ECE 645 – Information Inequality and Cramer-Rao Lower Bound

J. V. Krogmeier

April 4, 2014

Contents

1 Likelihood Functions: Example 2
2 Scalar Parameter 6
3 Example: Phase Estimation 13
4 Vector Parameter 17
1 Likelihood Functions: Example
Example Related to CRLB:

Intuitively, the more a pdf depends on the unknown parameter, the easier it ought to be to estimate it.

\[ Y = \Theta + N \]
\[ N \sim N(0, \sigma^2) \]
\[ \Theta \in \mathbb{R} \] the unknown param.

Then the family of pdfs which model the situation are

\[ f_\Theta(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (y-\Theta)^2 \right\} \]

Upon observing a particular value of \( Y \), say \( Y = y^*_k \), then the pdf

\[ f_\Theta(y^*_k) \]

viewed as a function of \( \Theta \) is called the likelihood. For example:

\[ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\cdot\frac{1}{\sqrt{2\pi}\sigma}} \]

\[ y^*_k - \sqrt{2}\sigma \quad y^*_k \quad y^*_k + \sqrt{2}\sigma \]
Intuitive: The narrower likelihood (smaller $\sigma^2$) has a smaller range of viable values for $\theta$ and hence should be easier to accurately estimate $\theta$. "Sharpness" of likelihood determines how accurately can estimate parameter.

Now for the example problem at hand a reasonable estimator for $\theta$ is

$$\hat{\theta}(y) = y$$

$$\Rightarrow E_\theta\{\hat{\theta}(Y)\} = \theta, \quad \text{Var}_\theta\{\hat{\theta}(Y)\} = \sigma^2$$

which bears out the intuition.

To connect with CRLB

$$\ln f_\theta(y) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(y-\theta)^2$$

$$\frac{\partial}{\partial \theta} \ln f_\theta(y) = \frac{1}{\sigma^2} (y-\theta)$$
\[ \frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(y) = -\frac{1}{\sigma^2} \quad \Rightarrow \quad -\frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(y) = \frac{1}{\sigma^2} \]

\[ \text{the curvature of the log-likelihood function} \]

Larger curvature \(\iff\) sharper \(\iff\) smaller \(\sigma^2\)

Thus in this case

\[ \text{Var}_{\theta} \{ \hat{\Theta}(Y) \} = \frac{1}{-\frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(y)} = \sigma^2 \]

\[ \text{Not a function of } Y=y \text{ in this case though would be in general. Thus should use average of curvature} \]

\[ -E_{\theta} \left\{ \frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(Y) \right\} \]
2 Scalar Parameter
Theorem (cont.)

Further more, if the following condition also holds:

5. \( \exists \theta \in \mathcal{A} \) such that \( \int_{\mathbb{R}} h(y) \phi_{\theta}(y) \, dy = \frac{2}{\sigma_0} \int_{\mathbb{R}} \phi_{\theta}(y) \, dy \)

Then \( I_0 \) can be computed via

\[
I_0 = - E_{\theta} \left\{ \frac{2}{\sigma_0} \log \left( \frac{\phi_{\theta}(y)}{\phi_{\theta}(y)} \right) \right\}.
\]

Proof

Differentiating

\[
E_{\theta} \left\{ \hat{\theta}(y) \right\} = \int_{\mathbb{R}} \hat{\theta}(y) \phi_{\theta}(y) \, dy
\]

with respect to \( \theta \) and using condition 4 we have

\[
\frac{2}{\sigma_0} E_{\theta} \left\{ \hat{\theta}(y) \right\} = \int_{\mathbb{R}} [\hat{\theta}(y) - \frac{2}{\sigma_0} \phi_{\theta}(y)] \, dy
\]

If we also apply 4 in the case \( h(y) = 1 \) we have

\[
\int_{\mathbb{R}} \frac{2}{\sigma_0} \phi_{\theta}(y) \, dy = \frac{2}{\sigma_0} \int_{\mathbb{R}} \phi_{\theta}(y) \, dy = \frac{2}{\sigma_0} \phi_{\theta}(y) = 0
\]

Therefore,

\[
\frac{2}{\sigma_0} E_{\theta} \left\{ \hat{\theta}(y) \right\} = \int_{\mathbb{R}} [\hat{\theta}(y) - E_{\theta} \left\{ \hat{\theta}(y) \right\}] \phi_{\theta}(y) \, dy
\]

Note:

\[
\frac{2}{\sigma_0} \log \phi_{\theta}(y) = \frac{1}{\phi_{\theta}(y)} \frac{2}{\sigma_0} \phi_{\theta}(y)
\]
Using the note, we write
\[
\frac{2}{\theta} E_0 \{ \hat{\theta}(\gamma) \} = \int \left[ \hat{\theta}(y) - E_0 \{ \hat{\theta}(\gamma) \} \right] \frac{2}{\theta} \log \frac{P_0(y)}{P_0(\gamma)} \, d\gamma
\]
\[
= E_0 \left\{ \left[ \hat{\theta}(y) - E_0 \{ \hat{\theta}(\gamma) \} \right] \frac{2}{\theta} \log \frac{P_0(y)}{P_0(\gamma)} \right\}
\]

Recall the Schartz inequality

Let \( U \) and \( V \) be random variables
\[
(E\{UV\})^2 \leq E\{U^2\} E\{V^2\}
\]
Also note \( |E\{UV\}| \leq E\{|UV|\} \)

Applying the Schwartz inequality to the above we have
\[
\left( \frac{2}{\theta} E_0 \{ \hat{\theta}(\gamma) \} \right)^2 \leq \frac{E_0 \left\{ (\hat{\theta}(y) - E_0 \{ \hat{\theta}(\gamma) \})^2 \right\}}{Var_0 \{ \hat{\theta}(\gamma) \}} \frac{E_0 \left\{ \frac{2}{\theta} \log \frac{P_0(y)}{P_0(\gamma)} \right\}}{E_0 \{ \frac{2}{\theta} \log \frac{P_0(y)}{P_0(\gamma)} \}^2}
\]
\[
\therefore \ Var_0 \{ \hat{\theta}(\gamma) \} \geq \frac{E_0 \left\{ \frac{2}{\theta} E_0 \{ \hat{\theta}(\gamma) \} \right\}^2}{I_0}
\]

For the last statement of the theorem compute
\[
\frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} = \frac{2}{\theta^2} \left[ \frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} \right]
\]
\[
= \left( \frac{2}{\theta^2} \left\{ \frac{1}{P_0(y)} \right\} \left( \frac{2}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) + \frac{1}{P_0(y)} \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) \right)
\]
\[
= \left( \frac{1}{P_0(y)} \right)^2 \left( \frac{2}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) + \frac{1}{P_0(y)} \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right)
\]
\[
= \left( \frac{2}{\theta^2} \left\{ \frac{1}{P_0(y)} \right\} \right)^2 \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) + \frac{1}{P_0(y)} \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right)
\]
\[
= \left( \frac{2}{\theta^2} \left\{ \frac{1}{P_0(y)} \right\} \right)^2 \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right) + \frac{1}{P_0(y)} \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right)
\]
\[
\frac{2}{\theta} \log \frac{P_0(y)}{P_0(\gamma)} = - \left( \frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} \right)^2 + \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right)
\]
\[
= - \left( \frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} \right)^2 + \left( \frac{3}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \right)
\]

Taking the expectation \( E_0 \{ \cdot \} \) of both sides of the above we find
\[
E_0 \left\{ \frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} \right\} = - I_0 + \int \frac{2}{\theta^2} \frac{P_0(y)}{P_0(\gamma)} \, dy
\]
\[
I_0 = - E_0 \left\{ \frac{2}{\theta^2} \log \frac{P_0(y)}{P_0(\gamma)} \right\}
\]

Notes:
- \( I_0 \) is known as Fisher's information.
- The inequality of the Thm is called the information inequality.
- Special case: \( E_0 \{ \hat{\theta}(\gamma) \} = \theta \) (unbiased) then the inf. inequality becomes
  \[
  Var_0 \{ \hat{\theta}(\gamma) \} \geq I_0
  \]
  the Cramer-Rao lower bound.
- Existence of an estimate achieving equality in CRB is not always possible.
Note on Fisher's Information
\[ I_0 = E_0 \left\{ \left( \frac{2}{\theta} \log p_\theta(y) \right)^2 \right\} = -E_0 \left\{ \frac{2}{\theta} \log p_\theta(y) \right\} \]
subject to "regularity" conditions.

In the case of an unbiased estimator \( \hat{\theta} \) of a real parameter \( \theta \), the information bound is the Cramér-Rao bound
\[ \text{Var}_\theta \{ \hat{\theta}(y) \} \geq \frac{1}{I_0} \quad \left( E_0 \{ \hat{\theta}(y) \} = \theta \right) \]

- More information (bigger \( I_0 \)) provided on the average by an observation \( y \) the smaller we might expect the variance of a good estimator. Because the bound \( 1/I_0 \) would be smaller.
- \( I_0 \) has a property expected of an information measure.

Suppose that \( y = [y_1, y_2, \ldots, y_n]^T \) where the \( y_k \) are iid with marginal density
\[ p_\theta^*(y) \]
Then the density of \( y \) would be \( p_\theta(y) = p_\theta^*(y_1) \cdots p_\theta^*(y_n) \) and
\[ \log p_\theta(y) = \sum_{i=1}^n \log p_\theta^*(y_i) \]

Now subject to regularity
\[ \frac{2}{\theta} \log p_\theta(y) = \sum_{i=1}^n \frac{2}{\theta} \log p_\theta^*(y_i) \]
shows the right side to be a sum of iid random variables each of zero mean. To see this
\[ \frac{2}{\theta} \log p_\theta(y) = \frac{2}{\theta} \log p_\theta(y) \Rightarrow \left( \frac{2}{\theta} \log p_\theta(y) \right) p_\theta(y) = \frac{2}{\theta} \log p_\theta(y) \]
and integrate over \( y \) using \( \theta \) to show
\[ E_\theta \left\{ \frac{2}{\theta} \log p_\theta(y) \right\} = 0. \]

Therefore
\[ I_0 = E_\theta \left\{ \left( \frac{2}{\theta} \log p_\theta(y) \right)^2 \right\} \]
\[ = \text{Var}_\theta \left\{ \frac{2}{\theta} \log p_\theta(y) \right\} \]
\[ = n \text{Var}_\theta \left\{ \frac{2}{\theta} \log p_\theta^*(y_i) \right\} \]
\[ = n E_\theta \left\{ \left( \frac{2}{\theta} \log p_\theta^*(y_i) \right)^2 \right\} \]
\[ = n \frac{\theta}{\theta} \]

This additivity property is basic to all measures of infromation.

Question: How sharp is CRLB? That is how nearly can it be attained?

Achieving the bound exactly (partial answer to above).

Subject to the usual regularity conditions, there exists an unbiased estimator whose variance attains the CRLB if and only if \( \frac{2}{\theta} \log p_\theta(y) \) can be expressed in the form
\[ \frac{2}{\theta} \log p_\theta(y) = a_\theta(\hat{\theta}) \left( \hat{\theta}(y) - \theta \right) \]

and in this case \( a_\theta(\theta) = I_\theta \).

To see this we need to go back to the proof of the theorem. The critical issue being the condition for equality in the Schwartz inequality.

\( U, V \) second order random variables
\[ (E[U^2])^2 \leq E[U] E[V^2] \]
with equality if and only if \( \lambda U + \mu V = 0 \) for some real numbers \( \lambda, \mu \) (not both zero).
In the proof (specialized to unbiased case) we had

$$1 = E_\theta \left\{ \left[ \frac{1}{2} (\hat{\theta}(Y) - \theta) \right] \log p_\theta(Y) \right\} = a(\theta) \text{Var}_\theta \left\{ \hat{\theta}(Y) \right\}$$

Applying the Schwarz inequality, we had

$$\left( E_\theta \left\{ \left[ \frac{1}{2} (\hat{\theta}(Y) - \theta) \right] \log p_\theta(Y) \right\} \right)^2 \leq \frac{1}{\text{Var}_\theta \left\{ \hat{\theta}(Y) \right\}} E_\theta \left\{ (\hat{\theta}(Y) - \theta)^2 \right\}$$

and equality holds if and only if there exist real numbers $\lambda_0$ and $\mu_0$ st (ie may dep. on $\theta$)

$$\lambda_0 (\hat{\theta}(Y) - \theta) + \mu_0 \frac{1}{2} \log p_\theta(Y) = 0$$

in

$$\frac{1}{2} \log p_\theta(Y) = -\frac{\lambda_0}{\mu_0} (\hat{\theta}(Y) - \theta)$$

[with prob $1$ under $p_\theta$]

$$= a(\theta) (\hat{\theta}(Y) - \theta)$$

Moreover, when the above holds

$$1 = \text{Var}_\theta \left\{ \hat{\theta}(Y) \right\} = a(\theta) \text{Var}_\theta \left\{ \hat{\theta}(Y) \right\} = a(\theta) / I_\theta$$

$$= a(\theta) = I_\theta.$$
Information Inequality for Exponential Families

Suppose $\Delta$ is an open interval and $\phi_0(y)$ is

$$p_0(y) = C(\theta) e^{g(\theta)Ty} h(y)$$

where $C, g, T,$ and $h$ are real valued functions. Also assume

- $g(\theta)$ exists
- $E_\theta\{T(Y)\} < \infty$
- $\frac{2}{\theta^2} \int \frac{e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy} = \frac{\int g'(\theta) e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy}$

Then the hypotheses 1 - 4 of the theorem hold and we need to do the computations for the bound.

Note $C(\theta) = \left(\int e^{g(\theta)Ty} h(y) dy\right)^{-1}$ since density must integrate to 1. Thus

$$\log p_0(y) = g(\theta)T(y) + \log h(y) - \log \left(\int e^{g(\theta)Ty} h(y) dy\right)$$

so differentiating

$$\frac{2}{\theta^2} \log p_0(y) = g'(\theta)T(y) - \frac{\int g'(\theta) e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy}$$

$$= g'(\theta)T(y) - g'(\theta) E_\theta\{T(Y)\}$$

$$= g'(\theta) \left[T(y) - E_\theta\{T(Y)\}\right]$$

Therefore

$$I_\theta = E_\theta\{(\frac{2}{\theta^2} \log p_0(y))^2\}$$

$$= (g'(\theta))^2 E_\theta\{(T(Y) - E_\theta\{T(Y)\})^2\}$$

$$= (g'(\theta))^2 \text{Var}_\theta\{T(Y)\}$$

Thus the information inequality is

$$\text{Var}_\theta\{T(Y)\} \geq \left(\frac{2 \theta^2 E_\theta\{\hat{\theta}(Y)\}}{(g'(\theta))^2 \text{Var}_\theta\{T(Y)\}}\right)^2$$

Achieving the bound

Consider $T(y)$ as an estimator of $\theta$. Then (is $\hat{\theta}(y) = T(y)$)

$$E_\theta\{T(Y)\} = \frac{\int T(y) e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy}$$

To put this into the bound we need to take the partial with respect to $\theta$.

$$\frac{2}{\theta} E_\theta\{T(Y)\} = \left(\frac{g'(\theta)\int T(y) e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy}\right) \left(\int e^{g(\theta)Ty} h(y) dy\right)$$

$$- \left(\frac{g'(\theta)\int T(y) e^{g(\theta)Ty} h(y) dy}{\int e^{g(\theta)Ty} h(y) dy}\right) \left(\int T(y) e^{g(\theta)Ty} h(y) dy\right)$$

$$\left(\int e^{g(\theta)Ty} h(y) dy\right)^2$$
\[ \frac{2}{\theta^2} E_{\theta} \{ T(\theta) \} = g'(\theta) E_{\theta} \{ T^2(\theta) \} - g'(\theta) (E_{\theta} \{ T(\theta) \})^2 \]

\[ = g'(\theta) \text{Var}_{\theta} \{ T(\theta) \} \]

Thus the lower bound in the information inequality becomes

\[ (g'(\theta))^2 \frac{\text{Var}_{\theta} \{ T(\theta) \}^2}{(g'(\theta))^2 \text{Var}_{\theta} \{ T(\theta) \}} = \text{Var}_{\theta} \{ T(\theta) \} \]

Conclude \( T(\theta) \) achieves the information lower bound.

Another way to look at this is that \( T(\theta) \) has minimum variance among the class of estimators \( \hat{\theta}(\cdot) \) which satisfy the hypotheses of the theorem. Of course, the only place where props of \( \hat{\theta}(\cdot) \) enter the hypotheses is in \( \mathbb{P} \).

\[ \frac{2}{\theta^2} E_{\theta} \{ \hat{\theta}(\cdot) \} = \int \hat{\theta}(y) \frac{2}{\theta^2} \mathbb{P}(y) \, dy \]

\[ \frac{2}{\theta^2} \mathbb{E}_{\theta} \left[ e^{g(\theta) \hat{T}(\theta)} h(y) \right] \]

\[ = \frac{g'(\theta) \mathbb{E}_{\theta} \{ e^{g(\theta) \hat{T}(\theta)} h(y) \} - g'(\theta) E_{\theta} \{ T(\theta) \} e^{g(\theta) \hat{T}(\theta) h(y)} \}}{\mathbb{E}_{\theta} \{ e^{g(\theta) \hat{T}(\theta) h(y)} \} - g'(\theta) \mathbb{E}_{\theta} \{ T(\theta) \} \mathbb{P}(y) \}} \]

\[ = g'(\theta) \text{Cov}_{\theta} \{ \hat{\theta}(\cdot), T(\theta) \} \]

\[ \frac{2}{\theta^2} \mathbb{E}_{\theta} \{ \hat{\theta}(\cdot) \} = g'(\theta) \left[ E_{\theta} \{ \hat{\theta}(\cdot) T(\theta) \} - E_{\theta} \{ \hat{\theta}(\cdot) \} E_{\theta} \{ T(\theta) \} \right] \]

\[ = g'(\theta) \text{Cov}_{\theta} \{ \hat{\theta}(\cdot), T(\theta) \} \]
3 Example: Phase Estimation
Example: Phase Estimation

\[ Y_k = A \cos(2\pi f_0 k + \phi) + N_k \quad k = 0, 1, \ldots, N-1 \]

\( A, f_0 \) assumed known

\( \phi \) the unknown parameter

\( N_k \sim N(0, \sigma^2) \) and iid

\[ f_\phi(y) = \frac{1}{(2\pi \sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{k=0}^{N-1} \left( y_k - A \cos(2\pi f_0 k + \phi) \right)^2 \right\} \]

\[ \frac{\partial}{\partial \phi} \ln f_\phi(y) = -\frac{1}{\sigma^2} \sum_{k=0}^{N-1} (y_k - A \cos(2\pi f_0 k + \phi)) A \sin(2\pi f_0 k + \phi) \]

\[ = -\frac{A}{\sigma^2} \sum_{k=0}^{N-1} \left[ y_k \sin(2\pi f_0 k + \phi) - \frac{A}{2} \sin(4\pi f_0 k + 2\phi) \right] \]

Taking another derivative:

\[ \frac{\partial^2}{\partial \phi^2} \ln f_\phi(y) = -\frac{A}{\sigma^2} \sum_{k=0}^{N-1} \left[ y_k \cos(2\pi f_0 k + \phi) - A \cos(4\pi f_0 k + 2\phi) \right] \]
Taking \(-E_\phi \left\{ \frac{3}{2} \phi^2 \ln f_\phi (\gamma) \right\}\)

\[
\begin{align*}
&= \frac{A}{\sigma^2} \sum_{k=0}^{N-1} \left[ A \cos^2 (2\pi f_0 k + \phi) - A \cos (4\pi f_0 k + 2\phi) \right] \\
&= \frac{A^2}{2\sigma^2} \sum_{k=0}^{N-1} \left[ \frac{1}{2} + \frac{1}{2} \cos (4\pi f_0 k + 2\phi) - \cos (4\pi f_0 k + 2\phi) \right] \\
&= \frac{N A^2}{2\sigma^2} + (-1) \frac{A^2}{2\sigma^2} \sum_{k=0}^{N-1} \cos (4\pi f_0 k + 2\phi) \\
&= \frac{N A^2}{2\sigma^2} \left[ 1 - \frac{1}{N} \sum_{k=0}^{N-1} \cos (4\pi f_0 k + 2\phi) \right] \\
&\approx 0 \quad \text{for } f_0 \text{ not near } 0 \text{ or } \frac{1}{2}
\end{align*}
\]
Therefore the CRLB for this problem is close to:

\[ \text{Var}_\theta \left\{ \hat{\phi}(Y) \right\} \geq \frac{2\sigma^2}{NA^2} \]

for an unbiased estimator \( \hat{\phi} \) of \( \phi \).

The bound is not attained since the form required of \( \frac{2}{\sigma^2} \ln \hat{f}_\phi(y) \) does not hold. An efficient estimator of the phase in this problem does not exist.

Still possible that an MVUE does exist.
4 Vector Parameter
Generalization of CRLB to vector parameter

Say \( \Theta \in \mathbb{R}^m \). The analog of the Fisher information is the \( m \times m \) information matrix \( I_{\Theta} \) defined as

\[
I_{\Theta} = E_{\Theta} \left( \frac{2}{\sigma_i \sigma_j} \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p_{\Theta}(Y) \right)
\]

subject to the same type of regularity conditions.

CRLB Theorem: Let \( \hat{\Theta} \) be an unbiased estimator of the vector parameter \( \Theta \). Then, subject to regularity conditions

\[
\text{Var}_{\Theta} \left( \hat{\Theta}(Y) \right) = E_{\Theta} \left( (\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T \right) \geq I_{\Theta}^{-1}
\]

Hence

\[
\text{Var}_{\Theta} \left( \hat{\Theta}(Y) \right) - I_{\Theta}^{-1}
\]

is positive semi-definite.

Proof: Let \( \Theta = [\Theta_1, \Theta_2, \ldots, \Theta_m]^T \) be an unbiased estimator of the vector parameter \( \Theta = [\Theta_1, \ldots, \Theta_m]^T \). Clearly

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{\Theta}_i(y) p_{\Theta}(y) \, dy = \Theta_i
\]

so assuming enough regularity to interchange integration and differentiation

\[
\delta_{ij} = \frac{2}{\sigma_i \sigma_j} \Theta_i = \int \hat{\Theta}_i(y) \frac{2}{\sigma_i \sigma_j} \log p_{\Theta}(y) \, dy = \int \hat{\Theta}_i(y) \frac{2}{\sigma_i \sigma_j} \log p_{\Theta}(y) \, dy
\]

this holding for \( 1 \leq i, j \leq m \).

Let \( \psi(Y, \Theta) \triangleq \nabla_\Theta \log p_{\Theta}(Y) \) be the usual gradient. Then the above shows:

1. \( E_{\Theta} \left( \hat{\Theta}(Y) \psi(Y, \Theta) \right) = I_{\Theta} (m \times m \text{ identity}) \)
2. \( E_{\Theta} \left( \psi(Y, \Theta) \right) = 0 \)
3. \( \text{Var}_{\Theta} \left( \psi(Y, \Theta) \right) = I_{\Theta} \)

Note that (3) implies that we may write (1) as

\[
\text{Cov}_{\Theta} \left( \hat{\Theta}(Y), \psi(Y, \Theta) \right) = I_{\Theta}
\]

Consider the variance matrix of the vector \( \left[ \hat{\Theta}_i \right] \) which is

\[
\left( \text{Var}_{\Theta} \left( \hat{\Theta}_i \right) \right) \text{ Cov}_{\Theta} \left( \hat{\Theta}_i, \psi \right) = \left( \text{Var}_{\Theta} \left( \hat{\Theta}_i \right) \right) \left( I_{\Theta} \right)
\]

The above must be positive semidefinite. If \( I_{\Theta}^{-1} \) exists

the following must also be pos. semidefinite:

\[
\left[ I_{\Theta}^{-1} \right] \left( \text{Var}_{\Theta} \left( \hat{\Theta}_i \right) \right) \left( I_{\Theta} \right) \left[ I_{\Theta} \right]^{-1} = \text{Var}_{\Theta} \left( \hat{\Theta}_i \right) - I_{\Theta}^{-1}
\]