mlogit : a R package for the estimation of the multinomial logit

Yves Croissant[1]

[1](LET University Lyon II)

UseR 2009
July, 9th 2009

Motivations

- the multinomial logit model is widely used to modelize the choice among a set of alternatives and R provide no function to estimate this model,
- mlogit enables the estimation of the basic multinomial logit model and provides the tools to manipulate the model,
- some extensions of the basic model (random parameter logit, heteroskedastic logit and nested logit) are also provided

Motivations

- the multinomial logit model is widely used to modelize the choice among a set of alternatives and R provide no function to estimate this model,
- mlogit enables the estimation of the basic multinomial logit model and provides the tools to manipulate the model,
- some extensions of the basic model (random parameter logit, heteroskedastic logit and nested logit) are also provided

Motivations

- the multinomial logit model is widely used to modelize the choice among a set of alternatives and R provide no function to estimate this model,
- mlogit enables the estimation of the basic multinomial logit model and provides the tools to manipulate the model,
- some extensions of the basic model (random parameter logit, heteroskedastic logit and nested logit) are also provided

Outline of the talk

1. Theoretical background
   - Discrete choice models
   - Logit models
2. Implementation
   - Data management
   - Estimation methods
   - Estimation functions
3. Examples

Random utility and decision rule

$$\begin{cases} U_1 &= \beta_1^\top x_1 + \epsilon_1 &= V_1 + \epsilon_1 \\ U_2 &= \beta_1^\top x_1 + \epsilon_2 &= V_2 + \epsilon_2 \\ &\vdots & \vdots \\ U_J &= \beta_J^\top x_J + \epsilon_J &= V_J + \epsilon_J \end{cases}$$

l chosen if :

$$\begin{cases} U_l - U_1 &= (V_l - V_1) + (\epsilon_l - \epsilon_1) > 0 \\ U_l - U_2 &= (V_l - V_2) + (\epsilon_l - \epsilon_2) > 0 \\ &\vdots \\ U_l - U_J &= (V_l - V_J) + (\epsilon_l - \epsilon_J) > 0 \end{cases}$$

Probability : general case

$$\begin{cases} \epsilon_1 & < & (V_I - V_1) + \epsilon_I \\ \epsilon_2 & < & (V_I - V_2) + \epsilon_I \\ & \vdots & \\ \epsilon_J & < & (V_I - V_J) + \epsilon_I \end{cases}$$

$$\begin{aligned} (\mathrm{P}_I \mid \epsilon_I) &= \mathrm{P}(U_I > U_1, \ldots, U_I > U_J) \\ &= F(\epsilon_1 < (V_I - V_1) + \epsilon_I, \ldots, \epsilon_J < (V_I - V_J) + \epsilon_I) \end{aligned}$$

$$\mathrm{P}_I = \int (\mathrm{P}_I \mid \epsilon_I) f_I(\epsilon_I) d\epsilon_I$$

$$\mathrm{P}_I = \int F((V_I - V_1) + \epsilon_I, \ldots, (V_I - V_J) + \epsilon_I) f_I(\epsilon_I) d\epsilon_I$$

Logit models

The marginal distribution of the error terms follows a Gumbel (or extreme value) distribution, which has the following cumulative and density functions :

$$F(\epsilon) = e^{-e^{-(\epsilon-\mu)/\theta}}$$

$$f(\epsilon) = \frac{1}{\theta} e^{-(\epsilon-\mu)/\theta} e^{-e^{-(\epsilon-\mu)/\theta}}$$

where μ is the location parameter and θ the scale parameter. If the observed part of utility contains an intercept, the location parameter is irrelevant. The mean is $\mu + \gamma\theta$ (where $\gamma = 0.577$ is the Euler-Macheroni constant) and the variance is $\theta\frac{\pi^2}{6}$

Typology of logit models

	multinomial	nested	heteroscedastic	mixed
independence	yes	no	yes	yes
homscedasticity	yes	yes	no	yes
identical parameters	yes	yes	yes	no

The multinomial logit model

$$\begin{aligned}
\mathrm{P}_l \mid \epsilon_l &= \mathrm{P}(U_l > U_1, \ldots, U_l > U_J) \\
&= F(\epsilon_1 < (V_l - V_1) + \epsilon_l, \ldots, \epsilon_J < (V_l - V_J) + \epsilon_l) \\
&= \prod_{k \neq l} e^{-e^{-(V_l - V_k + \epsilon_l)}}
\end{aligned}$$

because of the hypothesis of independance and homoscedasticity.

$$\begin{aligned}
\mathrm{P}_l &= \int (\mathrm{P}_l \mid \epsilon_l) f_l(\epsilon_l) d\epsilon_l \\
&= \int \prod_{k \neq l} e^{-e^{-(V_l - V_k + \epsilon_l)}} e^{-e^{-\epsilon_l}} d\epsilon_l
\end{aligned}$$

$$\mathrm{P}_l = \frac{e^{V_l}}{\sum_k e^{V_k}}$$

The probabilities that enter the log-likelihood has a closed form.

The heteroskedastic logit model

$$P_l \mid \epsilon_l = \prod_{j \neq i} e^{-e^{-\frac{(V_l - V_j + \epsilon_l)}{\theta_j}}}$$

$$P_l = \int_{-\infty}^{+\infty} \prod_{k \neq l} e^{-e^{-\frac{(V_l - V_k + \epsilon_l)}{\theta_k}}} \frac{1}{\theta_l} e^{-\frac{\epsilon_l}{\theta_l}} e^{-e^{-\frac{\epsilon_l}{\theta_l}}} d\epsilon_l$$

There is no closed form for this integral, but it can be writen :

$$P_l = \int_0^{+\infty} \prod_{k \neq l} e^{-e^{-\frac{(V_l - V_k + \theta_l \ln u)}{\theta_k}}} e^{-u} du$$

This integral has the form : $P_l = \int_0^{+\infty} G(u) e^{-u} du$ and can efficiently estimated using Gauss-Laguerre quadrature.

The nested logit model

Alternatives are grouped in different nests $n, m = 1 \ldots N$. The unobservable part of utilities still have marginal distributions which are Gumbell, but they are now correlated within nests :

$$\exp\left(-\sum_{n=1}^{N}\left(\sum_{k\in B_n} e^{-\epsilon_k/\lambda_n}\right)^{\lambda_n}\right)$$

It can be shown that the probability of choosing an alternative l in nest m is :

$$P_l = \frac{e^{V_l/\lambda_m}\left(\sum_{k\in B_m} e^{V_k/\lambda_m}\right)^{\lambda_m - 1}}{\sum_{n=1}^{N}\left(\sum_{k\in B_n} e^{V_k/\lambda_n}\right)^{\lambda_n}}$$

The mixed (or random parameters) logit model

The ϵ are assumed to be *iid*. But the parameters of the observed part of utility are now individual specific : $V_{li} = \beta_i^\top x_{li}$

$$\mathrm{P}_{li}|\beta_i = \frac{e^{V_{li}}}{\sum_k e^{V_{ki}}}$$

Some hypothesis are made about the distribution of the individual specific parameters: $\beta_i \mid f(\theta)$. The expected value of the probability is then :

$$\mathrm{E}(\mathrm{P}_{li}|\beta_i) = \int\int\ldots\int \frac{e^{V_{li}}}{\sum_k e^{V_{ki}}} f(\beta,\theta) d\beta$$

The dimension of the integral is the number of random parameters

Shaping the data

Like panel (or longitudinal) data, data may be stored in a "wide" or in a "long" format :

- in "wide" format, each row is a choice and each column is a variable for a specific alternative,
- in "long" format, each row is an alternative and each column is a variable.

with the `mlogit` package, data should be stored in "long" format. Raw data are reshaped using the `mlogit.data` function.

Shaping the data

Like panel (or longitudinal) data, data may be stored in a "wide" or in a "long" format :

- in "wide" format, each row is a choice and each column is a variable for a specific alternative,
- in "long" format, each row is an alternative and each column is a variable.

with the `mlogit` package, data should be stored in "long" format. Raw data are reshaped using the `mlogit.data` function.

Shaping the data

Like panel (or longitudinal) data, data may be stored in a “wide” or in a “long” format :

- in “wide” format, each row is a choice and each column is a variable for a specific alternative,
- in “long” format, each row is an alternative and each column is a variable.

with the `mlogit` package, data should be stored in “long” format. Raw data are reshaped using the `mlogit.data` function.

Shaping the data

Like panel (or longitudinal) data, data may be stored in a "wide" or in a "long" format :

- in "wide" format, each row is a choice and each column is a variable for a specific alternative,
- in "long" format, each row is an alternative and each column is a variable.

with the `mlogit` package, data should be stored in "long" format. Raw data are reshaped using the `mlogit.data` function.

Shaping the data : a "long" data.frame

```
R> library("mlogit")
R> data("ModeChoice", package = "Ecdat")
R> head(ModeChoice, 5)

  mode ttme invc invt gc hinc psize
1    0   69   59  100 70   35     1
2    0   34   31  372 71   35     1
3    0   35   25  417 70   35     1
4    1    0   10  180 30   35     1
5    0   64   58   68 68   30     2

R> Mo <- mlogit.data(ModeChoice, choice = "mode",
+      shape = "long", alt.levels = c("air",
+          "train", "bus", "car"))
```

```
R> head(Mo, 5)

         chid   alt  mode ttme invc invt gc
1.air       1   air FALSE   69   59  100 70
1.train     1 train FALSE   34   31  372 71
1.bus       1   bus FALSE   35   25  417 70
1.car       1   car  TRUE    0   10  180 30
2.air       2   air FALSE   64   58   68 68
        hinc psize
1.air     35     1
1.train   35     1
1.bus     35     1
1.car     35     1
2.air     30     2
```

Shaping the data : a “wide” data.frame

```
R> data("Heating", package = "Ecdat")
R> head(Heating, 2)

  idcase depvar  ic.gc  ic.gr  ic.ec  ic.er  ic.hp  oc.gc
1      1     gc 866.00 962.64 859.90 995.76 1135.5 199.69
2      2     gc 727.93 758.89 796.82 894.69  968.9 168.66
   oc.gr  oc.ec  oc.er  oc.hp income agehed rooms region
1 151.72 553.34 505.60 237.88      7     25     6 ncostl
2 168.66 520.24 486.49 199.19      5     60     5 scostl
     pb.gc    pb.gr    pb.ec    pb.er    pb.hp
1 4.336722 6.344846 1.554017 1.969462 4.773415
2 4.315961 4.499526 1.531639 1.839072 4.864200

R> Heat <- mlogit.data(Heating, varying = c(3:12,
+     17:21), choice = "depvar", shape = "wide")
```

```
R> head(Heat)

     chid alt idcase depvar income agehed rooms region
1.ec    1  ec      1  FALSE      7     25     6 ncostl
1.er    1  er      1  FALSE      7     25     6 ncostl
1.gc    1  gc      1   TRUE      7     25     6 ncostl
1.gr    1  gr      1  FALSE      7     25     6 ncostl
1.hp    1  hp      1  FALSE      7     25     6 ncostl
2.ec    2  ec      2  FALSE      5     60     5 scostl
          ic     oc       pb
1.ec  859.90 553.34 1.554017
1.er  995.76 505.60 1.969462
1.gc  866.00 199.69 4.336722
1.gr  962.64 151.72 6.344846
1.hp 1135.50 237.88 4.773415
2.ec  796.82 520.24 1.531639
```

Model formulae

Special formula class is provided to take into account that two kind of variables are used :

```
R> f <- logitform(mode ~ invc + invt | hinc)
R> f
```

```
mode ~ invc + invt | hinc
```

which can be updated :

```
R> update(f, . ~ . - invc + ttme | . - hinc + psize)
```

```
mode ~ invt + ttme | psize
```

Model matrix

```
R> X <- model.matrix(logitform(mode ~ invc + invt |
+     hinc), data = Mo)
R> head(X)
```

```
        alttrain altbus altcar invc invt alttrain:hinc
1.air          0      0      0   59  100             0
1.train        1      0      0   31  372            35
1.bus          0      1      0   25  417             0
1.car          0      0      1   10  180             0
2.air          0      0      0   58   68             0
2.train        1      0      0   31  354            30
        altbus:hinc altcar:hinc
1.air             0           0
1.train           0           0
1.bus            35           0
1.car             0          35
2.air             0           0
2.train           0           0
```

Model frame

```
R> mf <- model.frame(logitform(mode ~ invc + invt |
+     hinc), data = Mo)
R> head(mf)

         mode invc invt hinc (chid) (alt)
1.air   FALSE   59  100   35      1   air
1.train FALSE   31  372   35      1 train
1.bus   FALSE   25  417   35      1   bus
1.car    TRUE   10  180   35      1   car
2.air   FALSE   58   68   30      2   air
2.train FALSE   31  354   30      2 train
```

Maximum likelihood

Standard maximum likelihood techniques are used when the probabilities are integrals that have a closed form (multinomial and nested logit models).

The `maxLik` package, which unables the use of several optimisation routines, including Newton-Ralphson, BHHH and BFGS.

Analytical gradient is coded for all the model. More precisely, a matrix containing the contribution of every observation to the gradient is computed (usefull for BHHH).

Gaussian quadrature

For the heteroscedastic logit model, the probabilities can be writen:

$$P_l = \int_0^{+\infty} \prod_{k \neq l} e^{-e^{-\frac{(V_l - V_k + \theta_l \ln u)}{\theta_k}}} e^{-u} du$$

This integral has the form : $P_l = \int_0^{+\infty} f(u) e^{-u} du$ and can efficiently estimated using Gauss-Laguerre quadrature. $\int_0^{+\infty} f(u) e^{-u} du$ is approximated by $\sum_{r=1}^{R} f(u_r) w_r$ where u_r and w_r are respectively vectors of nodes and weights. These vectors are computed using the function `gauss.quad` of the package `statmod`. Very accurate approximation is obtained for R about 40.

Simulations

When the probabilities are multi-dimentional integrals with no closed form, simulations are used (*i.e.* mixed logit)

- use `runif` to generate pseudo random-draws from a uniform distribution, or use more deterministic methods like Halton's draws
- transform this random numbers with the quantile function of the required distribution.

ex: for the Gumbell distribution :

$$F(x) = e^{-e^{-x}} \Rightarrow F^{-1}(x) = -\ln(-\ln(x))$$

To obtain correlated random numbers, Cholesky decomposition is used

The mlogit function

This function unables the estimation of the multinomial logit model

```
R> args(mlogit)

function (formula, data, subset, weights, na.action, alt.subset = NULL,
    reflevel = NULL, estimate = TRUE, ...)
NULL
```

The first 5 arguments are standard. `alt.subset` unables the estimation of the model on a subset of alternatives. `reflevel` indicates which alternative is the reference, the one for which the coefficients are fixed to 0. With `estimate = FALSE`, no estimation is computed, but the model.frame is returned. The dots may include arguments to `mlogit.data` and `maxLik`

mlogit : an example

```
R> data("TravelMode", package = "AER")
R> mlogit(choice ~ travel + wait | income,
+      TravelMode, reflevel = "car",
+      alt.subset = c("train", "car",
+          "bus"), choice = "choice",
+      shape = "long", alt.var = "mode",
+      print.level = 3, iterlim = 10,
+      method = "bfgs")
```

Other estimation functions

- `hlogit` : heteroscedastic logit model: one further argument `R`, the number of evaluations of the function,
- `nlogit` : the nested logit: one further argumen `nests` which indicates the composition of the nests,
- `rlogit`, the random parameter logit model: further arguments include `rpar` (the random parameters and their distribution), `correlation` (a boolean which indicates whether the random parameters are correlated), `R` (the number of draws).

Other estimation functions

- `hlogit` : heteroscedastic logit model: one further argument `R`, the number of evaluations of the function,
- `nlogit` : the nested logit: one further argumen `nests` which indicates the composition of the nests,
- `rlogit`, the random parameter logit model: further arguments include `rpar` (the random parameters and their distribution), `correlation` (a boolean which indicates whether the random parameters are correlated), `R` (the number of draws).

Other estimation functions

- `hlogit` : heteroscedastic logit model: one further argument `R`, the number of evaluations of the function,
- `nlogit` : the nested logit: one further argumen `nests` which indicates the composition of the nests,
- `rlogit`, the random parameter logit model: further arguments include `rpar` (the random parameters and their distribution), `correlation` (a boolean which indicates whether the random parameters are correlated), `R` (the number of draws).

hlogit

```
R> data("TravelMode", package = "AER")
R> hl <- hlogit(choice ~ wait + travel +
+      vcost, TravelMode, shape = "long",
+      id.var = "individual", alt.var = "mode",
+      choice = "choice", print.level = 0,
+      method = "bfgs")
```

```
R> summary(hl)
Call:
hlogit(formula = choice ~ wait + travel + vcost, data = TravelMode,
    shape = "long", id.var = "individual", alt.var = "mode",
    choice = "choice", print.level = 3, method = "bfgs")

Frequencies of alternatives:
    air   train     bus     car
0.27619 0.30000 0.14286 0.28095

70 iterations, 0h:1m:27s
g'(-H)^-1g = 6.46E-06

Coefficients :
             Estimate  Std. Error t-value  Pr(>|t|)
alttrain   0.38199639  0.51723872  0.7385 0.4601923
altbus     0.29217716  0.50365468  0.5801 0.5618377
altcar    -1.60153629  0.74221321 -2.1578 0.0309446 *
wait      -0.04502942  0.00959419 -4.6934 2.687e-06 ***
travel    -0.00290908  0.00079689 -3.6505 0.0002617 ***
vcost     -0.01170644  0.00503115 -2.3268 0.0199763 *
sd.train   0.66909520  0.20913289  3.1994 0.0013772 **
sd.bus     0.30190771  0.12331875  2.4482 0.0143576 *
sd.car     0.46921466  0.26042473  1.8017 0.0715881 .
```

rlogit : revealed prference data

Data about fishing mode choice (used in Cameron and Trivedi)

```
R> data("Fishing", package = "mlogit")
R> Fish <- mlogit.data(Fishing, varying = c(4:11),
+     shape = "wide", choice = "mode", opposite = c("pr"))
R> rlf <- rlogit(mode ~ pr + ca, data = Fish, rpar = c(ca = "n"),
+     R = 100, halton = NA, print.level = 0, norm = "pr",
+     method = "bhhh")
```

R> summary(rlf)

```
Call:
rlogit(formula = mode ~ pr + ca, data = Fish, rpar = c(ca = "n"),
    R = 100, halton = NA, norm = "pr", print.level = 3, method = "bhhh")

Simulated maximum likelihood with 100 draws
20 iterations, 0h:0m:36s
Halton's sequences used
g'(-H)^-1g = 1.27E-08

Coefficients :
             Estimate  Std. Error   t-value     Pr(>|t|)
altboat     0.87026330 0.125546554  6.931798 4.155343e-12
altcharter 1.56022026 0.143989634 10.835643 0.000000e+00
altpier     0.30562456 0.114913999  2.659594 7.823495e-03
pr          0.02778602 0.001464047 18.978915 0.000000e+00
ca          0.46362417 0.158817775  2.919221 3.509074e-03
sd.ca       1.31157680 0.369803939  3.546682 3.901159e-04

log Likelihood : -1225

random coefficients

   Min.   1st Qu.   Median      Mean  3rd Qu. Max.
```

rlogit : stated preference data

Data about train tickets (Journal Of Applied Econometrics data archive)

```
R> data("Train", package = "Ecdat")
R> Train <- mlogit.data(Train, choice = "choice",
+     varying = 4:11, sep = "", alt.levels = c("ch1",
+         "ch2"), shape = "wide", opposite = c("price",
+         "change", "comfort", "time"))
```

- stated prefence data, four attributes (price, comfort, time and change),
- opposite is taken so that coefficients signs are positive,
- two tickets are proposed,
- panel data (each traveler answers about 10 questions)

```
R> rlt <- rlogit(choice ~ price + time + change +
+     comfort - 1, data = Train, rpar = c(change = "n",
+     comfort = "n", time = "n"), R = 20, halton = NA,
+     print.level = 0, id = "id", correlation = TRUE,
+     norm = "price", method = "bhhh")
```

```
R> summary(rlt)
Call:
rlogit(formula = choice ~ price + time + change + comfort - 1,
    data = Train, rpar = c(change = "n", comfort = "n", time = "n"),
    correlation = TRUE, id = "id", R = 20, halton = NA, norm = "price",
    print.level = 3, method = "bfgs")

Simulated maximum likelihood with 20 draws
80 iterations, 0h:0m:45s
Halton's sequences used
g'(-H)^-1g = 9.57E-08

Coefficients :
                   Estimate   Std. Error   t-value    Pr(>|t|)
price           0.002901567 0.0001299284 22.332048 0.000000000
time            0.066946023 0.0046011930 14.549710 0.000000000
change          1.151024508 0.1028259579 11.193910 0.000000000
comfort         2.601321445 0.1494953105 17.400689 0.000000000
change.change   0.096798289 0.0066620796 14.529741 0.000000000
change.comfort -0.286813604 0.1046027051 -2.741933 0.006107881
change.time     1.206042026 0.1274389943  9.463681 0.000000000
comfort.comfort 1.109500379 0.1046907353 10.597885 0.000000000
comfort.time    1.263928574 0.1218029550 10.376830 0.000000000
time.time       2.375357395 0.1689469626 14.059782 0.000000000
```

nlogit

```
R> data("TravelMode", package = "AER")
R> TravelMode$avincome <- with(TravelMode, income * (mode ==
+     "air"))
R> TravelMode$time <- with(TravelMode, travel + wait)/60
R> TravelMode$timeair <- with(TravelMode, time * I(mode ==
+     "air"))
R> TravelMode$income <- with(TravelMode, income/10)
R> nl <- nlogit(choice ~ time + timeair | income, TravelMode,
+     choice = "choice", shape = "long", alt.var = "mode",
+     print.level = 3, method = "bfgs", nest = list(public = c("train",
+         "bus"), other = c("air", "car")))
```

```
R> summary(nl)

Call:
nlogit(formula = choice ~ time + timeair | income, data = TravelMode,
    nest = list(public = c("train", "bus"), other = c("air",
        "car")), choice = "choice", shape = "long", alt.var = "mode",
    print.level = 3, method = "bfgs")

Frequencies of alternatives:
    air   train     bus     car
0.27619 0.30000 0.14286 0.28095

89 iterations, 0h:0m:8s
g'(-H)^-1g = 2.51E-10

Coefficients :
                Estimate Std. Error t-value  Pr(>|t|)
alttrain        -1.78590    2.30610 -0.7744 0.4386803
altbus          -2.78181    2.21062 -1.2584 0.2082544
altcar          -6.38296    2.67459 -2.3865 0.0170088 *
time            -1.30149    0.18350 -7.0924 1.318e-12 ***
timeair         -5.87837    0.80718 -7.2826 3.273e-13 ***
alttrain:income -0.83070    0.31355 -2.6494 0.0080638 **
altbus:income   -0.55447    0.32293 -1.7170 0.0859819 .
altcar:income   -0.36207    0.51361 -0.7050 0.4808378
```