Probit regression

- Like logistic regression, just the connection between the linear predictor η and $P(Y=1|X)$ is changed.

Contents

Print to PDF

Details

- Let $\eta_i = \eta_i(X_i, \beta) = \beta_0 + \sum_{j=1}^p \beta_j X_{ij}$ be our *linear predictor*.
- Probit model says:

$$P(Y=1|X) = \Phi(\eta) = \int_{-\infty}^{\eta} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz$$

- Likelihood for independent $Y_i|X_i$:

$$L(\beta|(X_1,Y_1),\dots,(X_n,Y_n)) = \prod_{i=1}^n \Phi(\eta_i)^{Y_i}(1-\Phi(\eta_i))^{1-Y_i}$$

- Or,

$$\log L(\beta) = \sum_{i=1}^n Y_i \log(\Phi(\eta_i)) + (1-Y_i)\log(1-\Phi(\eta_i))$$

Score

- Computing derivative

$$\begin{aligned}\nabla \log L(\beta) &= \sum_{i=1}^n \mathbf{X}_i \cdot \phi(\eta_i)\left(\frac{Y_i}{\Phi(\eta_i)} - \frac{1-Y_i}{1-\Phi(\eta_i)}\right) \\ &= \sum_{i=1}^n \mathbf{X}_i \cdot \phi(\eta_i)\left(\frac{Y_i}{\Phi(\eta_i)(1-\Phi(\eta_i))} - \frac{1}{1-\Phi(\eta_i)}\right)\end{aligned}$$

with

$$\phi(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

- Second derivative (more complicated) but (by link between expected 2nd derivative and variance of score):

$$E_\beta[\nabla^2 \log L(\beta)] = -\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i^T \cdot \frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1-\Phi(\eta_i))} = -\mathbf{X}^T\mathbf{W}\mathbf{X}$$

with

$$\mathbf{W} = \mathrm{diag}\left(\frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1-\Phi(\eta_i))}, 1 \le i \le n\right)$$

```
library(ISLR)
data(Default)
names(Default)
```

'default' · 'student' · 'balance' · 'income'

```
M = glm(default ~ student + balance + income, family=binomial(link="probit"), data=Default)
summary(M)
```

```
Call:
glm(formula = default ~ student + balance + income, family = binomial(link = "probit"),
    data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2226  -0.1354  -0.0321  -0.0044   4.1254

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -5.475e+00  2.385e-01 -22.960   <2e-16 ***
studentYes  -2.960e-01  1.188e-01  -2.491   0.0127 *
balance      2.821e-03  1.139e-04  24.774   <2e-16 ***
income       2.101e-06  4.121e-06   0.510   0.6101
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1583.2  on 9996  degrees of freedom
AIC: 1591.2

Number of Fisher Scoring iterations: 8
```

```r
linpred = predict(M)
D = model.matrix(M)
sum((linpred - D %*% coef(M))^2)
```

0

```r
EY = pnorm(linpred)
W = dnorm(linpred)^2 / (EY * (1 - EY))
Vi = t(D) %*% diag(W) %*% D
V = solve(Vi)
```

```r
V - vcov(M)
```

	(Intercept)	studentYes	balance	income
(Intercept)	-7.218128e-06	3.308719e-07	3.523911e-09	3.743661e-11
studentYes	3.308719e-07	-9.761490e-07	5.119793e-10	-2.743668e-11
balance	3.523911e-09	5.119793e-10	-2.344713e-12	7.620989e-15
income	3.743661e-11	-2.743668e-11	7.620989e-15	-1.316252e-15

A matrix: 4 × 4 of type dbl

```r
sqrt(sum((V - vcov(M))^2) / sum(V^2))
```

0.000118162851085932

```r
sqrt(diag(V))
```

(Intercept): 0.238456043563375 **studentYes:** 0.118817612111096 **balance:** 0.000113851249648238 **income:** 4.12071895479049e-06

Fitting the model

- The variance / covariance matrix of the score is also informative to *fit* the logistic regression model.

Newton-Raphson

- Iterative algorithm to find a 0 of the score (i.e. the MLE)
- Based on 2nd order Taylor expansion of $\log L(\beta)$.
- Given a *base point* $\tilde{\beta}$

$$\log L(\beta) = \log L(\tilde\beta) + \nabla \log L(\tilde\beta)^T(\beta-\tilde\beta) + \frac{1}{2}(\beta-\tilde\beta)^T \nabla^2 \log L(\tilde\beta)(\beta-\tilde\beta) + \dots$$

- Iterates successively maximize these 2nd order Taylor approximations

$$\begin{aligned}\hat\beta_{(t+1)} &= \text{argmax}_\beta \left[\log L(\hat\beta_{(t)}) + \nabla \log L(\hat\beta_{(t)})^T(\beta-\hat\beta_{(t)}) + \frac{1}{2}(\beta-\hat\beta_{(t)})^T \nabla^2 \log L(\hat\beta_{(t)})(\beta-\hat\beta_{(t)})\right] \\ &= \hat\beta_{(t)} - \nabla^2 \log L(\hat\beta_{(t)})^{-1} \nabla \log L(\hat\beta_{(t)})\end{aligned}$$

Fisher scoring

- Replaces $-\nabla^2 \log L(\hat\beta_{(t)})$ with *Fisher information*

$$-E_{\hat\beta_{(t)}}\left[\nabla^2 \log L(\hat\beta_{(t)})\right] = \text{Var}_{\hat\beta_{(t)}}\left[\nabla \log L(\hat\beta_{(t)})\right]$$

- Does not change anything for logistic regression.
- Algorithm becomes

$$\hat\beta_{(t+1)} = \hat\beta_{(t)} + \left(\text{Var}_{\hat\beta_{(t)}}\left[\nabla \log L(\hat\beta_{(t)})\right]\right)^{-1} \nabla \log L(\hat\beta_{(t)})$$

Algorithm

- Basic parts of algorithm

```
score = function(beta, D, Y) {
    eta = D %*% beta
    EY = pnorm(eta)
    d = dnorm(eta)
    return(t(D) %*% (d * (Y/(EY*(1-EY)) - 1 / (1-EY))))
}

information_matrix = function(beta, D, Y) {
    eta = D %*% beta
    EY = pnorm(eta)
    W = as.numeric(dnorm(eta)^2 / (EY * (1 - EY)))
    return(t(D) %*% (W * D))
}
```

- Now we look at iterative algorithm

```
beta_hat = rep(0, ncol(D))
Y = Default$default == "Yes"
for (i in 1:10) {
    cur_score = score(beta_hat, D, Y)
    cur_info = information_matrix(beta_hat, D, Y)
    beta_hat = beta_hat + solve(cur_info) %*% cur_score
}
```

```
beta_hat
```

(Intercept)	-5.475359e+00
studentYes	-2.959823e-01
balance	2.820782e-03
income	2.101338e-06

A matrix: 4 × 1 of type dbl

```
coef(M)
```

(Intercept): -5.47535136776302 **studentYes:** -0.295980628047736 **balance:** 0.00282077720882078 **income:** 2.10137534488741e-06

```
solve(cur_info)
```

	(Intercept)	studentYes	balance	income
(Intercept)	5.686147e-02	-1.383742e-02	-1.885129e-05	-6.768671e-07
studentYes	-1.383742e-02	1.411765e-02	-2.298291e-06	3.817505e-07
balance	-1.885129e-05	-2.298291e-06	1.296216e-08	-5.489556e-12
income	-6.768671e-07	3.817505e-07	-5.489556e-12	1.698036e-11

A matrix: 4 × 4 of type dbl

```
vcov(M)
```

	(Intercept)	studentYes	balance	income
(Intercept)	5.686850e-02	-1.383773e-02	-1.885472e-05	-6.769034e-07
studentYes	-1.383773e-02	1.411860e-02	-2.298791e-06	3.817771e-07
balance	-1.885472e-05	-2.298791e-06	1.296445e-08	-5.497050e-12
income	-6.769034e-07	3.817771e-07	-5.497050e-12	1.698164e-11

A matrix: 4 × 4 of type dbl