1 Theory

There are N agents in the sample. Agent i chooses K_i most preferred options from a personal choice set C_i and ranks them in the order of preference. L_i denotes i's preference list, whereas L_{ik} stands for k^{th} best item.

Agent's preferences are given by the utility function

$$U_{ijt} = X'_{ij}\beta_t + \varepsilon_{ijt}$$

Agent's utility depends on the vector of choice characteristics X_{ij}, idiosyncratic shocks ε_{ijt} and agent type t ("latent class"). The vector X_{ij} is observed in the data, while ε_{ijt} and t are not.

The idiosyncratic shocks are drawn from the standard Gumbel distribution.[1] These shocks are independent across agents and choices; they also don't depend on X_{ij} or t. The distribution of latent classes is paramaterized by α_t:

$$\omega_t = \Pr\{t_i = t\} = \frac{\exp \alpha_t}{\sum_{s=1}^{T} \exp \alpha_s}, \quad t = 1, \dots, T.$$

Without the loss of generality, α_1 is normalized to zero.

To allow for stratified sampling, let w_i denote a weight inversely proportional to the sampling rate used for i's subpopulation.[2] For instance, suppose that the sampling rates for males and females are 50% and 100% respectively. That is, the sample contains every female and every second male in the population. Then, $w_i = 2$ if i is male and $w_i = 1$ otherwise: every male in the sample represents two males in the population.

For each agent in the sample, the dataset includes

- The choice set, C_i,
- Covariates for all feasible choices, $X_i = [X_{ij}]_{j \in C_i}$,
- The ranked list of top K_i choices, L_i.

The unknown parameters are

- Preference coefficients, by latent class, $\beta = [\beta_1, \dots, \beta_T]$,
- The distribution of latent classes, $\alpha = [\alpha_1, \dots, \alpha_{T-1}]$.

[1] The distribution function for ε_{ijt} is $\exp(-\exp(-\varepsilon))$.

[2] The definition of w_i is similar to that of pweight in Stata.

1.1 Likelihood Function

Let $\delta_{ijt} = X'_{ij}\beta_t$. The log likelihood function is given by

$$\mathcal{L} = \sum_{i=1}^{N} w_i \ln \Pr\{L = L_i | X_i\} = \sum_{i=1}^{N} w_i \ln \left[\sum_{t=1}^{T} \omega_t \Pr\{L = L_i | X_i, t\} \right] \tag{1}$$

As one conditions on the covariates and the latent class, the likelihood function takes the standard "exploded logit" form:

$$\Pr\{L = L_i | X_i, t\} = \prod_{k=1}^{K_i} \frac{\exp \delta_{iL_{ik}t}}{\Delta_{it} + \sum_{m=k}^{K_i} \exp \delta_{iL_{im}t}} \tag{2}$$

where $\Delta_{it} = \sum_{j \in C_i \setminus L_i} \exp(\delta_{ijt})$.

1.2 Gradient Vector

First, note that the parametrization for ω implies

$$\frac{\partial \omega_s}{\partial \alpha_t} = \begin{cases} -\omega_s \omega_t, & \text{if } s \neq t \\ \omega_t(1 - \omega_t), & \text{otherwise} \end{cases}$$

This expression is used to obtain the derivatives of the loglikelihood function with respect to α:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \alpha_t} &= \sum_{i=1}^{N} \frac{w_i}{\Pr\{L = L_i | X_i\}} \left[\sum_{s=1}^{T} \frac{\partial \omega_s}{\partial \alpha_t} \Pr\{L = L_i | X_i, s\} \right] \\ &= \sum_{i=1}^{N} w_i \omega_t \left[\frac{\Pr\{L = L_i | x_i, t\}}{\Pr\{L = L_i | x_i\}} - 1 \right] \end{aligned} \tag{3}$$

For type-specific preference parameters β_t, the derivative is

$$\frac{\partial \mathcal{L}}{\partial \beta_t} = \sum_{i=1}^{N} \frac{w_i}{\Pr\{L = L_i | X_i\}} \left[\omega_t \frac{\partial}{\partial \beta_t} \Pr\{L = L_i | X_i, t\} \right] \tag{4}$$

In order to find the derivative of the conditional loglikelihood, it is convenient to switch to logarithms:

$$
\begin{aligned}
\frac{\partial}{\partial \beta_t} \ln \Pr\{L = L_i | X_i, t\} &= \frac{\partial}{\partial \beta_t} \ln \left[\prod_{k=1}^{K_i} \frac{\exp \delta_{iL_{ik}t}}{\Delta_{it} + \sum_{m=k}^{K_i} \exp \delta_{iL_{im}t}} \right] \\
&= \sum_{k=1}^{K_i} \frac{\partial}{\partial \beta_t} \left[\delta_{iL_{ik}t} - \ln \left(\Delta_{it} + \sum_{m=k}^{K_i} \exp \delta_{iL_{im}t} \right) \right] \\
&= \sum_{k=1}^{K_i} \left[\frac{\partial \delta_{iL_{ik}t}}{\partial \beta_t} - \frac{\frac{\partial \Delta_{it}}{\partial \beta_t} + \sum_{m=k}^{K_i} \frac{\partial \delta_{iL_{im}t}}{\partial \beta_t} \exp \delta_{iL_{im}t}}{\Delta_{it} + \sum_{m=k}^{K_i} \exp \delta_{iL_{im}t}} \right] \\
&= \sum_{k=1}^{K_i} \left[X_{iL_{ik}} - \frac{\sum_{j \in C_i \setminus L_i} X_{ij} \exp \delta_{ijt} + \sum_{m=k}^{K_i} X_{iL_{im}} \exp \delta_{iL_{im}t}}{\Delta_{it} + \sum_{m=k}^{K_i} \exp \delta_{iL_{im}t}} \right] \qquad (5)
\end{aligned}
$$

1.3 Notes on Computation

Roughly speaking, the algorithm works as follows:

1. Compute $\exp \delta_{ijt}$ for all i, t and $j \in C_i$. This step tends to account for a significant portion of the total computation time.

2. Loop over agents and types. For each agent-type pair (i, t):

   (a) Compute the sums over inferior choices, Δ_{it} and $\sum_{j \in C_i \setminus L_i} X_{ij} \exp \delta_{ijt}$.

   (b) Loop over choices in L_i going backwards from L_{iK_i} to L_{i1}. Compute the denominator in (2) by accumulating $\exp \delta_{iL_{im}t}$ on each step. Compute the numerator in (5) by accumulating $X_{iL_{im}} \exp \delta_{iL_{im}t}$. Use the results to find the contribution of each element k of L_i into (2) and (5), accumulate the product in (2) and the sum in (5).

3. Use $\Pr\{L = L_i | X_i, t\}$ its gradient found above to calculate the unconditional probabilities $\Pr\{L = L_i | X_i\}$.

4. Put everything together. Find the likelihood function in (1) and its gradient in (3) and (4).

Note that the loops in 2a and 2b run most efficiently if data on choices of each agent are stored contiguously. Unlisted part of the choice set ($C_i \setminus L_i$) should come first followed by the ranked choices, L_i, in the reverse order. This is the reason why the dataset has to be arranged in the memory in a certain way before the loglikelihood function and its derivatives can be calculated.

2 Usage

2.1 Exploded logit

```
[logl, grad] = explogit(beta, X, nskipped, nlisted, weight)
```

Name	Description	Dimensions	Type
	Arguments		
beta	Vector of logit coefficients,	M	double
X	Matrix of covariates. Rows correspond to covariates, columns correspond to agent-choice pairs.	$M \times C$	double
nskipped	Number of skipped choices for each agent,	N	uint16
nlisted	Number of listed choices for each agent,	N	uint16
weight	Optional argument. Agent i in the sample represents weight(i) agents in the population. Default: the vector of ones,	N	double
	Return values		
logl	Loglikelihood function,	1	double
grad	Gradient of logl.	M	double

M is the number of covariates in X_{ij}. C is the total number of agent-choice combinations: $C = \sum_{i=1}^{N} C_i$.

2.2 Latent class exploded logit

```
[logl, grad] = lcexplogit(beta, nclasses, X, nskipped, nlisted, weight)
```

Name	Description	Dimensions	Type
	Arguments		
beta	The first $T-1$ elements are $\alpha_2, \ldots, \alpha_T$. The remaining MT elements are $[\beta_1, \ldots, \beta_T]$,	$(M+1)T-1$	double
nclasses	T, the number of latent classes,	1	int
X	Matrix of covariates. Rows correspond to covariates, columns correspond to agent-choice pairs,	$M \times C$	double
nskipped	Number of skipped choices for each agent,	N	uint16
nlisted	Number of listed choices for each agent,	N	uint16
weight	Optional argument. Agent i in the sample represents weight(i) agents in the population. Default: the vector of ones,	N	double
	Return values		
logl	Loglikelihood function,	1	double
grad	Gradient of logl.	$(M+1)T-1$	double