Cramer-Rao in the multiparameter case

$\boldsymbol{\theta} = (\theta_1, \ldots \theta_k)$

Define the Score function $\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial}{\partial\theta_1} \log f(\boldsymbol{x}|\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial\theta_i} \log f(\boldsymbol{x}|\boldsymbol{\theta}) \end{bmatrix} = \nabla \log f(\boldsymbol{x}|\boldsymbol{\theta})$

Define the Fisher information $I(\boldsymbol{\theta}) = Cov\left[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})\right]$

We have as in the univariate case that $E\left[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})\right] = \mathbf{0}$ and $I(\boldsymbol{\theta}) = E\left[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})^T\right] =$ $-E\left[H(\boldsymbol{X}|\boldsymbol{\theta})\right]$ where $h_{ij} = \frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\log f(\boldsymbol{x}|\boldsymbol{\theta})$.

Let $\tau = \tau(\boldsymbol{\theta})$ be univariate and let $\nabla\tau(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial}{\partial\theta_1}\tau(\boldsymbol{\theta}) \\ \vdots \\ \frac{\partial}{\partial\theta_k}\tau(\boldsymbol{\theta}) \end{bmatrix}$

Theorem. For an estimator $W(\boldsymbol{X})$ with $E\left[W(\boldsymbol{X})\right] = \tau$, we have under similar regularity conditions as in the univariate case that $Var\left[W(\boldsymbol{X})\right] \geq (\nabla\tau(\boldsymbol{\theta}))^T (I(\boldsymbol{\theta}))^{-1} (\nabla\tau(\boldsymbol{\theta}))$.

Proof

$$\frac{\partial}{\partial\theta_i}\tau(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\int W(\boldsymbol{x}) f(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x} = \int W(\boldsymbol{x})\frac{\partial}{\partial\theta_i} f(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x} = \int W(\boldsymbol{x})\left(\frac{\partial}{\partial\theta_i}\log f(\boldsymbol{x},\boldsymbol{\theta})\right) f(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x} = E\left[W(\boldsymbol{X})S_i(\boldsymbol{X}|\boldsymbol{\theta})\right]$$

where $S_i(\boldsymbol{X}|\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i}\log f(\boldsymbol{X},\boldsymbol{\theta})$. This implies: $\nabla\tau(\boldsymbol{\theta}) = E\left[W(\boldsymbol{X})\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})\right]$.

Since $\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})$ is a vector we know introduce a scalar $U(\boldsymbol{X}|\boldsymbol{\theta}) = (\nabla\tau(\boldsymbol{\theta}))^T (I(\boldsymbol{\theta}))^{-1} \boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})$.
We obtain:

$$Cov\left[W(\boldsymbol{X}), U(\boldsymbol{X}|\boldsymbol{\theta})\right] = (\nabla\tau(\boldsymbol{\theta}))^T (I(\boldsymbol{\theta}))^{-1} E\left[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})W(\boldsymbol{X})\right] = (\nabla\tau(\boldsymbol{\theta}))^T (I(\boldsymbol{\theta}))^{-1} (\nabla\tau(\boldsymbol{\theta}))$$

and using that $Var\left[\boldsymbol{a}^T\boldsymbol{X}\right]=\boldsymbol{a}^T Cov\left[\boldsymbol{X}\right]\boldsymbol{a}$ we get

$$Var\left[U\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\right]=\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)^T\left(I\left(\boldsymbol{\theta}\right)\right)^{-1}\left(I\left(\boldsymbol{\theta}\right)\right)\left(I\left(\boldsymbol{\theta}\right)\right)^{-1}\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)=\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)^T\left(I\left(\boldsymbol{\theta}\right)\right)^{-1}\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)$$

From Cauchy Schwartz we then have that :

$$\left(Cov\left[W\left(\boldsymbol{X}\right),U\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\right]\right)^2\leq Var\left[W\left(\boldsymbol{X}\right)\right]Var\left[U\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\right]$$

or $\left[\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)^T\left(I\left(\boldsymbol{\theta}\right)\right)^{-1}\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)\right]^2\leq\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)^T\left(I\left(\boldsymbol{\theta}\right)\right)^{-1}\left(\nabla\tau\left(\boldsymbol{\theta}\right)\right)Var\left[W\left(\boldsymbol{X}\right)\right]$