Description   Remarks and examples   References   Also see

Description

In this introduction, we introduce you to the random utility model (RUM) formulation under which choice models are typically derived. We also discuss the independence of irrelevant alternatives (IIA) assumption. We tell you which models rely on this assumption and how the other models relax the assumption. Finally, we introduce you to maximum simulated likelihood (MSL) estimation, which is used by many of the choice model commands.

Remarks and examples

Remarks are presented under the following headings:

Random utility models

Choice models are typically derived under an assumption of utility-maximizing behavior by the decision maker (Train 2009). Say that the decision makers are enumerated as $i = 1, 2, \ldots, N$, each facing a choice among $a = 1, 2, \ldots, A$ alternatives. The decision makers derive a certain utility (such as a profit or other benefit) from each possible choice. The utility can be expressed as

$$U_{ia} = V_{ia} + \epsilon_{ia} \tag{1}$$

where U_{ia} is the utility of alternative a for the ith decision maker and V_{ia} is the observed component of the utility, typically modeled as a linear function of observed data vectors. The term ϵ_{ia} represents the unobserved components of the utility. The ϵ_{ia} are assumed to have a random distribution, the precise formulation of which depends on the choice model. This general model is called a random utility model.

When there is a discrete choice, the largest U_{ia} among the $a = 1, 2, \ldots, A$ alternatives gives the alternative chosen by the ith decision maker. When there are rank-ordered choices, the order of the U_{ia} corresponds to the ranks assigned by the decision maker.

For a discrete choice model, the probability that the ith decision maker picks alternative a is

$$\begin{aligned} P_{ia} &= \Pr(U_{ia} > U_{ib} \quad \text{for all } b \neq a) \\ &= \Pr(V_{ia} + \epsilon_{ia} > V_{ib} + \epsilon_{ib} \quad \text{for all } b \neq a) \\ &= \Pr(\epsilon_{ia} - \epsilon_{ib} > V_{ib} - V_{ia} \quad \text{for all } b \neq a) \end{aligned}$$

With the distribution of the $\boldsymbol{\epsilon}_i$ given by $f(\boldsymbol{\epsilon}_i)$, this probability can be written as

$$P_{ia} = \int I(\epsilon_{ia} - \epsilon_{ib} > V_{ib} - V_{ia} \quad \text{for all } a \neq b)\, f(\boldsymbol{\epsilon}_i)\, d\boldsymbol{\epsilon}_i \tag{2}$$

where $I(\cdot)$ is the indicator function equal to 1 when the expression inside the parentheses is true and 0 otherwise.

Alternative-specific variables and case-specific variables

Some observed measures are characteristics related to the alternative. For example, if alternatives are different modes of public transportation—bus, subway, or commuter train—one measure might be the cost of a ticket for each alternative. We call these measures alternative specific.

Other observed measures are characteristics of the decision maker alone, for example, his or her age or income. We call these measures case specific. Because case-specific measures may affect different alternatives in different ways, there is a not a single coefficient estimated for each case-specific measure, but rather $A-1$ of them, one for each alternative less one. The coefficients estimate relative differences among alternatives due to the case-specific variable.

In McFadden's choice model, the observed part of the utility is modeled as

$$V_{ia} = \mathbf{w}_{ia}\boldsymbol{\alpha} + \mathbf{z}_i\boldsymbol{\delta}_a + c_a$$

where $\boldsymbol{\alpha}$ are coefficients for $\mathbf{w}_{ia}$, a vector of k alternative-specific variables; $\boldsymbol{\delta}_a$ are coefficients (which vary by alternative) for $\mathbf{z}_i$, a vector of m case-specific variables; and c_a are alternative-specific intercepts.

Only $A-1$ alternative-specific intercepts c_a and $A-1$ coefficients $\boldsymbol{\delta}_a$ for the case-specific variables need to be estimated because only relative differences among the utilities U_{ia} matter. Hence, for a model with k alternative-specific variables and m case-specific variables, a total of $k + m(A-1) + A - 1$ coefficients will be estimated: k for the alternative-specific variables, $m(A-1)$ for the case-specific variables, and $A-1$ alternative-specific intercepts.

Independence of irrelevant alternatives

When the relative probabilities of two alternatives in the model do not depend on the characteristics of other alternatives, the model has the IIA property. In terms of the utilities of (1), only models in which errors are independent across alternatives have the IIA property. Thus, whether IIA is plausible hinges on whether the errors are independent over the alternatives or whether they might be correlated.

Stata has estimators for models that have IIA and for models that do not have IIA. For discrete choice models, multinomial logit (`mlogit`) and McFadden's choice model (`cmclogit`) have the IIA property. The mixed logit model (`cmmixlogit` and `cmxtmixlogit`) and the multinomial probit model (`cmmprobit`) allow you to explicitly model the correlations of the errors to fit models that do not have the IIA property. The nested logit model (`nlogit`) models nested alternatives and also does not impose IIA. For rank-ordered outcomes, the rank-ordered logit model (`cmrologit`) imposes IIA, and the rank-ordered probit model (`cmroprobit`) can fit models that do not impose IIA.

Let's illustrate IIA with an example. Consider a commuter who has the choice of either using a car or walking to get to work. A choice model gives the probabilities P_{car} and P_{walk}, the probabilities for the choices car and walking. Now suppose that a new bus line opens, and the commuter can now take a bus to work. Now the choice model has P_{bus} as well, the probability of taking a bus. Because the probabilities must sum to one, $P_{\text{car}} + P_{\text{walk}}$ must be smaller now. But what about their ratio $P_{\text{car}}/P_{\text{walk}}$? Should that change?

The property of IIA implies that the ratio $P_{\text{car}}/P_{\text{walk}}$ does not change when the new alternative of taking a bus becomes available. Say $P_{\text{car}}/P_{\text{walk}} = 2$ before the availability of taking a bus. So the model says a car is taken with probability $2/3$ and walking with probability $1/3$.

Now a bus is available, say, with $P_{\text{bus}} = 0.25$. IIA assumes that the existence of the choice of a bus does not change the relative appeal of taking a car over walking. So $P_{\text{car}}/P_{\text{walk}} = 2$ is still true, and now $P_{\text{car}} = 0.5$ and $P_{\text{walk}} = 0.25$. Discussions of whether these new predicted probabilities are valid hinge on whether the assumed independence is realistic.

Models that have the IIA property have the property because of the functional formulation of the model. Consider McFadden's choice model (McFadden 1974), which has IIA and is fit by the command `cmclogit` using conditional logistic regression. The probabilities in McFadden's choice model are given by

$$P_{ia} = \frac{e^{V_{ia}}}{\sum_{j=1}^{A} e^{V_{ij}}} \tag{3}$$

See *Methods and formulas* in [CM] **cmclogit** and *Methods and formulas* in [R] **clogit**.

The ratio for the probability of alternative a to the probability for alternative b is

$$\frac{P_{ia}}{P_{ib}} = \frac{e^{V_{ia}}}{e^{V_{ib}}} \tag{4}$$

The ratio is independent of the probabilities of any of the other alternatives, and hence, the McFadden's choice model satisfies IIA.

Regardless of whether the true model for our data has IIA, if we fit a McFadden's choice model to our data, the probabilities from the model will satisfy IIA. IIA is a mathematical consequence of the formulation of McFadden's choice model.

If the errors in (1) are correlated, the choice probabilities do not have the form of (3), and the ratio in (4) depends on the characteristics of other alternatives. If you do not want to assume independence of errors across alternatives at the outset, you can fit a model that estimates correlation parameters and test these parameters.

Estimators that do not assume IIA

The CM estimators for discrete choice models that do not assume IIA are `cmmprobit`, `cmmixlogit`, and its extension to panel data, `cmxtmixlogit`. For rank-ordered alternatives, `cmroprobit` fits models that do not assume IIA.

As described earlier, for McFadden's choice model, the utility is modeled as

$$\begin{aligned} U_{ia} &= V_{ia} + \epsilon_{ia} \\ &= \mathbf{w}_{ia}\boldsymbol{\alpha} + \mathbf{z}_i\boldsymbol{\delta}_a + c_a + \epsilon_{ia} \end{aligned} \tag{5}$$

where $\boldsymbol{\alpha}$ are coefficients for the alternative-specific variables $\mathbf{w}_{ia}$; $\boldsymbol{\delta}_a$ are coefficients for the case-specific variables $\mathbf{z}_i$; c_a are intercepts; and ϵ_{ia} are unobserved random variables, modeled as independent type I (Gumbel-type) extreme-value random variables.

In the mixed logit model fit by `cmmixlogit`, the utility is

$$\begin{aligned} U_{ia} &= V_{ia} + \epsilon_{ia} \\ &= \mathbf{x}_{ia}\boldsymbol{\beta}_i + \mathbf{w}_{ia}\boldsymbol{\alpha} + \mathbf{z}_i\boldsymbol{\delta}_a + c_a + \epsilon_{ia} \end{aligned} \tag{6}$$

where $\boldsymbol{\beta}_i$ are random coefficients that vary over individuals in the population and $\mathbf{x}_{ia}$ is a vector of alternative-specific variables. The other terms are as in McFadden's choice model.

The $\boldsymbol{\beta}_i$ are not directly estimated. They are assumed to have a particular distribution, and the parameters of the distribution are estimated. For example, if the $\boldsymbol{\beta}_i$ are assumed to have a multivariate normal distribution, $\boldsymbol{\beta}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the mixed logit model estimates $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The probability of the ith individual picking alternative a is derived from an integral over the distribution of $\boldsymbol{\beta}$, $f(\boldsymbol{\beta})$,

$$P_{ia} = \int P_{ia}(\boldsymbol{\beta}) f(\boldsymbol{\beta}) d\boldsymbol{\beta} \tag{7}$$

where

$$P_{ia}(\boldsymbol{\beta}) = \frac{e^{V_{ia}}}{\sum_{b=1}^{A} e^{V_{ib}}} \tag{8}$$

Although (8) looks like (3), it is (7) that gives the probabilities P_{ia}, and IIA is not satisfied because the random parameters cause the errors in (1) to be correlated. This correlation means that nothing similar to (4) is true.

In a mixed logit model, one or more alternative-specific variables are selected to have random coefficients. When the random coefficients are modeled with a nonzero variance, IIA is not true. When the variance is zero, IIA is true. Testing whether variance parameters are zero gives a model-based test of IIA.

The multinomial probit model fit by `cmmprobit` also does not assume IIA. Its utility is formulated as

$$U_{ia} = \mathbf{w}_{ia}\boldsymbol{\alpha} + \mathbf{z}_i\boldsymbol{\delta}_a + c_a + \xi_{ia} \tag{9}$$

where the random-error term $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, \dots, \xi_{iA})$ is distributed multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma}$. The other terms in the model are the same as they are in McFadden's choice model. Probabilities P_{ia} are computed using (2).

The ratio P_{ia}/P_{ib} is not independent of the other P_{ic} when the covariance matrix $\boldsymbol{\Sigma}$ is specified with nonzero correlation parameters for the alternatives. If, say, the errors for alternative b and c are correlated, a change in P_{ic} will cause a change in P_{ia}/P_{ib}. So IIA is not satisfied in this case.

`cmmixlogit` and `cmmprobit` are described in the introduction [CM] **Intro 5**. `cmroprobit` is covered in [CM] **Intro 6**, and `cmxtmixlogit` in [CM] **Intro 7**.

Maximum simulated likelihood

The integral in (2), for `cmmixlogit`, `cmxtmixlogit`, `cmmprobit`, and `cmroprobit` models, must be approximated because it has no closed-form solution. The integral is computed by simulation, and the estimation is said to be done using MSL.

Consistency of the MSL estimator requires that the number of points in the simulation be sufficiently large. More points will produce more precise estimates by reducing approximation error, at the cost of increased computation time. This should be kept in mind when fitting these models. Practically speaking, it means that when you have found a model that you consider final, you should increase the number of integration points. It also means that if your model is having a hard time converging, the first thing you should try to get the model to converge is increasing the number of integration points.

See *Setting the number of integration points* in [CM] **Intro 5** for examples and advice about setting the number of integration points.

See *Methods and formulas* in `cmmixlogit` and *Methods and formulas* in `cmmprobit` for further statistical details, and see Cameron and Trivedi (2005) for an introduction to MSL estimation.

References

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

McFadden, D. L. 1974. "Conditional logit analysis of qualitative choice behavior". In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. New York: Academic Press.

Train, K. E. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511805271.

Also see

[CM] **Intro 5** — Models for discrete choices