0.1 `netpoisson`: Network Poisson Regression for Event Count Proximity Matrix Dependent Variables

Use network Poisson regression analysis for a dependent variable that represents the number of events that occur during a fixed period of time as a proximity matrix (a.k.a. sociomatricies, adjacency matrices, or matrix representations of directed graphs).

Syntax

```
> z.out <- zelig(y ~ x1 + x2, model = "netpoisson", data = mydata)
> x.out <- setx(z.out)
> s.out <- sim(z.out, x = x.out)
```

Additional Inputs

In addition to the standard inputs, `zelig()` takes the following additional options for network poisson regression:

- `LF`: specifies the link function to be used for the network poisson regression. Default is `LF="log"`, but `LF` can also be set to `"sqrt"` by the user.

Examples

1. Basic Example

   Load the sample data (see `?friendship` for details on the structure of the network dataframe):

   ```
   > data(friendship)
   ```

   Estimate model:

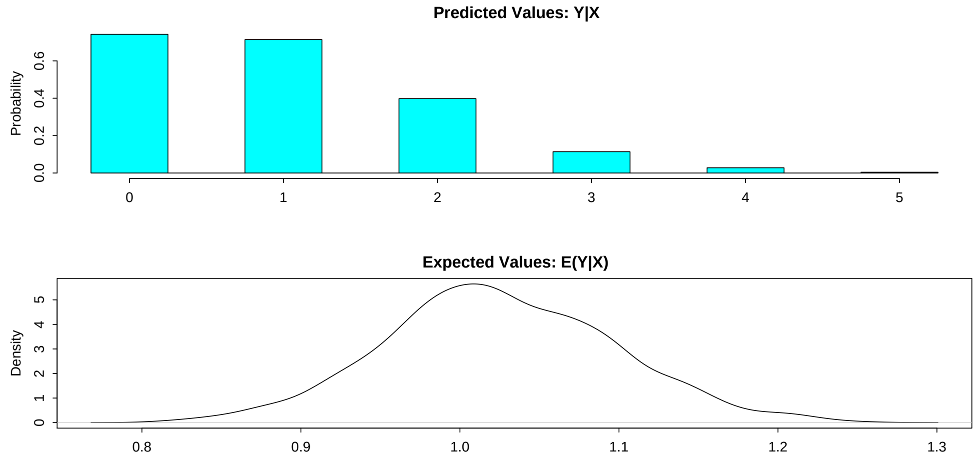
   ```
   > z.out <- zelig(count ~ advice + prestige + perpower, model = "netpoisson",
   +     data = friendship)
   > summary(z.out)
   ```

   Setting values for the explanatory variables to their default values:

   ```
   > x.out <- setx(z.out)
   ```

   Simulate fitted values.

   ```
   > s.out <- sim(z.out, x = x.out)
   > summary(s.out)
   > plot(s.out)
   ```



Model

The `netpoisson` model performs a Poisson regression of the proximity matrix $\mathbf{Y}$, a $m \times m$ matrix representing network ties, on a set of proximity matrices $\mathbf{X}$. This network regression model is directly analogous to standard Poisson regression element-wise on the appropriately vectorized matrices. Proximity matrices are vectorized by creating Y, a $m^2 \times 1$ vector to represent the proximity matrix. The vectorization which produces the Y vector from the $\mathbf{Y}$ matrix is performed by simple row-concatenation of $\mathbf{Y}$. For example, if $\mathbf{Y}$ is a 15×15 matrix, the $\mathbf{Y}_{1,1}$ element is the first element of Y, and the $\mathbf{Y}_{2,1}$ element is the second element of Y and so on. Once the input matrices are vectorized, standard Poisson regression is performed.

Let Y_i be the dependent variable, produced by vectorizing an event count proximity matrix, for observation i. Y_i is thus the number of independent events that occur during a fixed time period. This variable can take any non-negative integer.

- The Poisson distribution has *stochastic component*

$$Y_i \sim \text{Poisson}(\lambda_i),$$

  where λ_i is the mean and variance parameter.

- The *systematic component* is given by:

$$\lambda_i = \exp(x_i \beta).$$

  where x_i is the vector of explanatory variables and β is the vector of coefficients.

Quantities of Interest

The quantities of interest for the network Poisson regression are the same as those for the standard Poisson regression.

- The expected value (`qi$ev`) for the `netpoisson` model is the mean of simulations from the stochastic component,

$$E(Y) = \lambda_i = \exp(x_i \beta),$$

  given draws of β from its sampling distribution.

- The predicted value (`qi$pr`) is a random draw from the Poisson distribution defined by mean λ_i.

- The first difference (`qi$fd`) for the network Poisson model is defined as

$$FD = \Pr(Y|x_1) - \Pr(Y|x)$$

Output Values

The output of each Zelig command contains useful information which you may view. For example, you run `z.out <- zelig(y ~ x, model = "netpoisson", data)`, then you may examine the available information in `z.out` by using `names(z.out)`, see the coefficients by using `z.out$coefficients`, and a default summary of information through `summary(z.out)`. Other elements available through the `$` operator are listed below.

- From the `zelig()` output stored in `z.out`, you may extract:
  - `coefficients`: parameter estimates for the explanatory variables.
  - `fitted.values`: the vector of fitted values for the systemic component λ.
  - `residuals`: the working residuals in the final iteration of the IWLS fit.
  - `linear.predictors`: the vector of $x_i \beta$.
  - `aic`: Akaikeś Information Criterion (minus twice the maximized log-likelihood plus twice the number of coefficients).
  - `bic`: the Bayesian Information Criterion (minus twice the maximized log-likelihood plus the number of coefficients times log n).
  - `df.residual`: the residual degrees of freedom.
  - `df.null`: the residual degrees of freedom for the null model.
  - `zelig.data`: the input data frame if `save.data = TRUE`
- From `summary(z.out)`(as well as from `zelig()`), you may extract:
  - `mod.coefficients`: the parameter estimates with their associated standard errors, p-values, and t statistics.
  - `cov.scaled`: a $k \times k$ matrix of scaled covariances.
  - `cov.unscaled`: a $k \times k$ matrix of unscaled covariances.
- From the `sim()` output stored in `s.out`, you may extract:

- `qi$ev`: the simulated expected probabilities for the specified values of `x`.
- `qi$pr`: the simulated predicted values for the specified values of `x`.
- `qi$fd`: the simulated first differences in the expected probabilities simulated from `x` and `x1`.

How to Cite

To cite the *netpoisson* Zelig model:

> Skyler J. Cranmer. 2007. "netpoisson: Network Poisson Regression for Event Count Proximity Matrix Dependent Variables," in Kosuke Imai, Gary King, and Olivia Lau, "Zelig: Everyone's Statistical Software," `http://gking.harvard.edu/zelig`.

To cite Zelig as a whole, please reference these two sources:

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Zelig: Everyone's Statistical Software," `http://GKing.harvard.edu/zelig`.

> Kosuke Imai, Gary King, and Olivia Lau. 2007. "Toward A Common Framework for Statistical Analysis and Development," `http://gking.harvard.edu/files/abs/z-abs.shtml`.

See also

The network normal regression is part of the `netglm` package by Skyler J. Cranmer and is built using some of the functionality of the `sna` package by Carter T. Butts (Butts and Carley 2001).In addition, advanced users may wish to refer to `help(netpoisson)`. Sample data are fictional.

Bibliography

Butts, C. and Carley, K. (2001), "Multivariate Methods for Interstructural Analysis," Tech. rep., CASOS working paper, Carnegie Mellon University.