Fast Estimation and Selection of Autologistic Regression Models via Penalized Pseudo-Likelihood

Rao Fu[a], Andrew L. Thurman[b], Michelle M. Steen-Adams[c], Jun Zhu[d]

[a] *Department of Statistics, University of Wisconsin at Madison, Madison, WI 53706*
[b] *Department of Statistics, University of Wisconsin at Madison, Madison, WI 53706*
[c] *Department of Environmental Studies, University of New England, Biddeford, ME 04005*
[d] *Department of Statistics and Department of Entomology, University of Wisconsin at Madison, Madison, WI 53706*

Abstract

XXX

Key words: XXX

1. Introduction

2. Autologistic Regression Models

2.1. Autologistic Regression

2.2. Centered Model

3. Estimation and Selection of Autologistic Models

3.1. Maximum Pseudolikelihood

3.2. Variable Selection

3.3. Variance Estimation

Let $\boldsymbol{\mathcal{J}}(\boldsymbol{\beta}) = \frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \{\frac{\partial \ell_{\mathrm{p}}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\}'$. By arguments similar to Banerjee *et al.* (2004), the following central limit theory holds for the MPLE $\widehat{\boldsymbol{\beta}}_{\mathrm{p}}$:

$$\{\boldsymbol{\mathcal{J}}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}})\}^{-1/2} \boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}}) \{\widehat{\boldsymbol{\beta}}_{\mathrm{p}} - \boldsymbol{\beta}\} \rightarrow_d \mathcal{N}_p(0, \boldsymbol{I}_{p+1}). \quad (1)$$

Email addresses: `rfu7@wisc.edu` (Rao Fu), `athurman@wisc.edu` (Andrew L. Thurman), `msteenadams@une.edu` (Michelle M. Steen-Adams), `jzhu@stat.wisc.edu` (Jun Zhu)

Preprint submitted to Elsevier *October 1, 2012*

Therefore, an estimate of the variance of $\widehat{\boldsymbol{\beta}}_{\mathrm{p}}$ is

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}}) \approx \boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}})^{-1}\boldsymbol{\mathcal{J}}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}})\boldsymbol{\mathcal{I}}(\widehat{\boldsymbol{\beta}}_{\mathrm{p}})^{-1}. \tag{2}$$

For the MPPLE, we replace the MPLE $\widehat{\boldsymbol{\beta}}_{\mathrm{p}}$ in the variance formula (2) with the vector of non-zero entries of the MPPLE $\widehat{\boldsymbol{\beta}}_{\mathrm{pp}}$. The operations involved in the variance estimation are of dimension $(p+1)\times(p+1)$ and generally manageable. We will show by a simulation study that these variance estimates perform reasonably well for finite samples.

In particular under the uncentered model for the 0-1 coding of the response variable, the first-order and second-order derivatives of $\ell_{\mathrm{p}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i} Z_{i'})}{\{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i} Z_{i'})\}^2} \quad \text{and} \\
\boldsymbol{\mathcal{J}}(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{i'\sim i, i'=i} \left[\boldsymbol{x_i}\left\{Z_i - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i} Z_{i'})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i} Z_{i'})}\right\}\right] \\
&\qquad \left[\boldsymbol{x}_{i'}'\left\{Z_{i'} - \frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta} + \eta\sum_{i''\sim i'} Z_{i''})}{1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta} + \eta\sum_{i''\sim i'} Z_{i''})}\right\}\right]'.
\end{aligned}$$

Under the centered model for the 0-1 coding of the response variable, the first-order and second-order derivatives of $\ell_{\mathrm{p}}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\left[\boldsymbol{x}_i - \eta\sum_{i'\sim i}\frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})}{\{1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})\}^2}\boldsymbol{x}_{i'}\right]\left[\boldsymbol{x}_i - \eta\sum_{i'\sim i}\frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})}{\{1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})\}^2}\boldsymbol{x}_{i'}\right]' \\
&\quad \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i}(Z_{i'}-\mu_{i'}))}{\{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i}(Z_{i'}-\mu_{i'}))\}^2} - \sum_{i=1}^{n}\sum_{i'\sim i}\eta\left\{Z_i - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i''\sim i}(Z_{i''}-\mu_{i''}))}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i''\sim i}(Z_{i''}-\mu_{i''}))}\right\} \\
&\quad \frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta}) - \{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})\}^2}{\{1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})\}^3}\boldsymbol{x}_{i'}\boldsymbol{x}_{i'}' \\
\boldsymbol{\mathcal{J}}(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\sum_{i'\sim i, i'=i}\left[\left\{\boldsymbol{x}_i - \eta\sum_{i'\sim i}\frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta})}{(1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta}))^2}\boldsymbol{x}_{i'}\right\}\left\{Z_i - \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i}(Z_{i'}-\mu_{i'}))}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \eta\sum_{i'\sim i}(Z_{i'}-\mu_{i'}))}\right\}\right] \\
&\qquad \left[\left\{\boldsymbol{x}_{i'} - \eta\sum_{i''\sim i'}\frac{\exp(\boldsymbol{x}_{i''}'\boldsymbol{\beta})}{(1+\exp(\boldsymbol{x}_{i''}'\boldsymbol{\beta}))^2}\boldsymbol{x}_{i''}\right\}\left\{Z_{i'} - \frac{\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta} + \eta\sum_{i''\sim i'}(Z_{i''}-\mu_{i''}))}{1+\exp(\boldsymbol{x}_{i'}'\boldsymbol{\beta} + \eta\sum_{i''\sim i'}(Z_{i''}-\mu_{i''}))}\right\}\right]'.
\end{aligned}$$

Under the alternative ± 1 coding, the variance estimation can be obtained in analogy to (2)

$$\widehat{Var}(\hat{\tilde{\boldsymbol{\beta}}}_{\mathrm{p}}) \approx \boldsymbol{\mathcal{I}}(\hat{\tilde{\boldsymbol{\beta}}}_{\mathrm{p}})^{-1}\boldsymbol{\mathcal{J}}(\hat{\tilde{\boldsymbol{\beta}}}_{\mathrm{p}})\boldsymbol{\mathcal{I}}(\hat{\tilde{\boldsymbol{\beta}}}_{\mathrm{p}})^{-1}. \tag{3}$$

Under uncentered model ,the first-order and second-order derivatives of $\ell_{\mathrm{p}}(\tilde{\boldsymbol{\theta}})$ with respect to $\tilde{\boldsymbol{\beta}}$ are

$$\begin{aligned}
\boldsymbol{\mathcal{I}}(\tilde{\boldsymbol{\beta}}) &= \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i' \left[1 - \left\{\frac{\sinh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i'\sim i}\tilde{Z}_{i'})}{\cosh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i'\sim i}\tilde{Z}_{i'})}\right\}^2\right] \\
\boldsymbol{\mathcal{J}}(\tilde{\boldsymbol{\beta}}) &= \sum_{i=1}^{n}\sum_{i'\sim i, i'=i} \left[\boldsymbol{x}_i\left\{\tilde{Z}_i - \frac{\sinh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i'\sim i}\tilde{Z}_{i'})}{\cosh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i'\sim i}\tilde{Z}_{i'})}\right\}\right] \\
&\qquad \left[\boldsymbol{x}_{i'}\left\{\tilde{Z}_{i'} - \frac{\sinh(\boldsymbol{x}_{i'}'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i''\sim i'}\tilde{Z}_{i''})}{\cosh(\boldsymbol{x}_{i'}'\tilde{\boldsymbol{\beta}} + \tilde{\eta}\sum_{i''\sim i'}\tilde{Z}_{i''})}\right\}\right]'.
\end{aligned}$$

4. Simulation Study

4.1. Simulation Set-up

We conducted a simulation study to examine the finite-sample properties of the method developed in Sections 2–3. Consider an $m \times m$ square lattice, where $m = 15$ or 30, corresponding to sample sizes $n = 225$ or 900. For spatial dependence, the neighborhood structure is of the first order and the autoregression coefficient η is either 0.3 or 0.7, corresponding to weaker or stronger spatial dependence.

Let $\boldsymbol{u}_j = (u_{j1}, \ldots, u_{jn})'$ denote the jth covariate vector such that $\{u_{ji} : i = 1, \ldots, n\}$ is a Gaussian random field with mean 0 and an exponential covariance function

$$Cov(u_{ji}, u_{ji'}) \quad = \quad \sigma^2 \exp(-|i - i'|/\tau), \tag{4}$$

where we let the variance parameter be $\sigma^2 = 1$ and the range parameter be $\tau = 0.1$. To obtain cross-covariate correlation, let $\boldsymbol{u}_i = (u_{1i}, \ldots, u_{pi})'$ and $\boldsymbol{x}_i = \boldsymbol{A}\boldsymbol{u}_i$ for site i, where $\boldsymbol{A}\boldsymbol{A}' = [\rho^{|j-j'|}]_{j,j'=1}^{p}$ and $\rho = 0.4$.

Let $p = 10$ be the number of covariates. The regression coefficients is set to be $\boldsymbol{\beta} = (1, \beta_1, 1, 1, \mathbf{0}_7')'$, that is, 3 out of 10 coefficients are non-zero and the remaining 7 coefficients are zero. For the 0-1 coding, we adopt the notion of an average large-scale structure as the average of μ_i over all sites and covariates (Fan and Li, 2001). Let

$$\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} \mu_i = \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_i'\boldsymbol{\beta})} \tag{5}$$

The large-scale structure is considered to be weak when $\bar{\mu}$ is around 0.5 and strong otherwise. For the ± 1 coding, we define the average large-scale structure analogously as

$$\bar{\tilde{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{\mu}_i = \frac{1}{n}\sum_{i=1}^{n} \frac{\sinh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})}{\cosh(\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})} \tag{6}$$

In this case, the large-scale structure is considered to be weak if $\bar{\tilde{\mu}}$ is close to 0 but strong otherwise. Here, we let $\beta_1 = 1$ or 5, which corresponds to a stronger or weaker large-scale structure, respectively.

4.2. Simulation Results

Table 1 provides the results of variable selection for sample size $n = 225$ and 900 in terms of the average numbers of correctly identified zero-valued and non-zero regression coefficients. The true number of non-zero and zero regression coefficients are 3 and 7, respectively.

5. Data Example

Banerjee *et al.* (2004) did this ... The arguments are like this (Banerjee *et al.*, 2004). Equation (2) implies this...

6. Conclusions and Discussion

Acknowledgments

References

Banerjee, S., and Carlin, B.P., and Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall: Boca Raton.

Besag, J. (1972) Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society Series B*, **34**, 75–83.

Diggle, P.J. and Ribeiro, P.J. (2007) *Model-based Geostatistics*, New York: Springer.

Efron, B., Hastie, T., and Johnstone, I. and Tibshirani, R. (2004) Least angle regression (with discussion). *Annals of Statistics*, **32**, 407–499.

Fan, J. and Li, R. (2001) Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

Steen-Adams, M.M. (2005) Change on a Northern Wisconsin Landscape: Legacies of Human History. University of Wisconsin, Madison, Wisconsin, USA.

Table 1: Average number of correctly identified non-zero and zero regression coefficients when $\beta_1 = 5$ (weak large-scale structure) for uncentered and centered model, sample size $n = 225$ or 900 and antoregression coefficient $\eta = 0.3$ or 0.7.

	$\{\beta_j\}$	Number of non-zero estimates		Number of zero estimates	
Model	n	$\eta = 0.3$	$\eta = 0.7$	$\eta = 0.3$	$\eta = 0.7$
Uncentered	255	2.77	2.69	6.12	6.14
	900	3.00	3.00	6.81	6.88
Centered	255	2.75	2.67	6.21	6.17
	900	3.00	3.00	6.87	6.82

Figure 1: Box plot of the MPPLE $\widehat{\beta}_0$ (row 1), $\widehat{\beta}_1$ (row 2), $\widehat{\beta}_2$ (row 3) and $\widehat{\eta}$ (row 4) from the 100 simulations with sample size $n = 225$. Column (a): uncentered model with $\beta_1 = 5$ (weak large-scale structure); (b): uncentered model with $\beta_1 = 1$ (strong large-scale structure); (c): centered model with $\beta_1 = 5$ (weak large-scale structure); (d): centered model with $\beta_1 = 1$ (strong large-scale structure). The true values are $\beta_0 = 1$, $\beta_1 = 1$ or 5, $\beta_2 = 1$, $\beta_3 = 1$, and $\eta =$ 0.3 or 0.7. The box plot of $\widehat{\beta}_3$ is similar to $\widehat{\beta}_2$ but omitted to save space.