Solution to Exercise 10.12

⋆10.12 Let $\tilde{\boldsymbol{\theta}}$ denote any unbiased estimator of the k parameters of a parametric model fully specified by the loglikelihood function $\ell(\boldsymbol{\theta})$. The unbiasedness property can be expressed as the following identity:

$$\int L(\boldsymbol{y}, \boldsymbol{\theta})\,\tilde{\boldsymbol{\theta}}\,d\boldsymbol{y} = \boldsymbol{\theta}. \tag{10.105}$$

By using the relationship between $L(\boldsymbol{y}, \boldsymbol{\theta})$ and $\ell(\boldsymbol{y}, \boldsymbol{\theta})$ and differentiating this identity with respect to the components of $\boldsymbol{\theta}$, show that

$$\mathrm{Cov}_{\boldsymbol{\theta}}(\boldsymbol{g}(\boldsymbol{\theta}), (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})) = \mathbf{I},$$

where $\mathbf{I}$ is a $k \times k$ identity matrix, and the notation $\mathrm{Cov}_{\boldsymbol{\theta}}$ indicates that the covariance is to be calculated under the DGP characterized by $\boldsymbol{\theta}$.

Let $\boldsymbol{V}$ denote the $2k \times 2k$ covariance matrix of the $2k$–vector obtained by stacking the k components of $\boldsymbol{g}(\boldsymbol{\theta})$ above the k components of $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Partition this matrix into 4 $k \times k$ blocks as follows:

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{C} \\ \boldsymbol{C}^\top & \boldsymbol{V}_2 \end{bmatrix},$$

where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are, respectively, the covariance matrices of the vectors $\boldsymbol{g}(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ under the DGP characterized by $\boldsymbol{\theta}$. Then use the fact that $\boldsymbol{V}$ is positive semidefinite to show that the difference between $\boldsymbol{V}_2$ and $\boldsymbol{I}^{-1}(\boldsymbol{\theta})$, where $\boldsymbol{I}(\boldsymbol{\theta})$ is the (finite-sample) information matrix for the model, is a positive semidefinite matrix. **Hint:** Use the result of Exercise 7.11.

Since the right-hand side of equation (10.105) is the vector $\boldsymbol{\theta}$, its derivative with respect to the vector $\boldsymbol{\theta}$ must be a $k \times k$ identity matrix. Moreover, because $\ell(\boldsymbol{y}, \boldsymbol{\theta}) = \log L(\boldsymbol{y}, \boldsymbol{\theta})$,

$$\frac{\partial L(\boldsymbol{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = L(\boldsymbol{y}, \boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}).$$

Therefore, the derivative of the identity (10.105) is

$$\int L(\boldsymbol{y}, \boldsymbol{\theta})\,\tilde{\boldsymbol{\theta}}\boldsymbol{g}^\top(\boldsymbol{y}, \boldsymbol{\theta})\,d\boldsymbol{y} = \mathbf{I}. \tag{S10.25}$$

The left-hand side of (S10.25) is simply the covariance matrix of $\boldsymbol{g}(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\theta}}$, where we write $\boldsymbol{g}(\boldsymbol{\theta})$ for $\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{y})$, taken with respect to the distribution characterized by $\boldsymbol{\theta}$. But since $\mathrm{E}_{\boldsymbol{\theta}}\,\boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, it is also the covariance matrix of $\boldsymbol{g}(\boldsymbol{\theta})$ and $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Therefore, we see that

$$\mathrm{Cov}_{\boldsymbol{\theta}}\big(\boldsymbol{g}(\boldsymbol{\theta}), (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})\big) = \mathbf{I},$$

which is what we were required to show.

Now consider the $2k \times 2k$ matrix $\boldsymbol{V}$ defined in the question. We have just seen that the off-diagonal blocks $\boldsymbol{C}$ and $\boldsymbol{C}^\top$ are $k \times k$ identity matrices. The upper left-hand $k \times k$ block $\boldsymbol{V}_1$ is the covariance matrix of the gradient vector $\boldsymbol{g}(\boldsymbol{\theta})$, which, by definition (as we saw in Exercise 10.5), is the information matrix $\boldsymbol{I}(\boldsymbol{\theta})$. Therefore

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{I}(\boldsymbol{\theta}) & \mathbf{I} \\ \mathbf{I} & \boldsymbol{V}_2 \end{bmatrix}.$$

Since $\boldsymbol{V}$ is a covariance matrix, it must be positive semidefinite. This implies that its inverse must also be positive semidefinite, as must each of the diagonal blocks of the inverse. By the result of Exercise 7.11, the lower right-hand block of the inverse is

$$\big(\boldsymbol{V}_2 - \mathbf{I}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\mathbf{I}\big)^{-1} = \big(\boldsymbol{V}_2 - \boldsymbol{I}^{-1}(\boldsymbol{\theta})\big)^{-1}.$$

Since this matrix is positive semidefinite, the matrix $\boldsymbol{V}_2 - \boldsymbol{I}^{-1}(\boldsymbol{\theta})$ must also be positive semidefinite. Because this is the difference between the covariance matrix of an arbitrary unbiased estimator and the inverse of the information matrix, the Cramér-Rao result has been proved.