USC, Fall 2016, Economics 513

Lecture 15: Discrete Choice II: Discrete choice models

We consider agents who choose between a finite number of distinct alternatives.

Examples:

- Occupational choice
- Choice of brand and model of a car
- Choice of neighborhood

Two types of heterogeneity: Alternatives differ in observed and unobserved *attributes* and agents have different observed and unobserved *characteristics*.

Example: Occupational choice

- Attributes: Average and standard deviation of income, unemployment risk, occupational hazards, 'glamour' of occupation (unobserved).
- Characteristics: level of education, gender, 'ability' (unobserved).
- Characteristics may affect the choice set: Some professional occupations not accessible for individuals who lack the required education.

Economic model: Additive random utility model

Consider consumer who chooses between alternatives $i = 1, \dots, I$. We omit subscript for consumer.

u_i is the utility associated with choice i. For u_i we specify a simple model, the Additive Random Utility (ARUM) model $u_i = -v_i + \varepsilon_i$

$-v_i$ is the mean utility of alternative i . Later we make it dependent on observed (and unobserved) characteristics of the alternative and of the agent. The minus sign simplifies the notation in the sequel. An increase in v_i makes the alternative less attractive.

ε_i is a random variable with $\mathrm{E}(\varepsilon_i) = 0$.
Interpretations of ε_i.

- Optimization error.
- Unobserved attributes of alternative i, e.g. 'glamour' of occupation.
- Unobserved characteristics of the agent.

- Interactions of these two.

Repeated choice by the same agent between the same alternatives would help in settling the interpretation.

The key assumption is that we do not know the random components ε_i, but the decision maker does know them and takes them into account when choosing.

Interpretation affects the nature of the joint distribution of $\varepsilon_1, \ldots, \varepsilon_I$. If there are unobserved agent attributes these random variables will be correlated. We assume

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_I \end{pmatrix} \sim F(\varepsilon_1, \ldots, \varepsilon_I)$$

with F left unspecified for now. We assume that F is independent of $v_1, \ldots, v_I$. The joint pdf is $f(\varepsilon_1, \ldots, \varepsilon_I)$.

ARUM and choice

The agent chooses alternative i if and only if

$$u_i > u_j \text{ for } i \neq j = 1, \ldots, I \Leftrightarrow -v_i + \varepsilon_i > -v_j + \varepsilon_j \text{ for } i \neq j = 1, \ldots, I$$

Because the ε_i are not known to us, we cannot predict with certainty which alternative is chosen, but we can compute the probability that i is chosen, $p_i(v)$, from the joint distribution of $\varepsilon_1, \ldots, \varepsilon_I$.

$$p_i(v) = \Pr(\varepsilon_1 < v_1 - v_i + \varepsilon_i, \ldots, \varepsilon_I < v_I - v_i + \varepsilon_i) =$$

$$= \int_{-\infty}^{\infty} \int_{\infty}^{v_1 - v_i + \varepsilon_i} \int_{-\infty}^{v_I - v_i + \varepsilon_i} f(\varepsilon_1, \ldots, \varepsilon_I) \mathrm{d}\varepsilon_1 \ldots \mathrm{d}\varepsilon_I \mathrm{d}\varepsilon_i =$$

$$\int_{-\infty}^{\infty} \frac{\partial F}{\partial \varepsilon_i} (v_1 - v_i + \varepsilon_i, \ldots, \varepsilon_i, \ldots, v_I - v_i + \varepsilon_i) \mathrm{d}\varepsilon_i$$

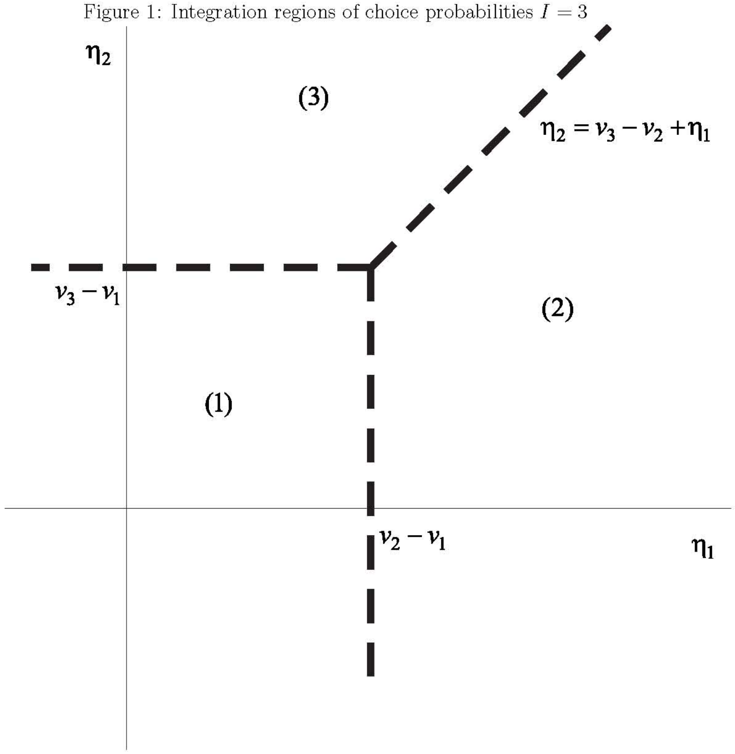
For $I = 3$

$$\begin{aligned} p_1(v) &= Pr(\varepsilon_2 < v_2 - v_1 + \varepsilon_1, \varepsilon_3 < v_3 - v_1 + \varepsilon_1) \\ p_2(v) &= Pr(\varepsilon_1 < v_1 - v_2 + \varepsilon_2, \varepsilon_3 < v_3 - v_2 + \varepsilon_2) \\ p_3(v) &= Pr(\varepsilon_1 < v_1 - v_3 + \varepsilon_3, \varepsilon_2 < v_2 - v_3 + \varepsilon_3) \end{aligned}$$

or if we define $\eta_1 = \varepsilon_2 - \varepsilon_1$ and $\eta_2 = \varepsilon_3 - \varepsilon_1$ (see figure 1)

$$\begin{aligned} p_1(v) &= \Pr(\eta_1 < v_2 - v_1, \eta_2 < v_3 - v_1) \\ p_2(v) &= Pr(\eta_1 > v_2 - v_1, \eta_2 < v_3 - v_2 + \eta_1) \\ p_3(v) &= Pr(\eta_1 < v_2 - v_3 + \eta_2, \eta_2 > v_2 - v_1) \end{aligned}$$

Figure 1: Integration regions of choice probabilities $I = 3$



Note that the choice probabilities can be expressed as the probability of an event involving the random variables $\varepsilon_2 - \varepsilon_1$ and $\varepsilon_3 - \varepsilon_1$ and the average utility differences $v_2 - v_1$ and $v_3 - v_1$, i.e. they involve only utility comparisons with alternative 1 that is chosen as the reference alternative (we can take any of the three alternatives as the reference alternative).

Multinomial logit

The choice probabilities have a closed-form only for special distributions of $\varepsilon_1, \ldots, \varepsilon_I$. Consider the case that the ε_i are independent and identically distributed with an Extreme Value distribution with pdf

$$f(\varepsilon_i) = e^{-\varepsilon_i} e^{-e^{-\varepsilon_i}}, -\infty < \varepsilon_i < \infty$$

and cdf

$$F(\varepsilon_i) = e^{-e^{-\varepsilon_i}}$$

We have $\mathrm{E}(\varepsilon_i) = \gamma = .5772$ which is Euler's constant. Also $\mathrm{Var}(\varepsilon_i) = \pi^2/6$.

For the choice probabilities (using the final expression above)

$$p_i(v) = \int_{-\infty}^{\infty} e^{-\varepsilon_i} e^{-e^{-\varepsilon_i}} \prod_{j=1,j\neq i}^{I} e^{-e^{-(v_j - v_i + \varepsilon_i)}} \mathrm{d}\varepsilon_i = \int_{-\infty}^{\infty} e^{-\varepsilon_i} e^{-e^{-\varepsilon_i}(1+\sum_{j=1,j\neq i}^{I} e^{v_i - v_j})} \mathrm{d}\varepsilon_i$$

Define

$$\lambda_i = \ln\left(1 + \sum_{j=1,j\neq i}^{I} e^{v_i - v_j}\right)$$

so that

$$p_i(v) = \int_{-\infty}^{\infty} e^{-\varepsilon_i} e^{-e^{-(\varepsilon_i - \lambda_i)}} \mathrm{d}\varepsilon_i$$

Change of variables to $\eta_i = \varepsilon_i - \lambda_i$ so that $\varepsilon_i = \eta_i + \lambda_i$

$$p_i(v) = e^{-\lambda_i} \int_{-\infty}^{\infty} e^{-\eta_i} e^{-e^{-\eta_i}} \mathrm{d}\eta_i = e^{-\lambda_i} = \frac{1}{1 + \sum_{j=1,j\neq i}^{I} e^{v_i - v_j}} = \frac{e^{-v_i}}{\sum_{j=1}^{I} e^{-v_j}}$$

The discrete choice model with these choice probabilities is called the Multinomial Logit (MNL) model.

Properties of MNL model

The MNL model is popular because the choice probabilities have a closed form expression that does not require numerical integration.

The main property that limits the usefulness of MNL (at least without changes): Independence of Irrelevant Alternatives (IIA)

$$\frac{p_i(v)}{p_j(v)} = e^{-(v_i - v_j)}$$

i.e. the ratio of any two choice probabilities depends only on the average utilities of these alternatives, and not on the average utilities of the other alternatives.

IIA and substitution of alternatives

IIA imposes restrictions on the substitution between alternatives. Substitution occurs if a *new alternative* is introduced, e.g. a new occupation emerges, or if an existing alternative becomes *more or less attractive*. This kind of changes is often the reason that the model is considered in the first place. An example is a bonus that increases the attractiveness of becoming a teacher.

New alternatives

New alternative: a variant on McFadden's red bus, blue bus example. We assume that initially there are 2 occupations that you choose from, 1=Dentist, and 2=Surgeon, that have the same average utility $-v_1 = -v_2$, so that

$$p_1(v) = \frac{e^{-v_1}}{e^{-v_1} + e^{-v_2}} = \frac{1}{2} = p_2(v)$$

Initially all surgeons are required to wear white coats in the operating room. This rule is changed and now you can choose to wear a blue coat (alternative 3). Obviously, $-v_2 = -v_3$, so that the probability of becoming a dentist is

$$p_1(v) = \frac{e^{-v_1}}{e^{-v_1} + e^{-v_2} + e^{-v_3}} = \frac{1}{3}$$

Conclusion: The probability of choosing to become a dentist has gone down, due to the introduction of a close substitute of the other alternative the blue coated surgeon. One would expect that the option to wear a blue coat would only reduce the probability of becoming a white coated surgeon. The reason is that although the options of a blue and white coated surgeon have the same average utilities, they have independent random utility components.
Changing the attractiveness of alternatives

Absolute and relative changes in attractiveness

$$\begin{aligned} \frac{\partial p_i}{\partial v_i}(v) &= -p_i(v)(1 - p_i(v)) \\ \frac{\partial p_i}{\partial v_j}(v) &= p_i(v)p_j(v) \end{aligned}$$

so that

$$\begin{aligned} \frac{\partial \ln p_i}{\partial v_i}(v) &= -(1 - p_i(v)) \\ \frac{\partial \ln p_i}{\partial v_j}(v) &= p_j(v) \end{aligned}$$

If alternative j becomes less attractive, then the relative change in the fraction that chooses i depends only on $p_j(v)$. So if few people choose j then the relative effect on the fraction that chooses i is small, even if i and j are close substitutes. If many people choose j making j less attractive has a big relative effect on the choice probability of i even if j and i are not substitutes.
If i is chosen by many the relative effect of a change in its attractiveness is small, independent of whether good substitutes are available.

Example

- Three restaurants: Expensive 1(C), Expensive 2 (L), and Fast food (B).
- Two attributes: price with $P_C = 95, P_L = 80$, and $P_B = 5$, and quality, with $Q_C = 10, Q_L = 9$, and $Q_B = 2$.

- Choice fractions, i.e. market shares $S_C = 0.10, S_L = 0.25$, and $S_B = 0.65$.
- Underlying MNL model utilities

$$u_i = -0.2P_i + 2Q_i + \varepsilon_i$$

- Let Expensive 2 go out of business. Predicted market shares of Expensive 1 and Fast food are $S_C = 0.13$ and $S_B = 0.87$.
- One would expect $S_C = 0.35$ and $S_B = 0.65$.

Nested Multinomial Logit (NMNL)

The key problem with IIA is that alternatives that are close, i.e. are close substitutes, have independent random components.

Therefore a solution is to make the random components of similar alternatives dependent.

Consider example with 4 alternatives. Choose the joint cdf of $\varepsilon_1, \ldots, \varepsilon_4$ as

$$F(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4) = e^{-\left(e^{-\frac{\varepsilon_1}{\sigma_1}} + e^{-\frac{\varepsilon_2}{\sigma_1}}\right)^{\sigma_1}} e^{-\left(e^{-\frac{\varepsilon_3}{\sigma_2}} + e^{-\frac{\varepsilon_4}{\sigma_2}}\right)^{\sigma_2}}$$

Note that $\varepsilon_1, \varepsilon_2$ are dependent, as are $\varepsilon_3, \varepsilon_4$. We have $\sigma_1 \approx 1 - \rho(\varepsilon_1, \varepsilon_2)$.

The choice probabilities are

$$p_{1,2}(v) = p_1(v) + p_2(v) = \frac{\left(e^{-\frac{v_1}{\sigma_1}} + e^{-\frac{v_2}{\sigma_1}}\right)^{\sigma_1}}{\left(e^{-\frac{v_1}{\sigma_1}} + e^{-\frac{v_2}{\sigma_1}}\right)^{\sigma_1} + \left(e^{\frac{-v_3}{\sigma_2}} + e^{-\frac{v_4}{\sigma_2}}\right)^{\sigma_2}}$$

and

$$p_{1|1,2}(v) = \frac{e^{-\frac{v_1}{\sigma_1}}}{e^{-\frac{v_1}{\sigma_1}} + e^{-\frac{v_2}{\sigma_1}}}$$

which has the MNL form. Hence

$$p_1(v) = p_{1|1,2}(v) p_{12}(v)$$

This ARUM model is called the Nested Multinomial Logit (NMNL) model. Reconsider the occupational choice example with alternatives 1 (dentist), 2 (white coated surgeon) and 3 (blue coated surgeon). Then

$$p_{2,3}(v) = \frac{2^\sigma e^{-v_2}}{2^\sigma e^{-v_2} + e^{-v_1}}$$

and

$$p_{2|2,3}(v) = \frac{1}{2}$$

If $v_1 = v_2$, then

$$p_{2,3}(v) = \frac{2^\sigma}{2^\sigma + 1}$$

Hence, if $\sigma \downarrow 0$, then $p_{2,3}(v) \to \frac{1}{2}$ which is the same as before the emergence of the blue coated surgeon.

Application

We study the choice of heating/cooling system for a house.

We have data on 250 newly built houses in California and the alternatives are

1. Gas central heat with cooling (74%)
2. Electric central resistance heat with cooling (1.5%)
3. Electric room resistance heat with cooling (.4%)
4. Electric heat pump, which provides cooling also (10%)
5. Gas central heat without cooling (9.6%)
6. Electric central resistance heat without cooling (.4%)
7. Electric room resistance heat without cooling (3.2%)

I will drop alternatives 3 and 6.

There is one household characteristic, annual income (in 1000$) and the attributes are the installation and operating costs of the heating/cooling systems (in $).

	Mean	Stand Dev
Inst Cost Gas with Cooling	3318	347
Inst Cost Elec with Cooling	3347	326
Inst Cost Heat Pump	1057	146
Inst Cost Gas without Cooling	2496	386
Inst Cost Elec without Cooling	1665	219
Oper Cost Gas with Cooling	474	46
Oper Cost Elec with Cooling	659	56
Oper Cost Heat Pump	151	23
Oper Cost Gas without Cooling	225	34
Oper Cost Elec without Cooling	411	43
Annual Household Income (1000$)	44.72	17.15

If i denotes the alternative and t the household, then we specify the average utility of alternative i for household t for instance as

$$v_{it} = \alpha_i + \beta_{i1} CI_{it} + \beta_{i2} CO_{it} + \beta_{i3} INC_t$$

with CI the installation cost and CO the operating cost. The MNL probabilities are

$$p_{it} = \frac{e^{v_{it}}}{\sum_{j=1}^{5} e^{v_{jt}}}$$

The MNL log likelihood is

$$\ln L = \sum_{t=1}^{n} \sum_{i=1}^{5} y_{it} \ln p_{it}$$

with $y_{it} = 1$ if and only if t chooses i. This log likelihood is maximized over the parameters to obtain the MLE.
The estimated MNL and NMNL models have

$$v_{it} = \alpha_i + \beta_1 CI_{it} + \beta_2 CO_{it} + \gamma_{i3} INC_t$$

Remember that we can add a constant to the utilities without changing the choices. For that reason we set $\alpha_1 = 0$ and $\gamma_{13} = 0$.

The MLE are reported in the table. The NMNL model nests alternatives 1 and 2 in one nest and 4 and 5 in another nest.

Parameter	MLE	Std Err	MLE	Std Err
α_2	-3.11	.98	-2.88	1.04
α_3	-1.45	.65	-1.61	.87
α_4	-1.63	.73	-1.80	.69
α_5	-2.80	.88	-2.97	.98
β_1	-.00355	.00156	-.00456	.00201
β_2	-.0256	.0122	-.0300	.0129
γ_2	.0035	.012	-.00038	.024
γ_3	-.0341	.0154	-.0411	.0198
γ_4	-.0012	.024	.013	.039
γ_5	.0023	.045	-.0035	.078
σ_1	-	-	.899	.211
σ_2	-	-	1.13	.321

Policy: \$500 subsidy for heat pump, i.e. lower the installation costs of a heat pump by \$500.

Effect on choice fractions

	Before	MNL	NMNL
Gas with cooling	.75	.70	.65
Elec with cooling	.016	.010	.006
Heat pump	.105	.180	.210
Gas no cooling	.097	.089	0.095
Elec no cooling	.032	.025	.030