The fmlogit Package: An Econometric Document

Xinde James Ji

Oct.10, 2016

This document provides theoretical documentations for the "fmlogit" package in R. Updates will be published at my github site: https://github.com/f1kidd/fmlogit. Any suggestions or concerns are welcomed[1]. For function usage and calls, please check directly with help(func_name) after loading the fmlogit package.

Motivation

Fractional multinomial responses, or multivariate share models, arises naturally in various occasions. For example, a municipality allocates its budgets to multiple departments, and we are interested in the proportion of the budgets that each department receives. Or, there are multiple candidates in a presendential election, and we are interested in the percentage of support for each candidate in each state.

The model is distinct in that 1) each of the response lies between 0 and 1, and 2) the share of all responses adds up to one. The fmlogit model utilizes the two distinct factors, and model it explicitly using a multinomial logit transformation on the response variables. If the true data generating process is multinomial fractions, or shares of multiple choices, then the fractional multinomial logit model is consistent and efficient, while other candidate models such as Dirichlet or Beta regression is not.

Econometric Model

The basis of this package is Papke and Wooldridge(1996)'s paper, in which they proposed a quasi-maximum likelihood(QMLE) estimator for fractional response variables. As their approach applies to binary response variables, here we expand it to a multinomial response variables with fractional structure.

We start by writing:[2]

$$E(y_{ij}|x_i) = G(x_i\beta_j)$$

for the j^{th} choice of the i^{th} obsevation , where G(.) is a know function satisfying 0<G(z)<1 for all $z \in \mathbb{R}$. Note that here we only allow for common covariates of x_i, and not for choice-specific attributes. Following the logit convention, G(.) is chosen to be the multinomial logit function, with the form:

$$G(z_j) = \frac{exp(z_j)}{\sum_{k=1}^{J} exp(z_k)}$$

And the multinomial likelihood function, is thus given by

$$ln(L_i) = \sum_{j=1}^{J} y_{ij} ln(G(x_i\beta_j))$$

with $\beta_1 = 0$, the baseline coefficient equal to zero. Papke and Wooldridge(1996) showed that the QMLE estimator of β, obtained by the the maximazation problem

$$argmax_\beta \sum_{i=1}^{N} ln(L_i)$$

is a consistent estimator for β if G(z) is the correct functional form for E(y|x).

[1] email: xji@vt.edu

[2] The demonstration below is in individual specific notation, but matrix notation is not hard to obtain from the individual specific notations. The actual function uses matrix calculation, which increases algorithm speed.

To estimate the standard error for the QMLE estimator, define $g(z_j) \equiv \partial G(z_j)/\partial z_j$, the partial derivative of the multinomial logit function with respect to choice j. Specifically, $g(z_j)$ has the following functional form:

$$g(z_j) = \frac{\hat{E}\hat{S} - \hat{E}^2}{\hat{S}^2}$$

where $\hat{E} = exp(x_i\beta_j)$, and $\hat{S} = \sum_{k=1}^{J} exp(x_i\beta_k)$.

A robust asymptotic standard error for $\hat{\beta}_j$ is given by the square root of the diagonal element of the following matrix:

$$\hat{A_j}^{-1}\hat{B}_j\hat{A_j}^{-1}$$

where

$$\hat{A} = \sum_{i=1}^{N} \frac{\hat{g}_{ij}^2 \mathbf{x}_i' \mathbf{x}_i}{\hat{G}_{ij}(1 - \hat{G}_{ij})}$$

$$\hat{B} = \sum_{i=1}^{N} \frac{\hat{u}_{ij}^2 \hat{g}_{ij}^2 \mathbf{x}_i' \mathbf{x}_i}{[\hat{G}_{ij}(1 - \hat{G}_{ij})]^2}$$

in which $\hat{u}_{ij}$ is the residual for the j^{th} choice of the i^{th} observation, given by $\hat{u}_{ij} = y_{ij} - G(x_i\beta_j)$. Specifically, $\hat{A}$ is the information matrix, which is not a consistent estimator itself, and $\hat{B}$ is a weight correction for A.

In most binary / multinomial response models, the convention is to treat one of the choices as a baseline. Here we apply the same logic, and treat j=1 as the baseline scenario. This implicitly generates a restriction that $\beta_1 = 0$, and all other betas are the marginal difference to the baseline case.

Partial Effects

Interpreting partial effects for limited dependent variables can be tricky, and this is especially true for multinomial logit models. The coefficients obtained in the regression model represents the logit-transformed odds ratio for that specific choice against the baseline choice, and should not be treated as the marginal or discrete effect obtained from the model. Instead, the modeller has to derive "partial effects" from the coefficient estimates, which is analogous to the coefficients in a regular linear model. The rest of the section provides detail information for deriving two types of partial effects: marginal effects and discrete effects.

Marginal Effects

Marginal effect is the counterpart of the coefficients of a continous variable in a linear model. It measures the effect of a marginal change of a continuous variable x_k on the choice variable y_j. It should be point out that the right hand size variable, x_k, will have J marginal effects, each on one choice variable y_j.

The marginal effect of the multinomial models actually has a very distinctive form:

$$ME_{jk} = \frac{\partial p_j}{x_k} = p_j(\beta_{kj} - \bar{\beta}_i)$$

where p_j is an 1*N vector of predicted probabilities for choice j, and $\bar{\beta}_i = \sum_{m=1}^{J} \beta_{km} p_m$ is the probability weighted average of βkm. This shows that the marginal effects among different individuals are actually different given different predicted probabilities of choice j.

Typically, two types of summary measures are used to illustrate the global average marginal effects. The first one is called marginal effects at the mean (MEM) in the code. In algebric form, this is represented as

$$MEM_{jk} = \bar{p}_j(\beta_{kj} - \bar{\beta}_i)$$

where $\bar{p}_j$ is the predicted value of choice j at the mean of all X covariates. Centering observations around the mean simplifies the calculation, however it ignores the potential heterogeneity in marginal effects, especially at the extreme values.

Another measure is called average marginal effects (AME). This can be written as:

$$AME_{jk} = \frac{1}{N}\sum_{i=1}^{N} p_j(\beta_{kj} - \bar{\beta}_i)$$

However, according to Greene(2003), there is no agreement as to which one is prefered. A more inclusive approach will be to plot the marginal effect of interest across all individuals. This is not provided in the function, but can certainly be implemented in a straightforward way in R.

Discrete Effects

Discrete effect is a little bit different from marginal effects. Instead of calculating the slope of the coefficients, discrete effect considers the impact of a discrete change in one covariates on the predicted outcome variables. This is especially useful for dummy variables, where calculating marginal effects does not make much sense.

The discrete effect has a straight-forward form. Consider a discrete change of a dummy variable k from 0 to 1. This is just

$$DE_{jk} = Pr(y = j|\mathbf{x}_{x_k=1}) - Pr(y = j|\mathbf{x}_{x_k=0})$$

The change in predicted value by setting $x_k = 1$ and $x_k = 0$.

Similar to the marginal effect case, we can calculate discrete effects at the mean (DEM) by predicting the outcome when all other covariates at the mean, or average discrete effects, which averages the predicted difference across all observations.

Standard Errors

Here we adopt the Krinsky-Robb method to compute standard errors for marginal and discrete effects. As oppose to the delta method commonly used in other programs such as Stata, Krinsky-Robb is a simulation-based method. The idea of Krinsky-Robb is that, to calculate the variance of a function $Var(f(\beta))$, we do the following step:

For i in 1:N, where N is a very large number,

1) Sample from the known distribution of β
2) For each of the sample, calculate $f(\beta)$
3) Take the empirical variance of $f(\beta)$.

And after sufficiently large sample size, the empirical variance converges to the theoretical variance.

Hypothesis Testing of Marginal and Discrete Effects

One of the major advantages of Krinsky-Robb over numerical Delta method is in hypothesis testing of those effects. The marginal and discrete effect are not normally distributed since the effect contains a multinomial logit distributed $p_i j$. So even though sample size is large, the central limit theorem does not really hold here. So knowing the standard error does not really help if we do not know the actual shape of the density. But that is out of the scope of the Delta method.

Krinsky-Robb solves this problem by providing empirical draws from the marginal effects. Hypothesis testing here will be very simple: say we test $H_0 : D_j = 0$. We just need to compare 0 with our N draws, and see if it falls out of the 95% mass. This is a major advantage we provide here comparing with Stata's fmlogit module.

Practical Concerns

Optimization Method

This function calls *maxLik()* in package *maxLik* to maximize the quasi-likelihood function. The *maxLik* function is a wrapper which provides several different maximization methods, including most *optim()* methods in the base package, as well as other useful methods such as BHHH(Berndt-Hall-Hall-Hausman). The choice of optimization method can create vastly different parameter estimates. Here it is recommended that either conjugate gradients(CG), or Berndt-Hall-Hall-Hausman(BHHH) to ensure convergence. In limited testing scenarios, BHHH typically has the best performace in terms of convergence for large datasets, while CG is faster in computation speed for smaller, easy to converge datasets.

Robust Standard Error

It is worth noting that the robust standard error created in this function is consistently lower than that created in Stata's fmlogit package, typically by about 20%. However, the robust SE here is a consistent estimator following Pakpe and Wooldridge(1996)'s ${\hat{A}_j}^{-1}\hat{B}_j{\hat{A}_j}^{-1}$ estimator, so it is recommended that the number should be used with causion.

References

Papke, L. E. and Wooldridge, J. M. (1996), Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J. Appl. Econ., 11: 619-632.

Wulff, Jesper N. "Interpreting Results From the Multinomial Logit Model Demonstrated by Foreign Market Entry." Organizational Research Methods (2014): 1094428114560024.

Greene, W. H. (2003). Econometric analysis (5th ed.). Upper Saddle River, NJ.: Prentice Hall