On Coding and Centering in the Autologistic Regression Model

Mark Wolters
Shanghai Center for Mathematical Sciences
Fudan University

IMS-APRM meeting
The Chinese University of Hong Kong

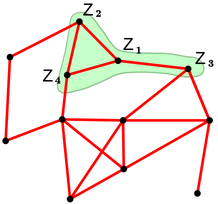
June 30, 2016

The talk in one slide

- Autologistic regression extends logistic regression to dependent responses.
- It's based on the autologistic model, a.k.a.:
  - Ising model
  - Boltzman machine
  - Quadratic exponential binary distribution.
- Physicists use $\{-1, 1\}$ coding.
- Statisticians us $\{0, 1\}$ coding.
- The physicists are right.
- If you've used autologistic regression, you were probably doing it wrong.

The autologistic model

The model with 0, 1 coding



It's a **Markov random field** model for binary random vector $\mathbf{Z}$.

Can be expressed in different ways:

- Joint PMF: $\Pr(\mathbf{Z} = \mathbf{z}) \propto \exp\left(Q(\mathbf{z})\right)$, where

$$\begin{aligned} Q(\mathbf{z}) &= \sum_{i \in \mathcal{V}} \alpha_i z_i + \sum_{(i,j) \in \mathcal{E}} \lambda_{ij} z_i z_j \\ &= \mathbf{z}^T \boldsymbol{\alpha} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\Lambda} \mathbf{z} \end{aligned}$$

pairwise parameter

unary parameter

- Conditional logit:

$$\text{logit}\left(\Pr(Z_i = 1 | \mathbf{Z}_{-i})\right) = \alpha_i + \sum_{j \sim i} \lambda_{ij} z_j$$

Note: often let $\boldsymbol{\Lambda} = \lambda \mathbf{A}$, where $\mathbf{A}$ is adjacency matrix.

Autologistic regression

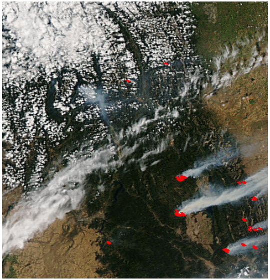
If Z_i is observed with covariates $\mathbf{x}_i$, let $\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta}$.

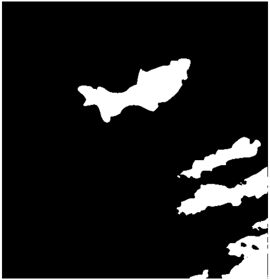
$$Q(\mathbf{z}) = \mathbf{z}^T\mathbf{X}\boldsymbol{\beta} + \frac{1}{2}\mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z}$$

$$\operatorname{logit}\left(\Pr(Z_i = 1|\mathbf{Z}_{-i})\right) = \mathbf{x}_i^T\boldsymbol{\beta} + \sum_{j\sim i}\lambda_{ij}z_j$$

Broadly applicable as an **extension of logistic regression**.

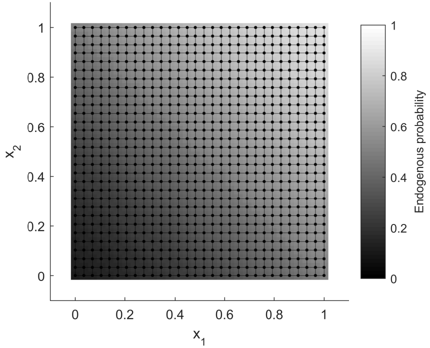
- Ecological modelling (spatial binary data)
- Image processing
- Dentistry
- ...





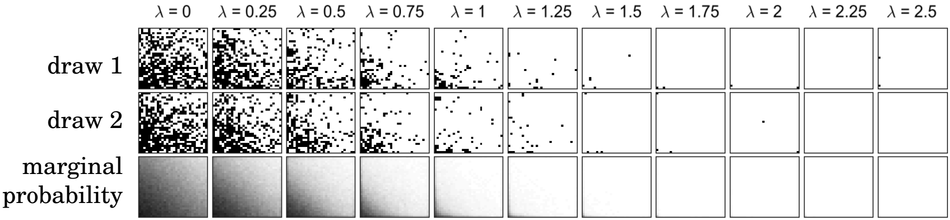
The centered autologistic model

Asymmetry of the standard 0, 1 model



- $n = 900$
- Regular grid on the unit square.
- Z_i's coordinates are (x_{i1}, x_{i2})
- Linear predictors are
$$\alpha_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$
- Set $\boldsymbol{\beta} = (-2, 2, 2)^T$
- Use $\mathbf{\Lambda} = \lambda \mathbf{A}$

Results of **perfect sampling** vs. λ:



The centered model (still coded 0, 1)

Standard $\{0,1\}$ model has $\boldsymbol{\beta}$ and λ strongly coupled.

Hard to interpret $\boldsymbol{\beta}$.

Caragea and Kaiser (2009) proposed a *centered* “parametrization” of the model:

- Let μ_j be the independence expectation of Z_j:

$$\mu_j = E[Z_j|\boldsymbol{\Lambda} = \mathbf{0}] = \frac{e^{\alpha_j}}{1 - e^{\alpha_j}}$$

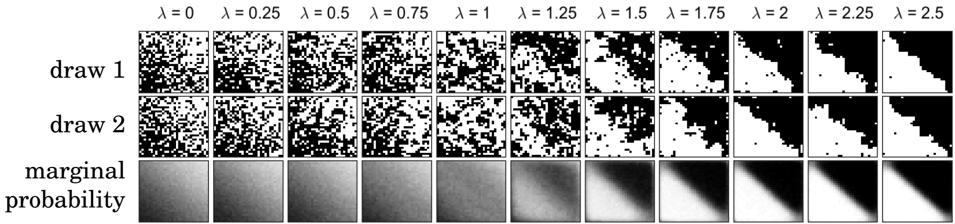
- Centered model then has:

$$\text{logit}\left(\Pr(Z_i = 1|\mathbf{Z}_{-i})\right) = \mathbf{x}_i^T \boldsymbol{\beta} + \sum_{j\sim i} \lambda_{ij}\left(z_j - \mu_j\right)$$

$$Q(\mathbf{z}) = \mathbf{z}^T \mathbf{X}\boldsymbol{\beta} - \mathbf{z}^T \boldsymbol{\Lambda}\boldsymbol{\mu} + \frac{1}{2}\mathbf{z}^T \boldsymbol{\Lambda}\mathbf{z}$$

But does it help?

Try the demonstration case with the centered, $\{0, 1\}$ model:



- Somewhat reasonable behaviour for $\lambda < 1$.
- Very undesirable behaviour with large λ.

An alternative solution:
change the coding

Binary variables

- *Bernoulli* random variables take values $\{0, 1\}$.
- *Binary* random variables are categorical.
  - We choose the coding.
  - Could be $\{0, 1\}$, $\{-1, 1\}$, or $\{\ell, h\}$.
- If $\mathbf{Z}$ has support $\{\ell, h\}^n$,

$$\mathbf{Y} = a\mathbf{Z} + b\mathbf{1}, \qquad \text{where} \quad a = \frac{H - L}{h - \ell}, \quad b = L - a\ell$$

  has support $\{L, H\}^n$.
- We shouldn't change coding without thinking...

Two ways to change the coding

Say $\mathbf{Z} \in \{\ell, h\}^n$, with PMF $f_{\mathbf{Z}}(\mathbf{z}) \propto g(\mathbf{z}; \boldsymbol{\theta})$.

But we want our model to use coding $\{L, H\}$.

The right way

$$\mathbf{Y} = a\mathbf{Z} + b\mathbf{1} \iff \mathbf{Z} = \tfrac{1}{a}\mathbf{Y} - \tfrac{b}{a}\mathbf{1}$$

$$\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= \Pr(\mathbf{Y} = \mathbf{y}) \\
&= \Pr(a\mathbf{Z} + b\mathbf{1} = \mathbf{y}) \\
&= f_{\mathbf{Z}}(\tfrac{1}{a}\mathbf{y} - \tfrac{b}{a}\mathbf{1}) \\
&\propto g(\tfrac{1}{a}\mathbf{y} - \tfrac{b}{a}\mathbf{1}; \boldsymbol{\theta}) \\
&\propto g(\mathbf{z}; \boldsymbol{\theta}).
\end{aligned}$$

The tempting way

Just plug in $\mathbf{y} = a\mathbf{z} + b\mathbf{1}$.

Let the parameter be $\boldsymbol{\theta}'$.

$$\begin{aligned}
f'_{\mathbf{Y}} &\propto g(\mathbf{y}; \boldsymbol{\theta}') \\
&\propto g(a\mathbf{z} + b\mathbf{1}; \boldsymbol{\theta}')
\end{aligned}$$

To achieve $f'_{\mathbf{Y}} = f_{\mathbf{Y}}$, we need $\boldsymbol{\theta}'$ to compensate for linear transformation of $\mathbf{z}$.

A general form of the model

Observation: maybe the asymmetry of the model is due to the coding?

- Derive the model for arbitrary $\{\ell, h\}$ coding, we find

$$\operatorname{logit}\left(\Pr(Z_i = h|\mathbf{Z}_{-i})\right) = (h-\ell)\left[\mathbf{x}_i^T\boldsymbol{\beta} + \sum_{j\sim i}\lambda_{ij}(z_j - \mu_j)\right]$$

where

$$\mu_j = \begin{cases} \mathbf{0} & \text{for a standard model} \\ \dfrac{\ell e^{\ell\alpha_i} + h e^{h\alpha_i}}{e^{\ell\alpha_i} + e^{h\alpha_i}} & \text{for a centered model} \end{cases}$$

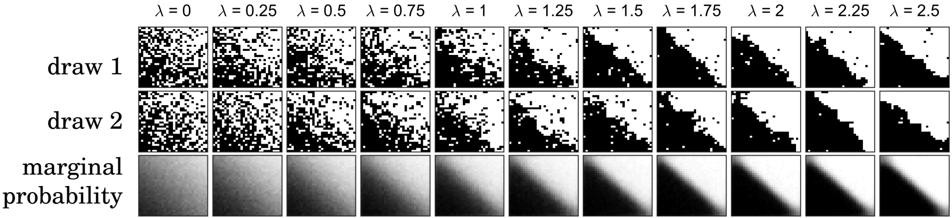
- Negpotential function:

$$Q(\mathbf{z}) = \mathbf{z}^T\mathbf{X}\boldsymbol{\beta} - \mathbf{z}^T\boldsymbol{\Lambda}\boldsymbol{\mu} + \frac{1}{2}\mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z}$$

- Of interest: **“plus/minus”** codings: $\{-h, h\}$.

Demonstration results with plus/minus coding

Standard model, $\{-1, 1\}$ coding:



- Finally see reasonable behaviour
- λ controls the balance between the unary part and the neighbour effect.

Theoretical results

Model equivalence

Different model **variants** have different choices of coding and centering.

Theorem 1: All **autologistic variants are equivalent** to any chosen standard model.

Theorem 2: **Autologistic regression variants are not equivalent**, in general.

- Say $f_1(\cdot; \mathbf{X}\boldsymbol{\gamma}, \boldsymbol{\Omega})$ is an ALR with coding $\{L, H\}$
- Say $f_2(\cdot; \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Lambda})$ is an ALR with coding $\{\ell, h\}$
- then f_2 and f_1 are equivalent iff $\boldsymbol{\beta}$ satisfies

$$\mathbf{X}\boldsymbol{\beta} - a^2\boldsymbol{\Omega}\boldsymbol{\mu}_{\mathbf{X}\boldsymbol{\beta}} = a\mathbf{X}\boldsymbol{\gamma} + a\boldsymbol{\Omega}(b\mathbf{1} - \boldsymbol{\mu}_{\mathbf{X}\boldsymbol{\gamma}})$$

- This is an overdetermined system in $\boldsymbol{\beta}$.

- So in regression case, changing coding and/or centering changes the distribution family!
  - Exception: standard models with plus/minus coding are all equivalent.

Large association behaviour

Theorem 3: **Only the standard, plus/minus models have reasonable large-association behaviour**.

- Say f is a standard model with coding $\{-h, h\}$.
- Let $p^*_{\mathbf{h}}$ and $p^*_{-\mathbf{h}}$ be the limiting probabilities of the two "saturated" states when $\lambda \to \infty$.
- then

$$p^*_{\mathbf{h}} = \frac{\exp\left(h\sum_{i=1}^{n}\alpha_i\right)}{\exp\left(h\sum_{i=1}^{n}\alpha_i\right)+\exp\left(-h\sum_{i=1}^{n}\alpha_i\right)} \quad \text{and} \quad p^*_{-h} = 1 - p^*_{\mathbf{h}}.$$

- And no other variants have more than one state with positive limiting probability in general.

Conclusions

Conclusions

- The ALR model is an example where "plugging in" different coding is not a trivial operation
  - Different codings $\Longleftrightarrow$ different distribution families.
- Coding like $\{-h, h\}$ is best
  - parameter interpretability
  - large-λ behaviour.
- Centered model is not necessary once you use $\{-h, h\}$ coding.
- What if you still want Bernoulli variables?
  - start with the $\{-h, h\}$ model
  - transform to $\{0, 1\}$ the "right" way.
  - You get

$$\operatorname{logit}(\Pr(Z_i = 1|\mathbf{Z}_{-i})) = \mathbf{x}_i^T \boldsymbol{\alpha} + \sum_{j\sim i} \lambda_{ij}(z_j - \tfrac{1}{2}),$$

which is a natural extension of logistic regression.