STA3000: Notation and definitions related to the likelihood function

Given a model for \( Y \) which assumes \( Y \) has a density \( f(y; \theta) \), \( \theta \in \Theta \subset \mathbb{R}^d \), we have the following definitions:

- **Observed likelihood function**
  \[ L(\theta; y) = c(y)f(y; \theta) \]

- **Log-likelihood function**
  \[ \ell(\theta; y) = \log\,L(\theta; y) = \log f(y; \theta) + a(y) \]

- **Score function**
  \[ u(\theta) = \frac{\partial \ell(\theta; y)}{\partial \theta} \]

- **Observed information function**
  \[ j(\theta) = -\frac{\partial^2 \ell(\theta; y)}{\partial \theta \partial \theta^T} \]

- **Expected information (in one observation)**
  \[ i(\theta) = E_\theta U(\theta)U(\theta)^T \] (called \( i_1(\theta) \) in CH)

When we have \( Y_i \) independent, identically distributed from \( f(y_i; \theta) \), then, denoting the observed sample \( y = (y_1, \ldots, y_n) \) we have:

- **Log-likelihood function**
  \[ \ell(\theta) = \ell(\theta; y) + a(y) \quad O_p(n) \]

- **Maximum likelihood estimate**
  \[ \hat{\theta} = \hat{\theta}(y) = \arg \sup_{\theta} \ell(\theta) \quad \theta + O_p(n^{-1/2}) \]

- **Score function**
  \[ U(\theta) = \ell'(\theta) = \sum U_i(\theta) = U_+(\theta) \quad O_p(n^{1/2}) \]

- **Observed information function**
  \[ j(\theta) = -\ell''(\theta) = -\ell(\theta; Y) \quad O_p(n) \]

- **Observed (Fisher) information**
  \[ j(\hat{\theta}) \]

- **Expected (Fisher) information**
  \[ i(\theta) = E_\theta \{U(\theta)U(\theta)^T\} = ni_1(\theta) \quad O(n), \]

where with the risk of some confusion we use the same notation. Sometimes the expected Fisher information is defined instead as \( i(\theta) = E_\theta \{-\partial U(\theta; Y)/\partial \theta^T\} \) (e.g. in BNC). In models for which we can interchange differentiation and integration in \( \int f(y; \theta)dy = 1 \), these are the same due to the Bartlett identities:

\[
\begin{align*}
E_\theta \{U(\theta)\} &= 0, \\
E_\theta \{U'(\theta)\} + E_\theta \{U^2(\theta)\} &= 0, \\
E_\theta \{U''(\theta)\} + 3E_\theta \{U(\theta)U'(\theta)\} + E_\theta \{U^3(\theta)\} &= 0,
\end{align*}
\]

and so on, where the result applies to vector \( \theta \), but as presented here is for scalar \( \theta \).

(In the vector setting the second derivative of \( U \) is a \( p \times p \times p \) array.)
First order asymptotic theory

The following results, to be derived later, are used for approximate inference based on the likelihood function:

1. \( \theta \) is a scalar
   \[
   \frac{1}{\sqrt{n}} U(\theta) / i_1^{1/2}(\theta) \xrightarrow{d} N(0, 1)
   \]
   standardized score statistic
   \[
   r_u = U(\theta) / j^{1/2}(\hat{\theta}) \xrightarrow{d} N(0, 1)
   \]
   \[
   \sqrt{n}(\hat{\theta} - \theta) i_1^{1/2}(\theta) = \frac{1}{\sqrt{n} i_1^{1/2}(\theta)} \{1 + o_p(1)\}
   \]
   standardized m.l.e.
   \[
   w(\theta) = 2\{\ell(\hat{\theta}) - \ell(\theta)\} = (\hat{\theta} - \theta)^2 i(\theta) \{1 + o_p(1)\}
   \]
   (log) likelihood ratio statistic
   \[
   w(\theta) \xrightarrow{d} \chi^2
   \]
   likelihood root

2. \( \theta \) a vector of length \( k \)
   \[
   \frac{1}{\sqrt{n}} \{U(\theta)\} \xrightarrow{d} N_k(0, i_1(\theta))
   \]
   standardized score statistic
   \[
   w_u = U(\theta)^T i(\theta)^{-1} U(\theta)
   \]
   \[
   \sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n} i_1(\theta)} U(\theta) \{1 + o_p(1)\}
   \]
   standardized m.l.e.
   \[
   w_e = (\hat{\theta} - \theta)^T i(\theta)(\hat{\theta} - \theta)
   \]
   likelihood ratio statistic
   \[
   w = 2\{\ell(\hat{\theta}) - \ell(\theta)\} = (\hat{\theta} - \theta)^T i(\theta)(\hat{\theta} - \theta) \{1 + o_p(1)\}
   \]
   \[
   w(\theta) \xrightarrow{d} \chi^2_k
   \]