1 Introduction

In this project, we are going to investigate the properties of different model structures to include latent class variables in discrete choice model. Specifically, we will compare models which regard latent class as the source of individual state with ones which regard latent class as the result of individual state.

Before we get into details, we first review the basic discrete choice models. Assume there are a set of options $\{1, \cdots, n\}$, and for individual $j \in \{1, \cdots, N\}$, the general property can be modeled with $x_S(j)$ and choice-specific property can be modeled with $x_I(i,j)$. We can then concatenate them into overall state vector x_{ij} for each option i. Then we can use a single coefficient vector β to model the choice behavior with the following model

$$p_{ij} = \frac{\exp(\beta^T x_{ij})}{\sum_i \exp(\beta^T x_{ij})}.$$

This is called multinomial logit model (MNL). If we want to make the coefficient β stochastic, we can use

$$p_{ij} = \int_\beta \frac{\exp(\beta^T x_{ij})}{\sum_i \exp(\beta^T x_{ij})} dF(\beta).$$

This is called mixture logit model (MMNL).

2 Classes as Results of States

Now consider latent class $k \in \{1, \cdots, K\}$, which is determined by global state $x = [x_S, x_I(1), \cdots, x_I(n)]$:

$$p(k|x_j) = \frac{\exp(\alpha_k^T x_j)}{\sum_k \exp(\alpha_k^T x_j)}.$$

where α_k are parameters to be determined. Then we can consider class-specific choice coefficient β_k:

$$p(i|x_j, k) = \frac{\exp(\beta_k^T x_{ij})}{\sum_i \exp(\beta_k^T x_{ij})}.$$

Next, we will discuss how to estimate these parameters for MNL and MMNL.

2.1 Multinomial Logit Model

The graphical representation of this model can be expressed as:

$$x \to k \to \beta \to I$$

The log likelihood can be expressed as

$$L = \sum_j \log(\sum_k p(k|x_j)p(I_j|j,k)).$$

in which $p(I_j|j,k) = \sum_i p(i|x_{ij},k)l(i|j)$ and l is the observed label. If we let

$$p_X(j) = \sum_k p(k|x_j)p(I_j|j,k)$$

$$p_x(k,j) = p(k|x_j)p(I_j|j,k)/p_X(j)$$

the gradient can be expressed as

$$\frac{\partial}{\partial \alpha_k} L = \sum_j \frac{1}{p_X(j)} (\sum_{k'} p(I_j|j,k') \frac{\partial}{\partial \alpha_k} p(k|x_j))$$

$$= \sum_j \frac{p(k|x_j)x_j}{p_X(j)} (p(I_j|j,k) - p_X(j))$$

$$= \sum_j (p_x(k,j) - p(k|x_j))x_j$$

and

$$\frac{\partial}{\partial \beta_k} L = \sum_j \frac{p(k|x_j)}{p(j)} \frac{\partial}{\partial \beta_k} p(I_j|j,k)$$

$$= \sum_j \frac{p(k|x_j)}{p(j)} (\sum_i l(i|j) \frac{\partial}{\partial \beta_k} p(i|x_{ij},k))$$

$$= \sum_j \frac{p(k|x_j)}{p(j)} (\sum_i l(i|j)(p(i|x_{ij},k)x_{ij} - \sum_{i'} p(i|x_{ij},k)p(i'|x_{i'j},k)x_{i'j}))$$

$$= \sum_j \frac{p(k|x_j)}{p(j)} p(I_j|j,k)(\sum_i l(i|j)x_{ij} - \sum_{i'} p(i'|x_{i'j},k)x_{i'j}))$$

$$= \sum_j p_x(k,j)(\sum_i (l(i|j) - p(i|x_{ij},k))x_{ij})$$

Then gradient based methods can be applied.

2.2 Multinomial Logit Model with Quadratic Class Probability

Now let us consider the exact same model as above but with a different class probability

$$p(k|x_j) = \frac{\exp(-(\alpha_k^T x_j)^2)}{\sum_k \exp(-(\alpha_k^T x_j)^2)}.$$

In almost the same way, the gradient can be expressed as

$$\frac{\partial}{\partial \alpha_k} L = \sum_j (p_x(k,j) - p(k|x_j))(-2\alpha_k^T x_j)x_j$$

and

$$\frac{\partial}{\partial \beta_k} L = \sum_j p_x(k,j)(\sum_i (l(i|j) - p(i|x_{ij},k))x_{ij})$$

Then gradient based methods can be applied.

2.3 Mixture Logit Model

The graphical representation of this model can be expressed as:

$$x \to k \to (\mu, \Sigma) \to \beta \to I$$

Now, the log likelihood can be expressed as

$$L = \sum_j \log(p_X(j)) = \sum_j \log(\sum_k p(k|x_j)E_{\beta_k}(p(I_j|j,\beta_k))).$$

2

Assume $\beta_k \sim N(\mu_k, \Sigma_k)$. Since we have semidefinite constraints on Σ_k which is hard to modeled in gradient method, we consider to use EM algorithm. Specifically, we consider two latent variables: k and β_k. The posterior of the joint distribution is:

$$
\begin{aligned}
h(k,j) &= p(k, \beta_k | I_j, x_j) \\
&= \frac{p(k|x_j)p(I_j|j,\beta_k)p(\beta_k)}{\sum_{k'} p(k'|x_j)E_{\beta_{k'}}(p(I_j|j,\beta_{k'}))}
\end{aligned}
$$

and the approximation becomes

$$
\begin{aligned}
L' &= \sum_{j,k} \int_{\beta_k} h(k,j)(\log p(k|x_j) + \log p(I_j|j,\beta_k) + \log p(\beta_k))d\beta_k \\
&= \sum_{j,k} \frac{E_{\beta_k}(p(k|x_j)p(I_j|j,\beta_k)(\log p(k|x_j) + \log p(I_j|j,\beta_k) + \log p(\beta_k)))}{\sum_k p(k|x_j)E_{\beta_k}(p(I_j|j,\beta_k))}
\end{aligned}
\tag{1}
$$

Then, we can then draw samples and use Monte-Carlo method to approximate the expectation $E_{\beta_k}(\cdot)$. Assume we take S samples for each j and β_k: $\beta_{k,j,1}, \cdots, \beta_{k,j,S}$. The new approximation becomes

$$
L'' = \sum_{j,k,s} \hat{p_x}(k,j,s)(\log p(k|x_j) + \log p(I_j|j,\beta_k) + \log p(\beta_k))
\tag{2}
$$

with

$$
\hat{p_x}(k,j,s) = \frac{p(k|x_j)p(I_j|j,\beta_{k,j,s})}{\sum_{k',s'} p(k'|x_j)p(I_j|j,\beta_{k',j,s'})}
$$

Now, for this approximated problem, we can calculate the optimum of μ and Σ directly:

$$
\begin{aligned}
\mu_k &= \frac{\sum_{j,s} \hat{p_x}(k,j,s)\beta_{j,k,s}}{\sum_{j,s} \hat{p_x}(k,j,s)} \\
\Sigma_k &= \frac{\sum_{j,s} \hat{p_x}(k,j,s)(\beta_{j,k,s} - \mu_k)(\beta_{j,k,s} - \mu_k)^T}{\sum_{j,s} \hat{p_x}(k,j,s)}
\end{aligned}
$$

but we need to apply gradient methods for α_k:

$$
\frac{\partial}{\partial \alpha_k} L'' = \sum_{j,s} (\hat{p_x}(k,j,s) - p(k|x_j))x_j
$$

3 Classes as Sources of States

Now consider latent class $k \in \{1, \cdots, K\}$ with occurence probability $p(k) = \pi_k$, which generates both global state $x = [x_S, x_I(1), \cdots, x_I(n)]$ with

$$
p(x_j|k) \sim N(\theta_k, \Lambda_k)
$$

and choice coefficient β_k:

$$
p(i|x_j, k) = \frac{\exp(\beta_k^T x_{ij})}{\sum_i \exp(\beta_k^T x_{ij})}.
$$

Again, since we have normal distribution now, we should consider EM algorithm. But to apply EM in an easy way, we need to change our objective function; rather than to optimize the conditional log likelihood $\sum_j \log p(I_j|x_j)$, we want to optimize the joint log likelihood $\sum_j \log p(I_j, x_j)$. This infact leads to some interesting discussion at the end of this section.

Again, next we will discuss how to estimate these parameters for MNL and MMNL.

3.1 Multinomial Logit Model

The graphical representation of this model can be expressed as:

$$k \rightarrow (x, \beta) \rightarrow I$$

The log likelihood can be expressed as

$$L = \sum_j \log(\sum_k p(k)p(x_j|k)p(I_j|j,k)).$$

in which $p(I_j|j,k) = \sum_i p(i|x_{ij},k)l(i|j)$ and l is the observed label. To apply EM algorithm, we consider posterior $h(k,j)$:

$$h(k,j) = p(k|I_j, x_j)$$
$$= \frac{p(I_j, x_j|k)p(k)}{\sum_{k'} p(I_j, x_j|k')p(k')}$$
$$= \frac{p(I_j|j,k)p(x_j|k)\pi_k}{\sum_{k'} p(I_j|j,k')p(x_j|k')\pi_{k'}}$$

So if we let

$$p_Y(j) = \sum_k \pi_k p(x_j|k)p(I_j|j,k)$$

$$p_y(k,j) = \pi_k p(x_j|k)p(I_j|j,k)/p_Y(j)$$

we will have $h(k,j) = p_y(k,j)$. The approximation problem is then

$$\max L' = \max \sum_j \sum_k p_y(k,j)(\log \pi_k + \log p(x_j|k) + \log p(I_j|j,k)) \tag{3}$$

The optimum of π_k, θ_k and Λ_k can be directly calculated:

$$\pi_k = \frac{1}{N}\sum_j p_y(k,j)$$

$$\theta_k = \frac{\sum_j p_y(k,j)x_j}{\sum_j p_y(k,j)}$$

$$\Lambda_k = \frac{\sum_j p_y(k,j)(x_j - \theta_k)(x_j - \theta_k)^T}{\sum_j p_y(k,j)}$$

and we need to apply gradient methods for β_k:

$$\frac{\partial}{\partial \beta_k}L' = \sum_j p_y(k,j)\frac{\partial}{\partial \beta_k}\log p(I_j|j,k)$$
$$= \sum_j p_y(k,j)(\sum_i (l(i|j) - p(i|x_{ij},k))x_{ij})$$

3.2 Mixture Logit Model

The graphical representation of this model can be expressed as:

$$k \rightarrow (x, \mu, \Sigma) \rightarrow \beta \rightarrow I$$

Now, the log likelihood can be expressed as

$$L = \sum_j (\log(p_Y(j))) = \sum_j \log(\sum_k \pi_k p(x_j|k)E_{\beta_k}p(I_j|j,\beta_k)).$$

4

Again, to apply EM algorithm, we consider two latent variables: k and β_k. The posterior of the joint distribution is:

$$h(k, j) = p(k, \beta_k | I_j, x_j)$$
$$= \frac{\pi_k p(x_j | k) p(I_j | j, \beta_k) p(\beta_k | k)}{\sum_{k'} \pi_{k'} p(x_j | k') E_{\beta_{k'}} (p(I_j | j, \beta_{k'}))}$$

and the approximation becomes

$$L' = \sum_{j,k} \int_{\beta_k} h(k, j)(\log \pi_k + \log p(x_j | k) + \log p(I_j | j, \beta_k) + \log p(\beta_k | k)) d\beta_k$$
$$= \sum_{j,k} \frac{E_{\beta_k} (\pi_k p(x_j | k) p(I_j | j, \beta_k)(\log p(k | x_j) + \log p(I_j | j, \beta_k) + \log p(\beta_k)))}{\sum_k \pi_k p(x_j | k) E_{\beta_k} (p(I_j | j, \beta_k))} \tag{4}$$

Then, we can then draw samples and use Monte-Carlo method to approximate the expectation $E_{\beta_k}(\cdot)$. Assume we take S samples for each j and β_k: $\beta_{k,j,1}, \cdots, \beta_{k,j,S}$. The new approximation becomes

$$L'' = \sum_{j,k,s} \hat{p_y}(k, j, s)(\log \pi_k + \log p(x_j | k) + \log p(I_j | j, \beta_k) + \log p(\beta_k)) \tag{5}$$

with

$$\hat{p_y}(k, j, s) = \frac{\pi_k p(x_j | k) p(I_j | j, \beta_{k,j,s})}{\sum_{k',s'} \pi_{k'} p(x_j | k') p(I_j | j, \beta_{k',j,s'})}$$

Now, for this approximated problem, we can calculate the optimum of all parameters directly:

$$\pi_k = \frac{1}{N} \sum_{j,s} \hat{p_y}(k, j, s)$$

$$\theta_k = \frac{\sum_{j,s} \hat{p_y}(k, j, s) x_j}{\sum_{j,s} \hat{p_y}(k, j, s)}$$

$$\Lambda_k = \frac{\sum_{j,s} \hat{p_y}(k, j, s)(x_j - \theta_k)(x_j - \theta_k)^T}{\sum_{j,s} \hat{p_y}(k, j, s)}$$

$$\mu_k = \frac{\sum_{j,s} \hat{p_y}(k, j, s) \beta_{j,k,s}}{\sum_{j,s} \hat{p_y}(k, j, s)}$$

$$\Sigma_k = \frac{\sum_{j,s} \hat{p_y}(k, j, s)(\beta_{j,k,s} - \mu_k)(\beta_{j,k,s} - \mu_k)^T}{\sum_{j,s} \hat{p_y}(k, j, s)}$$

4 Numerical Experiment

5 Reference