Discrete Choice Models

Li Zichao
15220162202173

April 14, 2019

Abstract

In this report I will summarize the basic concept of discrete choice models and three forms of probability: standart logit, nested logit and ordered logit.

1 Basic Concept

Instead of being continuous variables, the individual choices in classification problem are usually discrete alternatives. In discrete models, an agent(i.e., person, firm, decision-maker) faces a series of choices over time among a set of options. For instance, constituency chooses which of several presidential cantidates to elect; customer chooses which of several products to consume; a firm decides which technology to use in production. In the discrete case, the calculus method like F.O.C is not applicable for empirical analysis. To examine the "which one" problem, techniques such as logistic regression are developed.

Economists attributes agent's choice to several factors. Those factors, observed or unobserved by researchers collectively determine the agent' choice. Here we denote observed factors as x and unobserved factors as ϵ with density of $f(\epsilon)$. The behavioral process is a function as the form of $y = h(x, \epsilon)$, where y is the outcome of the choice process. We assume that with given x and ϵ, agent's choice is fully determined.

However ϵ is unobserved, we can only derive the probability for each particular outcome. The probability of y given x is the expected value of an indicator function that equal to 1 when the truly outcome is consistent with our prediction.

$$
\begin{aligned}
P(y|x) &= \text{Prob}(I[h(x,\epsilon) = y] = 1) \\
&= \int I[h(x,\epsilon) = y] f(\epsilon) \, \mathrm{d}\epsilon
\end{aligned} \tag{1}
$$

To calculate the probability, lets specify the behavioral process function as a utility function:

$$
U = \beta\prime x + \epsilon \tag{2}
$$

To simplify the process, we start from the binary case. An individual will take the action only if the utility of that choice is positive, namely $P = \int I[\beta' x + \epsilon > 0] f(\epsilon) \mathrm{d}\epsilon$. Generally, as probability is necessarily between zero and one, we assume that ϵ is distributed logistic[1] with a density function as $f(\epsilon) = e^{-\epsilon}/(1 + e^{-\epsilon})^2$.

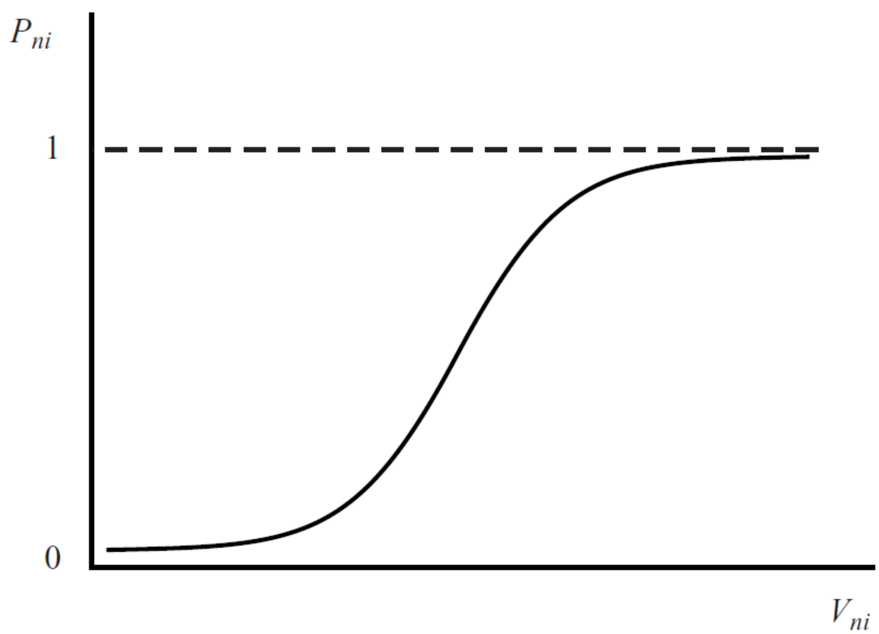

Figure 1: Graph of Logit Curve

[1]The logistic distribution is in shape of sigmoid.

Then the probability is:

$$
\begin{aligned}
p &= \int I[\beta\prime + x > 0] f(\epsilon) d\epsilon \\
&= \int I[\epsilon > -\epsilon' x] f(\epsilon) d\epsilon \\
&= \int_{\epsilon=-\beta' x}^{\infty} f(\epsilon) d\epsilon \\
&= 1 - F(-\beta' x) \ = \ 1 - \frac{1}{1+e^{\beta' x}} \\
&= \frac{e^{\beta' x}}{1+e^{\beta' x}}
\end{aligned} \tag{3}
$$

Above is the binary logit model. In the next section I will introduce other three prominent expressions for the probabilities including multinomial logit, nested logit and ordered logit.

2 Other Forms of Probability

2.1 Multinomial Logit

The multinomial Logit, as known as Standard logit model, is an extension of binary logit. Differently, the individual is no longer hesitating whether to take action or not, he will choose from a series of independent irrelecant alternatives[2].

$$
P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}} \tag{4}
$$

For any two alternatives i and k, the ratio of the logit probabilities is

$$
\begin{aligned}
\frac{P_{ni}}{P_{nk}} &= \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} \\
&= \frac{e^{V_{ni}}}{e^{V_{nk}}} \ = \ e^{V_{ni} - V_{nk}}
\end{aligned} \tag{5}
$$

The multinomial logit model have power as follows:

[2]**IIA**: independence of irrelevant alterantives exhbited in multinomial logistic model, with which the comparative probability between two choices will not affected by the properties of other classes

1. Logit can represent systematic taste variation that relates to observed characteristics.

2. The logit model implies proportional substitution across alternative.

3. If unobserved factors are independent over time in repeated choice situations, then logit can capture the dynamics of repeated choice, including state dependence.

2.2 Nested Logit

Standard logit requires that alternatives are independent with each other. If not, the nested logit model is appropriate for the classification problem. **IIA** holds within each nest but not across nests. For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests. The nested logit model is obtained by assuming that the vector of unobserved utility, $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nj})$ has cumulative distribution

$$exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_k} e^{-\epsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right) \tag{6}$$

The distribution is a type of **GEV**[3] distribution. The following choice probability for alternative $i \in B_k$:

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l}} \tag{7}$$

The parameter λ_k is a measure of the degree of independence in unobserved utility among the alternatives in nest k. A higher value of λ_k means greater independence and less correlation. The ratio of probabilities is the ratio of numerators:

$$\frac{P_{ni}}{P_{nm}} \frac{e^{V_{ni}/\lambda_k}\left(\sum_{j\in B_k} e^{V_{nj}/\lambda_k}\right)^{\lambda_k - 1}}{e^{V_{nm}/\lambda_l}\left(\sum_{j\in B_l} e^{V_{nj}/\lambda_l}\right)^{\lambda_l - 1}} \tag{8}$$

[3] Generalized extreme value(GEV) models constitute a large class of models that exhibit a variety of substitution pattern. The unifying attribute of these models is that the unobserved portions of utility for all alternatives are jointly distributed as a generalized extreme value.

2.3 Ordered Logit

Sometimes individual faces alternatives that are ordered, which is inconsistent with the assumption of independent error for each alternatives. For example, respondents may asked to provide rating of various kinds like:

1. Good
2. Fair
3. Bad

With ordered alternatives,one alternative is similar to those close to it and less similar to those further away. Assume the rate is decided by the satisfactory level(U) of agents, the decision is represented as:

- "Good" if $U > k_1$
- "Fair" if $k_1 > U > K_2$
- "Bad" if $K_2 > U$

The probability of "Bad" is

$$\begin{aligned} \text{Prob}(\text{``Bad``}) &= \text{Prob}(U < k_2) \\ &= \text{Prob}(\beta' x + \epsilon < k_2) \\ &= \text{Prob}(\epsilon < k_4 - \beta' x) \\ &= \frac{e^{k_2 - \beta' x}}{1 + e^{k_2 - \beta' x}} \end{aligned} \tag{9}$$

The probability of "Fair" is

$$\begin{aligned} \text{Prob}(\text{``}Fair\text{``}) &= \text{Prob}(k_2 < U < k_1) \\ &= \text{Prob}(k_2 - \beta' x + \epsilon < k_1 - \beta' x) \\ &= \frac{e^{k_1 - \beta' x}}{1 + e^{k_1 - \beta' x}} - \frac{e^{k_2 - \beta' x}}{1 + e^{k_2 - \beta' x}} \end{aligned} \tag{10}$$

References

[1] Discrete choice. https://en.wikipedia.org/wiki/Discrete_choice.

[2] Jiaming Mao. Classification. https://github.com/jiamingmao/data-analysis/blob/master/Lectures/Classification_and_Discrete_Choice_Models.pdf, 2019.

[3] Kenneth Train. Discrete choice methods with simulation, 2002.