Derivation of vector CRB

Magnus Jansson

2017/2019

Define the (vector) score function as

$$s(\boldsymbol{\theta}, \boldsymbol{x}) = \nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{x}; \boldsymbol{\theta}). \tag{1}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]$. Note that

$$\begin{aligned} \mathbb{E}[\boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x})] &= \int \Big(\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{x}; \boldsymbol{\theta}) \Big) p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x} \\ &= \int \nabla_{\boldsymbol{\theta}} p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x} \\ &\overset{(A_1)}{=} \nabla_{\boldsymbol{\theta}} \underbrace{\int p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x}}_{=1} = 0. \end{aligned} \tag{2}$$

That is, under the regularity condition (A_1), the score has mean zero. The covariance of $\boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x})$ is

$$\begin{aligned} \mathbb{E}[\boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x}) \boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x})^T] &= \mathbb{E}\Big[\Big(\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{x}, \boldsymbol{\theta}) \Big) \Big(\nabla_{\boldsymbol{\theta}} \ln p(\boldsymbol{x}, \boldsymbol{\theta}) \Big)^T \Big] \\ &\triangleq I(\boldsymbol{\theta}), \text{ the Fisher information matrix.} \end{aligned} \tag{3}$$

Let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ be any unbiased estimator of $\boldsymbol{\theta}$. That is,

$$\mathbb{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = 0 = \int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}) p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x}. \tag{4}$$

Computing the derivative on both sides for each θ_j, we obtain

$$\begin{aligned} \mathbf{0} &= \frac{\partial}{\partial \theta_j} \int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}) p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x} \\ &\overset{(A_1)}{=} \int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x} - \int \underbrace{(0, \ldots, 1, \ldots, 0)^T}_{1 \text{ at } j\text{-th position}} p(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x} \\ &= \int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Big(\frac{\partial}{\partial \theta_j} \ln p(\boldsymbol{x}; \theta) \Big) p(\boldsymbol{x}; \theta) d\boldsymbol{x} - (0, \ldots, 1, \ldots, 0)^T. \end{aligned} \tag{5}$$

Hence, under (A_1) we have

$$\underbrace{\mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x})^T]}_{\text{cross cov.}} = \mathbf{I}. \tag{6}$$

Now define the joint vector

$$\begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \\ \boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{x}) \end{bmatrix}, \tag{7}$$

which has mean zero. Its covariance matrix is

$$\mathbb{E}\left(\begin{bmatrix}\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\\ \boldsymbol{s}(\boldsymbol{\theta},\boldsymbol{x})\end{bmatrix}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T, \boldsymbol{s}^T(\boldsymbol{\theta},\boldsymbol{x})]\right) = \begin{bmatrix}\boldsymbol{C} & \mathbf{I}\\ \mathbf{I} & \mathbf{I}(\boldsymbol{\theta})\end{bmatrix} \triangleq \boldsymbol{Q}, \tag{8}$$

where $\boldsymbol{C} \triangleq \mathbb{E}\big[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\big]$.
$\boldsymbol{Q}$ can be block diagonalized as

$$\underbrace{\begin{bmatrix}\mathbf{I} & -\mathbf{I}^{-1}(\boldsymbol{\theta})\\ 0 & \mathbf{I}\end{bmatrix}}_{\boldsymbol{T}}\underbrace{\begin{bmatrix}\boldsymbol{C} & \mathbf{I}\\ \mathbf{I} & \mathbf{I}(\boldsymbol{\theta})\end{bmatrix}}_{\boldsymbol{Q}}\underbrace{\begin{bmatrix}\mathbf{I} & 0\\ -\mathbf{I}^{-1}(\boldsymbol{\theta}) & \mathbf{I}\end{bmatrix}}_{\boldsymbol{T}^T} = \begin{bmatrix}\boldsymbol{C}-\mathbf{I}(\boldsymbol{\theta})^{-1} & 0\\ 0 & \mathbf{I}\end{bmatrix}. \tag{9}$$

$\boldsymbol{Q}$ is a covariance matrix and hence positive semidefinite. Then so must $\boldsymbol{C}-\mathbf{I}^{-1}(\boldsymbol{\theta})$ be:

$$\boldsymbol{C}-\mathbf{I}^{-1}(\boldsymbol{\theta}) \succeq \mathbf{0} \quad \text{or} \quad \boldsymbol{C} \succeq \mathbf{I}^{-1}(\boldsymbol{\theta}) \tag{10}$$

which is the CRB! Looking at the diagonal elements, we have

$$\boldsymbol{C}_{ii} = \operatorname{var}(\hat{\theta}_i - \theta_i) \geq [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}. \tag{11}$$

That is, the variance of a single parameter is lower bounded by the corresponding diagonal element of the inverse Fisher information matrix.

The estimator is said to be statistically efficient if

$$\boldsymbol{C} = \mathbf{I}^{-1}(\boldsymbol{\theta}). \tag{12}$$

This happens if (and essentially only if) $\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} = \mathbf{I}^{-1}(\boldsymbol{\theta})\boldsymbol{s}(\boldsymbol{\theta},\boldsymbol{x})$ (compare with the diagonalization operation in (9)) or

$$\boldsymbol{s}(\boldsymbol{\theta},\boldsymbol{x}) = \mathbf{I}(\boldsymbol{\theta})\big(\hat{\boldsymbol{\theta}}(\boldsymbol{x})-\boldsymbol{\theta}\big),$$

which is the CRB condition for equality.

Further note that if

$$A_2\colon \int \frac{\partial^2}{\partial\theta_i\partial\theta_j}p(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}\int p(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x} = 0 \tag{13}$$

then

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \mathbb{E}\big[\big\{\nabla_{\boldsymbol{\theta}}\ln p(\boldsymbol{x};\boldsymbol{\theta})\big\}\big\{\nabla_{\boldsymbol{\theta}}\ln p(\boldsymbol{x};\boldsymbol{\theta})\big\}^T\big]\\ &= -\big[\mathbb{E}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\boldsymbol{x};\boldsymbol{\theta})\big].\end{aligned} \tag{14}$$

Proof. We have

$$\frac{\partial}{\partial\theta_i}\ln p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{p(\boldsymbol{x};\boldsymbol{\theta})}\frac{\partial p(\boldsymbol{x};\boldsymbol{\theta})}{\partial\theta_i},$$

which leads to

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{p(\boldsymbol{x};\boldsymbol{\theta})}\frac{\partial^2}{\partial\theta_i\partial\theta_j}p(\boldsymbol{x};\boldsymbol{\theta}) - \underbrace{\frac{1}{p(\boldsymbol{x};\boldsymbol{\theta})^2}\frac{\partial}{\partial\theta_i}p(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}p(\boldsymbol{x};\boldsymbol{\theta})}_{\frac{\partial}{\partial\theta_i}\ln p(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\ln p(\boldsymbol{x};\boldsymbol{\theta})}. \tag{15}$$

Multiplying both sides with $p(\boldsymbol{x};\boldsymbol{\theta})$ and taking the integral over $\boldsymbol{x}$, i.e. performing $\mathbb{E}[\cdot]$, gives us the desired result. □