relation (8.56) with respect to the elements of $\boldsymbol{\theta}$, interchanging the order of the operations of differentiation and integration, and taking the limit as $n \to \infty$. We omit discussion of the regularity conditions necessary for this to be admissible and proceed directly to write down the result of differentiating the j^{th} component of (8.56) with respect to the i^{th} component of $\boldsymbol{\theta}$:

$$\lim_{n\to\infty} \int_{\mathcal{Y}^n} L^n(\boldsymbol{y}^n, \boldsymbol{\theta}) \frac{\partial \ell^n(\boldsymbol{y}^n, \boldsymbol{\theta})}{\partial \theta_i}\, \hat{\theta}_j(\boldsymbol{y}^n)\, d\boldsymbol{y}^n = \delta^i_j. \tag{8.57}$$

The right-hand side of this equation is the Kronecker delta, equal to 1 when $i = j$ and equal to 0 otherwise. Equation (8.57) can be rewritten as

$$\lim_{n\to\infty} E_\theta\left(n^{-1/2} \frac{\partial \ell^n(\boldsymbol{y}^n, \boldsymbol{\theta})}{\partial \theta_i}\, n^{1/2}\big(\hat{\theta}_j - \theta_j\big)\right) = \delta^i_j, \tag{8.58}$$

where we have put in some powers of n to ensure that the quantities which appear in the expression have probability limits of order unity. We have also subtracted θ_j from $\hat{\theta}_j$; this was possible because $E_\theta\big(D_\theta\, \ell(\boldsymbol{\theta})\big) = \mathbf{0}$, and hence θ_j times $E_\theta\big(D_\theta\, \ell(\boldsymbol{\theta})\big)$ is also equal to zero.

Expression (8.58) can be written without any limiting operation if we use the limiting distributions of the gradient $D_\theta\, \ell$ and the vector $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. Let us introduce a little more notation for the purposes of discussing limiting random variables. We make the definitions

$$\boldsymbol{s}^n(\boldsymbol{\theta}) \equiv n^{-1/2} \boldsymbol{g}(\boldsymbol{y}^n, \boldsymbol{\theta}), \quad \boldsymbol{s}(\boldsymbol{\theta}) \equiv \operatorname*{plim}_{n\to\infty}{}_\theta\, \boldsymbol{s}^n(\boldsymbol{\theta}), \tag{8.59}$$

$$\hat{\boldsymbol{t}}^n(\boldsymbol{\theta}) \equiv n^{1/2}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\big), \text{ and } \hat{\boldsymbol{t}}(\boldsymbol{\theta}) \equiv \operatorname*{plim}_{n\to\infty}{}_\theta\, \hat{\boldsymbol{t}}^n(\boldsymbol{\theta}). \tag{8.60}$$

Thus $\boldsymbol{s}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{t}}(\boldsymbol{\theta})$ are k–vectors with typical elements $s_i(\boldsymbol{\theta})$ and $\hat{t}_j(\boldsymbol{\theta})$, respectively. The former is the limiting value of $n^{-1/2}$ times a typical element of the gradient of $\ell(\boldsymbol{y}, \boldsymbol{\theta})$, while the latter is the limiting value of $n^{1/2}$ times a typical element of the difference between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$. The notation is intended to be mnemonic, $\boldsymbol{s}(\boldsymbol{\theta})$ corresponding to the *score* vector and $\hat{\boldsymbol{t}}(\boldsymbol{\theta})$ corresponding to *theta hat*. In this convenient new notation, expression (8.58) becomes

$$E_\theta\big(\hat{\boldsymbol{t}}(\boldsymbol{\theta}) \boldsymbol{s}^\top(\boldsymbol{\theta})\big) = \mathbf{I}_k, \tag{8.61}$$

where $\mathbf{I}_k$ is simply the $k \times k$ identity matrix.

It is not generally true for *any* consistent estimator that the plim in (8.60) exists or, if it does, is not zero. The class of estimators for which it exists and is nonzero is called the class of **root-n consistent estimators**. As we discussed in Chapter 5, this means that the rate of convergence, as $n \to \infty$, of the estimator $\hat{\boldsymbol{\theta}}$ to the true value $\boldsymbol{\theta}$ is the same as the rate of convergence of $n^{-1/2}$ to zero. The existence of a nonzero plim in (8.60) clearly implies just that, and we have already shown that the ML estimator is root-n consistent.