Fisher Information in MLE

Chao Cheng

November 15, 2022

Contents

Let i.i.d sample $x_1, \cdots, x_n \sim F_x$. The p.d.f is $f_x(x|\theta)$ where θ is the parameter and we use θ_0 to denote the value of underlying parameter θ. Then the log-likelihood

$$l = \log L = \sum_{i=1}^{n} \log f(x_i|\theta).$$

1 Score function

The score function is the first derivative (the gradient) of log-likelihood

$$S(\boldsymbol{x}|\theta) = \frac{\partial l}{\partial \theta} \tag{1}$$

The MLE $\hat{\theta}$ makes $S(\boldsymbol{x}|\theta)$ equals to zero:

$$S\left(\boldsymbol{x}\middle|\hat{\theta}\right) = 0.$$

Also under some regularity conditions

$$\mathrm{E}_{\theta_0}\left(S(\boldsymbol{x}|\theta_0)\right) = \int \frac{\partial l}{\partial \theta} \cdot f(\boldsymbol{x}|\theta_0)\,\mathrm{d}\boldsymbol{x} = \int \sum_{i=1}^{n} \frac{\partial \log f(x_i|\theta_0)}{\partial \theta} \prod_{i=1}^{n} f(x_i|\theta_0)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

Note that

$$\begin{aligned}
&\int \frac{\partial \log f(x_1|\theta_0)}{\partial \theta} \prod_{i=1}^{n} f(x_i|\theta_0)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
=&\int \frac{\partial f(x_1|\theta_0)}{\partial \theta} \cdot \frac{1}{f(x_1|\theta_0)} \prod_{i=1}^{n} f(x_i|\theta_0)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
=&\int \left(\int \frac{\partial f(x_1|\theta_0)}{\partial \theta}\mathrm{d}x_1\right) \mathrm{E}_{\theta_0}\left(S(\boldsymbol{x}|\theta_0)\right) \prod_{i=2}^{n} f(x_i|\theta_0)\,\mathrm{d}x_2 \cdots \mathrm{d}x_n \\
=&\frac{\partial}{\partial \theta} \int f(x_1|\theta_0)\,\mathrm{d}x_1 \cdot \int \prod_{i=2}^{n} f(x_i|\theta_0)\,\mathrm{d}x_2 \cdots \mathrm{d}x_n \\
=&0.
\end{aligned}$$

Therefore

$$\mathrm{E}_{\theta_0}\left(S\left(\boldsymbol{x}|\theta_0\right)\right) = 0.$$

2 Fisher information

Fisher information (matrix) is the second order moment of the score function

$$I\left(\theta\right) = \mathrm{E}_{\theta_0}\left(S\left(\boldsymbol{x}|\theta\right) S\left(\boldsymbol{x}|\theta\right)^{\top}\right). \tag{2}$$

Since $\mathrm{E}_{\theta_0}\left(S\left(\boldsymbol{x}|\theta_0\right)\right) = 0$, the fisher information is the covariance matrix of $S\left(\boldsymbol{x}|\theta_0\right)$

$$I\left(\theta_0\right) = \mathrm{Var}_{\theta_0}\left(S\left(\boldsymbol{x}|\theta_0\right)\right).$$

Also it can be shown that

$$I\left(\theta_0\right) = -\mathrm{E}\left(\left.\frac{\partial^2 l}{\partial\theta^2}\right|_{\theta=\theta_0}\right),$$

and

$$I^{\star}\left(\theta_0\right) = I\left(\theta_0\right)/n,$$

where $I^{\star}\left(\theta_0\right)$ is the fisher info for only one single sample. Additionally

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} N\left(0, I^{\star}\left(\theta_0\right)^{-1}\right).$$

References