Let $Y_1, \dots, Y_n$ be iid. Suppose $Y_i \sim N(\mu, \sigma^2)$

$$f_{Y_i}(y_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$$

Let $Y = (Y_1, \dots, Y_n)$

$$f_Y(y|\underbrace{\mu,\sigma^2}_{\equiv\theta}) = \prod_{i=1}^{n} f_{Y_i}(y_i|\mu,\sigma^2)$$

$$= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

$$\mathcal{L}(\theta|y) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$

Ask the question: which θ is most likely to have produced the data y?

i.e. $\hat{\theta}_{ML} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \mathcal{L}(\theta|y) = \underset{\theta\in\Theta}{\operatorname{argmax}}\, \frac{1}{n}\log\mathcal{L}(\theta|y)$

Here, $\Theta = \{(\mu,\sigma^2) : |\mu| < +\infty, \sigma^2 \geq 0\}$

$$\log\frac{1}{n}\mathcal{L}(\theta|Y) = \frac{1}{n}\left(-\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

$$= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\underbrace{\log f_{Y_i}(y_i|\mu,\sigma^2)}_{iid} \xrightarrow{P} E\left[\log f_{Y_i}(y_i|\mu,\sigma^2)\right]$$

$$\hat{\theta}_{ML} = \begin{bmatrix}\hat{\mu}_{ML}\\ \hat{\sigma}^2_{ML}\end{bmatrix} = \begin{bmatrix}\frac{1}{n}\sum_{i=1}^{n}Y_i\\ \frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{\mu}_{ML})^2\end{bmatrix}$$

Suppose $Y_i = X_i'\beta + \varepsilon_i$ where $\varepsilon_i|X_i \sim N(0,\sigma_\varepsilon^2)$

$\Rightarrow (Y_i - X_i'\beta)|X_i \sim N(0,\sigma^2)$

$\Rightarrow Y_i|X_i \sim N(X_i'\beta, \sigma^2)$

$$\Rightarrow f_\varepsilon(\varepsilon_i | X_i, \sigma_\varepsilon^2, \beta) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left\{-\frac{(Y_i - X_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}$$

$$\parallel$$

$$f_Y(\{Y_i\} | \{X_i\}, \sigma_\varepsilon^2, \beta)$$

$$\log \mathcal{L}(\sigma_\varepsilon^2, \beta | \{Y_i\}\{X_i\}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}(Y_i - X_i'\beta)^2$$

$$\Rightarrow \hat{\theta}_{ML} = \begin{bmatrix} \hat{\beta}_{ML} \\ \hat{\sigma}^2_{ML} \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'Y \\ \frac{1}{n}(Y - X\hat{\beta})'(Y - X\hat{\beta}) \end{bmatrix} = \begin{bmatrix} (X'X)^{-1}X'Y \\ \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n} \end{bmatrix}$$

Here, $\hat{\sigma}^2_{ML} = \frac{n-k}{n}\hat{\sigma}^2_{OLS}$. The bias goes to 0 as $n \to \infty$

Binary Logit model

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

Suppose $Y_i^* = X_i'\beta + \varepsilon_i$ latent variable

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

$$\begin{aligned} \Rightarrow \Pr[Y_i = 1 | X_i] &= \Pr[X_i'\beta + \varepsilon_i > 0 | X_i] \\ &= \Pr[\varepsilon_i > -X_i'\beta | X_i] \\ &= \Pr[-\varepsilon_i \le X_i'\beta | X_i] \\ &= F_{-\varepsilon_i}(X_i'\beta) \text{ where } -\varepsilon_i | X_i \sim \Lambda \\ &= \frac{\exp\{X_i'\beta\}}{1 + \exp\{X_i'\beta\}} \cdot \text{ Logit} \end{aligned}$$

Suppose $-\varepsilon_i | X_i \sim N(0,1)$

$\Pr[Y_i = 1 | X_i] = \Phi(X_i'\beta)$ Probit.