the regularity conditions needed for the ML estimates $\hat{\boldsymbol{\beta}}$ to be consistent and asymptotically normal, with asymptotic covariance matrix given by the inverse of the information matrix in the usual way. See, for example, Gouriéroux and Monfort (1981). In the case of the logit model, the first-order conditions (15.10) simplify to

$$\sum_{t=1}^{n}\big(y_t - \Lambda(\boldsymbol{X}_t\hat{\boldsymbol{\beta}})\big)X_{ti} = 0, \;\; i = 1, \ldots, k,$$

because $\lambda(x) = \Lambda(x)\big(1 - \Lambda(x)\big)$. Notice that conditions (15.10) look just like the first-order conditions for weighted least squares estimation of the nonlinear regression model

$$y_t = F(\boldsymbol{X}_t\boldsymbol{\beta}) + e_t, \tag{15.11}$$

with weights given by

$$\Big(F(\boldsymbol{X}_t\boldsymbol{\beta})\big(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\big)\Big)^{-1/2}.$$

This makes sense, since the variance of the error term in (15.11) is

$$\begin{aligned} E(e_t^2) &= E\big(y_t - F(\boldsymbol{X}_t\boldsymbol{\beta})\big)^2 \\ &= F(\boldsymbol{X}_t\boldsymbol{\beta})\big(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\big)^2 + \big(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\big)\big(F(\boldsymbol{X}_t\boldsymbol{\beta})\big)^2 \\ &= F(\boldsymbol{X}_t\boldsymbol{\beta})\big(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\big). \end{aligned}$$

Thus one way to obtain ML estimates of any binary response model is to apply iteratively reweighted nonlinear least squares to (15.11) or to whatever nonlinear regression model is appropriate if the index function is not $\boldsymbol{X}_t\boldsymbol{\beta}$. For most models, however, this is generally not the best approach, and a better one is discussed in the next section.

Using the fact that ML is equivalent to a form of weighted NLS for binary response models, it is obvious that the asymptotic covariance matrix for $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ must be

$$\Big(\frac{1}{n}\boldsymbol{X}^{\top}\boldsymbol{\Psi}(\boldsymbol{\beta}_0)\boldsymbol{X}\Big)^{-1},$$

where $\boldsymbol{X}$ is an $n \times k$ matrix with typical row $\boldsymbol{X}_t$ and typical element X_{ti}, and $\boldsymbol{\Psi}(\boldsymbol{\beta})$ is a diagonal matrix with typical diagonal element

$$\Psi(\boldsymbol{X}_t\boldsymbol{\beta}) = \frac{f^2(\boldsymbol{X}_t\boldsymbol{\beta})}{F(\boldsymbol{X}_t\boldsymbol{\beta})\big(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\big)}. \tag{15.12}$$

The numerator reflects the fact that the derivative of $F(\boldsymbol{X}_t\boldsymbol{\beta})$ with respect to β_i is $f(\boldsymbol{X}_t\boldsymbol{\beta})X_{ti}$, and the denominator is simply the variance of e_t in (15.11). In the logit case, $\Psi(\boldsymbol{X}_t\boldsymbol{\beta})$ simplifies to $\lambda(\boldsymbol{X}_t\boldsymbol{\beta})$.