Logit model

$Y_i^* = X_i'\beta + \varepsilon_i$ latent variable $= -\varepsilon_i | X_i \sim \Lambda$

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

$$\begin{aligned} \Pr[Y_i = 1 | X_i] &= \Pr[X_i'\beta + \varepsilon_i > 0 \mid X_i] \\ &= \Pr[\varepsilon_i > -X_i'\beta \mid X_i] \\ &= \Pr[-\varepsilon_i \leq X_i'\beta \mid X_i] \\ &= \Lambda(X_i'\beta) \\ &= \frac{\exp\{X_i'\beta\}}{1+\exp\{X_i'\beta\}} \end{aligned}$$

$$\Rightarrow \Pr[Y_i = 0 | X_i] = \frac{1}{1+\exp\{X_i'\beta\}}$$

$$f(y_i | X_i, \beta) = \left(\frac{\exp\{X_i'\beta\}}{1+\exp\{X_i'\beta\}}\right)^{y_i}\left(\frac{1}{1+\exp\{X_i'\beta\}}\right)^{1-y_i}$$

$$\Rightarrow f(\{y_i\} | \{X_i\}, \beta) = \left(\frac{\exp\{X_i'\beta\}}{1+\exp\{X_i'\beta\}}\right)^{\sum_{i=1}^n y_i}\left(\frac{1}{1+\exp\{X_i'\beta\}}\right)^{\sum_{i=1}^n (1-y_i)}$$

$$\Rightarrow \log L(\beta | \{Y_i\}, \{X_i\}) = \sum_{i=1}^n \left[Y_i \left(\frac{\exp\{X_i'\beta\}}{1+\exp\{X_i'\beta\}}\right) + (1-Y_i)\left(\frac{1}{1+\exp\{X_i'\beta\}}\right)\right]$$

The FOCs are:

$$\sum_{i=1}^n \left(Y_i - \frac{\exp\{X_i'\hat\beta_{ML}\}}{1+\exp\{X_i'\hat\beta_{ML}\}}\right) X_i = 0$$

Goal: What is the asymptotic distribution of $\hat{\theta}_n^{ML}$?

i.e. $\sqrt{n}\,(\hat{\beta}_n^{ML} - \underbrace{\beta_0}_{\text{true value}}) \xrightarrow{d} ?$ in the above distribution.

where $\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q_0(\theta)$.

For example: Suppose $-Q_0(\beta) = E[(Y_i - E[Y_i | X_i])^2]$
$= E[(Y_i - X_i'\beta)^2]$

$\Rightarrow \beta_0 = \operatorname{argmax} Q_0(\beta)$
$= (E[X_i X_i'])^{-1} E[X_i Y_i]$

We want to show that $\hat{\theta}_n^{ML} \xrightarrow{a.s.} \theta_n$

- This will imply $\hat{\theta}_n^{ML} \xrightarrow{P} \theta_n$

On $\xrightarrow{P}$ vs $\xrightarrow{as}$: The difference between

1. winning almost \$1M ($\xrightarrow{a.s.}$)
2. almost winning \$1M ($\xrightarrow{P}$)

We will show that $\sqrt{n}\,(\hat{\theta}_n^{ML} - \theta_0) \xrightarrow{d} N(0, \Lambda_{ML})$

- It turns out that $\Lambda_{ML} = [I_1(\theta_0)]^{-1}$

$L(\theta | \{Y_i\}, \{X_i\}) = \prod_{i=1}^{n} f(Y_i; X_i, \theta)$

$\log L(\theta | \{Y_i\}, \{X_i\}) = \sum_{i=1}^{n} \ln f(Y_i; X_i, \theta)$

$\hat{\theta}_{ML} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \log L(\theta | \{Y_i\}, \{X_i\})$

For the logit model

$$\log L(\theta \mid \{Y_i\}, \{X_i\}) = \sum_{i=1}^{n} \left[Y_i X_i' \theta - \ln(1+\exp\{X_i'\theta\}) \right]$$

$$E[Y_i \mid X_i] = \frac{\exp\{X_i'\theta_0\}}{1+\exp\{X_i'\theta_0\}}$$

The FOCs say that

$$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^{n} \underbrace{\Big(Y_i - \overbrace{\frac{\exp\{X_i'\hat{\theta}_{ML}\}}{1+\exp\{X_i'\hat{\theta}_{ML}\}}}^{E[Y_i \mid X_i]} \Big)}_{\text{"residual"}} X_i = 0$$

$$\frac{1}{n} \log L = \frac{1}{n} \sum_{i=1}^{n} \log f(Y_i \mid X_i, \theta)$$

If $E[\ln f(Y_i \mid X_i, \theta)]$ exists, by WLLN

$$\frac{1}{n} \log L \xrightarrow{P} E[\ln f(Y_i \mid X_i, \theta)]$$
$$= \int \ln f(y_i \mid X_i; \theta)\, f(y_i \mid X_i, \underbrace{\theta_0}_{\text{true value}})\, dy$$

Let $W_i = (Y_i, X_i)$. If $E[|g(W_i, \theta)|] < +\infty$, then by SLLN,

$$\frac{1}{n} \sum_{i=1}^{n} g(W_i, \theta) \xrightarrow{a.s.} E[g(W_i, \theta)]$$

Thm: (USLLN) Let $W_1, \dots, W_n$ be iid. Let $g(W;\theta)$ be continuous in θ with $\theta \in \Theta$ and Θ compact. If

$$E\left[\sup_{\theta \in \Theta} |g(W_i;\theta)|\right] < +\infty, \text{ then}$$

$\frac{1}{n} \sum_{i=1}^{n} g(W_i;\theta) \xrightarrow{as} E[g(W_i, \theta)]$ uniformly in θ, and

$E\left[\sup_{\theta \in \Theta} |g(W_i, \theta)|\right]$ is continuous in θ.