The Stata Journal (2004)
4, Number 3, pp. 265–273

Understanding the multinomial-Poisson transformation

Paulo Guimarães
Medical University of South Carolina

Abstract. There is a known connection between the multinomial and the Poisson likelihoods. This, in turn, means that a Poisson regression may be transformed into a logit model and vice versa. In this paper, I show the data transformations required to implement this transformation. Several examples are used as illustrations.

Keywords: st0069, Poisson regression, logit, conditional logit

1 Introduction

Following McFadden's (1974) conditional logit model, the probability that individual i selects choice j among a set of J_i alternatives is given by

$$P_{ij} = \frac{\exp(\beta' \mathbf{x}_{ij})}{\sum_{j=1}^{J_i} \exp(\beta' \mathbf{x}_{ij})}$$

where β is a vector of unknown parameters and the $\mathbf{x}_{ij}$ are covariates that may change with individual, choice, or both. This logit formulation is quite general and is statistically equivalent to the multinomial logit model, when covariates are restricted to characteristics of the individual, as well as to the conditional logistic model with one case and (possibly) multiple controls (see [R] **clogit**). If we let d_{ij} be an indicator variable that takes the value one if individual i selects choice j and zero otherwise, we can write the log-likelihood function for the conditional logit model as

$$\mathrm{LL}_{\mathrm{cl}} = \sum_{i=1}^{N} \sum_{j=1}^{J_i} d_{ij} \ln P_{ij} \tag{1}$$

where N is the total number of individuals. Several authors (e.g., Palmgren [1981], Baker [1994], Lang [1996], and Guimarães, Figueiredo, and Woodward [2003]) have shown that a multinomial likelihood can be transformed into a Poisson likelihood with additional parameters. Indeed, maximization of the following Poisson likelihood

$$\mathrm{LL}_P = \sum_{i=1}^{N} \sum_{j=1}^{J_i} - \exp(\alpha_i + \beta' \mathbf{x}_{ij}) + d_{ij}(\alpha_i + \beta' \mathbf{x}_{ij}) - \log(d_{ij}!) \tag{2}$$

where the α_is are additional parameters per individual will yield exactly the same estimates for β and the same asymptotic variance–covariance matrix as the maximization

 st0069

of the conditional logit likelihood in (1). Because the Poisson log likelihood is a less complex function, in some particular settings it may be simpler to maximize (2) instead of (1), despite the penalty incurred by the additional number of parameters that need to be estimated.

In this article, we show how to manipulate the data in Stata 8 to implement this transformation. Because this approach relies on a simple set of rules, it is easy to understand. Researchers may be interested in learning this relation for two reasons: it may simplify the estimation problem at hand, and it may help interpretation of results derived from different approaches.

2 Implementing the multinomial-Poisson transformation

To fit the conditional logit model in Stata, one needs to lay out the data (ignoring identifier variables) in a particular way

$$\mathbf{A} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{X}_1 \\ \mathbf{y}_2 & \mathbf{X}_2 \\ \dots & \dots \\ \mathbf{y}_N & \mathbf{X}_N \end{bmatrix}$$

where each (block) row is a group of observations, such that $\mathbf{y}_i$ is a vector with one element set to one and the remaining equal to zero and $\mathbf{X}_i$ is a matrix containing the different covariates for each choice (see [R] **clogit**). The multinomial-Poisson transformation discussed above implies that if we expand our set of covariates to include a dummy variable per group, as in

$$\mathbf{B} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{X}_1 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{y}_2 & \mathbf{X}_2 & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{y}_N & \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \end{bmatrix}$$

then a Poisson regression of the $\mathbf{y}$ variable on all other covariates will yield the same coefficients for the $\mathbf{X}$ variables as the conditional logit model. In this case, the Poisson regression requires the estimation of an additional $N-1$ coefficients and the use of a substantially larger dataset (note that one dummy would have to be dropped for identification purposes). However, if the information for the $\mathbf{X}$ covariates is identical for groups of observations, the data may be collapsed before applying the Poisson regression. To see how this can be done, consider the situation where $\mathbf{X}_1 = \mathbf{X}_2$. In this case, we could condense the submatrix consisting of the first two (block) rows of $\mathbf{B}$ using a single dummy variable to represent both groups, as in

$$\mathbf{B1} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{X}_1 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{y}_2 & \mathbf{X}_1 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$

or, even better, dropping one group and modifying the dependent variable to be the sum of the dependent variables for both groups:

$$\mathbf{B2} = \begin{bmatrix} \mathbf{y}_1 + \mathbf{y}_2 & \mathbf{X}_1 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$$

Now if we apply a Poisson regression to the data matrix **B** where the first two (block) rows are replaced by either **B1** or **B2**, we still obtain the same estimates for the coefficients on the **X** variables, but we will be using a smaller dataset. We can extend this logic to collapse all groups with identical sets of covariates **X**, and depending on the type of data, we could possibly achieve a substantial reduction in the size of the data matrix. Note that we can also "travel" backwards and transform any Poisson regression into a conditional logit model. In this latter case, we could possibly use dummy variables in the set of covariates of the Poisson regression to define artificial choice sets and, thus, establish a relation with alternative conditional logit model specifications. We next provide several examples of applications of the multinomial-Poisson transformation using online datasets from the *Stata 8 Reference Manuals*.

3 Examples

3.1 Example 1

Consider the dataset used in the *Stata Base Reference Manual* to illustrate the use of the Poisson regression command. The data are loaded with the command

```
. use http://www.stata-press.com/data/r8/airline, clear
```

and we can run the following Poisson regression:

```
. poisson injuries n XYZowned, nolog
Poisson regression                                Number of obs   =          9
                                                  LR chi2(2)      =      18.46
                                                  Prob > chi2     =     0.0001
Log likelihood = -22.681583                       Pseudo R2       =     0.2892

------------------------------------------------------------------------------
    injuries |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |   11.61203   2.971555     3.91   0.000     5.787893    17.43617
    XYZowned |    .650753   .3878742     1.68   0.093    -.1094665    1.410973
       _cons |   .2992084   .5131932     0.58   0.560    -.7066318    1.305049
------------------------------------------------------------------------------
```

We can expand the data to be as in matrix **A**:

```
. expand _N
(72 observations created)
. by airline, sort: gen choiceid=_n
. gen y=(choiceid==airline)
. gsort choiceid -y
. by choiceid: gen weight=sum(y*injuries)
```

If we now run a Poisson regression on this dataset and include dummy variables for each block row of the data, we obtain

```
. xi: poisson y n XYZowned i.choiceid [fweight=weight], nolog
i.choiceid        _Ichoiceid_1-9      (naturally coded; _Ichoiceid_1 omitted)
Poisson regression                                Number of obs   =        576
                                                  LR chi2(10)     =      18.46
                                                  Prob > chi2     =     0.0477
Log likelihood = -195.39419                       Pseudo R2       =     0.0451

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |   11.61203   2.971555     3.91   0.000     5.787893    17.43617
    XYZowned |    .650753   .3878742     1.68   0.093    -.1094665    1.410973
_Ichoiceid_2 |  -4.77e-17   .4834938    -0.00   1.000    -.9476304    .9476304
_Ichoiceid_3 |  -6.28e-17   .4834938    -0.00   1.000    -.9476304    .9476304
_Ichoiceid_4 |   6.38e-17   .3788676     0.00   1.000    -.7425669    .7425669
_Ichoiceid_5 |  -1.57e-16   .4494666    -0.00   1.000    -.8809383    .8809383
_Ichoiceid_6 |   9.63e-17   .5838742     0.00   1.000    -1.144372    1.144372
_Ichoiceid_7 |   1.73e-17   .6513389     0.00   1.000    -1.276601    1.276601
_Ichoiceid_8 |   9.03e-17   1.044466     0.00   1.000    -2.047116    2.047116
_Ichoiceid_9 |   6.50e-17   .6513389     0.00   1.000    -1.276601    1.276601
       _cons |  -3.859675   .5819376    -6.63   0.000    -5.000251   -2.719098
------------------------------------------------------------------------------
```

Despite now having 576 observations, we still obtain the same results for the variables `n` and `XYZowned`. But there is no need to add the dummy variables because the $\mathbf{X_i}$ are identical for all blocks of data:

```
. poisson y n XYZowned [fweight=weight], nolog
Poisson regression                                Number of obs   =        576
                                                  LR chi2(2)      =      18.46
                                                  Prob > chi2     =     0.0001
Log likelihood = -195.39419                       Pseudo R2       =     0.0451

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |   11.61203   2.971555     3.91   0.000     5.787893    17.43617
    XYZowned |    .650753   .3878742     1.68   0.093    -.1094665    1.410973
       _cons |  -3.859675   .5131932    -7.52   0.000    -4.865515   -2.853835
------------------------------------------------------------------------------
```

Given that the data are as in matrix $\mathbf{A}$, we can obtain the same estimates for the coefficients of the variables `n` and `XYZowned` by means of a conditional logit model:

```
. clogit y n XYZ [fweight=weight], group(choiceid) nolog
Conditional (fixed-effects) logistic regression   Number of obs   =        576
                                                  LR chi2(2)      =      18.46
                                                  Prob > chi2     =     0.0001
Log likelihood = -131.39419                       Pseudo R2       =     0.0656

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           n |   11.61203   2.971555     3.91   0.000     5.787892    17.43618
    XYZowned |    .650753   .3878742     1.68   0.093    -.1094665    1.410973
------------------------------------------------------------------------------
```

Suppose now that our interest was restricted to the coefficient on the variable `n`. In this case, we can use the different levels of the variable `XYZowned` to define artificial choice sets for the conditional logit model. We can load the dataset again

```
. use http://www.stata-press.com/data/r8/airline, clear
```

and reshape it to make it look like matrix **A**:

```
. by XYZowned, sort: gen nchoice=_N
. by XYZowned: gen choiceid=_n
. expand nchoice
(36 observations created)
. by XYZowned choiceid,sort: gen groupid=_n
. egen group2id=group(XYZowned groupid)
. gen y=(choiceid==groupid)
. gsort XYZowned group2id -y
. by XYZowned group2id: gen weight=sum(y*injuries)
. sort XYZowned groupid choiceid
```

Now, whether we apply the Poisson regression

```
. poisson y n XYZowned [fweight=weight], nolog
Poisson regression                                Number of obs   =        330
                                                  LR chi2(2)      =      23.21
                                                  Prob > chi2     =     0.0000
Log likelihood = -157.36986                       Pseudo R2       =     0.0687
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
n	11.61203	2.971555	3.91	0.000	5.787893	17.43617
XYZowned	1.589023	.3878742	4.10	0.000	.8288031	2.349242
_cons	-3.529433	.5131932	-6.88	0.000	-4.535273	-2.523593

or fit the discrete-choice model using `XYZowned` to define the choice set

```
. clogit y n XYZ [fweight=weight], group(group2id)
note: XYZowned omitted due to no within-group variance.
Iteration 0:   log likelihood = -99.602856
Iteration 1:   log likelihood = -93.435036
Iteration 2:   log likelihood = -93.369897
Iteration 3:   log likelihood = -93.369864
Conditional (fixed-effects) logistic regression   Number of obs   =        330
                                                  LR chi2(1)      =      17.65
                                                  Prob > chi2     =     0.0000
Log likelihood = -93.369864                       Pseudo R2       =     0.0864
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
n	11.61203	2.971554	3.91	0.000	5.787895	17.43617

we still obtain the same estimates for the coefficient (and standard deviation) on `n`.

3.2 Example 2

Consider now the following dataset, which was used as an example for the `clogit` command in the *Stata Base Reference Manual*

```
. use http://www.stata-press.com/data/r8/choice, clear
```

and fit the following discrete-choice model:

```
. clogit choice dealer, group(id) nolog
Conditional (fixed-effects) logistic regression   Number of obs   =        885
                                                  LR chi2(1)      =     126.68
                                                  Prob > chi2     =     0.0000
Log likelihood = -260.74988                       Pseudo R2       =     0.1954

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dealer |   .0962276   .0090957    10.58   0.000     .0784004    .1140548
------------------------------------------------------------------------------
```

Because this data is already arranged as in matrix **A**, we can apply a Poisson regression, as long as we introduce a dummy variable for each one of the choices:

```
. xi: poisson choice dealer i.id, nolog
i.id              _Iid_1-295          (naturally coded; _Iid_1 omitted)
Poisson regression                                Number of obs   =        885
                                                  LR chi2(295)    =     126.68
                                                  Prob > chi2     =     1.0000
Log likelihood = -555.74988                       Pseudo R2       =     0.1023

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dealer |   .0962276   .0090957    10.58   0.000     .0784004    .1140548
      _Iid_2 |   .1486226   1.414251     0.11   0.916    -2.623258    2.920503
  (output omitted)
    _Iid_295 |  -.0375944     1.4143    -0.03   0.979    -2.809571    2.734382
       _cons |  -2.243873   1.007488    -2.23   0.026    -4.218513   -.2692327
------------------------------------------------------------------------------
```

This amounts to adding 295 dummy variables. However, the data can be collapsed if there are blocks of data identical across choices. We can reduce the dataset in the following manner

```
. by id: gen nn=_n
. by id: gen unique=dealer[1]+100*dealer[2]+10000*dealer[3]
. collapse (sum) choice, by (nn unique dealer)
. by nn: gen nnn=_n
```

and we can now apply a Poisson regression to the collapsed dataset

```
. xi: poisson choice dealer i.nnn, nolog
i.nnn            _Innn_1-34         (naturally coded; _Innn_1 omitted)

Poisson regression                                Number of obs   =        102
                                                  LR chi2(34)     =     143.46
                                                  Prob > chi2     =     0.0000
Log likelihood =  -309.7318                       Pseudo R2       =     0.1880

------------------------------------------------------------------------------
      choice |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dealer |   .0962276   .0090957    10.58   0.000     .0784004    .1140548
     _Innn_2 |  -.3251537   .5012864    -0.65   0.517    -1.307657    .6573496
  (output omitted)
    _Innn_34 |  -.8581789   .5092949    -1.69   0.092    -1.856379    .1400208
       _cons |    .277042   .3621829     0.76   0.444    -.4328234    .9869073
------------------------------------------------------------------------------
```

and still obtain the same estimate for the coefficient associated with the variable `dealer`.

3.3 Example 3

Our final example illustrates a situation in which the multinomial-Poisson relation is particularly useful. Let us load the dataset used to illustrate the application of the `xtpoisson` command in the *Stata Cross-Sectional Time-Series Reference Manual*:

```
. use http://www.stata-press.com/data/r8/ships, clear

. drop if accident==.|op_75_79==.|co_65_69==.|co_70_74==.|co_75_79==.
(6 observations deleted)
```

We can estimate a fixed-effect Poisson regression by adding dummy variables for each "individual", as in

```
. xi: poisson accident op_75_79 co_65_69 co_70_74 co_75_79 i.ship, nolog
i.ship           _Iship_1-5         (naturally coded; _Iship_1 omitted)

Poisson regression                                Number of obs   =         34
                                                  LR chi2(8)      =     475.45
                                                  Prob > chi2     =     0.0000
Log likelihood = -118.47588                       Pseudo R2       =     0.6674

------------------------------------------------------------------------------
    accident |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    op_75_79 |   .2928003   .1127466     2.60   0.009      .071821    .5137796
    co_65_69 |   .5824489   .1480547     3.93   0.000     .2922671    .8726308
    co_70_74 |   .4627844    .151248     3.06   0.002     .1663437    .7592251
    co_75_79 |  -.1951267   .2135749    -0.91   0.361    -.6137258    .2234724
    _Iship_2 |    1.79572   .1666196    10.78   0.000     1.469151    2.122288
    _Iship_3 |  -1.252763   .3273268    -3.83   0.000    -1.894312   -.6112142
    _Iship_4 |  -.9044563   .2874597    -3.15   0.002    -1.467867   -.3410457
    _Iship_5 |  -.1462833   .2351762    -0.62   0.534    -.6072202    .3146537
       _cons |   1.308451   .1972718     6.63   0.000     .9218049    1.695096
------------------------------------------------------------------------------
```

or we can use the conditional fixed-effects estimator of Hausman, Hall, and Griliches (1984):

```
. xtpoisson accident op_75_79 co_65_69 co_70_74 co_75_79, i(ship) fe nolog
Conditional fixed-effects Poisson regression    Number of obs      =        34
Group variable (i): ship                        Number of groups   =         5

                                                Obs per group: min =         6
                                                               avg =       6.8
                                                               max =         7

                                                Wald chi2(4)       =     30.48
Log likelihood  = -104.83697                    Prob > chi2        =    0.0000
```

accident	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
op_75_79	.2928003	.1127466	2.60	0.009	.071821	.5137796
co_65_69	.5824489	.1480547	3.93	0.000	.2922671	.8726308
co_70_74	.4627844	.151248	3.06	0.002	.1663437	.7592251
co_75_79	-.1951267	.2135749	-0.91	0.361	-.6137258	.2234724

However, we can think of all the "dummy" variables for individuals as defining artificial choice sets. Rearranging the data to be as matrix **A**, we can make the data suitable for application of the conditional logit estimator, thus avoiding the estimation of the coefficients for all the "individual" dummy variables. The data are transformed by doing

```
. sort ship
. by ship: gen nchoice=_N
. by ship: gen choiceid=_n
. expand nchoice
(198 observations created)
. by ship choiceid, sort: gen groupid=_n
. egen group2id=group(ship groupid)
. gen y=(choiceid==groupid)
. gsort ship group2id -y
. by ship group2id: gen weight=sum(y*accident)
. sort ship groupid choiceid
. drop if weight==0
(55 observations deleted)
```

We are now ready to apply the conditional logit estimator:

```
. clogit y op_75_79 co_65_69 co_70_74 co_75_79 [fweight=weight],
> group(group2id) nolog
Conditional (fixed-effects) logistic regression   Number of obs   =       2460
                                                  LR chi2(4)      =      31.62
                                                  Prob > chi2     =     0.0000
Log likelihood = -671.99869                       Pseudo R2       =     0.0230
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
op_75_79	.2928003	.1127466	2.60	0.009	.071821	.5137796
co_65_69	.5824489	.1480547	3.93	0.000	.2922671	.8726308
co_70_74	.4627844	.151248	3.06	0.002	.1663437	.7592251
co_75_79	-.1951267	.2135749	-0.91	0.361	-.6137258	.2234724

4 Conclusion

In this paper, we have shown how the multinomial-Poisson transformation may be used in practice. We have used several examples to show the data manipulations required in Stata to estimate equivalent Poisson and logit regressions. Understanding the data arrangement required for each model helps students and practitioners identify which equivalent approaches may be used for estimation and, thus, whether it is advantageous to consider the application of this transformation.

5 References

Baker, S. G. 1994. The multinomial-Poisson transformation. *The Statistician* 43: 495–504.

Guimarães, P., O. Figueiredo, and D. Woodward. 2003. A tractable approach to the firm location decision problem. *Review of Economic and Statistics* 85(1): 201–204.

Hausman, J., B. Hall, and Z. Griliches. 1984. Economic models for count data with an application to the patents-R&D relationship. *Econometrica* 52: 909–938.

Lang, J. 1996. On the comparison of multinomial and Poisson log-linear models. *Journal of the Royal Statistical Society*, Series B, 58: 253–266.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, ed. P. Zerembka, 105–142. New York: Academic Press.

Palmgren, J. 1981. The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika* 68: 563–566.

About the Author

Paulo Guimarães is a research assistant professor at the Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina. He holds a PhD in Economics and a PostDoc in Biostatistics. He taught previously at the University of Porto and Minho (Portugal).