Automultinomial Vignette

Stephen Berg

March 27, 2018

This vignette explains the installation and use of the R package automultinomial. The automultinomial package is designed to be used for regressions similar to logistic or multinomial logit regression. However, unlike ordinary logistic and multinomial logit models, the auto-logistic (multinomial) model includes an autocorrelation parameter to account for spatial dependence between observations.

The organization of this document is as follows:

1. Description of the problem automultinomial solves
2. A data example with binary response (2 response categories)
3. A data example with 3 response categories
4. Technical appendix-a comparison of different parameterizations of the 2-category autologistic model

1 The problem automultinomial solves

Consider a data problem where covariates (the independent variables) and categorical responses (the dependent variables) are observed on a spatial grid or lattice. To describe the setup in automultinomial, we will frequently use the notations

- $i = 1, ..., n$ to index the n observation sites on the spatial grid
- $K \geq 2$ to denote the number of possible values taken by the response
- p to denote the number of covariates observed at each site
- $x_i \in \mathbb{R}^p$ to denote the values of the covariates at site i
- $z_i = 1, ..., K$ to denote the categorical response at site i
- $\mathbf{z}$ to denote the vector of categorical responses for all the sites
- z_{-i} to denote the entire response vector $\mathbf{z}$, *except* for site i
- $X = [x_1, x_2, ..., x_n]^T \in \mathbb{R}^{n\times p}$ to denote the design matrix
- $\beta \in \mathbb{R}^{p\times(K-1)}$ to denote the coefficient matrix (when $K = 2$, the $\beta \in \mathbb{R}^p$ is a vector of length p, as in logistic regression)
- β_k to denote the k-th column of *beta*.
- γ to denote the autocorrelation parameter

We will assume that for each site i, a collection N_i of neighboring sites is known. If the data are collected in a square lattice fashion, then a natural neighborhood setup is to set N_i to be the Up, Down, Left, and Right neighbors of site i (sites at the boundary may have less than 4 neighbors). We will assume $i \notin N_i$ for each i. It is also allowable for sites to have no neighbors. In this case $N_i = \phi$, where ϕ denotes the empty set. We will use the notation $i' \sim i$ to indicate that site i' is a neighbor of site i. More formally $i' \sim i \iff i' \in N_i$.

In the automultinomial package, we will require that the neighborhoods are *symmetric*: if $i' \in N_i$, then necessarily $i \in N_{i'}$ as well.

We will also define an *energy* function $H(\mathbf{z}|\beta,\gamma)$. When $K = 2$ and the two categories are $k = 1$ and $k = 2$, then β is just equal to a vector β with length p, and

$$H(\mathbf{z}|\beta,\gamma) = \sum_{i=1}^{n} x_i^T \beta I(z_i = 2) + \gamma \sum_{i=1}^{n} \sum_{i'\sim i, i'>i} \sum_{k=1}^{2} I(z_i = z_{i'} = k) \tag{1}$$

In general, for $K \geq 2$ we define $H(\mathbf{z}|\beta,\gamma)$ by

$$H(\mathbf{z}|\beta,\gamma) = \sum_{i=1}^{n}\sum_{k=1}^{K-1} x_i^T\beta_k I(z_i = k+1) + \gamma\sum_{i=1}^{n}\sum_{i'\sim i, i'>i}\sum_{k=1}^{K} I(z_i = z_{i'} = k) \tag{2}$$

With this preamble accomplished, we can define the probability model automultinomial seeks to estimate:

$$p(\mathbf{z}|\beta,\gamma) = \frac{\exp\{H(\mathbf{z}|\beta,\gamma)\}}{\sum_{\mathbf{z}'}\exp\{H(\mathbf{z}'|\beta,\gamma)\}} \tag{3}$$

The term $\sum_{\mathbf{z}'}$ indicates a sum over all possible categorical responses for the entire dataset.

Some intuition regarding (3) can be gained as follows: when $\gamma = 0$, then $\gamma\sum_{i=1}^{n}\sum_{i'\sim i,i'>i}\sum_{k=1}^{2} I(z_i = z_{i'} = k) = 0$ so that the z_i are completely independent. In this case the autologistic (multinomial) model is the same as an ordinary logistic (multinomial logit) model. On the other hand, when $\gamma > 0$, then response configurations $\mathbf{z}$ where $z_i = z_{i'}$ for many neighbor pairs $i \sim i'$ becomes more likely. In this way, γ incorporates positive spatial correlation into the responses. When $\gamma < 0$, we expect neighboring $z_i, z_{i'}$ to *disagree* more frequently than if the z_i were independent, and the $\gamma < 0$ case will in practice be less common.

Estimating β and γ

The automultinomial package follows the pseudolikelihood approach of (Besag, 1974) to estimate the parameters β, γ. We briefly explain the procedure and its motivation below.

When γ is known to be 0, then the responses z_i are independent, and we can use common maximum likelihood techniques such as logistic or multinomial regression to estimate β. On the other hand, when spatial correlation is present and we use the automultinomial framework to account for it, then we need to estimate β and γ jointly. Unfortunately, maximum likelihood is computationally infeasible due the well known computational intractability of the denominator of (3) when $\gamma \neq 0$.

Fortunately, a proposal of (Besag, 1974) provides a consistent estimation procedure, maximum pseudolikelihood, for the parameters of the model in equation (3). In maximum pseudolikelihood, we maximize

$$\ell_{PL} = \sum_{i=1}^{n}\log\{p(z_i|z_{-i},\beta,\gamma)\} \tag{4}$$

over β and γ. The expression $p(z_i|z_{-i},\beta,\gamma)$ refers to the conditional density of z_i given the sites at all of the other grid locations. A short calculation based on (3) shows that for $K = 2$

$$p(z_i = 1|z_{-i},\beta,\gamma) = \frac{\exp\{\gamma\sum_{i'\sim i} I(z_{i'} = 1)\}}{\exp\{\gamma\sum_{i'\sim i} I(z_{i'} = 1)\} + \exp\{x_i^T\beta I(z_i = 2) + \gamma\sum_{i'\sim i} I(z_{i'} = 2)\}} \tag{5}$$

$$p(z_i = 2|z_{-i},\beta,\gamma) = \frac{\exp\{x_i^T\beta I(z_i = 2) + \gamma\sum_{i'\sim i} I(z_{i'} = 2)\}}{\exp\{\gamma\sum_{i'\sim i} I(z_{i'} = 1)\} + \sum_{k=1}^{K-1}\exp\{x_k^T\beta I(z_i = 2) + \gamma\sum_{i'\sim i} I(z_{i'} = 2)\}} \tag{6}$$

For $K \geq 2$ response categories, we have

$$p(z_i = 1|z_{-i},\beta,\gamma) = \frac{\exp\{\gamma\sum_{i'\sim i} I(z_{i'} = 1)\}}{\exp\{\gamma\sum_{i'\sim i} I(z_{i'} = 1)\} + \sum_{k=1}^{K-1}\exp\{x_k^T\beta_k I(z_i = k+1) + \gamma\sum_{i'\sim i} I(z_{i'} = k+1)\}} \tag{7}$$

and for $k > 1$,

$$p(z_i = k | z_{-i}, \beta, \gamma) = \frac{\exp\{x_i^T \beta_k I(z_i = k+1) + \gamma \sum_{i' \sim i} I(z_{i'} = k+1)\}\}}{\exp\{\gamma \sum_{i' \sim i} I(z_{i'} = 1)\} + \sum_{k=1}^{K-1} \exp\{x_k^T \beta_k I(z_i = k+1) + \gamma \sum_{i' \sim i} I(z_{i'} = k+1)\}} \tag{8}$$

In the automultinomial package, equations (5), (6), (7), and (8) and plugged into (4), and (4) is optimized over β and γ using the function optim.

Data example 1: K=2 response categories

Here, we will demonstrate how to simulate data using automultinomial, how to fit data using automultinomial, and how to analyze the output.

Simulating data

Generating simulated data using the function drawSamples().

```
library(automultinomial2)
#10 predictors
p=10

#n times n grid
n=40

#make grid and adjacency matrix
latticeGraph=igraph::make_lattice(c(n,n))
A=igraph::get.adjacency(latticeGraph)

#set coefficient values
beta=matrix(rnorm(p),ncol=1)*0.3
beta
```

```
##              [,1]
##  [1,]  0.07913056
##  [2,]  0.07541560
##  [3,] -0.14740529
##  [4,] -0.17342691
##  [5,] -0.12233061
##  [6,] -0.77628435
##  [7,] -0.18246107
##  [8,]  0.10971084
##  [9,]  0.01288815
## [10,]  0.12543183
```

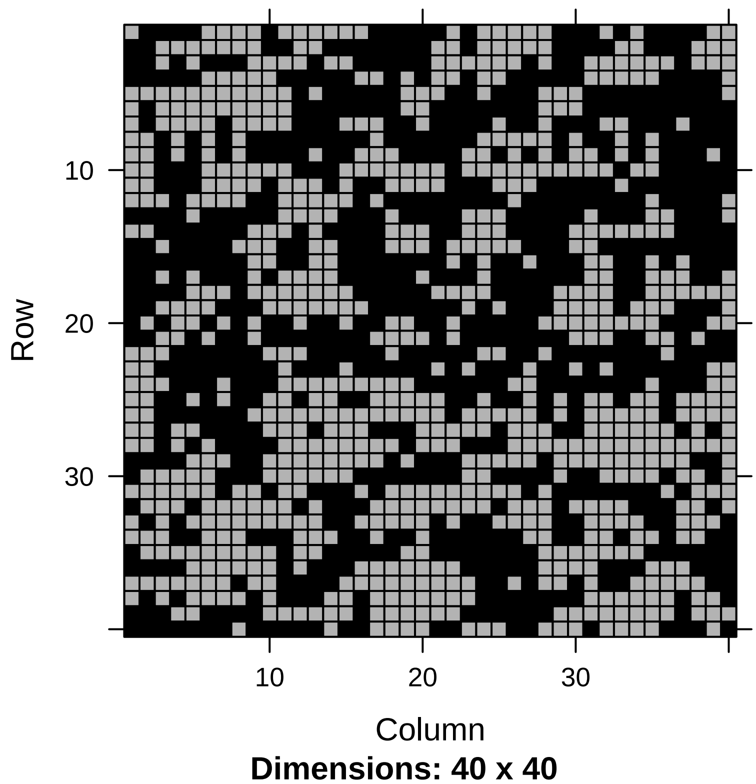
```
#set covariate values
X=matrix(rnorm(n^2*p),ncol=p)

#set the correlation parameter value (0.5 is a moderate amount of spatial correlation)
gamma=0.5

#use drawSamples to simulate data with parameters beta and gamma by Gibbs sampling
y=drawSamples(beta,gamma,X,A,nSamples = 1)
```

```
## Burn-in samples
## 100  burn-in samples so far
## 200  burn-in samples so far
## 300  burn-in samples so far
## Drawing samples
```

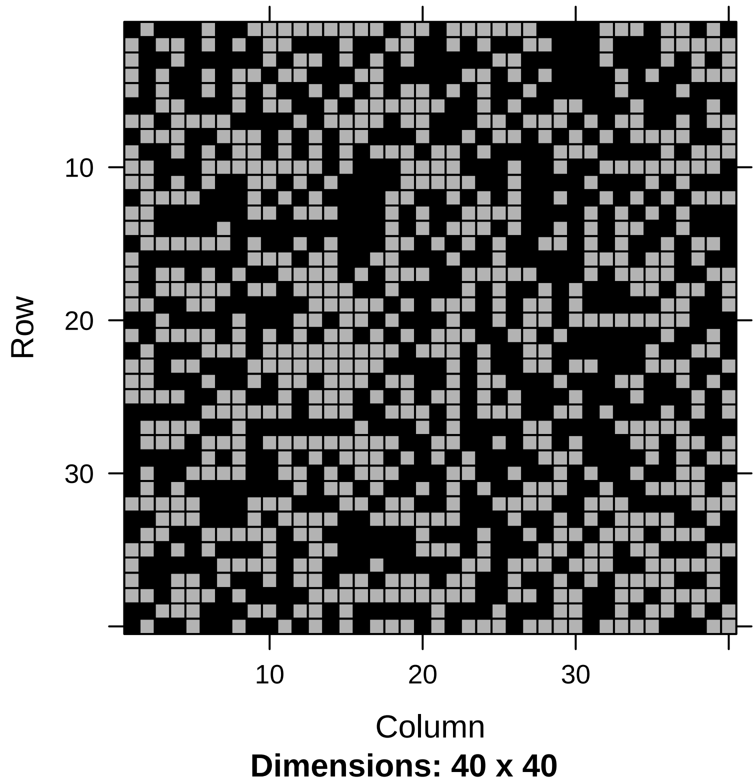
```
## Burn-in samples
## 100  burn-in samples so far
## 200  burn-in samples so far
## 300  burn-in samples so far
## Drawing samples
```

Below is a plot of the responses on the grid. We can see "clumping" of the responses due to the positive autocorrelation parameter γ

```
par(mfrow=c(2,1))
Matrix::image(Matrix::Matrix(matrix(y,ncol=n)))
```




```
Matrix::image(Matrix::Matrix(matrix(y2,ncol=n)))
```



Fitting an autologistic model to the data (K=2 categories)

We fit an autologistic model using the MPLE function. First, we will use confidence intervals based on the asymptotic distribution of the pseudolikelihood estimator.

```
#responses must be input as a factor
y=factor(y)
fit=automultinomial2::MPLE(X=X,y=y,A=A,ciLevel=0.99,method="asymptotic")
```

```
## Starting model fitting
## Model fitting done, starting variance estimation
## Creating asymptotic confidence intervals
```

```
##
##
## Table: Summary with confidence intervals
##
##      2 vs. 1                   gamma
## ---  -----------------------  -------------------
## 1    0.044 (-0.114,0.202)     0.507 (0.41,0.604)
## 2    0.072 (-0.088,0.232)
## 3    -0.134 (-0.289,0.02)
## 4    -0.079 (-0.238,0.08)
## 5    -0.056 (-0.219,0.107)
## 6    -0.698 (-0.878,-0.519)
## 7    -0.109 (-0.266,0.047)
```

```
## 8    0.125 (-0.036,0.287)
## 9    0.063 (-0.099,0.225)
## 10   0.189 (0.03,0.348)
##
##
## Table: Summary with p-values
##
##      2 vs. 1          gamma
## ---  ---------------  ----------
## 1    0.044 (0.472)    0.507 (0)
## 2    0.072 (0.244)
## 3    -0.134 (0.025)
## 4    -0.079 (0.199)
## 5    -0.056 (0.375)
## 6    -0.698 (0)
## 7    -0.109 (0.072)
## 8    0.125 (0.046)
## 9    0.063 (0.319)
## 10   0.189 (0.002)
```