Notes and homework on latent class logit

Consider an individual who associates a random utility $u_j = v_j + \epsilon_j$ with alternative j. Suppose the individual selects an alternative from each of n choice sets, $C_1, \ldots, C_n$. Let $j_k \in C_k$ denote the item the individual chooses from set C_k, $k = 1, \ldots, n$. Assuming a logit model, the probability that the individual chooses alternative $j_k \in C_k$ is

$$p_{jk} = \frac{e^{v_{j_k}}}{\sum_{l \in C_k} e^{v_l}}.$$

The likelihood (joint probability) of observing the individual's choices is

$$\ell = \prod_{j_k \in C_k} p_{j_k}.$$

Next, consider not one individual but a set I of individuals. Let ℓ_i denote the likelihood for individual $i \in I$. Each ℓ_i has the same form as ℓ above, except that the choice sets and the choices from the sets can differ across people. Then the joint likelihood across people is

$$L = \prod_{i \in I} \ell_i.$$

Suppose $v_j = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$. Then L is a function of $\beta_0, \ldots, \beta_m$. The maximum likelihood estimates $\hat{\beta}_0, \ldots, \hat{\beta}_m$ are obtained by maximizing L over the parameters $\beta_0, \ldots, \beta_m$. This is a convex optimization problem, which can be solved in polynomial time.

Next, suppose s is an unobserved segment (latent class). Let S denote the set of latent classes. Let

$$v_{js} = \beta_{0s} + \beta_{1s} x_1 + \cdots + \beta_{ms} x_m$$

denote the deterministic utility component for segment $s \in S$, where $\beta_{0s}, \ldots, \beta_{ms}$ are parameters that can differ across the segments. That is, instead of assuming that all individuals have the same preferences, we allow segments of individuals who have the same preferences. Then we can obtain an expression ℓ_{is} for the likelihood of observing individual i's choices, assuming that he or she belongs to segment s. It is the same expression as we obtained above for ℓ_i, expect that we use the segment-level parameters, $\beta_{0s}, \ldots, \beta_{ms},$, in its expression. Suppose π_s is the probability that a randomly chosen individual belongs to segment s. Then the unconditional likelihood for individual i is

$$\ell_i = \sum_{s \in S} \pi_s \ell_{is}.$$

As before, the joint likelihood across individuals is

$$L = \prod_{i \in I} \ell_i.$$

The difference is that ℓ_i is now a weighted sum of individual i's likelihood values across segments, where the weights are the segment membership probabilities π_s. We can estimate the π_s and the segment-level parameters $\beta_{0s}, \ldots, \beta_{ms}$, for all $s \in S$, by maximizing the likelihood function L. This is a nonlinear optimization problem, but not a convex optimization problem. We can solve it by using a Newton method, but the solution can be a local optimum. Since a globally optimum solution cannot be assured, it is useful to choose the solution with the highest likelihood value that is obtained by using different starting points for the optimization procedure.

EXAMPLE 1. Consider two individuals, $i = 1$ and $i = 2$. Suppose individual 1 chooses alternative 1 from each of the sets $C_1 = \{1,2\}$ and $C_2 = \{1,3\}$; and individual 2 select alternative 3 from $C_2 = \{1,3\}$ and alternative 2 from $C_3 = \{2,3\}$.

Let $u_{js} = v_{js} + \epsilon_{js}$ denote the utility of alternative j for segment s. Consider a latent class logit model with $|S| = 2$ segments. Then the likelihood function is $L = \ell_1 \ell_2$, where

$$\ell_1 = \pi_1 \left(\frac{e^{v_{11}}}{e^{v_{11}} + e^{v_{21}}} \cdot \frac{e^{v_{11}}}{e^{v_{11}} + e^{v_{31}}} \right) + (1 - \pi_1) \left(\frac{e^{v_{12}}}{e^{v_{12}} + e^{v_{22}}} \cdot \frac{e^{v_{12}}}{e^{v_{12}} + e^{v_{32}}} \right)$$

$$\ell_2 = \pi_1 \left(\frac{e^{v_{31}}}{e^{v_{21}} + e^{v_{31}}} \cdot \frac{e^{v_{21}}}{e^{v_{21}} + e^{v_{31}}} \right) + (1 - \pi_1) \left(\frac{e^{v_{32}}}{e^{v_{22}} + e^{v_{32}}} \cdot \frac{e^{v_{22}}}{e^{v_{22}} + e^{v_{32}}} \right).$$

Suppose each alternative is described using two attributes, x_1 and x_2. Let x_{jk} denote the value of alternative j on attribute k. We can then substitute

$$v_{1s} = \beta_{0s} + \beta_{1s} x_{11} + \beta_{2s} x_{22}$$

$$v_{2s} = \beta_{0s} + \beta_{1s} x_{21} + \beta_{2s} x_{22}$$

$$v_3 = \beta_{0s} + \beta_{1s} x_{31} + \beta_{2s} x_{32}$$

in the likelihood function (note that each x_{jk} is a numerical value). We obtain the parameter estimates by maximizing the likelihood function L over the variables π_1, β_{11}, β_{12}, β_{21} and β_{22}.

The preceding procedure gives us estimates of $\hat{\pi}_s$. We can obtain individual-level estimates of segment membership using Bayes rule. Let $p(s|i)$ denote the probability of membership in segment s given (the choice data) for individual i. According to Bayes rule,

$$p(s|i) = \frac{p(i|s)p(s)}{p(i)} = \frac{\ell_{is}\pi_s}{\sum_{s \in S} \ell_{is}\pi_s}.$$

In this expression, $p(s) = \pi_s$ is the (prior) probability that an individual belongs to segment s and $p(i|s) = \ell_{is}$ is the likelihood (joint probability) of observing the choices made by individual i assuming that he or she is a member of segment s. The estimates $\hat{p}(s|i)$ of the posterior membership probabilities can be obtained from estimates of ℓ_{is} and π_s.

We can use the posterior segment membership probability for individual i to estimate the probability that individual i chooses alternative $j \in C$:

$$p(j|i) = p(j|s)p(s|i)$$

where $p(j|s)$ is the probability that an individual in segment s chooses alternative $j \in C$ and $p(s|i)$ is the probability that individual i belongs to segments s.

E-M algorithm. Suppose we have some estimates $\hat{p}(s|i)$ of the membership probabilities in segment s given the data for individual i. Compute their mean across individuals for each segment:

$$\hat{p}(s) = \frac{1}{|I|} \sum_i \hat{p}(s|i)$$

where $|I|$ is the number of individuals in the sample. Then $\hat{\pi}_s = \hat{p}(s)$ is an estimate of the probabilities of segment membership, π_s. Fix the values of $\hat{\pi}_s$. Then maximize L to obtain new estimates of $\beta_0, \ldots, \beta_m$. Repeat, until the solution converges to within a pre-specified tolerance level. The only thing left is to specify the initial values of $\hat{p}(s|i)$. We can begin with any arbitrary values (for example, $1/|S|$, or 0-1 values obtained by assigning each individual arbitrarily to a single segment). The procedure can be summarized as the following algorithm.

1. Specify initial values for $\hat{p}(s|i)$ for all $i \in I$ and $s \in S$.
2. Compute
$$\hat{p}(s) = \frac{1}{|I|} \sum_i \hat{p}(s|i), \forall s \in S.$$
3. Set $\pi_s = \hat{p}(s), \forall s \in S$.
4. Maximize $\ln(L)$ over the parameters $\beta_{0s}, \ldots, \beta_{ms}$, $\forall s \in S$ (maximizing $\ln(L)$ is equivalent to maximizing L).
5. Given the estimates of $\beta_{0s}, \ldots, \beta_{ms}$ for all $s \in S$, recompute ℓ_{is} for each individual $i \in I$ and segment $s \in S$.
6. Compute
$$\hat{p}(s|i) = \frac{\ell_{is}\pi_s}{\sum_{s \in S} \ell_{is}\pi_s}.$$
7. Stop if the increase in the value of $\ln(L)$ is within a tolerance ϵ. Otherwise, return to step 2.

Homework. The homework is due in class on Friday, Oct. 27, 2017. Please typeset your answer to each question.

Objective: The purpose of the homework is to use data on choices of tablet computers to estimate and validate a latent class logit model.

Data: The data is in the attached file data_ipad.csv. It contains data on choices made by 137 subjects. Each subject evaluated 15 choice sets. Thus, the file contains data on $137 \times 15 = 2055$ choice sets. Each choice set had three alternatives. A subject's task was to choose one alternative from a choice set. Each alternative was described using the following attributes: brand name, screen size, size of hard drive, RAM, battery life and price. The data in the attached file uses dummy variable coding to describe the product profiles. The reference levels used for the dummy coding are: Nexus, 7" screen, 16GB HD, 1GB RAM, 7-hour battery, \$169 price. Since the data is already coded using dummy variables, you should use it directly for estimation, without introducing additional reference level.

Column 1 of the data file (consumer_id) identifies each of the 137 subjects. Column 2 of the data file (choice_set_id) identifies each of 2055 choice sets. Column 3 (alternative_id_in_set) identifies the three alternatives in a choice set. Column 4 identifies the id of the alternative chosen from the choice set. The remaining columns contain the dummy variable coding for each alternative.

Modeling: For each individual, randomly select 14 choice sets for estimation and keep one choice set for holdout validation. Use the estimation data to fit (1) an aggregate model with no segments (you have already done this in the first homework), and (2) a latent-class models with 2, 3, 4 and 5 segments. Use the holdout data for model validation. You may, but do not have to, use the E-M algorithm; instead, you can directly maximize the likelihood function over all parameters using a nonlinear optimization routine. Make sure to use multiple (say, 5) random starting values for the parameters. Select the solution with the largest likelihood value.

Questions: Answer the following six questions. Questions 1-4 are based on your data analysis. Questions 5 and 6 are conceptual and do not require data analysis.

1. Select the number of segments by using two criteria:
   (i) BIC (Bayesian Information Criterion), which is more conservative than the AIC criterion we discussed in class. The BIC value is equal to $-2\ln(\hat{L}) - K\ln(N)$ where $\hat{L}$ is the maximum likelihood value, K is the number of parameters and N is the total number of observations (the number of choice sets across individuals used for estimation). A solution is preferred if it has a lower BIC value.
   (ii) the interpretability of the results (how different are the preferences across segments).
2. Interpret the results. How big are different segments; which attribute levels are significantly different from the "base" levels; and how do preferences differ across segments.
3. Calculate the posterior probabilities of segment membership. How well separated are the membership probabilities?
4. Explain how you can use the holdout data to validate the model. Perform the validation. Interpret your results.
5. Suppose you obtained choice data for a new person, for whom you do not know segment membership probabilties. How would you predict the choice probabilities for this person?
6. Suppose you had information on customer demographics. How could you use it to model segment membership?