Inference 2, 2023, lecture 3

Rolf Larsson

November 7, 2023

Outline of the course

Chap. 3.2 Fisher information (continued):

- More on the Fisher Information
- The multivariate case

More on the Fisher Information

Recall: under regularity conditions, we have

Definition (3.2)

For every $\mathbf{x} \in \mathcal{A}$, we define the **score function** as the derivative of the log-likelihood function, i.e.

$$V(\theta; \mathbf{x}) = l'(\theta; \mathbf{x}) = \frac{\partial}{\partial \theta} \log\{L(\theta; \mathbf{x})\}.$$

Definition (3.3)

The **Fisher information** is defined by

$$I_{\mathbf{X}}(\theta) = \mathrm{Var}_{\theta}\{V(\theta; \mathbf{X})\}.$$

$$I_{\mathbf{X}}(\theta) = -\mathrm{E}_{\theta}\{l''(\theta; \mathbf{X})\} = -\mathrm{E}_{\theta}\left[\frac{\partial^2}{\partial \theta^2} \log\{p(\mathbf{X}; \theta)\}\right].$$

More on the Fisher Information

Theorem (3.3)

Let X and Y be independent random variables. If $I_X(\theta)$ and $I_Y(\theta)$ are the Fisher informations contained in X and Y, respectively, then the Fisher information contained in (X, Y) is given by

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta).$$

More on the Fisher Information

Corollary (3.1)

Let $\mathbf{X} = (X_1, ..., X_n)$*, where* $X_1, ..., X_n$ *are independent and distributed as* X*, that has Fisher information* $i(\theta) = I_X(\theta)$*. Then,*

$$I_{\mathbf{X}}(\theta) = nI_X(\theta).$$

Example 1:

- Let $\mathbf{X} = (X_1, ..., X_n)$, where $X_1, ..., X_n$ are independent and distributed as X, which is exponential with intensity β.
- Verify by direct calculation that $I_{\mathbf{X}}(\beta) = nI_X(\beta)$.

More on the Fisher Information

Theorem (3.4)

Let X be a random variable with distribution P_θ. Suppose that another parametrization is given by $\theta = h(\xi)$ where h is differentiable. Then, the information contained in X about ξ is given by

$$I_X^*(\xi) = I_X\{h(\xi)\}\{h'(\xi)\}^2.$$

Here, $I_X(\theta)$ is the information that $X \sim \mathrm{P}_\theta$ contains about θ and $I_X^*(\xi)$ is the information that $X \sim \mathrm{P}_{h(\xi)}$ contains about ξ.

More on the Fisher Information

Example 1':

- Let $\mathbf{X} = (X_1, ..., X_n)$, where $X_1, ..., X_n$ are independent and distributed as X, which is exponential with intensity β.
- Let $\beta = 1/\mu$, where $\mu = E(X)$.
- Calculate the Fisher information in the μ parametrization
  1. by direct calculation.
  2. by using the Fisher information in the β parametrization and theorem 3.4.

More on the Fisher Information

Theorem (3.5)

Let $\mathbf{X}$ *be a random variable with* $\mathbf{X} \sim \mathrm{P}_\theta$*.*
Let $T(\mathbf{X})$ *be any statistic. Suppose that regularity conditions 1-3 hold. Then,*

$$I_{T(\mathbf{X})}(\theta) \leq I_{\mathbf{X}}(\theta),$$

where $I_{T(\mathbf{X})}$ *is computed with respect to the distribution of* $T(\mathbf{X})$

The multivariate case

Recall the regularity conditions when the parameter θ is a scalar:

1. The distributions $\{\mathrm{P} : \theta \in \Theta\}$ have common support, so that the set
$$\mathcal{A} = \{\mathbf{x} : p(\mathbf{x}; \theta) > 0\}$$
is independent of θ.
2. The parameter space Θ is an open interval (finite or infinite).
3. For any $\mathbf{x} \in \mathcal{A}$ and all $\theta \in \Theta$, the derivative $\partial p(\mathbf{x}; \theta)/\partial\theta$ exists and is finite.

Now, assume that $\theta = (\theta_1, ..., \theta_k)'$ is k-dimensional.

Modify conditions 2 and 3 as follows:

2'. The parameter space $\theta \subseteq \mathcal{R}^k$ is an open set.

3'. For all $\mathbf{x} \in \mathcal{A}$ the likelihood function has finite partial derivatives.

The multivariate case

Under regularity conditions 1, 2', 3':

Definition (3.5)

For all $\mathbf{x} \in \mathcal{A}$, the vector of partial derivatives (w.r.t. the parameters) of the log likelihood function

$$V(\theta; \mathbf{x}) = \left(\frac{\partial}{\partial \theta_1} l(\theta; \mathbf{x}), ..., \frac{\partial}{\partial \theta_k} l(\theta; \mathbf{x}) \right)'$$

is called the **score function** or score vector.

The multivariate case

Definition (3.5)

The $k \times k$ matrix

$$I_{\mathbf{X}}(\theta) = \mathrm{Cov}_\theta\{V(\theta; \mathbf{X})\}$$

is called the **Fisher information matrix**. The element in row j and column r is given by

$$I_{\mathbf{X}}(\theta)_{jr} = \mathrm{E}_\theta\left\{\frac{\partial}{\partial\theta_j} l(\theta; \mathbf{X}) \frac{\partial}{\partial\theta_r} l(\theta; \mathbf{X})\right\}.$$

The multivariate case

Regularity condition 4': For all $\mathbf{x} \in \mathcal{A}$ the likelihood function has second order partial derivatives and for all $\theta \in \Theta$ and $j, r = 1, ..., k$,

$$\frac{\partial^2}{\partial\theta_j\partial\theta_r}\int_{\mathcal{A}} f(\mathbf{x};\theta)d\mathbf{x} = \int_{\mathcal{A}} \frac{\partial^2}{\partial\theta_j\partial\theta_r} f(\mathbf{x};\theta)d\mathbf{x}$$

in the continuous case, and similarly in the discrete case.

Theorem (3.6)

Suppose regularity conditions 1, 2', 3' and 4' hold.
Let $J(\theta;\mathbf{X})$ (the observed Fisher information matrix) be a $k \times k$ matrix with elements

$$J(\theta;\mathbf{X})_{jr} = -\frac{\partial^2}{\partial\theta_j\partial\theta_r} l(\theta;\mathbf{X}), \quad j, r = 1, ..., k.$$

Then the Fisher information satisfies $I_{\mathbf{X}}(\theta) = \mathrm{E}_\theta\{J(\theta;\mathbf{X})\}$.

The multivariate case

Example 2:

- Suppose $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent and distributed as X
- and that we have a sample $\mathbf{x} = (x_1, ..., x_n)$.
- Calculate the Fisher information matrix of $X \sim N(\mu, \sigma^2)$ (both parameters unknown).

The multivariate case

Example 3:

- Suppose $\mathbf{X} = (X_1, ..., X_n)$ where the X_i are independent and distributed as X
- and that we have a sample $\mathbf{x} = (x_1, ..., x_n)$.
- Let X be multinomial such that the probability function is $p(k) = p_k > 0$ for $k = 1, 2, 3, 4$ and $p(k) = 0$ otherwise.
- Calculate the Fisher information matrix of X.

News of today

- For independent random variables, the Fisher information is additive.
- The Fisher information for a transformed parameter is given by a simple formula.
- A statistic does not contain more (Fisher) information than the original sample does.
- The multivariate case:
    - Score function
    - Fisher information