⋆11.5 Consider the latent variable model

$$y_t^\circ = \beta_1 + \beta_2 x_t + u_t, \quad u_t \sim \mathrm{N}(0,1),$$
$$y_t = 1 \text{ if } y_t^\circ > 0, \;\; y_t = 0 \text{ if } y_t^\circ \leq 0.$$

Suppose that $x_t \sim \mathrm{N}(0,1)$. Generate 500 samples of 20 observations on (x_t, y_t) pairs, 100 assuming that $\beta_1 = 0$ and $\beta_2 = 1$, 100 assuming that $\beta_1 = 1$ and $\beta_2 = 1$, 100 assuming that $\beta_1 = -1$ and $\beta_2 = 1$, 100 assuming that $\beta_1 = 0$ and $\beta_1 = 2$, and 100 assuming that $\beta_1 = 0$ and $\beta_2 = 3$. For each of the 500 samples, attempt to estimate a probit model. In each of the five cases, what proportion of the time does the estimation fail because of perfect classifiers? Explain why there were more failures in some cases than in others.

Repeat this exercise for five sets of 100 samples of size 40, with the same parameter values. What do you conclude about the effect of sample size on the perfect classifier problem?

11.6 Suppose that there is quasi-complete separation of the data used to estimate the binary response model (11.01), with a transformation function F such that $F(-x) = 1 - F(x)$ for all real x, and a separating hyperplane defined by the parameter vector $\boldsymbol{\beta}^\bullet$. Show that the upper bound of the loglikelihood function (11.09) is equal to $-n_b \log 2$, where n_b is the number of observations for which $\boldsymbol{X}_t\boldsymbol{\beta}^\bullet = 0$.

11.7 The contribution to the loglikelihood function (11.09) made by observation t is $y_t \log F(\boldsymbol{X}_t\boldsymbol{\beta}) + (1 - y_t)\log\bigl(1 - F(\boldsymbol{X}_t\boldsymbol{\beta})\bigr)$. First, find G_{ti}, the derivative of this contribution with respect to β_i. Next, show that the expectation of G_{ti} is zero when it is evaluated at the true $\boldsymbol{\beta}$. Then obtain a typical element of the asymptotic information matrix by using the fact that it is equal to $\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \mathrm{E}(G_{ti}G_{tj})$. Finally, show that the asymptotic covariance matrix (11.15) is equal to the inverse of this asymptotic information matrix.

11.8 Calculate the Hessian matrix corresponding to the loglikelihood function (11.09). Then use the fact that minus the expectation of the asymptotic Hessian is equal to the asymptotic information matrix to obtain the same result for the latter that you obtained in the previous exercise.

⋆11.9 Plot $\Upsilon_t(\boldsymbol{\beta})$, which is defined in equation (11.16), as a function of $\boldsymbol{X}_t\boldsymbol{\beta}$ for both the logit and probit models. For the logit model only, prove that $\Upsilon_t(\boldsymbol{\beta})$ achieves its maximum value when $\boldsymbol{X}_t\boldsymbol{\beta} = 0$ and declines monotonically as $|\boldsymbol{X}_t\boldsymbol{\beta}|$ increases.

11.10 The file **participation.data**, which is taken from Gerfin (1996), contains data for 872 Swiss women who may or may not participate in the labor force. The variables in the file are:

y_t	Labor force participation variable (0 or 1).
I_t	Log of nonlabor income.
A_t	Age in decades (years divided by 10).
E_t	Education in years.
nu_t	Number of children under 7 years of age.
no_t	Number of children over 7 years of age.
F_t	Citizenship dummy variable (1 if not Swiss).