ECE531 Screencast 2.4: Fisher Information for Vector Parameters

D. Richard Brown III

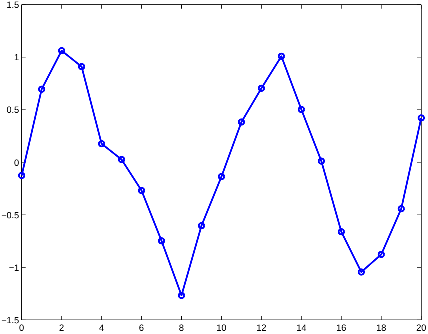
Worcester Polytechnic Institute

Vector Parameter Estimation Problems

In many problems, we have more than one parameter that we would like to estimate. For example,

$$Y_k = a\cos(\omega k + \phi) + W_k \text{ for } k = 0, 1, \ldots, n-1$$

where $a > 0$, $\phi \in (-\pi, \pi)$, and $\omega \in (0, \pi)$ are all non-random parameters and $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. In this problem $\theta = [a, \phi, \omega]$.



Fisher Information Matrix

Recall, in the scalar parameter case, the Fisher information was motivated by a computation of the mean squared relative slope of the likelihood function:

$$I(\theta) \quad := \quad \mathrm{E}\left[\left(\frac{\frac{\partial}{\partial\theta} p_Y(y\,;\,\theta)}{p_Y(y\,;\,\theta)}\right)^2\right] = \int_{\mathcal{Y}} \left(\frac{\partial}{\partial\theta} \ln p_Y(Y\,;\,\theta)\right)^2 p_Y(y\,;\,\theta)\,dy$$

In multiparameter problems, we are now concerned with the relative steepness of the likelihood function with respect to each of the parameters. A natural choice (assuming that all of the required derivatives exist) would be

$$I(\theta) = \mathrm{E}\left[(\nabla_\theta \ln p_Y(Y\,;\,\theta))\,(\nabla_\theta \ln p_Y(Y\,;\,\theta))^\top\right] \in \mathbb{R}^{m\times m}$$

where ∇_x is the gradient operator defined as

$$\nabla_x f(x) \quad := \quad \left[\frac{\partial}{\partial x_0} f(x), \ldots, \frac{\partial}{\partial x_{m-1}} f(x)\right]^\top.$$

Fisher Information Matrix

Let $p := p_Y(Y\,;\,\theta)$. The Fisher information matrix is then

$$I(\theta) = \begin{bmatrix} \mathrm{E}\left[\frac{\partial}{\partial\theta_0}\ln p \cdot \frac{\partial}{\partial\theta_0}\ln p\right] & \dots & \mathrm{E}\left[\frac{\partial}{\partial\theta_0}\ln p \cdot \frac{\partial}{\partial\theta_{m-1}}\ln p\right] \\ \mathrm{E}\left[\frac{\partial}{\partial\theta_1}\ln p \cdot \frac{\partial}{\partial\theta_0}\ln p\right] & \dots & \mathrm{E}\left[\frac{\partial}{\partial\theta_1}\ln p \cdot \frac{\partial}{\partial\theta_{m-1}}\ln p\right] \\ \vdots & \ddots & \vdots \\ \mathrm{E}\left[\frac{\partial}{\partial\theta_{m-1}}\ln p \cdot \frac{\partial}{\partial\theta_0}\ln p\right] & \dots & \mathrm{E}\left[\frac{\partial}{\partial\theta_{m-1}}\ln p \cdot \frac{\partial}{\partial\theta_{m-1}}\ln p\right] \end{bmatrix}$$

Note that the ijth element of the Fisher information matrix is given as

$$\boxed{I_{ij}(\theta) = \mathrm{E}\left[\frac{\partial}{\partial\theta_i}\ln p_Y\{Y\,;\,\theta\} \cdot \frac{\partial}{\partial\theta_j}\ln p_Y\{Y\,;\,\theta\}\right]}$$

hence we can say that $I(\theta)$ is symmetric.

Fisher Information Matrix

When the second derivatives all exist, we can write

$$\frac{\partial^2}{\partial\theta_i\theta_j}\ln p_Y(y\,;\,\theta) = \frac{\frac{\partial^2}{\partial\theta_i\theta_j}p_Y(y\,;\,\theta)}{p_Y(y\,;\,\theta)} - \frac{\frac{\partial}{\partial\theta_i}p_Y(y\,;\,\theta)}{p_Y(y\,;\,\theta)}\frac{\frac{\partial}{\partial\theta_j}p_Y(y\,;\,\theta)}{p_Y(y\,;\,\theta)}$$

and, under the theorem's assumptions, we can write

$$\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\theta_j}\ln p_Y(y\,;\,\theta)\right] = -\mathrm{E}\left[\frac{\partial}{\partial\theta_i}\ln p_Y(y\,;\,\theta)\cdot\frac{\partial}{\partial\theta_j}\ln p_Y(y\,;\,\theta)\right] = -I_{ij}(\theta).$$

Hence, we can say that

$$\boxed{I_{ij}(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\theta_j}\ln p_Y(y\,;\,\theta)\right]}$$

This expression is often more convenient to compute than the former expression for $I_{ij}(\theta)$.

Fisher Information Matrix

Under the conditions of the theorem

$$\begin{aligned} \mathrm{E}\left[\frac{\partial}{\partial\theta_i}\ln p_Y\{Y\,;\,\theta\}\right] &= \int_{\mathcal{Y}} \frac{\frac{\partial}{\partial\theta_i}p_Y(y\,;\,\theta)}{p_Y(y\,;\,\theta)} p_Y(y\,;\,\theta)\,dy \\ &= \frac{\partial}{\partial\theta_i}\int_{\mathcal{Y}} p_Y(y\,;\,\theta)\,dy = 0 \end{aligned}$$

Hence

$$I_{ij}(\theta) = \mathrm{cov}\left\{\frac{\partial}{\partial\theta_i}\ln p_Y\{Y\,;\,\theta\}, \frac{\partial}{\partial\theta_j}\ln p_Y\{Y\,;\,\theta\}\right\}.$$

The Fisher information matrix $I(\theta)$ is a covariance matrix and is invertible if the unknown parameters are linearly independent.

Example: Fisher Information Matrix of Signal in AWGN

Many problems require the estimation of unknown signal parameters in additive white Gaussian noise. The observations in this case can be modeled as

$$Y_k \quad = \quad s_k(\theta) + W_k \text{ for } k = 0, 1, \dots, n-1$$

where $s_k(\theta) : \Lambda \mapsto \mathbb{R}$ is a deterministic signal with an unknown **vector** parameter θ and where $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$. We assume σ^2 is a known parameter and that all of the regularity conditions are satisfied.

To compute the ijth element of the Fisher information matrix, we can write

$$\frac{\partial^2}{\partial\theta_i\theta_j} \ln p_Y(Y\,;\,\theta) = \frac{1}{\sigma^2} \sum_{k=0}^{n-1} \left\{ [Y_k - s_k(\theta)] \frac{\partial^2}{\partial\theta_i\theta_j} s_k(\theta) - \left(\frac{\partial}{\partial\theta_i} s_k(\theta)\right) \left(\frac{\partial}{\partial\theta_j} s_k(\theta)\right) \right\}$$

Since $\mathrm{E}[Y_k] = s_k(\theta)$, the ijth element of the FIM can be written as

$$I_{ij}(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\theta_j} \ln p_Y(Y\,;\,\theta)\right] = \frac{1}{\sigma^2} \sum_{k=0}^{n-1} \left(\frac{\partial}{\partial\theta_i} s_k(\theta)\right) \left(\frac{\partial}{\partial\theta_j} s_k(\theta)\right)$$