Better Autologistic Regression

Mark Wolters

Shanghai Center for
Mathematical Sciences

Fudan University

ICSA-Canada Symposium, Vancouver
August 19, 2017

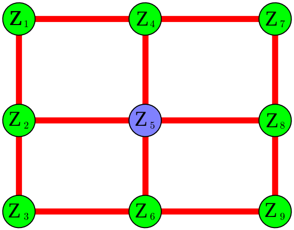
Overview

What is the talk about?

- When you do ***autologistic regression***,
  you must make certain ***implementation decisions***.
- These decisions ***seem trivial***,
  but they are actually ***very important***.
- This fact is ***not explored in the literature*** to date.
- The ***best version*** of the model is ***not*** the one commonly used.

The talk is based on a paper (hopefully) soon to appear in
Frontiers in Applied Mathematics and Statistics

The Autologistic Model



Let $\mathbf{Z}$ be a vector of dichotomous random variables.

Autologistic model (Besag JRSSB 1972, 1974):

- It's a ***Markov Random Field***
- An ***undirected graph***, Adjacency matrix $\mathbf{A}$
- PMF:

$$f_{\mathbf{Z}}(\mathbf{z}) \propto \exp\left(\underbrace{\mathbf{z}^T\boldsymbol{\alpha}}_{\text{unary term}} + \underbrace{\frac{1}{2}\mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z}}_{\text{pairwise term}}\right)$$

- Conditional logit, $\pi_i = \Pr(Z_i = \text{high} \mid \text{neighbours})$:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha_i + \sum_{j\sim i} \lambda_{ij} z_j$$

- Let $\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta}$ ➡ autologistic *regression* (ALR)
- Let $\boldsymbol{\Lambda} = \lambda\mathbf{A}$ ➡ "simple" form of the model

other names

Ising model

QEB distribution

Boltzmann machine

Three ALR models: ***traditional***, ***centered***, and ***symmetric***.

1. The TRADITIONAL model: $\mathbf{Z} \in \{0,1\}^n$

general form	simple form
$f_{\mathbf{Z}}(\mathbf{z}) \propto \exp(\mathbf{z}^T\mathbf{X}\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z})$	$f_{\mathbf{Z}}(\mathbf{z}) \propto \exp(\mathbf{z}^T\mathbf{X}\boldsymbol{\beta} + \frac{\lambda}{2}\mathbf{z}^T\mathbf{A}\mathbf{z})$
$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} + \sum_{j\sim i}\lambda_{ij}z_j$	$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} + \lambda\sum_{j\sim i}z_j$

linear predictor:
n variables, p coefficients

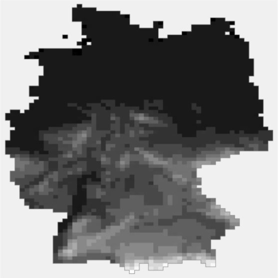
Application areas:

- Spatial binary data
- Image segmentation
- Graph– or network– structured data

Example 1

H. vulgaris data
(Carl & Kühn, 2007, Ecological Modeling; Bardos et al. 2015 arXiv)

$\mathbf{z}$:
presence/absence



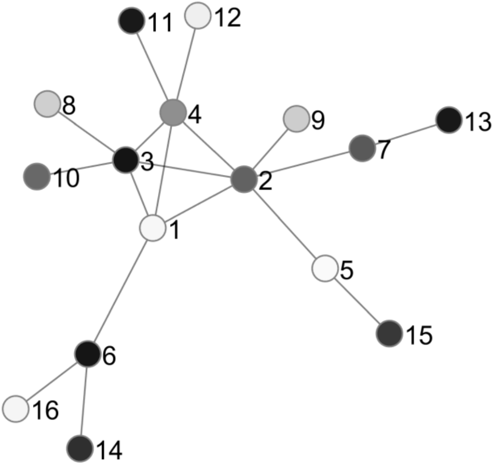
$\mathbf{x}_1$:
altitude

$\Pr(Z_i = 1 | x_i)$,
logistic regression

Example 2

Network regression, preferential attachment models (two cases)

Node shading =
marginal probability



1st Decision: Centering

- Traditional model has a problem
  - Fix β, increase λ, you will find $Z = 1$ everywhere.
  - Why? Because $\sum_{j \sim i} z_j$ is never negative.
- Caragea & Kaiser (2009, JABES) "***centered parametrization***"

2. The CENTERED model (Z $\in \{0, 1\}^n$)

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{j \sim i} \lambda_{ij}(z_j - \mu_j), \quad \text{where} \quad \mu_j = \frac{e^{\mathbf{x}_j^T \boldsymbol{\beta}}}{1 - e^{\mathbf{x}_j^T \boldsymbol{\beta}}}$$

- μ_j is the independence expectation of the Z_j

2nd Decision: Coding

- The responses are ***categorical***. Don't have to use $\{0,1\}$ coding.
  - Statisticians: $\{0,1\}$
  - Ising model (physics): $\{-1,+1\}$
  - Image processing: either $\{0,1\}$ or $\{-1,1\}$
- In general, could use $\{\ell, h\}$.
- If $\mathbf{Z} \in \{\ell, h\}^n$, then $\left(\frac{H-L}{h-\ell}\right)(\mathbf{Z} - \ell\mathbf{1}) + L\mathbf{1} \in \{L, H\}^n$
- But autologistic models with different codings are obtained by ***plugging different numbers into the same PMF***.

3. The SYMMETRIC model

- Just the standard model, with $\mathbf{Z} \in \{-h, h\}^n$
- No centering
- Coding symmetric around 0

Some Questions

A ***variant***: A specific combination of coding and centering choices.

- All variants have independence when $\mathbf{\Lambda} = \mathbf{0}$
- Natural interpretation: a trade-off between individual and neighbourhood effects as λ_{ij}'s increase.
  - $\mathbf{x}_i^T \boldsymbol{\beta}$ controls the "endogenous" tendency of Z_i
  - Larger $\lambda_{ij} \longleftrightarrow$ more likely $Z_i = Z_j$

Some questions

Q: Are all of the variants equivalent?

Q: Do they all adhere to the natural interpretation?

Q: If differences exist, does it matter?

Result: Model Equivalence

equivalent: parameter settings always exist that give the same PMF under two variants.

Are autologistic models equivalent?

- ***yes***

Are autologistic *regression* models equivalent?

- Centered and standard models: ***no***
- Centered models, different codings: ***no***
- Symmetric models, different h values: ***yes***

➡ Many variants, all called "autologistic regression models," are ***actually different, non-nested distribution families***.

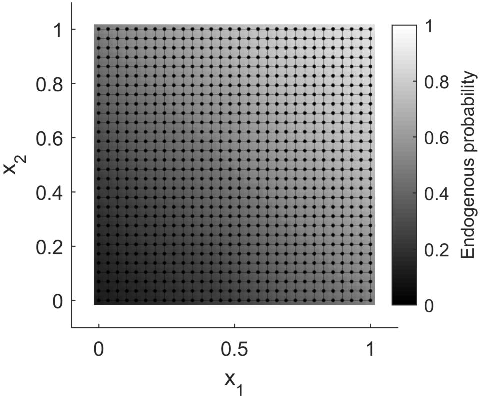
Result: Limiting Behaviour, Parameter Interpretation

"Simple" model. Let λ increase.

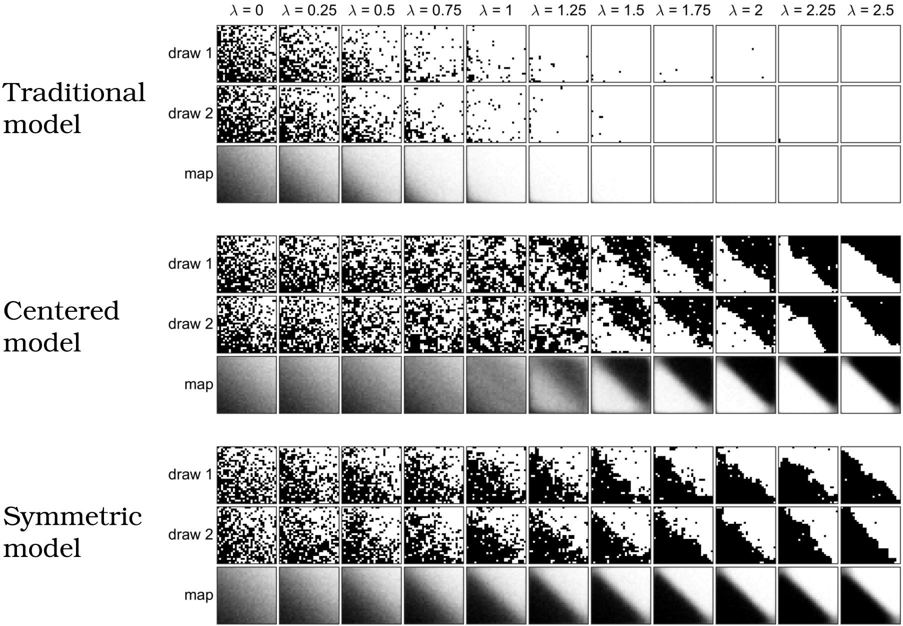
- Centered variants behave ***counterintuitively*** when λ large.
- Symmetric variants are the ***only ones*** with reasonable behaviour as $\lambda \to \infty$.

Example:

- Two predictors + intercept
- Predictors are spatial coordinates
- Square lattice
- $\boldsymbol{\beta} = (-2, 2, 2)^T$



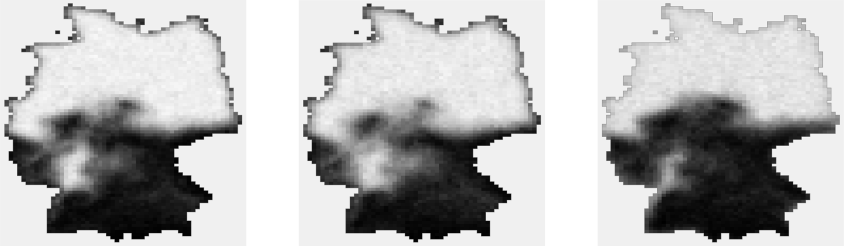
Result: Limiting Behaviour, Parameter Interpretation



Does it Matter?

H. vulgaris fitted models

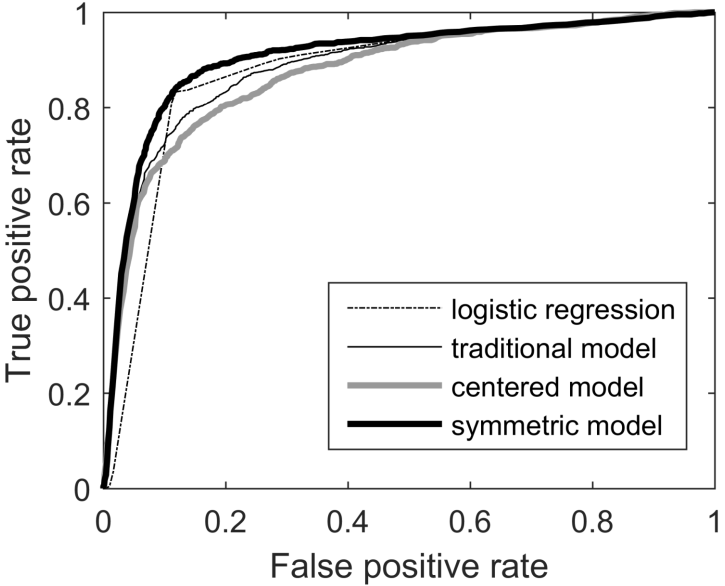
traditional centered symmetric



	β_0 (intercept)		β_1 (altitude)		λ (association)	
Model	$\hat{\beta}_0$ (SE)	impact	$\hat{\beta}_1$ (SE)	impact	$\hat{\lambda}$ (SE)	impact
logistic	2.78 (0.10)	0.37	−0.79 (0.028)	0.48	–	–
traditional	−2.12 (0.22)	0.44	−0.16 (0.026)	0.39	1.43 (0.066)	0.48
centered	−1.74 (0.31)	0.34	−0.17 (0.040)	0.34	1.51 (0.050)	0.47
symmetric	0.50 (0.11)	0.40	−0.13 (0.029)	0.44	1.43 (0.071)	0.27

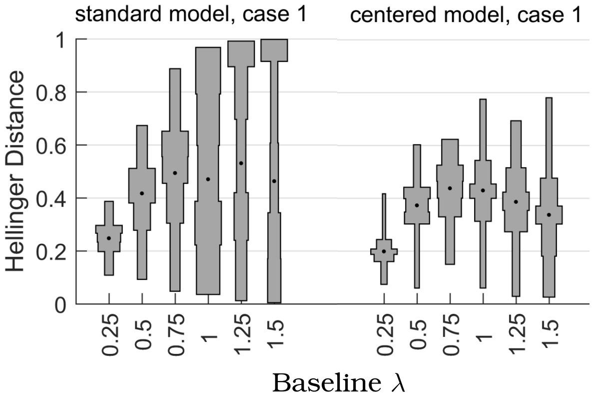
Does it Matter?

H. vulgaris
ROC curves



Does it Matter?

- Network regression example, $n = 16$
- Linear predictor: $\beta_0 + \beta_1 x_i$, with $x_i \sim N(0, 1)$
- Baseline model: symmetric model, $\boldsymbol{\beta} = [0 \quad 1]^T$, fixed λ.
- Find the traditional & centered models with ***minimum Hellinger distance to the baseline***.



Recommendations

- The symmetric model, with $Z_i \in \{-h, h\}$:
  - Is the only one that's easy to interpret
  - Is the only one without pathologies

  We should use it unless there's a good reason to do otherwise.

- There's no reason to use centering
  - Changing the coding resolves the problem with the standard model, in a simpler way.

- If you still want Bernoulli RVs:
  - Start with symmetric model, $\mathbf{Z} \in \{-h, h\}^n$
  - Let $\mathbf{Y} = \frac{1}{2h}\mathbf{Z} + \frac{1}{2}\mathbf{1}$, do proper transformation of variables
  - You will get $Y_i \in \{0, 1\}$, and

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \mathbf{x}_i^T \boldsymbol{\gamma} + \sum_{j \sim i} \omega_{ij} \left(y_j - \frac{1}{2}\right)$$