Fisher information matrix

Canonical name	FisherInformationMatrix
Date of creation	2013-03-22 14:30:15
Last modified on	2013-03-22 14:30:15
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	14
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62H99
Classification	msc 62B10
Classification	msc 62A01
Synonym	information matrix
Defines	Fisher information
Defines	information
Defines	Cramer-Rao inequality
Defines	Cramer-Rao lower bound

Given a statistical model $\{f_{\mathbf{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})\}$ of a random vector $\mathbf{X}$, the *Fisher information matrix*, I, is the variance of the score function U. So,

$$I = \operatorname{Var}[U].$$

If there is only one parameter involved, then I is simply called the *Fisher information* or *information* of $f_{\mathbf{X}}(\boldsymbol{x} \mid \theta)$.

Remarks

- If $f_{\mathbf{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$ belongs to the exponential family, $I = \mathrm{E}\left[U^{\mathrm{T}}U\right]$. Furthermore, with some regularity conditions imposed, we have

$$I = -\,\mathrm{E}\left[\frac{\partial U}{\partial \boldsymbol{\theta}}\right].$$

- As an example, the normal distribution, $N(\mu, \sigma^2)$, belongs to the exponential family and its log-likelihood function $\ell(\boldsymbol{\theta} \mid x)$ is

$$-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2},$$

  where $\boldsymbol{\theta} = (\mu, \sigma^2)$. Then the score function $U(\boldsymbol{\theta})$ is given by

$$\left(\frac{\partial \ell}{\partial \mu}, \frac{\partial \ell}{\partial \sigma^2}\right) = \left(\frac{x-\mu}{\sigma^2}, \frac{(x-\mu)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right).$$

  Taking the derivative with respect to $\boldsymbol{\theta}$, we have

$$\frac{\partial U}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial U_1}{\partial \mu} & \dfrac{\partial U_2}{\partial \mu} \\[2ex] \dfrac{\partial U_1}{\partial \sigma^2} & \dfrac{\partial U_2}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{-1}{\sigma^2} & -\dfrac{x-\mu}{\sigma^4} \\[2ex] -\dfrac{x-\mu}{\sigma^4} & \dfrac{1}{2\sigma^4} - \dfrac{(x-\mu)^2}{\sigma^6} \end{pmatrix}.$$

  Therefore, the Fisher information matrix I is

$$-\,\mathrm{E}\left[\frac{\partial U}{\partial \boldsymbol{\theta}}\right] = \frac{1}{2\sigma^4}\begin{pmatrix} 2\sigma^2 & 0 \\ 0 & -1 \end{pmatrix}.$$

- Now, in linear regression model with constant variance σ^2, it can be shown that the Fisher information matrix I is

$$\frac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}\mathbf{X},$$

  where $\mathbf{X}$ is the design matrix of the regression model.

- In general, the Fisher information meansures how much "information" is known about a parameter θ. If T is an unbiased estimator of θ, it can be shown that
$$\mathrm{Var}\left[T(X)\right] \geq \frac{1}{I(\theta)}$$
This is known as the *Cramer-Rao inequality*, and the number $1/I(\theta)$ is known as the *Cramer-Rao lower bound*. The smaller the variance of the estimate of θ, the more information we have on θ. If there is more than one parameter, the above can be generalized by saying that
$$\mathrm{Var}\left[T(X)\right] - I(\boldsymbol{\theta})^{-1}$$
is positive semidefinite, where I is the Fisher information matrix.