and

$$\int_{\mathscr{X}} \frac{\partial^2 \hat{p}(\mathbf{x}|\boldsymbol{\varphi})}{\partial (\boldsymbol{\varphi})_{[i]} \partial (\boldsymbol{\varphi})_{[j]}} d\mathbf{x} = 0 \;\; \forall\, i, j \in \{1, 2, \cdots, M\} \tag{7.130}$$

It is important to note that since due to the *regularity assumption* 1, $\hat{p}(\mathbf{x}|\boldsymbol{\varphi})$ is a $\mathfrak{C}^3$ continuous function in the interval $[\boldsymbol{\varphi}, \boldsymbol{\varphi} + \Delta\boldsymbol{\varphi}]$, then based on Definition 24.19 and Property 24.6 all its derivatives up to the third derivative are bounded. Therefore, this property is implied and need not be listed.[7]

Given the above *regularity conditions*, Equation 7.126 may be simplified as follows,

$$\mathscr{D}_{KL}(\boldsymbol{\varphi} \to \hat{\boldsymbol{\varphi}}) = \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{M} (\boldsymbol{\mathscr{I}}_F)_{[i][j]} (\Delta\boldsymbol{\varphi})_{[i]} (\Delta\boldsymbol{\varphi})_{[j]} \tag{7.131}$$

where $\boldsymbol{\mathscr{I}}_F$ is the *Fisher information matrix* and its elements are given by the following definition,

$$(\boldsymbol{\mathscr{I}}_F)_{[i][j]} \stackrel{\Delta}{=} \int_X \hat{p}(\mathbf{x}|\boldsymbol{\varphi}) \left(\frac{1}{\hat{p}(\mathbf{x}|\boldsymbol{\varphi})} \frac{\partial \hat{p}(\mathbf{x}|\boldsymbol{\varphi})}{\partial (\boldsymbol{\varphi})_{[i]}} \right) \left(\frac{1}{\hat{p}(\mathbf{x}|\boldsymbol{\varphi})} \frac{\partial \hat{p}(\mathbf{x}|\boldsymbol{\varphi})}{\partial (\boldsymbol{\varphi})_{[j]}} \right) d\mathbf{x} \tag{7.132}$$

Using Equation 7.124, we may write the *Fisher information matrix* in terms of the *log-likelihood* as follows,

$$(\boldsymbol{\mathscr{I}}_F)_{[i][j]} = \int_X \hat{p}(\mathbf{x}|\boldsymbol{\varphi}) \left(\frac{\partial \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi}))}{\partial (\boldsymbol{\varphi})_{[i]}} \right) \left(\frac{\partial \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi}))}{\partial (\boldsymbol{\varphi})_{[j]}} \right) d\mathbf{x} \tag{7.133}$$

Equation 7.133 may be seen as the *expected value* of the product of partial derivatives of the *log-likelihood*, namely,

$$(\boldsymbol{\mathscr{I}}_F)_{[i][j]} = \mathscr{E}\left\{ \left(\frac{\partial \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi}))}{\partial (\boldsymbol{\varphi})_{[i]}} \right) \left(\frac{\partial \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi}))}{\partial (\boldsymbol{\varphi})_{[j]}} \right) \right\} \tag{7.134}$$

Then the *Fisher information matrix* may be written in matrix form as follows,

$$\boldsymbol{\mathscr{I}}_F = \mathscr{E}\left\{ \left(\nabla_{\boldsymbol{\varphi}} \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi})) \right) \left(\nabla_{\boldsymbol{\varphi}} \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi})) \right)^T \right\} \tag{7.135}$$

where $\nabla_{\boldsymbol{\varphi}} \ln(\hat{p}(\mathbf{x}|\boldsymbol{\varphi}))$ is known as the *Fisher score* or *score statistic* – see 10.1.

Also, we may write the *Kullback-Leibler divergence* of Equation 7.131 in matrix form as,

---

[7] *Cramér* [4] and *Kullback* [11] include these conditions as a part of the *second regularity condition*, but aside from having a role in clarity and completeness, they do not technically need to be specified as conditions, since they are implied.