Home Work 2

Due: October 26, before 9:00 PM

1 Nested Logit Model

Read chapter 4 of Kenneth Train's book to answer this question. Suppose you are interested in using a nested logit model to capture the choice process for a set of four tuna brands. The model can be setup to contain two nests. The first nest $k = 1$ is the oil nest and contains choice alternatives with oil in it. The alternatives belonging to the oil nest include 1) Starkist-Oil and 2) Chicken-of-Sea-Oil. The second nest $k = 2$ is the water nest and contains the choice alternatives 3) Starkist-Water and 4) Chicken-of-Sea-Water.

Assume that a single alternative specific variable $Price_{nj}$ is available to describe the price faced by consumer n for alternative j. In addition, the data contains a nest-specific variable $slip_k$ that contains the slipperiness of the alternatives in the nest k. Note that all alternatives within the nest will have the same slipperiness.

Use the fact that the probablity of choice of a given alternative for a nested logit can be decomposed into the product of two probabilities — the probability of choosing the nest for the alternative and the probability of choosing the alternative, given the nest (see section 4.2.3 of Kenneth Train's book).

1. Draw the nesting tree structure for this model.

2. Write the expression for the inclusive value for each of the two nests.

3. Write the expression for the probability of choosing Nest 1 (see equation 4.4). In doing this, make sure you write the entire specification of the systematic utility for this probability, i.e., write W_{n1} in terms of the appropriate variables.

4. Write the expression for the probability of choosing Alternative 3, given Nest 2, i.e., specify the form of Equation 4.5 in Train's book.

2 Finite Mixture of Poisson's Regression

In this exercise you are required to estimate a finite mixture Poisson regression model. Our interest here is in modeling count data using a number of explanatory variables. In this application the counts pertain to the number of packages of tuna that are bought on each purchase occasion by consumers. This is the dependent variable. We have a single explanatory variable, the price per package.

The purpose of this exercise is to enhance your ability to formulate likelihoods and estimate them with real data. In order to do this exercise you need to understand what is a Poisson distribution and why we need to do a Poisson regression rather than a ordinary regression.

Poisson Distribution The Poisson distribution is useful for modeling count data (the number of events that occur on an observation). The dependent variable is discrete and is incremented by integers. The probability mass function for a Poisson distribution is given by

$$P(Y = y) = \frac{\lambda^y \exp(-\lambda)}{y!}, \tag{1}$$

where $\lambda > 0$ is called the Poisson rate and is interpreted as the rate at which events happen. The mean and variance of the random variable Y are both given by λ and therefore the Poisson distribution is not very realistic for most data sets. However, it is good enough for a start and we will use it.

Poisson Regression When interest is in finding out what determines how many events happen, then the rate λ can be written in terms of independent variables. As $\lambda > 0$, $\log(\lambda)$ is modeled as a linear combination of the independent variables. We therefore can write $\lambda = \exp(\beta_1 + \beta_2 \text{ Price})$.

In this exercise, do the following

1. Formulate the log-likelihood for the a single segment Poisson regression
2. Program it (Use the program for the multinomial logit or regression as a template)
3. Give me the parameter estimates and the standard errors associated with each parameter.
4. Interpret the results.

5. Repeat the above steps for a two-segment finite mixture. Use the finite mixture of regressions program or the finite mixture of logits programs that are attached to this email, as a template.

6. Compare the one-segment solution with the two-segment solution and see which one is better. Use the BIC statistic for choosing among the models. (BIC=Maximum Log-Likelihood $-k * \log(N)/2$, where k is the number of parameters in the model and N is the total number of observations in the data).

You need the attached files for doing the exercise. The file **mixturePoisson.csv** contains the data. It has 2777 observations from 200 individuals and three columns. The first column is the id for the individual. The second column is the dependent variable and the third column contains the price.

The R file, **finiteMixtureRegression.R** gives you a program that estimates a two segment finite mixture model for a linear regression with an intercept and a slope parameter. Use this file as a template and modify it to obtain the results for the two-segment Poisson regression model. Notice that the R function *dpois* can be handy here. The linear regression is based on the **mixturereg.csv** data file.