Independence of Irrelevant Alternatives Assumptions (IIA) and some Extentions

Zihan Zhang*

April 14, 2019

Abstract

In this note, I would like to concentrate on the topic about Independence of Irrelevant Alternatives(IIA) assumptions, which plays a crucial role in model selections when doing classifications. Thus, based on what we have learned from class, I would like to illustrate what IIA is, the method of testing it and some reliable models that could be applied if IIA failed.

1 What is IIA?

Generally speaking, Independence of Irrevelant Alternatives assumption suggests that the preference of an individual between item A and item B does not change when another item X alters if IIA assumption holds, which means the outcome of a choice between two items only depend on the features of these two items and it is not affected by the features or characteristics of other items or variables.

To have a better understanding of the implications of IIA, it is better to look at the formula of the multinomial logit model. According a standard multinomial logit model with 3 categories (J = 3), we can express the probabilities as:

$$Pr(y_i = 1 | \mathbf{x_i}) = \frac{1}{1 + exp(\mathbf{x'_i}\beta_2) + exp(\mathbf{x'_i}\beta_3)}$$
$$Pr(y_i = 2 | \mathbf{x_i}) = \frac{exp(\mathbf{x'_i}\beta_2)}{1 + exp(\mathbf{x'_i}\beta_2) + exp(\mathbf{x'_i}\beta_3)}$$
$$Pr(y_i = 3 | \mathbf{x_i}) = \frac{exp(\mathbf{x'_i}\beta_3)}{1 + exp(\mathbf{x'_i}\beta_2) + exp(\mathbf{x'_i}\beta_3)}$$

where $\mathbf{x}_{\mathbf{i}}$ represents the vector of features of individual *i* and $\mathbf{x}'_{\mathbf{i}}\beta_{\mathbf{j}} = \sum_{k=0}^{K} \beta_{jk} x_{ik}$.

^{*2016} WISE IUEC, Student ID: 15220162202518

Therefore, we calculate the relevance between these categories (which is called odds) by:

$$log \frac{Pr(y_i = j | \mathbf{x}_i)}{Pr(y_i = k | \mathbf{x}_i)} = \mathbf{x}'_i(\beta_j - \beta_k)$$

which is in linear form and please notice that I have normalized $\beta_1 = 0$.

From the formula, it is evident that the odds is only affected by option j and k, which is the implication of IIA assumption (at least in our classification topic and logit model).

2 Method of Testing IIA

As the previous chapter mentioned, the multinomial requires the independence of irrevelent allternatives to hold, making the result to be reliable. Hence, in order to test the IIA assumption, Hausman and McFadden (1984) proposed a generalized method with underlying principle of removing one alternatives, estimating the model again and comparing the results.

Specifically, Hausman-McFadden test requires us to estimate a model containing all the categories with J alternatives and $\hat{\beta}_{\mathbf{J}}$ as coefficients. After that, we have to estimate the model with one alternative deleted, which means it conprise (J-1) categories and $\widehat{\beta}_{\mathbf{J}-\mathbf{1}}$ as coefficients. Besides, we also need to calculate the variance and covariance matrix, which could be donated as $\mathbf{V}_{\mathbf{J}}$ and $\mathbf{V}_{\mathbf{J}-\mathbf{1}}$ respectively.

Hence, we are able to calculate the Hausman-McFadden statistics by

$$H_1 = [\widehat{\beta_{\mathbf{J}}} - \widehat{\beta_{\mathbf{J}-1}}][\widehat{\mathbf{V}_{\mathbf{J}-1}} - \widehat{\mathbf{V}_{\mathbf{J}}}][\widehat{\beta_{\mathbf{J}}} - \widehat{\beta_{\mathbf{J}-1}}]$$

It follows the Chi-Square distribution with the degree of freedom of 1.

Actually, if the categories J is large, we can delete more than one alternatives and do this test following the procedures above. In that circumstance, the H-M statistics ought to be:

$$H_k = [\widehat{\beta_{\mathbf{J}}} - \widehat{\beta_{\mathbf{J}-\mathbf{k}}}][\widehat{\mathbf{V}_{\mathbf{J}-\mathbf{k}}} - \widehat{\mathbf{V}_{\mathbf{J}}}][\widehat{\beta_{\mathbf{J}}} - \widehat{\beta_{\mathbf{J}-\mathbf{k}}}]$$

where k manifest the number of alternatives we delete in the H-M test, while the degree of freedom tend to be k.

For the example of transportation discussed in class, we can test it in R and the result is as follows.

Data	transport.long	
Chi-square	degree of freedom	P-value
1.8773	3	0.5983
lternative hypothesis: IIA is rejected		

3 Nested logit model

Sometimes the IIA assumption fails, the multinomial logit model fails correspondingly and we have to try other models. In essence, the multinomial probit model that we have learned in class could be an alternative, however, it suffers from the slow speed of calculation. Commonly it takes relatively long time to get access to the result. Hence, in this section I would like to introduce another optional model, which is the nested logit model.

Nested logit model deals with the problems that IIA fails. It divided all the alternatives into some subsets (also called nests) and those alternatives within the same subsets or nests are allowed to be similiar and correlated. The correlation between alternatives are shown in the following figure. The left one shows multinomial logit model while the right one shows the nested logit model.



where A_i and B_i indicates the alternatives that individual *i* has.

For classification with nested logit model, first we estimate for the alternatives in Multinomial logit model within the lower nests, which is represented as B in the above figure. Notice that we have to omit all the variables (denote as \mathbf{Z}) that could be the same for alternatives within the nests.

Take the example of mode of transportation that we have discussed in class to illustrate, suppose households are choosing their transportations from bus subway and car. the relationships for the alternatives should be:



Hence, we can donete these probabilities as follows:

$$Pr(Bus|PT) = \frac{exp(\mathbf{x}'\beta_{\mathbf{bus}})}{exp(\mathbf{x}'\beta_{\mathbf{bus}}) + exp(\mathbf{x}'\beta_{\mathbf{subway}})}$$
$$Pr(Subway|PT) = 1 - Pr(Bus|PT)$$
$$P(Car) = \frac{exp(\mathbf{x}'\beta_{\mathbf{car}})}{exp(\mathbf{x}'\beta_{\mathbf{car}}) + exp(V_{PT})}$$
where $V_{PT} = \phi log(exp(\mathbf{x}'\beta_{\mathbf{bus}}) + exp(\mathbf{x}'\beta_{\mathbf{subway}})) + \sum_{k} \alpha_{k}Z_{k}$
$$P(PT) = 1 - P(Car)$$

Therefore, according to the conditional probability, we are able to compute the probability of various choices and do the classification.

$$P(Bus) = P(Bus|PT)P(PT)$$
$$P(Subway) = P(Subway|PT)P(PT)$$