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 $I(\boldsymbol{\theta}) = Cov[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})]$

We have as in the univariate case that $E[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})] = \mathbf{0}$ and $I(\boldsymbol{\theta}) = E\left[\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})\boldsymbol{S}(\boldsymbol{X}|\boldsymbol{\theta})^T\right] =$ $-E[H(\boldsymbol{X}|\boldsymbol{\theta})]$ 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 $E[W(\boldsymbol{X})] = \tau$, we have under similar regularity conditions as in the univariate case that $Var[W(\boldsymbol{X})] \geq (\nabla \tau(\boldsymbol{\theta}))^T (I(\boldsymbol{\theta}))^{-1} (\nabla \tau(\boldsymbol{\theta}))$.

Proof

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

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

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:

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

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

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

From Cauchy Schwartz we then have that :

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

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

Let $X_1, \ldots, X_n$ be iid with pdf/pmf $f\left(x|\boldsymbol{\theta}\right)$ and $Cov\left[\boldsymbol{S}\left(X_i|\boldsymbol{\theta}\right)\right] = \boldsymbol{I}(\boldsymbol{\theta})$. Then $Cov\left[\boldsymbol{S}\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\right] = n\boldsymbol{I}(\boldsymbol{\theta})$. If $\hat{\boldsymbol{\theta}}_n$ is the MLE of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}$, then

$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{D} N\left(\boldsymbol{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})\right)$ or $Cov\left(\hat{\boldsymbol{\theta}}_n\right) \approx \frac{1}{n} I^{-1}(\boldsymbol{\theta})$ or $Cov\left[\boldsymbol{S}\left(\boldsymbol{X}|\boldsymbol{\theta}\right)\right]^{-1}$.