(1)\}$. We can also show the same result for $i(\psi,\hat\lambda_\psi)$, and for $j(\psi,\hat\lambda_\psi)$, and for $j(\hat\psi,\hat\lambda)$.

For result (8), we start from (1) to get

$$U_\psi(\psi,\lambda)^T\{i_{\psi\psi}(\psi,\lambda)\}^{-1}U_\psi(\psi,\lambda)\xrightarrow{d}\chi^2_q. \tag{14}$$

We now write

$$\begin{aligned} U_\psi(\psi,\hat\lambda_\psi) &= U_\psi(\psi,\lambda)+\ell_{\psi\lambda}(\psi,\lambda)(\hat\lambda_\psi-\lambda)+O_p(1) && (15)\\ &= U_\psi(\psi,\lambda)-j_{\psi\lambda}(\psi,\lambda)(\hat\lambda_\psi-\lambda)+O_p(1). && (16) \end{aligned}$$

It is important to note that both the first two terms are $O_p(\sqrt{n})$.

In the model with ψ fixed, we have again from (2) that

$$(\hat\lambda_\psi-\lambda)=i^{-1}_{\lambda\lambda}(\psi,\lambda)U_\lambda(\psi,\lambda)\{1+o_p(1)\}; \tag{17}$$

using this in (16), and $j_{\psi\lambda}(\psi,\lambda)=i_{\psi\lambda}(\psi,\lambda)\{1+o_p(1)\}$, as well as $i^{\psi\psi}(\psi,\hat\lambda_\psi)=i^{\psi\psi}(\psi,\lambda)\{1+o_p(1)\}$ gives

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

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

$$\begin{pmatrix}Y\\X\end{pmatrix}\sim N\left(\begin{matrix}\mu_Y\\ \mu_X\end{matrix},\begin{pmatrix}\Sigma_{YY} & \Sigma_{YX}\\ \Sigma_{XY} & \Sigma_{XX}\end{pmatrix}\right) \tag{18}$$

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

To get (10), we write

$$\begin{aligned}
w(\psi) &= 2\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \lambda)\} - 2\{\ell(\psi, \hat{\lambda}_\psi) - \ell(\psi, \lambda)\} \\
&= U(\psi, \lambda)^T i^{-1}(\psi, \lambda) U(\psi, \lambda)\{1 + o_p(1)\} - U_\lambda(\psi, \lambda)^T i_{\lambda\lambda}^{-1}(\psi, \lambda) U_\lambda(\psi, \lambda)\{1 + o_p(1)\} \\
&= \{U_\psi(\theta) - i_{\psi\lambda}(\theta) i_{\lambda\lambda}^{-1}(\theta) U_\lambda(\theta)\}^T i^{\psi\psi}(\theta)\{U_\psi(\theta) - i_{\psi\lambda}(\theta) i_{\lambda\lambda}^{-1}(\theta) U_\lambda(\theta)\}\{1 + o_p(1)\}
\end{aligned}$$

and using the multivariate normal result above. The second equality follows from (3) and (2), and the formula for the inverse of a partitioned matrix.

There are several choices for the variance part w_u and w_e, as noted above: $i(\theta), i(\hat{\theta}), i(\hat{\theta}_\psi)$, and the versions based on observed information. I've used the versions defined in CH: $i(\hat{\theta})$ for the maximum likelihood estimator, and $i(\psi, \hat{\lambda}_\psi)$ for the score function. The score function is usually only used for a single value of ψ, typically something like 0, and used when the full MLE is too difficult to compute. Thus it is easier to use the 'null' version in the variance. Further analysis of the asymptotic results using higher order terms suggests, as in the scalar parameter case, that the observed Fisher information at the MLE is preferred. If we make this change, and in the special case that $k = 1$, we have 3 analogues to our scalar parameter pivotal quantities:

$$\begin{aligned}
r_{\mathrm{P}}(\psi) &= \operatorname{sign}(\hat{\psi} - \psi)\sqrt{[2\{\ell_{\mathrm{P}}(\hat{\psi}) - \ell_{\mathrm{P}}(\psi)\}]}, \\
r_e(\psi) &= (\hat{\psi} - \psi) j_{\mathrm{P}}^{1/2}(\hat{\psi}), \\
r_u(\psi) &= \ell'_{\mathrm{P}}(\psi) j_{\mathrm{P}}^{-1/2}(\hat{\psi}),
\end{aligned}$$

all approximately standard normal under the model. These approximations follow from the above results; note that

$$j_{\mathrm{P}}(\psi) = -\ell''_{\mathrm{P}}(\psi) = j_{\psi\psi}(\psi, \hat{\lambda}_\psi) - j_{\psi\lambda}(\psi, \hat{\lambda}_\psi) j_{\lambda\lambda}^{-1}(\psi, \hat{\lambda}_\psi) j_{\lambda\psi}(\psi, \hat{\lambda}_\psi) = \{j^{\psi\psi}(\psi, \hat{\lambda}_\psi)\}^{-1}.$$

References

[BNC] Barndorff-Nielsen & Cox (1994). *Inference and Asymptotics.* Ch. 3

[SM] Davison (2003). *Statistical Models* Ch. 4.4-4.6.

[CH] Cox & Hinkley (1974). *Theoretical Statistics.* Ch. 9.2, 9.3.