Notes on Multinomial Model

Max Leung

January 16, 2022

1 General Multinomial Model

$$y_i \in \{1, \cdots, m\}$$

Define m dummy variables

$$y_{ij} = \begin{cases} 1 & y_i = j \\ 0 & y_i \neq j \end{cases} \qquad \text{for } j \in \{1, \cdots, m\}$$

Model p_{ij}

$$p_{ij} := Pr(y_i = j|\boldsymbol{x}_i) = F_j(\boldsymbol{x}_i, \boldsymbol{\beta}) \qquad \text{for } j \in \{1, \cdots, m\}$$

where $\sum_{j=1}^m p_{ij} = 1$

1.1 ML Estimation

Log likelihood function

$$f(y_i) = p_{i1}^{y_{i1}} \cdots p_{im}^{y_{im}} = \Pi_{j=1}^m p_{ij}^{y_{ij}}$$

$$\begin{aligned} ln[L_N(\boldsymbol{\beta})] &= ln[\Pi_{i=1}^N f(y_i)] && \text{assume independence} \\ &= \sum_{i=1}^N ln[f(y_i)] \\ &= \sum_{i=1}^N ln[\Pi_{j=1}^m F_j(\boldsymbol{x}_i, \boldsymbol{\beta})^{y_{ij}}] \\ &= \sum_{i=1}^N \sum_{j=1}^m y_{ij} ln[F_j(\boldsymbol{x}_i, \boldsymbol{\beta})] \end{aligned}$$

Gradient Vector

$$\begin{aligned} \frac{\partial ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} &= \frac{\partial \sum_{i=1}^N \sum_{j=1}^m y_{ij} ln[F_j(\boldsymbol{x}_i, \boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^N \sum_{j=1}^m y_{ij} \frac{\partial ln[F_j(\boldsymbol{x}_i, \boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^N \sum_{j=1}^m y_{ij} \frac{1}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})} \frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \end{aligned}$$

FOC

$$\sum_{i=1}^N \sum_{j=1}^m \frac{y_{ij}}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})} \frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} |_{\widehat{\boldsymbol{\beta}}} = \mathbf{0}$$

1.2 Consistency of MLE

It requires the correct specification of $F_j(\boldsymbol{x}_i, \boldsymbol{\beta})$

1.3 Asymptotic Distribution of MLE

$$
\begin{aligned}
\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &\to_d N(\mathbf{0}, [\boldsymbol{I}(\boldsymbol{\beta}_0)]^{-1}) \\
&= N(\mathbf{0}, [\sum_{i=1}^{N}\sum_{j=1}^{m}\{F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)^{-1}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0} - \frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}\}]^{-1})
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{I}(\boldsymbol{\beta}_0) &:= -\mathbb{E}[\frac{\partial^2 lnL_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}|\boldsymbol{x}_i] = \mathbb{E}[\frac{\partial lnL_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \cdot \frac{\partial lnL_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}|\boldsymbol{x}_i] \\
&= -\mathbb{E}[\frac{\partial \sum_{i=1}^{N}\sum_{j=1}^{m}\frac{y_{ij}}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}|\boldsymbol{x}_i] \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\mathbb{E}[\frac{\partial \frac{y_{ij}}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}|\boldsymbol{x}_i] \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\mathbb{E}[y_{ij}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})^{-1}}{\partial \boldsymbol{\beta}'} + \frac{y_{ij}}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|\boldsymbol{x}_i]|_{\boldsymbol{\beta}_0} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\mathbb{E}[-y_{ij}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}F_j(\boldsymbol{x}_i, \boldsymbol{\beta})^{-2}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \frac{y_{ij}}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|\boldsymbol{x}_i]|_{\boldsymbol{\beta}_0} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\{-\mathbb{E}[y_{ij}|\boldsymbol{x}_i]\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}F_j(\boldsymbol{x}_i, \boldsymbol{\beta})^{-2}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \frac{\mathbb{E}[y_{ij}|\boldsymbol{x}_i]}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\}|_{\boldsymbol{\beta}_0} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\{-F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}F_j(\boldsymbol{x}_i, \boldsymbol{\beta})^{-2}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \frac{F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}\frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\}|_{\boldsymbol{\beta}_0} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\{-F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}_0}F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)^{-2}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0} + \frac{F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)}{F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)}\frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}\} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m}\{-F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)^{-1}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0} + \frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}\} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{m}\{F_j(\boldsymbol{x}_i, \boldsymbol{\beta}_0)^{-1}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0} - \frac{\partial^2 F_j(\boldsymbol{x}_i, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}|_{\boldsymbol{\beta}_0}\}
\end{aligned}
$$

1.4 McFadden's Pseudo-R^2

$$
\begin{aligned}
R^2_{multi} &= 1 - \frac{lnL_{fit}}{lnL_0} \\
&= \frac{lnL_0 - lnL_{fit}}{lnL_0} \\
&= \frac{lnL_0 - lnL_{fit}}{lnL_0 - 0} \\
&= \frac{lnL_0 - lnL_{fit}}{lnL_0 - lnL_{max}} \\
&= \frac{lnL_{fit} - lnL_0}{lnL_{max} - lnL_0}
\end{aligned}
$$

$$
L_{max} = \Pi_{i=1}^{N}\Pi_{j=1}^{m}p_{ij}^{y_{ij}} = \Pi_{i=1}^{N}0^0 \cdots 1^1 \cdots 0^0 = \Pi_{i=1}^{N}1 \cdots 1 \cdots 1 = \Pi_{i=1}^{N}1 = 1
$$

1.5 Special Case: Conditional Logit Model

if $F_j(\boldsymbol{x}_i, \boldsymbol{\beta}) = F_{softmax}(\boldsymbol{x}'_{ij}\boldsymbol{\beta})$

$$p_{ij} := Pr(y_i = j|\boldsymbol{x}_i) = F_{softmax}(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) = \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})} \qquad \text{for } j \in \{1, \cdots, m\}$$

1.5.1 Gradient Vector

$$\begin{aligned}
\frac{\partial p_{ij}}{\partial \boldsymbol{\beta}} &= \frac{\partial \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})}}{\partial \boldsymbol{\beta}} \\
&= \frac{\partial exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}(\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}))^{-1} + \frac{\partial(\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}))^{-1}}{\partial \boldsymbol{\beta}} exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) \\
&= exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})\boldsymbol{x}_{ij}(\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}))^{-1} - (\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}))^{-2}[\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})\boldsymbol{x}_{il}] exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}) \\
&= \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})}\boldsymbol{x}_{ij} - \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{(\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}))^2}[\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})\boldsymbol{x}_{il}] \\
&= \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})}\boldsymbol{x}_{ij} - \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta})}{\sum_{l=1}^{m} exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})}[\sum_{l=1}^{m} \frac{exp(\boldsymbol{x}'_{il}\boldsymbol{\beta})}{\sum_{k=1}^{m} exp(\boldsymbol{x}'_{ik}\boldsymbol{\beta})}\boldsymbol{x}_{il}] \\
&= p_{ij}\boldsymbol{x}_{ij} - p_{ij}[\sum_{l=1}^{m} p_{il}\boldsymbol{x}_{il}] \\
&= p_{ij}\boldsymbol{x}_{ij} - p_{ij}\bar{\boldsymbol{x}}_i \qquad \bar{\boldsymbol{x}}_i = \textstyle\sum_{l=1}^{m} p_{il}\boldsymbol{x}_{il} \\
&= p_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)
\end{aligned}$$

$$\frac{\partial ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{j=1}^{m} \frac{y_{ij}}{p_{ij}}\frac{\partial p_{ij}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N}\sum_{j=1}^{m} \frac{y_{ij}}{p_{ij}} p_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i) = \sum_{i=1}^{N}\sum_{j=1}^{m} y_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)$$

1.5.2 Asymptotic Distribution

$$\begin{aligned}
\frac{\partial^2 ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{m} \partial y_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)}{\partial \boldsymbol{\beta}'} = \sum_{i=1}^{N}\sum_{j=1}^{m} \frac{\partial y_{ij}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)}{\partial \boldsymbol{\beta}'} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m} \frac{\partial y_{ij}\bar{\boldsymbol{x}}_i}{\partial \boldsymbol{\beta}'} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m} \frac{\partial y_{ij}\sum_{l=1}^{m} p_{il}\boldsymbol{x}_{il}}{\partial \boldsymbol{\beta}'} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m} y_{ij}\sum_{l=1}^{m} \frac{\partial p_{il}}{\partial \boldsymbol{\beta}'}\boldsymbol{x}_{il} \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{m} y_{ij}\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il} - \bar{\boldsymbol{x}}_i)'\boldsymbol{x}_{il} \\
&= -\sum_{i=1}^{N}[\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il} - \bar{\boldsymbol{x}}_i)\boldsymbol{x}'_{il}]\sum_{j=1}^{m} y_{ij} \\
&= -\sum_{i=1}^{N}\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il} - \bar{\boldsymbol{x}}_i)\boldsymbol{x}'_{il} \qquad \textstyle\sum_j y_{ij} = 1 \\
&= -\sum_{i=1}^{N}\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{il} - \bar{\boldsymbol{x}}_i)'
\end{aligned}$$

The last line is due to $\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il}-\bar{\boldsymbol{x}}_i)\bar{\boldsymbol{x}}_i' = \sum_{l=1}^{m}(p_{il}\boldsymbol{x}_{il} - p_{il}\bar{\boldsymbol{x}}_i)\bar{\boldsymbol{x}}_i' = (\sum_{l=1}^{m} p_{il}\boldsymbol{x}_{il} - [\sum_{l=1}^{m} p_{il}]\bar{\boldsymbol{x}}_i)\bar{\boldsymbol{x}}_i' = (\bar{\boldsymbol{x}}_i - 1\cdot\bar{\boldsymbol{x}}_i)\bar{\boldsymbol{x}}_i' = \mathbf{0}$
Thus, the asymptotic variance is $[\boldsymbol{I}(\boldsymbol{\beta}_0)]^{-1} := [-\mathbb{E}(\frac{\partial^2 ln[L_N(\boldsymbol{\beta})]}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}|\boldsymbol{x}_i)]^{-1} = [\sum_{i=1}^{N}\sum_{l=1}^{m} p_{il}(\boldsymbol{x}_{il}-\bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{il}-\bar{\boldsymbol{x}}_i)']^{-1}$

1.6 Special Case: Multinomial Logit Model

if $F_j(\boldsymbol{x}_i,\boldsymbol{\beta}) = F_{softmax}(\boldsymbol{x}_i'\boldsymbol{\beta}_j)$

$$p_{ij} := Pr(y_i = j|\boldsymbol{x}_i) = F_{softmax}(\boldsymbol{x}_i'\boldsymbol{\beta}_j) = \frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)} \qquad \text{for } j \in \{1,\cdots,m\}$$

1.6.1 Gradient Vector

$$\begin{aligned}
\frac{\partial p_{ij}}{\partial\boldsymbol{\beta}_j} &= \frac{\partial \frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}}{\partial\boldsymbol{\beta}_j} \\
&= \frac{\partial exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\partial\boldsymbol{\beta}_j}(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-1} + \frac{\partial(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-1}}{\partial\boldsymbol{\beta}_j} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j) \\
&= exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\boldsymbol{x}_i(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-1} - (\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-2}[\sum_{l=1}^{m}\frac{\partial exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}{\partial\boldsymbol{\beta}_j}] exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j) \\
&= exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\boldsymbol{x}_i(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-1} - (\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-2} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\boldsymbol{x}_i exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j) \\
&= \frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}\boldsymbol{x}_i - (\frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)})^2\boldsymbol{x}_i \\
&= p_{ij}\boldsymbol{x}_i - p_{ij}^2\boldsymbol{x}_i
\end{aligned}$$

$$\begin{aligned}
\frac{\partial p_{ij}}{\partial\boldsymbol{\beta}_k} &= \frac{\partial \frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}}{\partial\boldsymbol{\beta}_k} \\
&= exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\frac{\partial(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-1}}{\partial\boldsymbol{\beta}_k} \\
&= -exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-2}\sum_{l=1}^{m}\frac{\partial exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}{\partial\boldsymbol{\beta}_k} \\
&= -exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)(\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l))^{-2} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_k)\boldsymbol{x}_i \\
&= -\frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}\frac{exp(\boldsymbol{x}_i'\boldsymbol{\beta}_k)}{\sum_{l=1}^{m} exp(\boldsymbol{x}_i'\boldsymbol{\beta}_l)}\boldsymbol{x}_i \\
&= -p_{ij}p_{ik}\boldsymbol{x}_i
\end{aligned}$$

Thus,

$$\frac{\partial p_{ij}}{\partial \boldsymbol{\beta}_l} = 1(l=j)p_{ij}\boldsymbol{x}_i - p_{ij}p_{il}\boldsymbol{x}_i = p_{ij}(1(l=j) - p_{il})\boldsymbol{x}_i$$

$$\begin{aligned}
\frac{\partial ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}_l} &= \sum_{i=1}^{N}\sum_{j=1}^{m}\frac{y_{ij}}{p_{ij}}\frac{\partial p_{ij}}{\partial \boldsymbol{\beta}_l} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{m}\frac{y_{ij}}{p_{ij}}p_{ij}(1(l=j) - p_{il})\boldsymbol{x}_i \\
&= \sum_{i=1}^{N}\boldsymbol{x}_i\sum_{j=1}^{m}y_{ij}(1(l=j) - p_{il}) \\
&= \sum_{i=1}^{N}\boldsymbol{x}_i[-p_{il}\sum_{j\neq l}^{m}y_{ij} + y_{il}(1-p_{il})] \\
&= \sum_{i=1}^{N}\boldsymbol{x}_i[-p_{il}\sum_{j\neq l}^{m}y_{ij} + y_{il} - y_{il}p_{il}] \\
&= \sum_{i=1}^{N}\boldsymbol{x}_i[-p_{il}\sum_{j=1}^{m}y_{ij} + y_{il}] \\
&= \sum_{i=1}^{N}\boldsymbol{x}_i[y_{il} - p_{il}] && \text{As } \textstyle\sum_{j=1}^{m} y_{ij} = 1
\end{aligned}$$

1.6.2 Asymptotic Distribution

$$\begin{aligned}
\frac{\partial^2 ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_k'} &= \frac{\partial \sum_{i=1}^{N}\boldsymbol{x}_i[y_{ij} - p_{ij}]}{\partial \boldsymbol{\beta}_k'} \\
&= -\sum_{i=1}^{N}\frac{\partial p_{ij}}{\partial \boldsymbol{\beta}_k'}\boldsymbol{x}_i \\
&= -\sum_{i=1}^{N}p_{ij}(1(k=j) - p_{ik})\boldsymbol{x}_i\boldsymbol{x}_i'
\end{aligned}$$

Thus, the asymptotic variance is $[\boldsymbol{I}(\boldsymbol{\beta}_0)]^{-1} := [-\mathbb{E}(\frac{\partial^2 ln[L_N(\boldsymbol{\beta})]}{\partial \boldsymbol{\beta}_j \partial \boldsymbol{\beta}_k'}|\boldsymbol{x}_i)]^{-1} = [\sum_{i=1}^{N}p_{ij}(1(k=j) - p_{ik})\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}$

1.7 Special Case: Universal Logit Model

if $F_j(\boldsymbol{x}_i, \boldsymbol{\beta}) = F_{softmax}(ln[V_{ij}])$

$$p_{ij} := Pr(y_i = j|\boldsymbol{x}_i) = F_{softmax}(V_{ij}) = \frac{exp(ln[V_{ij}])}{\sum_{l=1}^{m}exp(ln[V_{il}])} = \frac{V_{ij}}{\sum_{l=1}^{m}V_{il}} \qquad \text{for } j \in \{1, \cdots, m\}$$

2 Additive Random Utility Model

Suppress i,

$$U_j = V_j + \varepsilon_j \qquad j \in \{1, \cdots, m\}$$

$$\begin{aligned}
p_j &:= Pr(y = j|\boldsymbol{x}) \\
&= Pr(U_j > U_k \ \forall k \neq j|\boldsymbol{x}) \\
&= Pr(V_j + \varepsilon_j > V_k + \varepsilon_k \ \forall k \neq j|\boldsymbol{x}) \\
&= Pr(V_j - V_k > \varepsilon_k - \varepsilon_j \ \forall k \neq j|\boldsymbol{x}) \\
&= Pr(\varepsilon_k - \varepsilon_j < V_j - V_k \ \forall k \neq j|\boldsymbol{x}) \\
&= Pr(\tilde{\varepsilon}_{kj} < -\tilde{V}_{kj} \ \forall k \neq j|\boldsymbol{x}) \\
&= F_j(-\tilde{V}_{kj} \ \forall k \neq j) \\
&= \int_{-\infty}^{-\tilde{V}_{1j}} \cdots \int_{-\infty}^{-\tilde{V}_{j-1,j}} \int_{-\infty}^{-\tilde{V}_{j+1,j}} \cdots \int_{-\infty}^{-\tilde{V}_{m,j}} f(\tilde{\varepsilon}_{1j}, \cdots, \tilde{\varepsilon}_{j-1,j}, \tilde{\varepsilon}_{j+1,j}, \cdots, \tilde{\varepsilon}_{mj}) \partial\tilde{\varepsilon}_{1j} \cdots \partial\tilde{\varepsilon}_{j-1,j} \partial\tilde{\varepsilon}_{j+1,j} \cdots \partial\tilde{\varepsilon}_{mj}
\end{aligned}$$

2.1 Special Case: Conditional Logit Model, Multinomial Logit Model

If $\varepsilon_k \ \forall k$ are independent and all follow Type 1 Extreme Value Distribution (or log Weibull Distribution). It can be shown $p_j := Pr(y = j|\boldsymbol{x}) = \frac{exp(V_j)}{\sum_{l=1}^{m} exp(V_l)}$ by direct integration method. If $V_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta}$, it is Conditional Logit Model. If $V_{ij} = \boldsymbol{x}_i'\boldsymbol{\beta}_j$, it is Multinomial Logit Model.

2.2 Special Case: Generalized Extreme Value Model (McFadden ,1978)

Nested Logit Model is a special case of GEV model. Conditional Logit and Multinomial Logit Models are special case of Nested Logit Model.

If $\boldsymbol{\varepsilon}$ follows (multivariate) GEV distribution i.e., $F(\boldsymbol{\varepsilon}) = F(\varepsilon_1, \cdots, \varepsilon_m) = exp[-G(e^{-\varepsilon_1}, \cdots, e^{-\varepsilon_m})]$. If $G(.)$ satisfies certain conditions, then $p_j := Pr(y = j|\boldsymbol{x}) = e^{V_j} \frac{G_j(e^{-V_1}, \cdots, e^{-V_m})}{G(e^{-V_1}, \cdots, e^{-V_m})}$ where G_j is derivative of the j argument.

2.2.1 Nested Logit Model

$$F(\varepsilon_{11}, \cdots, \varepsilon_{1K_1}; \cdots; \varepsilon_{J1}, \cdots, \varepsilon_{JK_J}) = exp[-G(e^{\varepsilon_{11}}, \cdots, e^{\varepsilon_{1K_1}}; \cdots; e^{\varepsilon_{J1}}, \cdots, e^{\varepsilon_{JK_J}})]$$

$$p_{jk} := Pr(y_{jk} = 1|\boldsymbol{x}) = e^{V_{jk}} \frac{G_{jk}(e^{-V_{11}}, \cdots, e^{-V_{1K_1}}; \cdots; e^{-V_{J1}}, \cdots, e^{-V_{JK_J}})}{G(e^{-V_{11}}, \cdots, e^{-V_{1K_1}}; \cdots; e^{-V_{J1}}, \cdots, e^{-V_{JK_J}}))}$$

Let $Y_{jk} = e^{V_{jk}}$ and specify

$$G(\boldsymbol{Y}) = G(Y_{11}, \cdots, Y_{1N_1}; \cdots; Y_{J1}, \cdots, Y_{JN_J}) = \sum_{j=1}^{J} a_j (\sum_{k=1}^{K_j} Y_{jk}^{1/\rho_j})^{\rho_j}$$

We have

$$
\begin{aligned}
G_{j'k'} &= \frac{\partial G(\boldsymbol{Y})}{\partial Y_{j'k'}} \\
&= \frac{\partial \sum_{j=1}^{J} a_j (\sum_{k=1}^{K_j} Y_{jk}^{1/\rho_j})^{\rho_j}}{\partial Y_{j'k'}} \\
&= a_{j'} \frac{\partial (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}}}{\partial Y_{j'k'}} \\
&= a_{j'} \rho_{j'} (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}-1} \sum_{k=1}^{K_{j'}} \frac{\partial Y_{j'k}^{1/\rho_{j'}}}{\partial Y_{j'k'}} \\
&= a_{j'} \rho_{j'} (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}-1} \frac{\partial Y_{j'k'}^{1/\rho_{j'}}}{\partial Y_{j'k'}} \\
&= a_{j'} \rho_{j'} (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}-1} (1/\rho_{j'}) Y_{j'k'}^{1/\rho_{j'}-1} \\
&= a_{j'} (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}-1} Y_{j'k'}^{1/\rho_{j'}-1}
\end{aligned}
$$

$$
Y_{j'k'} G_{j'k'} = a_{j'} (\sum_{k=1}^{K_{j'}} Y_{j'k}^{1/\rho_{j'}})^{\rho_{j'}-1} Y_{j'k'}^{1/\rho_{j'}}
$$

$$
\begin{aligned}
p_{jk} &= Y_{jk} \frac{G_{jk}}{G(\boldsymbol{Y})} \\
&= \frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j - 1} Y_{jk}^{1/\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}}
\end{aligned}
$$

$$
\begin{aligned}
p_j := \sum_{k=1}^{K_j} p_{jk} &= \sum_{k=1}^{K_j} \frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j - 1} Y_{jk}^{1/\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}} \\
&= \frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j - 1} \sum_{k=1}^{K_j} Y_{jk}^{1/\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}} \\
&= \frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}}
\end{aligned}
$$

$$
\begin{aligned}
p_{k|j} = \frac{p_{jk}}{p_j} &= \frac{\frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j - 1} Y_{jk}^{1/\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}}}{\frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j}}{\sum_{m=1}^{J} a_m (\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}}} \\
&= \frac{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j - 1} Y_{jk}^{1/\rho_j}}{a_j (\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j}} \\
&= \frac{Y_{jk}^{1/\rho_j}}{\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j}}
\end{aligned}
$$

Substitute back

$$
\begin{aligned}
(e^{V_{jl}})^{1/\rho_j} &= (e^{\boldsymbol{z}_j'\boldsymbol{\alpha}+\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j})^{1/\rho_j} \\
&= e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j+\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j} \\
&= e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}
\end{aligned}
$$

$$
\begin{aligned}
\sum_{l=1}^{K_j}(e^{V_{jl}})^{1/\rho_j} &= \sum_{l=1}^{K_j} e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j} \\
&= e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} \sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j} \\
&= e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} exp(ln[\sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}]) \\
&= e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} exp(l_j)
\end{aligned}
$$

$$
\begin{aligned}
(\sum_{l=1}^{K_j}(e^{V_{jl}})^{1/\rho_j})^{\rho_j} &= (exp(\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j)exp(l_j))^{\rho_j} \\
&= (exp(\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j + l_j))^{\rho_j} \\
&= exp(\boldsymbol{z}_j'\boldsymbol{\alpha} + \rho_j l_j)
\end{aligned}
$$

$$
\begin{aligned}
p_j &= \frac{a_j(\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j})^{\rho_j}}{\sum_{m=1}^{J} a_m(\sum_{l=1}^{K_m} Y_{ml}^{1/\rho_m})^{\rho_m}} \\
&= \frac{a_j(\sum_{l=1}^{K_j} e^{V_{jl}^{1/\rho_j}})^{\rho_j}}{\sum_{m=1}^{J} a_m(\sum_{l=1}^{K_m} e^{V_{ml}^{1/\rho_m}})^{\rho_m}} \\
&= \frac{a_j exp(\boldsymbol{z}_j'\boldsymbol{\alpha} + \rho_j l_j)}{\sum_{m=1}^{J} a_m exp(\boldsymbol{z}_m'\boldsymbol{\alpha} + \rho_m l_m)}
\end{aligned}
$$

$$
\begin{aligned}
p_{k|j} &= \frac{Y_{jk}^{1/\rho_j}}{\sum_{l=1}^{K_j} Y_{jl}^{1/\rho_j}} \\
&= \frac{e^{V_{jk}^{1/\rho_j}}}{\sum_{l=1}^{K_j} e^{V_{jl}^{1/\rho_j}}} \\
&= \frac{e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}_j/\rho_j}}{\sum_{l=1}^{K_j} e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}} \\
&= \frac{e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}_j/\rho_j}}{e^{\boldsymbol{z}_j'\boldsymbol{\alpha}/\rho_j} \sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}} \\
&= \frac{e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}_j/\rho_j}}{\sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
p_{jk} &= p_{k|j}p_j \\
&= \frac{e^{\boldsymbol{x}_{jk}'\boldsymbol{\beta}_j/\rho_j}}{\sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}} \frac{a_j exp(\boldsymbol{z}_j'\boldsymbol{\alpha} + \rho_j l_j)}{\sum_{m=1}^{J} a_m exp(\boldsymbol{z}_m'\boldsymbol{\alpha} + \rho_m l_m)}
\end{aligned}
$$

where $I_j = ln[\sum_{l=1}^{K_j} e^{\boldsymbol{x}_{jl}'\boldsymbol{\beta}_j/\rho_j}]$ is log sum
It can be shown that $\rho_j = \sqrt{1 - Corr[\varepsilon_{jk}, \varepsilon_{jl}]}$. $\rho_j = 1 \iff Corr[\varepsilon_{jk}, \varepsilon_{jl}] = 0$ i.e., Multinomial Logit Model

2.2.2 FIML Estimation of Nested Logit Model

$$\begin{aligned}
f(\boldsymbol{y}_i) &= \Pi_{j=1}^{J}\Pi_{k=1}^{K_j} p_{ijk}^{y_{ijk}} \\
&= \Pi_{j=1}^{J}\Pi_{k=1}^{K_j} (p_{ik|j}p_{ij})^{y_{ijk}} \\
&= \Pi_{j=1}^{J}\Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}} p_{ij}^{y_{ijk}} \\
&= \Pi_{j=1}^{J}\Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}} \Pi_{k=1}^{K_j} p_{ij}^{y_{ijk}} \\
&= \Pi_{j=1}^{J}\Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}} \cdot p_{ij}^{\sum_{k=1}^{K_j} y_{ijk}} \\
&= \Pi_{j=1}^{J} p_{ij}^{\sum_{k=1}^{K_j} y_{ijk}} \Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}} \\
&= \Pi_{j=1}^{J} p_{ij}^{y_{ij}} \Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}}
\end{aligned}$$

$$\begin{aligned}
ln[L(\boldsymbol{\alpha}, \boldsymbol{\beta}_j, \rho_j)] &= ln[\Pi_{i=1}^{N} f(\boldsymbol{y}_i)] \qquad \text{assume independence of i} \\
&= ln[\Pi_{i=1}^{N}\Pi_{j=1}^{J}\Pi_{k=1}^{K_j} p_{ik|j}^{y_{ijk}} p_{ij}^{y_{ijk}}] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} ln[p_{ik|j}^{y_{ijk}} p_{ij}^{y_{ijk}}] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} \{y_{ijk} ln[p_{ik|j}] + y_{ijk} ln[p_{ij}]\} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} y_{ijk} ln[p_{ik|j}] + \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} y_{ijk} ln[p_{ij}] \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} y_{ijk} ln[p_{ik|j}] + \sum_{i=1}^{N}\sum_{j=1}^{J} ln[p_{ij}] \underbrace{\sum_{k=1}^{K_j} y_{ijk}}_{y_{ij}} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} y_{ijk} ln[\frac{e^{\boldsymbol{x}'_{ijk}\boldsymbol{\beta}_j/\rho_j}}{\sum_{l=1}^{K_j} e^{\boldsymbol{x}'_{ijl}\boldsymbol{\beta}_j/\rho_j}}] + \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln[\frac{exp(\boldsymbol{z}'_{ij}\boldsymbol{\alpha} + \rho_j l_{ij})}{\sum_{m=1}^{J} exp(\boldsymbol{z}'_{im}\boldsymbol{\alpha} + \rho_m l_{im})}]
\end{aligned}$$

2.2.3 LIML Estimation of Nested Logit Model

It is less efficient compared to FIML
First Stage
Estimate $\boldsymbol{\beta}_j/\rho_j$ with maximizing log likelihood $\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K_j} y_{ijk} ln[\frac{e^{\boldsymbol{x}'_{ijk}\boldsymbol{\beta}_j/\rho_j}}{\sum_{l=1}^{K_j} e^{\boldsymbol{x}'_{ijl}\boldsymbol{\beta}_j/\rho_j}}]$

Second Stage
Estimate $\boldsymbol{\alpha}$ and ρ_j with maximizing log likelihood $\sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij} ln[\frac{exp(\boldsymbol{z}'_{ij}\boldsymbol{\alpha}+\rho_j l_{ij})}{\sum_{m=1}^{J} exp(\boldsymbol{z}'_{im}\boldsymbol{\alpha}+\rho_m l_{im})}]$

Get $\boldsymbol{\beta}_j$ by $\boldsymbol{\beta}_j/\rho_j \cdot \rho_j$

2.3 Random Parameters Logit Model

If $V_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_i$ i.e., Conditional Logit Model with heterogeneous beta.

$$U_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_i + \varepsilon_{ij}$$

where

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})$$

It can be written as

$$\begin{aligned} U_{ij} &= \boldsymbol{x}'_{ij}(\boldsymbol{\beta} + \boldsymbol{u}_i) + \varepsilon_{ij} \\ &= \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \underbrace{(\boldsymbol{x}'_{ij}\boldsymbol{u}_i + \varepsilon_{ij})}_{v_{ij}} \end{aligned} \qquad \boldsymbol{u}_i \sim N(\boldsymbol{0}, \Gamma_{\boldsymbol{\beta}})$$

$Cov(v_{ij}, v_{ik}) = Cov(\boldsymbol{x}'_{ij}\boldsymbol{u}_i + \varepsilon_{ij}, \boldsymbol{x}'_{ik}\boldsymbol{u}_i + \varepsilon_{ik}) = Cov(\boldsymbol{x}'_{ij}\boldsymbol{u}_i, \boldsymbol{u}'_i\boldsymbol{x}_{ik}) = \boldsymbol{x}'_{ij}\Gamma_{\boldsymbol{\beta}}\boldsymbol{x}_{ij}$ for $j \neq k$ i.e., Conditional Logit Model with correlated error v_{ij} across alternative.

2.3.1 Maximum Simulated Estimation

$$\begin{aligned} p_{ij} := Pr(y_i = j|\boldsymbol{x}_i) &= \int Pr(y_i = j, \beta_i|\boldsymbol{x}_i, \boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})d\boldsymbol{\beta}_i \\ &= \int Pr(y_i = j|\boldsymbol{\beta}_i, \boldsymbol{x}_i, \boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})\phi(\boldsymbol{\beta}_i|\boldsymbol{x}_i, \boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})d\boldsymbol{\beta}_i \\ &= \int Pr(y_i = j|\boldsymbol{\beta}_i, \boldsymbol{x}_i)\phi(\boldsymbol{\beta}_i|\boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})d\boldsymbol{\beta}_i \\ &= \int \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_i)}{\sum_{l=1}^m exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}_i)}\phi(\boldsymbol{\beta}_i|\boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})d\boldsymbol{\beta}_i \\ &= \mathbb{E}[\frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_i)}{\sum_{l=1}^m exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}_i)}] \end{aligned}$$

Simulate p_{ij} by drawing $\boldsymbol{\beta}_i$ from $\phi(\boldsymbol{\beta}_i|\boldsymbol{\beta}, \Gamma_{\boldsymbol{\beta}})$ S times

$$\widehat{p}_{ij} = S^{-1}\sum_{s=1}^{S} \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_i^s)}{\sum_{l=1}^m exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}_i^s)}$$

Thus, the log simulated likelihood function is

$$\begin{aligned} ln[\widehat{L}] &= ln[\Pi_{i=1}^N \Pi_{j=1}^m \hat{p}_{ij}^{y_{ij}}] \\ &= \sum_{i=1}^N \sum_{j=1}^m y_{ij} ln[\hat{p}_{ij}] \\ &= \sum_{i=1}^N \sum_{j=1}^m y_{ij} ln[S^{-1}\sum_{s=1}^S \frac{exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_i^s)}{\sum_{l=1}^m exp(\boldsymbol{x}'_{il}\boldsymbol{\beta}_i^s)}] \end{aligned}$$

3 References

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications