W03 LOGIT PROBIT

Probit !!

$\boxed{y_i \in \{0, 1\}}$ $\quad y_i \in \{-1, 1\}$

hidden utility

y_i^* – utility of alternative [1] [smiles]
utility of alternative [0] is 0. [smiles]

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

$$y_i^* = \beta_1 + \beta_2 \cdot x_i + u_i$$

is not observable

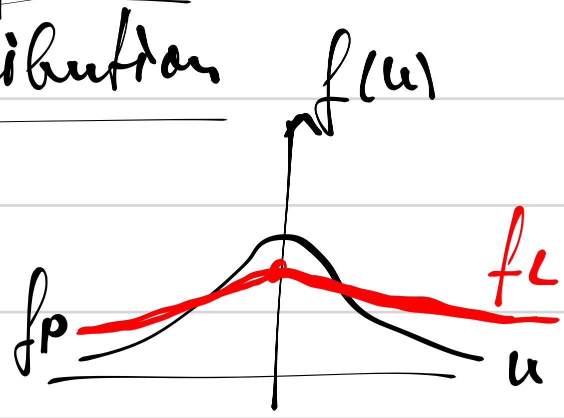
$$y^* = X \cdot \beta + u \qquad y^* = \begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix} \qquad X = \begin{bmatrix} \cdots \end{bmatrix}$$

$$u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

PROBIT: $(u_i | X) \sim N(0; 1)$ indep.

LOGIT: $(u_i | X) \sim$ logistic distribution

PROBIT: $f_p(u) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{u^2}{2}\right)$



LOGIT: $f_L(u) = \frac{\exp(u)}{1+\exp(u)} \cdot \frac{1}{1+\exp(u)}$ [Good ex.]

logistic distr. is similar $N(0; \frac{\pi^2}{3})$ $\quad E(u_i) = 0 \quad Var(u_i) = \frac{\pi^2}{3}$

properties.

1 LOGIT model discriminative (not generative)

you need x_{new} to predict $\hat{y}_{new}$
you can't generate x_{new}
LOGIT has no assumptions about the distribution of regressors.

2 Sometimes LOGIT estimation fails.
[esp-lly with many predictors]
message: "~ 0 or 1 probability"

y_i	x_i
0	1
1	2
1	3

→ no problems with OLS
$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_i$
...
→ let's try logit

obtain the loss function for the LOGIT model
LOSS = − log likelihood
max log likelihood
min loss

$$\ln\left[P(y_1=0|X)\cdot P(y_2=1|X)\cdot P(y_3=1|X)\right] =$$

$$= \ln P(y_1=0|X) + \ln \underline{P(y_2=1|X)} + \ln P(y_3=1|X)$$

$$P(y_i=1|X) = P(y_i^* > 0|X) = P(\beta_1+\beta_2\cdot x_i + u_i > 0|X)$$

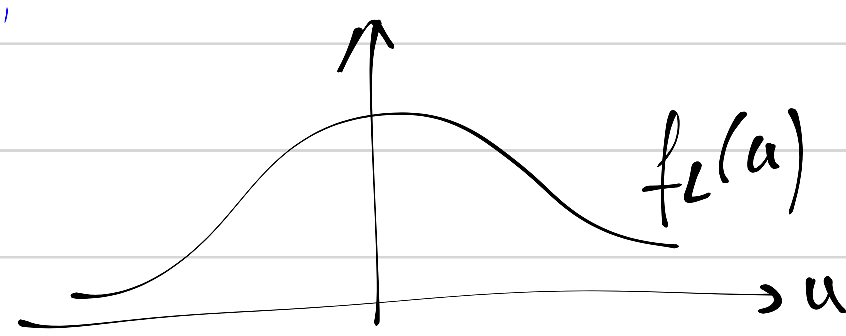
$$= P(u_i > -\beta_1 - \beta_2 x_i|X)$$

Logistic th: $f(u)=f(-u)$

$$f(u) = \frac{\exp(u)}{1+\exp(u)} \cdot \frac{1}{1+\exp(u)}$$

$\cdot\exp(u)\exp(u)$

$$f(-u) = \frac{\exp(-u)}{1+\exp(-u)} \cdot \frac{1}{1+\exp(-u)} = \frac{\exp(u)\cdot 1}{(\exp(u)+1)(\exp(u)+1)} = f(u)$$



due to symmetry

$$= P(u_i > -\beta_1 - \beta_2 x_i \,|\, X) = P(u_i < \beta_1 + \beta_2 \cdot x_i \,|\, X) =$$

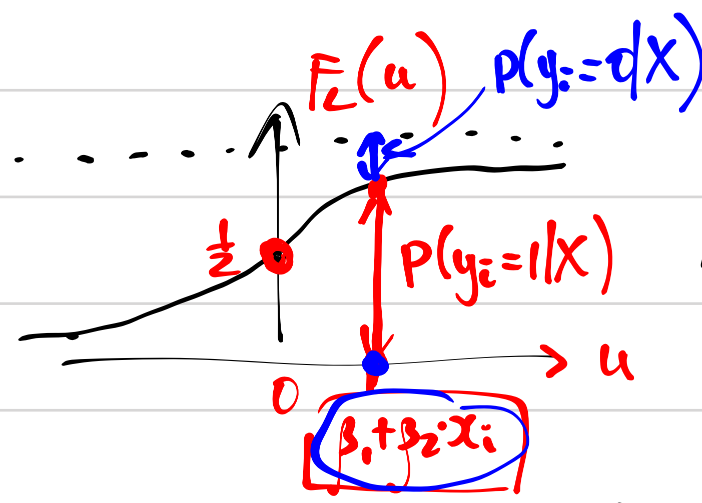
$$= F(\beta_1 + \beta_2 \cdot x_i) = P(y_i = 1 \,|\, X) =$$

$P(y_i = 0 \,|\, X) = 1 - F(\beta_1 + \beta_2 \cdot x_i)$

logistic RV.

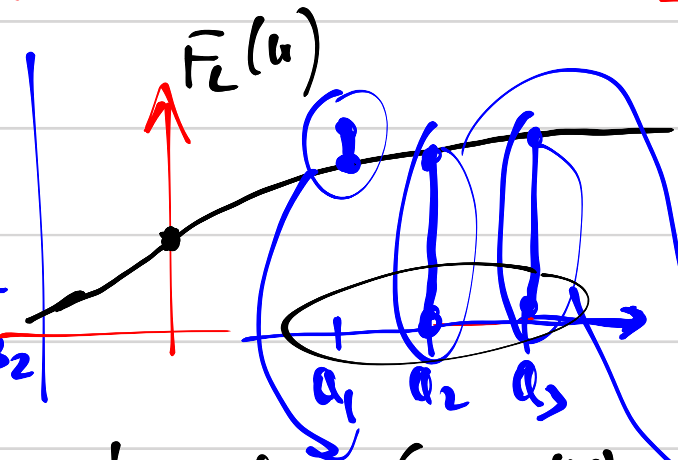
$$f_L(u) = \frac{\exp(u)}{(1+\exp(u))^2}$$

$f = \frac{\partial F}{\partial u}$



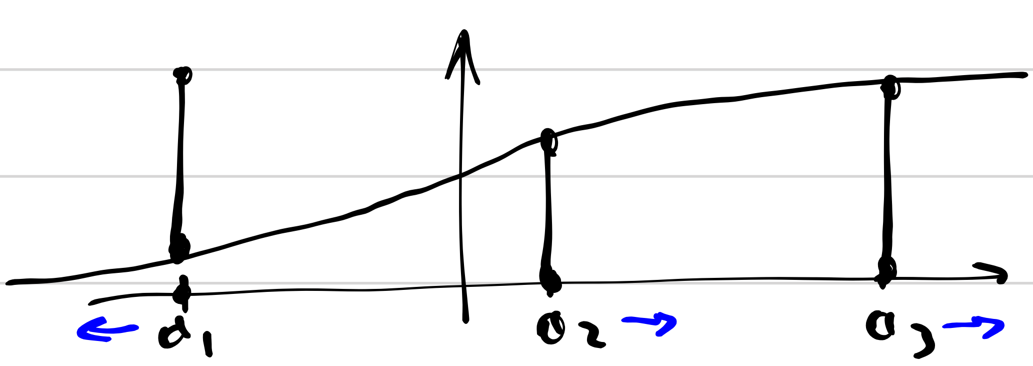
$$F_L(u) = \frac{\exp(u)}{1+\exp(u)} = \frac{\exp(\beta_1+\beta_2 \cdot x_i)}{1+\exp(\beta_1+\beta_2 \cdot x_i)}$$

y_i	x_i	a_i
0	1	$\beta_1+\beta_2$
1	2	$\beta_1+2\beta_2$
1	3	$\beta_1+3\beta_2$



$a_i = \beta_1 + \beta_2 \cdot x_i$

$$\ln \text{likelihood} = \ln P(y_1 = 0 \,|\, X) + \ln P(y_2 = 1 \,|\, X) + \ln P(y_3 = 1 \,|\, X) \to \max_{\beta_1, \beta_2}$$



$\max_{\beta_1, \beta_2} \ln \text{likelihood} = \ln 1 + \ln 1 + \ln 1$

"perfect separation case"

$\hat{\beta}_2 \to +\infty$

you see big values of $\hat{\beta}_1$, $\hat{\beta}_2$

$\hat{\beta}_2 \approx 100$ or more

Solutions:

* reduce the number of predictors.
* introduce penalty in the likelihood function.

logit in statsmodels. honest LOGIT with max likelihood

risk of no sol.n

logit in sklearn penalized LOGIT

no risk of no sol.n

no CI, no hypoth. test.

by the way:

$$\text{ln Likelihood} = \sum_{i=1}^{n} l_i =$$

$$l_i = \begin{cases} \ln(1-F(\beta_1+\beta_2\cdot x_i)) & \text{if } y_i=0 \\ \ln F(\beta_1+\beta_2\cdot x_i) & \text{if } y_i=1 \end{cases}$$

$$F(u) = \frac{\exp(u)}{1+\exp(u)}$$

$$= \sum_{i=1}^{n} y_i \cdot \ln F(\beta_1+\beta_2\cdot x_i) + (1-y_i)\cdot \ln(1-F(\beta_1+\beta_2 x_i))$$

* A small funny fact !! $y_i \in \{0;1\}$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{P}(y_i=1 \mid X)$$

$F_i = F(\beta_1+\beta_2\cdot x_i)$

$f_i = f(\beta_1+\beta_2\cdot x_i)$

$$\frac{\partial \ln Lik}{\partial \beta_1} = \sum\left[y_i\cdot\frac{1}{F_i}\cdot f_i + (1-y_i)\cdot\frac{1}{1-F_i}\cdot(-f_i)\right] =$$

$$= \sum\left[y_i\cdot(1-F_i) + (1-y_i)\cdot(-1)\cdot F_i\right] = 0.$$

$$F = \frac{\exp(u)}{1+\exp(u)} \qquad f = \frac{\exp(u)}{(1+\exp(u))^2} \qquad \frac{f}{F} = \frac{1}{1+\exp(u)} = 1-F$$

$$1-F = \frac{1}{1+\exp(u)} \qquad \frac{f}{1-F} = \frac{\exp(u)}{1+\exp(u)} = F$$

$$\sum\left[y_i \cdot (1-\hat{F}_i) + (1-y_i)\cdot(-1)\cdot \hat{F}_i\right] = 0. \qquad [\hat{\beta}_1, \hat{\beta}_2]$$

$$\sum y_i - \cancel{\sum y_i \hat{F}_i} - \sum \hat{F}_i + \cancel{\sum y_i \hat{F}_i} = 0.$$

$$\sum y_i = \sum \hat{F}_i \quad !!$$

Example

y_i	$\hat{F}_i$
0	3/5
1	3/5
0	3/5
1	3/5
1	3/5

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* < 0 \end{cases}$$

logit: $y_i^* = \beta_1 + u_i$

[on const]

---

* how to generalize to many altern-s?

multinomial logit