Solution to Exercise 11.1

$\star$**11.1** Consider the contribution made by observation t to the loglikelihood function (11.09) for a binary response model. Show that this contribution, and hence (11.09) itself, is globally concave with respect to $\boldsymbol{\beta}$ if the function F is such that $F(-x) = 1 - F(x)$, and if it, its derivative f, and its second derivative f' satisfy the condition

$$f'(x)F(x) - f^2(x) < 0 \tag{11.88}$$

for all real finite x.

Show that condition (11.88) is satisfied by both the logistic function $\Lambda(\cdot)$, defined in (11.07), and the standard normal CDF $\Phi(\cdot)$.

The contribution made by observation t is

$$y_t \log F(\boldsymbol{X}_t\boldsymbol{\beta}) + (1 - y_t)\log\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr). \tag{S11.01}$$

When $y_t = 1$, the vector of first derivatives of (S11.01) with respect to $\boldsymbol{\beta}$ is

$$\frac{f(\boldsymbol{X}_t\boldsymbol{\beta})}{F(\boldsymbol{X}_t\boldsymbol{\beta})}\boldsymbol{X}_t,$$

and the matrix of second derivatives is therefore

$$\frac{f'(\boldsymbol{X}_t\boldsymbol{\beta})F(\boldsymbol{X}_t\boldsymbol{\beta}) - f^2(\boldsymbol{X}_t\boldsymbol{\beta})}{F^2(\boldsymbol{X}_t\boldsymbol{\beta})}\boldsymbol{X}_t\boldsymbol{X}_t^{\top}. \tag{S11.02}$$

When condition (11.88) is satisfied, this is a negative number times the positive semidefinite matrix $\boldsymbol{X}_t\boldsymbol{X}_t^{\top}$.

When $y_t = 0$, the first derivative of (S11.01) with respect to $\boldsymbol{\beta}$ is

$$\frac{-f(\boldsymbol{X}_t\boldsymbol{\beta})}{1 - F(\boldsymbol{X}_t\boldsymbol{\beta})}\boldsymbol{X}_t,$$

and the second derivative is therefore

$$\frac{-f'(\boldsymbol{X}_t\boldsymbol{\beta})\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr) - f^2(\boldsymbol{X}_t\boldsymbol{\beta})}{\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr)^2}\boldsymbol{X}_t\boldsymbol{X}_t^{\top}.$$

Since $F(-x) = 1 - F(x)$, it follows that $f(x) = f(-x)$ and $f'(x) = -f'(-x)$. Therefore, this second derivative can be rewritten as

$$\frac{f'(-\boldsymbol{X}_t\boldsymbol{\beta})F(-\boldsymbol{X}_t\boldsymbol{\beta}) - f^2(-\boldsymbol{X}_t\boldsymbol{\beta})}{F^2(-\boldsymbol{X}_t\boldsymbol{\beta})}\boldsymbol{X}_t\boldsymbol{X}_t^{\top}. \tag{S11.03}$$

When condition (11.88) is satisfied, this is also a negative number times the positive semidefinite matrix $\boldsymbol{X}_t\boldsymbol{X}_t^{\top}$.

The contribution to the Hessian by observation t is either (S11.02) or (S11.03). In either case, it is a negative number times a positive semidefinite matrix. Therefore, this contribution must be negative semidefinite, which implies that expression (S11.01) is globally concave, although not strictly so. Since the loglikelihood function is the sum of n contributions, the Hessian is the sum of n of these negative semidefinite matrices, and it is therefore also negative semidefinite. Therefore, the loglikelihood function must be globally concave. In practice, of course, the Hessian is almost always negative definite.

For the logistic function,

$$\begin{aligned}F(x) &= \Lambda(x),\\ f(x) &= \Lambda(x)\Lambda(-x), \text{ and}\\ f'(x) &= \Lambda(x)\Lambda^2(-x) - \Lambda^2(x)\Lambda(-x).\end{aligned}$$

Therefore,

$$\begin{aligned}f'(x)F(x) - f^2(x) &= \Lambda^2(x)\Lambda^2(-x) - \Lambda^3(x)\Lambda(-x) - \Lambda^2(x)\Lambda^2(-x)\\ &= -\Lambda^3(x)\Lambda(-x) = -\lambda(x)\Lambda^2(x).\end{aligned}$$

Since this is minus the logistic PDF times the square of the logistic CDF, it must be negative, and we see that condition (11.88) is satisfied.

For the standard normal distribution,

$$F(x) = \Phi(x),\;\; f(x) = \phi(x), \text{ and } f'(x) = -x\,\phi(x).$$

It is clear that, if $x \geq 0$, condition (11.88) is satisfied, since we are then subtracting a positive number from a nonpositive one. But it is not so obvious what happens when $x < 0$.

Using the fact that $f'(x) = -x\,\phi(x)$, condition (11.88) for the standard normal distribution becomes $-x\,\phi(x)\Phi(x) - \phi^2(x) < 0$. Because $\phi(x) > 0$ for all finite x, this condition can be simplified to

$$a(x) \equiv -x\,\Phi(x) - \phi(x) < 0. \tag{S11.04}$$

As $x \to -\infty$, the function $a(x)$ defined in (S11.04) tends to 0, because $\phi(x)$ does so and $\Phi(x)$ tends to 0 faster than x tends to $-\infty$. Now consider the derivative of $a(x)$, which is

$$a'(x) = -\Phi(x) - x\,\phi(x) + x\,\phi(x) = -\Phi(x) < 0.$$

Since $a(-\infty) = 0$ and its derivative is always negative, $a(x)$ must be negative for all finite x. This implies that condition (S11.04) holds, which in turn implies that condition (11.88) holds for the standard normal distribution.