Since $\exp(\ell_t(\boldsymbol{y}^t, \boldsymbol{\theta}))$ is, for the DGP characterized by $\boldsymbol{\theta}$, the density of y_t conditional on $\boldsymbol{y}^{t-1}$, this last equation, along with the definition (10.26), gives

$$\mathrm{E}_{\boldsymbol{\theta}}\big(G_{ti}(\boldsymbol{y}^t, \boldsymbol{\theta}) \,|\, \boldsymbol{y}^{t-1}\big) = 0 \tag{10.29}$$

for all $t = 1, \ldots, n$ and $i = 1, \ldots, k$. The notation "$\mathrm{E}_{\boldsymbol{\theta}}$" here means that the expectation is being taken under the DGP characterized by $\boldsymbol{\theta}$. Taking unconditional expectations of (10.29) yields the desired result. Summing (10.29) over $t = 1, \ldots, n$ shows that $\mathrm{E}_{\boldsymbol{\theta}}(g_i(\boldsymbol{y}, \boldsymbol{\theta})) = 0$ for $i = 1, \ldots, k$, or, equivalently, that $\mathrm{E}_{\boldsymbol{\theta}}(\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta})) = \mathbf{0}$.

In addition to the conditional expectations of the elements of the matrix $\boldsymbol{G}(\boldsymbol{y}, \boldsymbol{\theta})$, we can compute the covariances of these elements. Let $t \neq s$, and suppose, without loss of generality, that $t < s$. Then the covariance under the DGP characterized by $\boldsymbol{\theta}$ of the ti^{th} and sj^{th} elements of $\boldsymbol{G}(\boldsymbol{y}, \boldsymbol{\theta})$ is

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\theta}}\big(G_{ti}(\boldsymbol{y}^t, \boldsymbol{\theta})G_{sj}(\boldsymbol{y}^s, \boldsymbol{\theta})\big) &= \mathrm{E}_{\boldsymbol{\theta}}\Big(\mathrm{E}_{\boldsymbol{\theta}}\big(G_{ti}(\boldsymbol{y}^t, \boldsymbol{\theta})G_{sj}(\boldsymbol{y}^s, \boldsymbol{\theta})\big) \,|\, \boldsymbol{y}^t\big)\Big) \\ &= \mathrm{E}_{\boldsymbol{\theta}}\Big(G_{ti}(\boldsymbol{y}^t, \boldsymbol{\theta})\mathrm{E}_{\boldsymbol{\theta}}\big(G_{sj}(\boldsymbol{y}^s, \boldsymbol{\theta}) \,|\, \boldsymbol{y}^t\big)\Big) = 0. \end{aligned} \tag{10.30}$$

The step leading to the second line above follows because $G_{ti}(\cdot)$ is a deterministic function of $\boldsymbol{y}^t$, and the last step follows because the expectation of $G_{sj}(\cdot)$ is zero conditional on $\boldsymbol{y}^{s-1}$, by (10.29), and so also conditional on the subvector $\boldsymbol{y}^t$ of $\boldsymbol{y}^{s-1}$. The above proof shows that the covariance of the two matrix elements is also zero conditional on $\boldsymbol{y}^t$.

The Information Matrix and the Hessian

The covariance matrix of the elements of the t^{th} row $\boldsymbol{G}_t(\boldsymbol{y}^t, \boldsymbol{\theta})$ of $\boldsymbol{G}(\boldsymbol{y}, \boldsymbol{\theta})$ is the $k \times k$ matrix $\boldsymbol{I}_t(\boldsymbol{\theta})$, of which the ij^{th} element is $\mathrm{E}_{\boldsymbol{\theta}}(G_{ti}(\boldsymbol{y}^t, \boldsymbol{\theta})G_{tj}(\boldsymbol{y}^t, \boldsymbol{\theta}))$. As a covariance matrix, $\boldsymbol{I}_t(\boldsymbol{\theta})$ is normally positive definite. The sum of the matrices $\boldsymbol{I}_t(\boldsymbol{\theta})$ over all t is the $k \times k$ matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) \equiv \sum_{t=1}^{n} \boldsymbol{I}_t(\boldsymbol{\theta}) = \sum_{t=1}^{n} \mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{G}_t^{\top}(\boldsymbol{y}^t, \boldsymbol{\theta})\boldsymbol{G}_t(\boldsymbol{y}^t, \boldsymbol{\theta})\big), \tag{10.31}$$

which is called the **information matrix**. The matrices $\boldsymbol{I}_t(\boldsymbol{\theta})$ are the **contributions** to the information matrix made by the successive observations.

An equivalent definition of the information matrix, as readers are invited to show in Exercise 10.5, is $\boldsymbol{I}(\boldsymbol{\theta}) \equiv \mathrm{E}_{\boldsymbol{\theta}}(\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta})\boldsymbol{g}^{\top}(\boldsymbol{y}, \boldsymbol{\theta}))$. In this second form, the information matrix is the expectation of the **outer product of the gradient** with itself; see Section 1.4 for the definition of the outer product of two vectors. Less exotically, it is just the covariance matrix of the score vector. As the name suggests, and as we will see shortly, the information matrix is a measure of the total amount of information about the parameters in the sample. The requirement that it should be positive definite is a condition