CH 9  THE CRAMER RAO INEQUALITY

RECALL WEIGHTED LEAST SQUARES

\[ y(u) = \Theta^T \varphi(u) + \varepsilon(u) \quad u = 1, \ldots, N \]

\[ \hat{\Theta}_{ML} = \arg \min_{\Theta} \frac{1}{2} (y - \Phi \Theta)^T \Sigma^{-1} (y - \Phi \Theta) \]

\[ = J^T \Sigma^{-\frac{1}{2}} y \quad J = \Sigma^{-\frac{1}{2}} \Phi \]

\[ E \left\{ \hat{\Theta}_{ML} \right\} = \Theta \Rightarrow \text{UNBIASED} \]

\[ E \left\{ (\hat{\Theta}_{ML} - \Theta) (\hat{\Theta}_{ML} - \Theta)^T \right\} = \text{COV} \hat{\Theta} = J^T \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} J = J^T J \]

\[ \varphi(u) = \begin{pmatrix} \varphi(u_1) \\ \vdots \\ \varphi(u_N) \end{pmatrix} \]

\[ \varepsilon = (\varepsilon_1, \ldots, \varepsilon_N) \]
\[ \begin{align*} 
\text{Cov} \quad \Theta & = (\phi^T \Sigma^{-1} \phi)^{-1} (\phi^T \Sigma^{-\frac{1}{2}} \mathbf{y} \phi) (\phi^T \Sigma^{-\frac{1}{2}} \phi)^{-1} \\
& = (\phi^T \Sigma^{-\frac{1}{2}} \phi)^{-1} \\
& = \Theta^{-1} \\
\end{align*} \]

\[
\frac{1}{2} \text{tr} (\mathbf{y} - \phi \Theta)^T \Sigma^{-1} (\mathbf{y} - \phi \Theta) = \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} (\mathbf{y} - \phi \Theta) \right\|_2^2
\]

\[= \frac{1}{2} \left\| \Sigma^{-\frac{1}{2}} \mathbf{y} - \Sigma^{-\frac{1}{2}} \phi \Theta \right\|_2^2
\]

\[= \frac{1}{2} \left\| \left( \Sigma^{-\frac{1}{2}} \mathbf{y} - \overline{J} \Theta \right) \right\|_2^2
\]

\[\text{arg min } g(\Theta) = J^T \Sigma^{-\frac{1}{2}} \mathbf{y}
\]

\[\frac{\partial}{\partial \Theta^2} = \phi^T \Sigma^{-1} \phi \quad \text{JUST A COINCIDENCE?}
\]
QUESTION: GIVEN

a) ANY ESTIMATOR \( \hat{\Theta}(Y) \) ( THAT IS UNBIASED)

b) A PDF \( f(\Theta, Y) \) (AND TRUE PARAMETER \( \Theta_0 \))

WHAT IS THE COVARIANCE OF \( \hat{\Theta} \)?

PARTIAL ANSWER (CRAMER-RAO-LOWER BOUND)

\[
\text{COV} \Theta \preceq M^{-1} \quad \iff \quad \text{COV} \Theta - M^{-1} \preceq 0
\]

WHERE \( M \) IS THE "FISHER INFORMATION MATRIX"

\[
M = \left[ \mathbb{E} \left\{ \frac{\partial^2}{\partial \Theta^2} \left( - \log f(\Theta, Y) \right) \right\} \right]_{\Theta = \Theta_0}
\]

[Cf. LJUNG AZ]

"NO ESTIMATOR CAN HAVE LOWER COVARIANCE THAN \( M \)."
EXAMPLE: GAUSSIAN NOISE, LINEAR MODEL, INDEPENDENT

\[ \Sigma = \begin{bmatrix} G(1)^2 & 0 \\ 0 & G(N)^2 \end{bmatrix} \]

\[ y = \begin{pmatrix} y(1) \\ \vdots \\ y(N) \end{pmatrix} \]

\[ f(\Theta, y) = \prod_{k=1}^{N} \text{CONST.} \exp\left(-\frac{1}{2} \left( \frac{y(\phi(k)) - \Theta^T \phi(k)}{G(k)} \right)^2 \right) \]

\[ -\log f(\Theta, y) = \frac{1}{2} \sum_{h=1}^{N} \left( \frac{y(\phi(h)) - \Theta^T \phi(h)}{G(\phi)} \right)^2 + \text{CONST} \]

\[ = \frac{1}{2} (Y - \Phi \Theta)^T \Sigma^{-1} (Y - \Phi \Theta) + \text{CONST} \]

\[ \frac{\partial^2 (-\log f(\Theta, y))}{\partial \Theta^2} = \frac{1}{2} \Theta^T \Phi^T \Sigma^{-1} \Phi \Theta + \text{LINEAR} + \text{CONSTANT} \]

\[ = \Phi^T \Sigma^{-1} \Phi = M \quad \text{"FISHER INFORMATION MATRIX"} \]
Conclusion (from Example)

No estimator $\hat{\theta}(y)$ can be better than max. likelihood estimator.

(Because)

$$\text{Cov} \hat{\theta} \leq M^{-1} = \text{Cov} \hat{\Theta}_{ML} = (\Phi^T \Sigma^{-1} \Phi)^{-1}$$

For nonlinear least squares

Hessian matrix of neg. log-likelihood is estimate for Fisher-information matrix.