Probit Models

Math 463, Spring 2017, University of Oregon

David A. Levin

University of Oregon

May 22, 2017

- When variables are standardized, all you need is the correlation matrix to estimate coefficients.
- To obtain an estimate of σ^2, note that $\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$, then

$$n - 1 = \|\boldsymbol{Y}\|^2 = \sum_{j=1}^{k} \hat{\beta}_j^2 \|\boldsymbol{x}_j\|^2 + 2 \sum_{1 \le i < j \le k} \hat{\beta}_i \hat{\beta}_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + \|\boldsymbol{e}\|^2$$

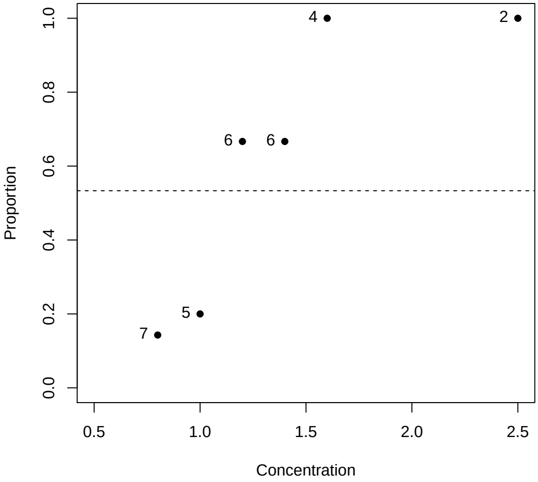
  This can be solved for $\|\boldsymbol{e}\|^2 = \mathrm{RSS}$, which is what is needed in $\hat{\sigma}^2 = \mathrm{RSS}/(n-p)$.
- Note that $\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = (n-1)\, r(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

- Linear predictor:
$$\eta = \boldsymbol{X}\boldsymbol{\beta}$$
- Response variable: $Y_i \in \{0, 1\}$.
- Model:
$$\mathbb{P}(Y_i = 1 \mid \boldsymbol{X}) = \Phi(\eta),$$
  Φ is the Normal CDF.
- The Y_i are independent.

```
> library(DAAG)
> anestot <- aggregate(anesthetic[,c("move","nomove")]
+                      ,by=list(conc=anesthetic$conc),FUN=sum)
> anestot$total <- apply(anestot[,c("move","nomove")],1,sum)
> anestot$prop <- anestot$nomove/anestot$total
> anestot
```

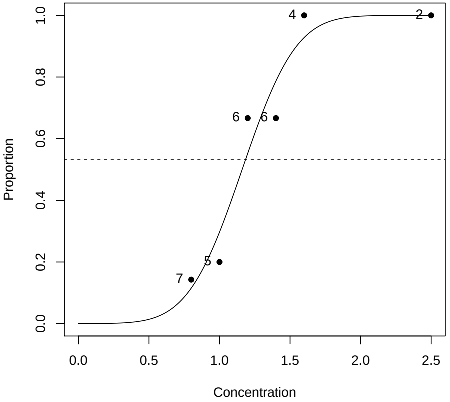
```
  conc move nomove total      prop
1  0.8    6      1     7 0.1428571
2  1.0    4      1     5 0.2000000
3  1.2    2      4     6 0.6666667
4  1.4    2      4     6 0.6666667
5  1.6    0      4     4 1.0000000
6  2.5    0      2     2 1.0000000
```

```
> plot(prop~conc, data=anestot,xlab="Concentration",ylab="Proportion",
+      xlim=c(0.5,2.5),ylim=c(0,1),pch=16)
> with(anestot,
+      {text(conc,prop,paste(total),pos=2)
+        abline(h=sum(nomove)/sum(total), lty=2)})
```


4
2
6
6
5
7
Proportion
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0
2.5
Concentration

```
> fit1 = glm(nomove~conc, family=binomial(link="probit"),
+            data=anesthetic)
> summary(fit1)

Call:
glm(formula = nomove ~ conc, family = binomial(link = "probit"),
    data = anesthetic)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.75889  -0.75326   0.00291   0.69198   2.07819

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.858      1.317  -2.929   0.0034 **
conc           3.324      1.112   2.989   0.0028 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 41.455  on 29  degrees of freedom
Residual deviance: 27.701  on 28  degrees of freedom
AIC: 31.701

Number of Fisher Scoring iterations: 6
```

```
> x = data.frame(conc=seq(0,2.5,0.01)); eta = predict(fit1,newdata=x)
> plot(prop~conc, data=anestot,xlab="Concentration",ylab="Proportion",
+       xlim=c(0,2.5),ylim=c(0,1),pch=16)
> with(anestot,
+       {text(conc,prop,paste(total),pos=2)
+         abline(h=sum(nomove)/sum(total), lty=2)})
> lines(x$conc,pnorm(eta))
```



Some MLE theory

- Likelihood function for probit: Let $\eta_i = \sum_{j=1}^{k} \beta_j x_{i,j}$.

$$L(\boldsymbol{\beta} \mid y_1, \ldots, y_n) = \prod_{i=1}^{n} \Phi(\eta_i)^{y_i} (1 - \Phi(\eta_i))^{1-y_i}.$$

- Can estimate $\boldsymbol{\beta}$ by finding parameters which make $L(\boldsymbol{\beta} \mid y_1, \ldots, y_n)$ as large as possible.
- No nice closed form expression for MLE, but can find numerically.
- There is asymptotic theory for MLE: Under certain assumptions $I^{-1/2}(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ tends to a multivariate Normal distribution,

$$I(\boldsymbol{\beta}) = -\sum_{k=1}^{n} \mathbb{E}[D^2 \ell_k(\boldsymbol{\beta})].$$

  Here, $D^2 \ell_k$ is the matrix with $(i,j)-th$ entry $\frac{\partial}{\partial \beta_j} \frac{\partial}{\partial \beta_i} \ell_k$.
- Use *observed information* $\sum_{k=1}^{n} D^2 \ell_k(\hat{\boldsymbol{\beta}})$.
- This is where the standard errors come from.

Latent Variables

- Note that if $V \sim N(0,1)$, then by symmetry of the Normal distribution,

$$\mathbb{P}(\eta + V > 0) = \mathbb{P}(V > -\eta) = \mathbb{P}(V \leq \eta) = \Phi(\eta)$$

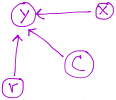
- Thus, we can think of the data as arising from the model

$$Y_i = \begin{cases} 1 & \text{if } \eta_i + V_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- This is the same as saying that each Y_i is the result of a coin toss with probability $\Phi(\eta_i)$ of "heads".
- We can now formulate our assumption: we need that $\boldsymbol{X}$ is independent of $\boldsymbol{V}$, the vector of V_i's.
- Given this, we have *asymptotically* unbiased estimator of $\boldsymbol{\beta}$.
- Note that the latent variables $\boldsymbol{V}$ are unobservable.

Example

- Y is an indicator a student graduates, $\boldsymbol{X}$ are covariates, and C is the indicator that the student went to catholic school.
- Path diagram:



$$Y_i = \begin{cases} 1 & c\alpha + X_i\beta + V_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

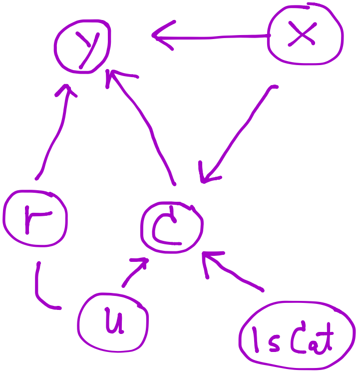
Is C_i independat of V_i?

- $$Y_i = \begin{cases} 1 & \text{if } \boldsymbol{X\beta} + \alpha C_i + V_i > 0 \\ 0 & \text{else} \end{cases}$$

- Danger: C is *endogenous*, as it is correlated with V.

Example

Add a variable IsCat



Two Equations!

Table VI. Maximum Likelihood Estimates of *HIGH SCHOOL GRADUATE* and *COLLEGE ENTRANT* Bivariate Probit Model Using *CATHOLIC RELIGION* as an Instrument

		MLE estimates of bivariate probit model				
Model	Other variables[b] in X_i	Coefficient on *CATHOLIC SCHOOL*	Marginal effect[c]	Average treatment effect	ρ	2SLS estimate of coefficient on *CATHOLIC SCHOOL*
High School Graduate[a]						
(1)		0.777	0.117	0.130		0.096[d]
		(0.056)	(0.014)	(0.007)		(0.008)
(2)		0.859	0.133	0.141	–0.053	0.127
		(0.115)	(0.022)	(0.014)	(0.067)	(0.024)
(3)	10th Grade Test Score and Test Missing	0.678	0.078	0.114	0.028	0.103
		(0.126)	(0.018)	(0.017)	(0.072)	(0.024)
(4)	State Effects	0.911	0.142	0.144	–0.050	0.114
		(0.121)	(0.027)	(0.015)	(0.072)	(0.024)
(5)	10th Grade Test Score, Test Missing, and State Effects	0.746	0.124	0.121	0.025	0.134
		(0.132)	(0.028)	(0.016)	(0.077)	(0.030)
College Entrant[a]						
(6)		0.384	0.144	0.132		0.137[d]
		(0.032)	(0.012)	(0.011)		(0.011)
(7)		0.288	0.109	0.098	0.067	0.148
		(0.079)	(0.033)	(0.028)	(0.049)	(0.030)
(8)	10th Grade Test Score and Test Missing	0.211	0.078	0.064	0.124	0.098
		(0.083)	(0.034)	(0.026)	(0.052)	(0.024)
(9)	State Effects	0.341	0.110	0.115	0.056	0.092
		(0.084)	(0.032)	(0.029)	(0.053)	(0.024)
(10)	10th Grade Test Score, Test Missing, and State Effects	0.277	0.071	0.082	0.113	0.098
		(0.090)	(0.026)	(0.027)	(0.046)	(0.028)

Asymptotic standard errors are in parentheses.

a. Models (1) and (6) are single-equation estimates from Table III. To estimate models (4), (5), (9), and (10), we deleted all states with no Catholic school students. The high school completion and college entrance models contain 10,120 and 8470 observations, respectively. Both models contain data from twenty states. Models (1), (2), and (3) contain 13,294 observations, and models (6), (7), and (8) contain 10,983 observations.

b. Other exogenous variables include those listed in Table III

c. Marginal effects are calculated for the individual defined in Table III

Assumptions for Nature

- Response schedule for Graduation: c is in principle manipulable (can send kid to type of school). Other covariates are personal characteristics which are hard to change. (Placeholder variables in response schedule should be possible to change via intervention.)

$$Y_{i,c} = 1\{\alpha c + \boldsymbol{X}_i \boldsymbol{\beta} + V_i > 0\} \tag{1}$$

- Catholic status, C_i, is determined via

$$C_i = 1\{\text{Iscat}_i a + \boldsymbol{X}_i \boldsymbol{b} + U_i > 0\} \tag{2}$$

- Choose Iscat_i and $\boldsymbol{X}_i$.
- Choose (U_i, V_i) independent of Iscat_i and $\boldsymbol{X}_i$.
- Observe the value C_i in (2); if $C_i = 1$ send to catholic school, otherwise public school.
- Set c in (1) to C_i; observe Y_i. If $Y_i = 1$ then make student graduate, otherwise prevent student.
- Reveal $\text{IsCat}_i, \boldsymbol{X}_i, C_i, Y_i$
- Trash U_i and V_i, they aren't needed anymore.

If these assumptions are true, we can estimate α and draw conclusion about the effect of Catholic schools on graduation.

Some questions

- Why not just make large cross-tab? I.e., break down by covariates and compare catholic schools and public schools within each subgroup?
- Are the coefficients the same for all groups? (Interactions?)
- Should IsCat be among the covariates in the equation for Y_i? (Exclusion restriction)
- What about relationship of other covariates with the latent variable (which represents unmeasured characteristics like intelligence, motivation, parental attitudes)?
- Others?
- Is the paper convincing?

Another recent example of using bivariate probit (with some wrinkles):
http://www.journals.uchicago.edu/doi/full/10.1017/S0022381614000127

- Suppose distribution of Y belongs to an exponential family:

$$f(y;\gamma) = \exp[(y\gamma + b(\gamma)/a(\phi) + c(y,\phi)]$$

- Let $\theta = \mathbb{E}(Y)$; write $\gamma = h(\theta)$.
- Link function $\eta_i = g(\theta_i)$