# Tutorial 10: Maximum Likelihood Estimation and Tobit Model

Hammad Shaikh

March 25, 2021

# Maximum Likelihood Estimation Introduction

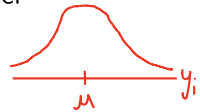
↳ Goal is to find a $\theta$ value so that it is most likely that we observe is drawn from $F_\theta$. MLE requires a parametric assumption on how data is generated.

① Parametric distributional assump. on how data is generated

- Let $y_i \sim F_\theta$ where $\theta$ is unobserved parameter

↳ Ex. $y_i \sim N(\mu, 1)$ ⇒

$\mu$ is unobs. param

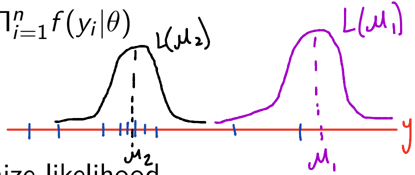



② Derive the likelihood function

- Given data $(y_1, y_2, \ldots, y_n)$ is observed, we define the likelihood of $\theta$ : $L(\theta|y_1, y_2, \ldots, y_n) = \Pi_{i=1}^{n} f(y_i|\theta)$

↳ Assume data is IID

↳ Ex. $y_i \overset{iid}{\sim} N(\mu, 1), i = 1, 2, \ldots 10$




⇒ $L(\mu_2) > L(\mu_1)$

③ max likelihood

- Estimate $\hat{\theta}_{mle}$ by selecting $\theta$ to maximize likelihood

↳ $\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\ L(\theta|y_1, y_2, \ldots, y_n) = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f(y_i|\theta)$

↳ Find $\theta$ so that $\frac{d \log(L)}{\ldots} = 0$

$d\theta$

## Estimate Probit Model using MLE (Similar for logit)

① Bernoulli assumption

- $Y_i \in \{0, 1\}$ is Bernoulli distributed with parameter $p$

$\hookrightarrow f(y_i|p) = p^{y_i}(1-p)^{1-y_i} \Rightarrow Pr(y_i=1) = p^1(1-p)^0 = p$

② Likelihood ($\beta_1$)

- Probit model: $Pr(Y_i = 1|X_i) = \Phi(\beta_1 X_i) = P_i(X_i)$

$\hookrightarrow f(y_i|X_i, P_i) = P_i(X_i)^{y_i}(1-P_i(X_i))^{1-y_i} = \Phi(\beta_1 X_i)^{y_i} \cdot (1-\Phi(\beta_1 X_i))^{1-y_i}$

- Likelihood function of $\beta_1$ given $Y_1, \ldots, Y_n$: (IID)

$$L(\beta_1|Y_1, \ldots, Y_n) = \Pi_{i=1}^N[\Phi(\beta_1 X_i)]^{(\sum_{i=1}^N Y_i)}[1-\Phi(\beta_1 X_i)]^{(N-\sum_{i=1}^N Y_i)}$$

③ Max Likelihood

$\hookrightarrow \hat{\beta}_1^{MLE} = \underset{\beta_1}{\operatorname{argmax}}\, L(\beta_1|Y_1, \ldots, Y_n) \Rightarrow \dfrac{d \log(L(\beta_1|Y_1, \ldots, Y_n))}{d\beta_1} = 0$

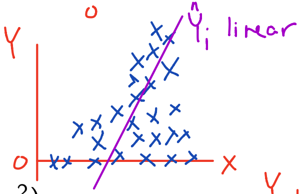
# Tobit Model for $\underbrace{\text{Corner Solutions}}$ Introduction

$\max_{c_i \geq 0} U(c_i)$ s.t. $c_i \leq w_i + e_i$ $\Rightarrow$ Corner soln $c_i = 0$


- Outcome is non-negative and a large number of 0s
  - ↳ Alcohol Consumption (0 for non-drinkers)
  - ↳ No. of Ciggaretes smoked (0 for non-smokers)

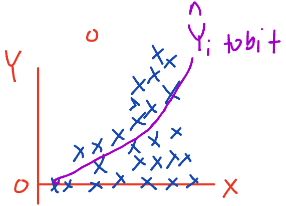
- Linear model may not be ideal
  - ↳ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ $\Rightarrow$

$\Rightarrow$ problem $\hat{Y}_i$ could be less than 0




- Tobit: $Y_i = (x_i\beta + u_i)I(x_i\beta + u_i \geq 0)$, $u_i \sim N(0, \sigma^2)$

$$= \begin{cases} X_i\beta + u_i, & X_i\beta + u_i \geq 0 \\ 0, & X_i\beta + u_i < 0 \end{cases} = \max\{0, \underbrace{X_i\beta + u_i}_{y^*}\}$$



## Tobit Model Estimation

$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \Rightarrow E(Y_i|X_i) = \beta_0 + \beta_1 X_i \Rightarrow \hat{Y}_i = \hat{E}(Y_i|X_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i$

* $E(Y|X_i)$ is a non-linear & non-negative function

- $E(Y_i|X_i) = \underbrace{Pr(Y_i = 0|X_i) * 0}_{0} + Pr(Y_i > 0|X_i)E(Y_i|X_i, Y_i > 0)$ $= Pr(Y_i > 0|X_i)\, E(Y_i|X_i, Y_i > 0)$

  $Y \geq 0$

  $Pr(Y_i > 0|X_i) = Pr(X_i\beta + u_i > 0) = Pr(u_i > -X_i\beta)$ $= 1 - Pr(u_i < -X_i\beta) = \Phi(\frac{X_i\beta}{\sigma})$

  $u_i \sim N(0, \sigma^2)$

- Tobit log-likelihood function:

$$log(L(\beta, \sigma; y, x)) = \sum_{i:y_i=0} log \left[\underbrace{1 - \Phi(\frac{x_i\beta}{\sigma})}_{Pr(Y_i=0|X_i)}\right] + \sum_{i:y_i>0} log \left[\underbrace{\frac{1}{\sigma}\phi(\frac{y_i - x_i\beta}{\sigma})}_{f(y_i|X_i, y_i>0)}\right] \Rightarrow \hat{\beta}_{MLE} = \underset{\beta}{\operatorname{argmax}} \log(L(\beta))$$

- Partial effect: $\frac{dE(y_i|x_i)}{dx_i} = \beta\Phi(\frac{x_i\beta}{\sigma})$

  ↳ Similar to probit/logit, do no interpret $\hat{\beta}$ on their own as partial effects