Fisher Information Matrix and Multivariate Cramér-Rao Inequality

Multivariate Analysis
Stat 541
Dr Scott

September 28, 2017

Definitions

- log-likelihood: ($\mathbf{X}$ is the data matrix)

$$\ell(\boldsymbol{\theta}|\mathbf{X}) = \log L(\boldsymbol{\theta}|\mathbf{X}) = \log f(\mathbf{X}|\boldsymbol{\theta})\,.$$

- score function:

$$\mathbf{s} = \mathbf{s}(\mathbf{X}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\mathbf{X}, \boldsymbol{\theta})\,.$$

- Fisher information matrix:

$$\mathbf{F} = cov(\mathbf{s}, \mathbf{s}) = E(\mathbf{s}\,\mathbf{s}^T) = -E\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}\right) \quad \text{(will show)}\,.$$

Prop: $\mathbf{t} = \mathbf{t}(\mathbf{X}, \boldsymbol{\theta}) \Longrightarrow E(\mathbf{s}\mathbf{t}^T) = \frac{\partial}{\partial \boldsymbol{\theta}} E(\mathbf{t}^T) - E\left(\frac{\partial \mathbf{t}^T}{\partial \boldsymbol{\theta}}\right)$.

Pf: Now $E(\mathbf{t}^T) = \int \mathbf{t}^T f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \int L \cdot \mathbf{t}^T \, d\mathbf{X}$

$$\begin{aligned}
\therefore \quad \frac{\partial}{\partial \boldsymbol{\theta}} E(\mathbf{t}^T) &= \int L \cdot \frac{\partial \mathbf{t}^T}{\partial \boldsymbol{\theta}} \, d\mathbf{X} + \int \frac{\partial L}{\partial \boldsymbol{\theta}} \mathbf{t}^T d\mathbf{X} \\
&= E\left(\frac{\partial \mathbf{t}^T}{\partial \boldsymbol{\theta}}\right) + \int L \frac{\partial \log L}{\partial \boldsymbol{\theta}} \mathbf{t}^T d\mathbf{X} \\
&= E\left(\frac{\partial \mathbf{t}^T}{\partial \boldsymbol{\theta}}\right) + \int L \cdot \mathbf{s}\,\mathbf{t}^T d\mathbf{X} \\
&= E\left(\frac{\partial \mathbf{t}^T}{\partial \boldsymbol{\theta}}\right) + E(\mathbf{s}\,\mathbf{t}^T) \quad \square .
\end{aligned}$$

Cor: $E(\mathbf{s}) = \mathbf{0}$. Pf: $\mathbf{t} = \mathbf{c} \Rightarrow E(\mathbf{s}\mathbf{t}^T) = E(\mathbf{s}\mathbf{c}^T) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{c}^T - E\mathbf{0} = \mathbf{0}$.

Covariance of Score Function

$$
\begin{aligned}
cov(\mathbf{s}) &= E(\mathbf{s}\,\mathbf{s}^T) \triangleq \mathbf{F}, \text{ the Fisher Information Matrix} \\
&= \frac{\partial}{\partial \boldsymbol{\theta}} E(\mathbf{s}^T) - E\left(\frac{\partial\,\mathbf{s}^T}{\partial \boldsymbol{\theta}}\right) \\
&= \frac{\partial}{\partial \boldsymbol{\theta}}\,\mathbf{0}^T - E\left(\frac{\partial}{\partial \boldsymbol{\theta}} \cdot \frac{\partial\,\ell}{\partial\,\boldsymbol{\theta}^T}\right) \\
&= -E\left(\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\theta}^T}\right) \qquad \text{as was to be shown.}
\end{aligned}
$$

Cor: If $E(\mathbf{t}) = \boldsymbol{\theta}$ is an unbiased estimator, then $E(\mathbf{s}\,\mathbf{t}^T) = \mathbf{I}_p$.

Pf:
$$
\begin{aligned}
E(\mathbf{s}\,\mathbf{t}^T) &= \frac{\partial}{\partial\,\boldsymbol{\theta}} E(\mathbf{t}^T) - E\left(\frac{\partial\,\mathbf{t}^T}{\partial\,\boldsymbol{\theta}}\right) \\
&= \frac{\partial}{\partial \boldsymbol{\theta}}\,\boldsymbol{\theta}^T - E(\mathbf{0}), \text{ since } \mathbf{t} \text{ is an estimator } \mathbf{t} \neq \mathbf{t}(\boldsymbol{\theta}) \\
&= \mathbf{I}_p\,.
\end{aligned}
$$

Cramér-Rao Lower Bound

Prop: If $E\,[\mathbf{t}(\mathbf{X})] = \boldsymbol{\theta}\,,$ then $V(\mathbf{t}) \geq \mathbf{F}^{-1}\,.$

Pf: Define $\alpha = \mathbf{a}^T\mathbf{t}$ and $\gamma = \mathbf{c}^T\mathbf{s}$; then,

$$\begin{aligned}
cov(\alpha,\gamma) &= \mathbf{a}^T cov(\mathbf{t},\mathbf{s})\mathbf{c} = \mathbf{a}^T\mathbf{I}_p\mathbf{c} = \mathbf{a}^T\mathbf{c} \\
var(\alpha) &= \mathbf{a}^T V(\mathbf{t})\mathbf{a} \\
var(\gamma) &= \mathbf{c}^T V(\mathbf{s})\mathbf{c} = \mathbf{c}^T\mathbf{F}\mathbf{c} \\
corr^2(\alpha,\gamma) &= \frac{(\mathbf{a}^T\mathbf{c})^2}{V(\alpha)\cdot V(\gamma)} = \frac{(\mathbf{a}^T\mathbf{c})^2}{\mathbf{a}^T V(\mathbf{t})\mathbf{a}\cdot\mathbf{c}^T\mathbf{F}\mathbf{c}} \leq 1\,.
\end{aligned}$$

Raleigh quotient (cf Izenman (3.39)): For positive definite $\mathbf{B}$,

$$\max_{\mathbf{x}^T\mathbf{B}\mathbf{x}=1} \frac{\mathbf{a}^T\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = \mathbf{a}^T\mathbf{B}^{-1}\mathbf{a} \quad \text{when } \mathbf{x}^* = \frac{\mathbf{B}^{-1}\mathbf{a}}{(\mathbf{a}^T\mathbf{B}^{-1}\mathbf{a})^{1/2}}$$

CRLB (Continued)

$$\text{Recall} \quad corr^2(\alpha, \gamma) = \frac{(\mathbf{a}^T\mathbf{c})^2}{\mathbf{a}^T V(\mathbf{t})\mathbf{a} \cdot \mathbf{c}^T\mathbf{F}\mathbf{c}} \leq 1\,.$$

The inequality holds even if we maximize the quotient over $\mathbf{c}$:

$$\begin{aligned}
\max_{\mathbf{c}^T\mathbf{F}\mathbf{c}=1} \quad \frac{(\mathbf{a}^T\mathbf{c})^2}{\mathbf{a}^T V(\mathbf{t})\mathbf{a} \cdot \mathbf{c}^T\mathbf{F}\mathbf{c}} &= \frac{1}{\mathbf{a}^T V(\mathbf{t})\mathbf{a}} \cdot \max_{\mathbf{c}} \frac{(\mathbf{a}^T\mathbf{c})^2}{\mathbf{c}^T\mathbf{F}\mathbf{c}} \\
&= \frac{1}{\mathbf{a}^T V(\mathbf{t})\mathbf{a}} \cdot \mathbf{a}^T\mathbf{F}^{-1}\mathbf{a} \\
&\leq 1 \\
\iff \quad \mathbf{a}^T\mathbf{F}^{-1}\mathbf{a} &\leq \mathbf{a}^T V(\mathbf{t})\mathbf{a} \\
\iff \quad \mathbf{a}^T \left(V(\mathbf{t}) - \mathbf{F}^{-1}\right)\mathbf{a} &\geq 0\,;
\end{aligned}$$

that is, the matrix $V(\mathbf{t}) - \mathbf{F}^{-1}$ is positive semi-definite.

CRLB (Continued)

We proved

$$V(\mathbf{t}) \geq \mathbf{F}^{-1}$$

for any unbiased estimator $\mathbf{t}$ of $\boldsymbol{\theta}$.

Conclusion: The Fisher Info Matrix inverse gives a lower bound on the uncertainty possible for all unbiased estimators.

Rao proved an extension of Fisher's univariate result:

$$\sqrt{n} \cdot (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \rightarrow AN(\mathbf{0}_p, \mathbf{F}^{-1})\,;$$

that is, under some regularity conditions, maximum likelihood estimators are asymptotically unbiased *and* asymptotically minimum variance. (Impressive.)