Fisher Information

- Fisher information is the variance of the score function
- It tells us how much information Y carries about the parameters of the distribution that models Y
- The asymptotic variance of the MLE is the inverse of the Fisher information
- The Cramér-Rao lower bound, an important result about related to minimum variance unbiased estimators, is given in terms of the Fisher information

Cramér-Rao inequality

Suppose that $Y_1, Y_2, \dots, Y_n$ are i.i.d. with pdf $f(\boldsymbol{y};\theta)$. Subject to regularity conditions on $f(\boldsymbol{y};\theta)$, we have that for any unbiased estimator $\tilde{\theta}$ for θ,

$$\boxed{var(\tilde{\theta}) \geq I_\theta^{-1}}$$

where

$$\boxed{\boxed{I_\theta} = E\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right].}$$

Fisher information about θ

This inequality gives the lower bound for the variance of unbiased estimators, under regularity conditions (see next slide).
Essentially, this says that, under these conditions, all unbiased estimator for θ have variance at least as big as the inverse of the Fisher information about θ.

Precise statement of the regularity conditions is somewhat technical. But, in broad terms, we require that:

- The probability density f is sufficiently many times continuously differentiable
- The support of Y does not depend on θ

These are needed so that we can exchange the order of integration (with respect to $\boldsymbol{y}$) and differentiation (with respect to θ). That is, we can perform operations like the following:

$$\frac{\partial^2}{\partial\theta^2}\int f(\theta;\boldsymbol{y})d\boldsymbol{y} = \int \frac{\partial^2}{\partial\theta^2} f(\theta;\boldsymbol{y})\, d\boldsymbol{y}$$

Fisher information

I_θ is known as the Fisher information about θ in the observations.

$$I_\theta = \text{var}\left[S(\theta; y) \right]$$ i.e. variance of the score

$$= E\left[\left(\frac{\partial \ell}{\partial \theta} \right)^2 \right]$$ under regularity conditions

If $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$, then I_θ is the Fisher information matrix, with dimensions $k \times k$. The ij^{th} element of I_θ is given by

$$[I_\theta]_{ij} = E\left[\frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j} \right].$$

Alternative form

Under the same regularity conditions as for the Cramér-Rao inequality:

$$\boxed{I_\theta = -E\left[\frac{\partial^2 \ell}{\partial \theta^2}\right].}$$

In the case of $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^T$,

$$[I_\theta]_{ij} = -E\left[\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\right].$$

Proof of alternative form

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\frac{\partial \ell}{\partial \theta}\right)$$

$$= \frac{\partial}{\partial \theta}\left(\frac{\partial \log L}{\partial \theta}\right)$$

$$= \frac{\partial}{\partial \theta}\left(\frac{1}{L}\,\frac{\partial L}{\partial \theta}\right)$$

$$= \frac{\left(\frac{\partial^2}{\partial \theta^2} L\right) L - \left(\frac{\partial L}{\partial \theta}\right)\left(\frac{\partial L}{\partial \theta}\right)}{L^2}$$

$$= \frac{\frac{\partial^2 L}{\partial \theta^2}}{L} - \left[\boxed{\frac{\left(\frac{\partial L}{\partial \theta}\right)}{L}}\right]^2$$

$$= \frac{1}{L}\frac{\partial^2 L}{\partial \theta^2} - \left[\frac{\partial \log L}{\partial \theta}\right]^2$$

quotient rule:

$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$$

In our case, let $f = \frac{\partial \ell}{\partial \theta}$ and $g = L$.

Then $f' = \frac{\partial^2 \ell}{\partial \theta^2}$ and $g' = \frac{\partial L}{\partial \theta}$.

$$\frac{\partial \log L}{\partial \theta} = \boxed{\frac{1}{L}\left(\frac{\partial L}{\partial \theta}\right)}$$

Proof of alternative form

$$E\left[-\frac{\partial^2 \ell}{\partial \theta^2}\right] = E\left[-\frac{1}{L}\frac{\partial^2 L}{\partial \theta^2} + \left(\frac{\partial \log L}{\partial \theta}\right)^2\right]$$

$$= -\boxed{E\left[\frac{1}{L}\frac{\partial^2 L}{\partial \theta^2}\right]} + E\left[\left(\frac{\partial \log L}{\partial \theta}\right)^2\right]$$

$$E\left[\frac{1}{L}\frac{\partial^2 L}{\partial \theta^2}\right] = \int_{-\infty}^{\infty} \frac{1}{\cancel{L}}\frac{\partial^2 L}{\partial \theta^2} \cancel{L}\, dy$$

$$= \int_{-\infty}^{\infty} \frac{\partial^2 L}{\partial \theta^2}\, dy$$

$$= \frac{\partial^2}{\partial \theta^2} \underbrace{\int_{-\infty}^{\infty} L\, dy}_{1}$$

under regularity conditions

as we are integrating a density over its support

$$= \frac{\partial^2}{\partial \theta^2} 1$$

$$= 0$$

$$\therefore \quad E\left[-\frac{\partial^2 \ell}{\partial \theta^2}\right] = E\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = E\left[S(\theta;y)^2\right]$$

Example 5.7

Suppose $y_1, y_2, \ldots, y_n$ are $i.i.d.\, Po(\lambda)$ observations. Find the Fisher information about λ.

Recall $\ell(\lambda; y) = -n\lambda + \left(\sum_{i=1}^{n} y_i\right)\log\lambda + \log \prod_{i=1}^{n}\left(\frac{1}{y_i!}\right)$

$$\frac{\partial \ell}{\partial \lambda} = -n + \frac{1}{\lambda}\left(\sum_{i=1}^{n} y_i\right)$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{1}{\lambda^2}\left(\sum_{i=1}^{n} y_i\right)$$

$$\therefore \quad I_\lambda = E\left[-\frac{\partial^2 \ell}{\partial \lambda^2}\right]$$

$$= E\left[\frac{1}{\lambda^2}\left(\sum_{i=1}^{n} Y_i\right)\right]$$

$$= \frac{1}{\lambda^2}\sum_{i=1}^{n} E(Y_i)$$

Recall $Y_i \sim Po(\lambda)$, then $E[Y_i] = \lambda$

$$= \frac{1}{\lambda^2}(n\lambda)$$

$$= \frac{n}{\lambda}$$

Example 5.8

Suppose $y_1, y_2, \ldots, y_n$ are $i.i.d.\, N(\mu, \sigma^2)$ observations with σ^2 known. Find the Fisher information about μ.

Recall $\ell(\mu; y) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu)$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(-1) = -\frac{n}{\sigma^2}$$

$$\therefore \quad I_\mu = E\left[-\frac{\partial^2 \ell}{\partial \mu^2}\right] = E\left[\frac{n}{\sigma^2}\right] = \frac{n}{\sigma^2}$$

Example 5.9

Suppose $y_1, y_2, \ldots, y_n$ are $i.i.d.\, N(\mu, \sigma^2)$ observations where both $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ are unknown. Find the Fisher information matrix.

Recall $\ell(\mu, \sigma^2; y) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$

$$S(\mu, \sigma^2; y) = \begin{bmatrix} \frac{\partial \ell}{\partial \mu} \\ \frac{\partial \ell}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2 \end{bmatrix}$$

Observe that:

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2} = \frac{\partial^2 \ell}{\partial \sigma^2 \partial \mu} = \frac{\partial}{\partial \sigma^2}\left[\frac{\partial \ell}{\partial \mu}\right] = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - \mu)$$

$$\frac{\partial^2 \ell}{\partial \sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(y_i - \mu)^2$$

Example 5.9 Solution

$$E\left[-\frac{\partial^2 \ell}{\partial \mu^2}\right] = \frac{n}{\sigma^2}$$

$E[Y_i] = \mu$ as $Y_i \sim N(\mu, \sigma^2)$

$$E\left[-\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2}\right] = E\left[\frac{1}{\sigma^4}\sum_{i=1}^{n}(Y_i - \mu)\right] = \frac{1}{\sigma^4}\sum_{i=1}^{n}\left[E(Y_i) - \mu\right] = \frac{1}{\sigma^4}\sum_{i=1}^{n}(\mu - \mu) = 0$$

$$\begin{aligned}
E\left[-\frac{\partial^2 \ell}{\partial \sigma^4}\right] &= E\left[\frac{1}{\sigma^6}\sum_{i=1}^{n}(Y_i - \mu)^2\right] - \frac{n}{2\sigma^4} \\
&= \frac{1}{\sigma^6}E\left[\sigma^2\sum_{i=1}^{n}\left(\boxed{\frac{Y_i - \mu}{\sigma}}\right)^2\right] - \frac{n}{2\sigma^4} \\
&= \frac{1}{\sigma^4}E\left[\sum_{i=1}^{n} Z_i^2\right] - \frac{n}{2\sigma^4} \\
&= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\
&= \frac{n}{2\sigma^4}
\end{aligned}$$

$Z_i \sim N(0,1)$

So $\sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$

$E\left[\sum_{i=1}^{n} Z_i^2\right] = n$

$$\therefore\ I_\theta = \begin{bmatrix} E\left[-\frac{\partial^2 \ell}{\partial \mu^2}\right] & E\left[-\frac{\partial^2 \ell}{\partial \mu \partial \sigma^2}\right] \\ E\left[-\frac{\partial^2 \ell}{\partial \sigma^2 \partial \mu}\right] & E\left[-\frac{\partial^2 \ell}{\partial \sigma^4}\right] \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$