Reducing Estimation Bias when fitting Autologistic Models using the Pseudo-likelihood Approach

Yi-Chen Zhang
zhang318@msu.edu

Chad Babcock
babcoc76@msu.edu

Yingjie Li
liyingj1@stt.msu.edu

Abstract

In this article we look at the maximum pseudo-likelihood approach to autologistic model parameter estimation. Little is understood about the characteristics of maximum pseudo-likelihood estimated autologistic model parameters. Since the maximum pseudo-likelihood method is so widely used, it is important to assess and, if present, correct for the bias in parameters estimated using maximum pseudo-likelihood procedures. We examine two bias reduction techniques, (1) a bias corrective procedure introduced by Cox & Snell [6] and (2) a bias preventative procedure introduced by Firth [8]. Simulation results show both bias reduction techniques are effective at minimizing parameter estimation bias in the presence of stronger spatial dependence.

I. Introduction

In ecological, epidemiological, graphical and many other scientific fields, researchers often use binary data collected on a lattice to garner insight about their domain of interest. Whether studying the spatial distribution of plant and animal species, the spread of disease, or developing image classification techniques, gridded binary data are prevalent in the scientific research community [1, 2, 10, 11, 14, 16]. The autologistic model, is a spatial model useful for regressing latticed binary responses in the presence of spatial autocorrelation [7]. First introduced by Besag in 1972 [5], the autologistic model has been studied heavily ever sense [3, 12, 13]. Extreme difficulties in the maximum likelihood estimation process have led to the development of various other methods to estimate autologistic model parameters, including maximum pseudo-likelihood, Monte Carlo maximum likelihood, and Bayesian techniques [4, 12, 17]. In this article we focus our attention on the maximum pseudo-likelihood approach, which is the most prevalent estimation technique in the applied sciences due to its ease of computation. Little is understood about the characteristics of maximum pseudo-likelihood estimated autologistic model parameters [9]. Since the maximum pseudo-likelihood method is so widely used, it is important to assess and, if present, correct for the bias in parameters estimated using maximum pseudo-likelihood procedures.

In this article we examine two bias reduction techniques, (1) a bias corrective procedure introduced by Cox & Snell [6] and (2) a bias preventative procedure introduced by Firth [8]. The rest of this article evolves as follows, a brief introduction to the autologistic model and pseudo-likelihood, followed by a explanation of Cox and Snell's and Firth's bias reduction techniques, followed by a simulation study showing the performance of the two bias reduction procedures. We end the article with a discussion about the performance of the bias reduction strategies, and indications for future research.

II. Autologistic Models

Let $Y_i \in \{0,1\}$ be the binary response variable from grid cell i on a lattice where $i = 1, 2, ..., n$ and n is the total number of grid cells. Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ be the response vector containing all n grid cell responses. Let $\boldsymbol{Y}_{-i} = (Y_1, ..., Y_{i-1}, Y_{i+1}, ..., Y_n)^T$ be all these responses from the grid except grid cell i. Here we are going to specify a first order neighborhood structure which consists of the first four nearest neighbors on the lattice. Let $\boldsymbol{\mathcal{N}}_i$ be the set of neighboring grid cells to grid cell i, and let $j \in \boldsymbol{\mathcal{N}}_i$ denote that grid cell j is a neighbor to grid cell i.

$$p(Y_i|Y_{-i}) = p(Y_i|Y_j : j \in \boldsymbol{\mathcal{N}}_i) \tag{1}$$

Equation (1) describes the conditional probability distribution of Y_i given it's neighbors Y_j. Further, we assume that $p(Y_i|Y_j : j \in \boldsymbol{\mathcal{N}}_i)$ is Bernoulli distributed the success probability π_i. So,

$$\pi_i = p(Y_i = 1|Y_j : j \in \boldsymbol{\mathcal{N}}_i) \tag{2}$$

π_i depends on Y_j through a logit link function like so,

$$\log \frac{\pi_i}{1-\pi_i} = \boldsymbol{X}_i^T \boldsymbol{\beta} + \sum_{j \in \boldsymbol{\mathcal{N}}_i} \eta_{ij} Y_j, \tag{3}$$

where $\boldsymbol{X}_i$ is a vector of regression coefficients from grid cell i and $\boldsymbol{\beta}$ is a corresponding vector of regression parameters. η_{ij} is the parameter that controls how influential the response at grid cell j is on the response at grid cell i. We are going to specify a constant and isotropic spatial dependence structure. This means all neighbors (j) have equal influence on the outcome of Y_i and that the influence of neighbors does not change as we move across the spatial domain. Let $\eta_{ij} = \eta \mathbf{1}_{\{j \in \boldsymbol{\mathcal{N}}_i\}}$ with $\eta > 0$ This allows us to rewrite equation (3) as

$$\log \frac{\pi_i}{1-\pi_i} = \boldsymbol{X}_i^T \boldsymbol{\beta} + \eta \sum_{j \in \boldsymbol{\mathcal{N}}_i} Y_j, \tag{4}$$

The full conditional distribution is then,

$$p(Y_i|Y_{-i}; \boldsymbol{\theta}) = p(Y_i|Y_j : j \in \boldsymbol{\mathcal{N}}_i; \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + \eta \sum_{j \in \boldsymbol{\mathcal{N}}_i} Y_j)}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + \eta \sum_{j \in \boldsymbol{\mathcal{N}}_i} Y_j)} \tag{5}$$

Where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \eta)^T$. The pseudo-likelihood is defined in [7] as the product of the full conditional probabilities for all grid cells so the log pseudo-likelihood(l_{PL}) for the autologistic model with a constant spatial dependence parameter is

$$l_{PL}(\boldsymbol{\theta}|\boldsymbol{Y}) = \sum_{i=1}^{n} \log \frac{\exp(Y_i(\boldsymbol{X}_i^T \boldsymbol{\beta} + \eta \sum_{j \in \boldsymbol{\mathcal{N}}_i} Y_j))}{1 + \exp(\boldsymbol{X}_i^T \boldsymbol{\beta} + \eta \sum_{j \in \boldsymbol{\mathcal{N}}_i} Y_j)} \tag{6}$$

The maximization of equation (6) yields maximum pseudo-likelihood estimates for $\boldsymbol{\theta}$.

Although (6) is not the true log likelihood, Besag [4] showed that the maximum pseudo-likelihood estimate converges almost surely to the maximum likelihood estimate as the number of observations goes to infinity. Hence we choose to maximize the pseudo-likelihood rather than the true likelihood to estimate $\boldsymbol{\theta}$ and apply bias reduction strategies.

III. Bias Correction Methods

In a regular model with a p-dimensional parameter θ the asymptotic bias of the maximum likelihood estimate $\hat{\theta}$ may be written as:

$$b(\theta) = \frac{b_1(\theta)}{n} + \frac{b_2(\theta)}{n^2} + ... \tag{7}$$

where $b(\theta) = \hat{\theta} - \theta$ and n is usually interpreted as the number of observations. We examine two approaches to reduce the bias of an estimator. The first method follows Cox & Snell [6], which gives an estimate of the bias of order n^{-1}. Then the bias-reduced estimate is calculated by $\hat{\theta} - \frac{b_1(\hat{\theta})}{n}$. The second method is introduced by Firth [8], which modifies the score equation to get the maximum likelihood estimate. Both methods try to eliminate or reduce the first-order of bias $O(n^{-1})$. We will introduce more fully these two methods in sections i and ii.

i. Cox & Snell Method

According to Cox & Snell [6], for independent but not necessarily identically distributed observations $Y_i (i = 1, 2, ...n)$, the log likelihood is:

$$L(\beta) = \sum_{i=1}^{n} \log p_i(Y_i, \beta), \tag{8}$$

where $p_i(Y_i, \boldsymbol{\beta})$ is the p.d.f. of Y_i, and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$. To get the maximum likelihood estimation, usually we expand the maximum likelihood equation $L^{'}(\boldsymbol{\beta}) = 0$ to its first order:

$$L^{'}(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) L^{''}(\boldsymbol{\beta}) = 0 \tag{9}$$

To obtain a more refined answer, we replace (9) by the second order equation:

$$L^{'}(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) L^{''}(\boldsymbol{\beta}) + \frac{1}{2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2 L^{'''}(\boldsymbol{\beta}) = 0 \tag{10}$$

Since $\boldsymbol{\beta}$ is a p dimension vector, (10) is actually a set of equations and the r-th equation is:

$$\frac{\partial L}{\partial \beta_r} + \sum_{s=1}^{p} \frac{\partial^2 L}{\partial \beta_r \partial \beta_s}(\hat{\beta}_s - \beta_s) + \frac{1}{2} \sum_{t,u=1}^{p} (\hat{\beta}_t - \beta_t) \frac{\partial^3 L}{\partial \beta_r \partial \beta_t \partial \beta_u}(\hat{\beta}_u - \beta_u) = 0 \tag{11}$$

where $r = 1, 2, ...p$. Define:

$$\begin{aligned} U_r &= \sum_{i=1}^{n} \frac{\partial \log p_i(Y_i, \boldsymbol{\beta})}{\partial \beta_r}, \quad V_{rs} = \sum_{i=1}^{n} \frac{\partial^2 \log p_i(Y_i, \boldsymbol{\beta})}{\partial \beta_r \partial \beta_s}, \\ W_{rst} &= \sum_{i=1}^{n} \frac{\partial^3 \log p_i(Y_i, \boldsymbol{\beta})}{\partial \beta_r \partial \beta_s \partial \beta_t} \\ I_{rs} &= E[-V_{rs}], \ J_{r,st} = E[U_r V_{st}], \ K_{rst} = E[W_{rst}] \end{aligned} \tag{12}$$

where I is actually the information matrix. By some approximately calculation, we can get estimates of the biases, which are approximately of order n^{-1}. The bias of the r-th parameter is:

$$b_r = E(\hat{\beta}_r - \beta_r) = (I_r^{-1}) \cdot \boldsymbol{R}^T, \tag{13}$$

where $\boldsymbol{R} = (R_1, R_2, ...R_p)$ and $R_r = \sum_{t,u=1}^{p}(I_{tu}^{-1}(J_{r,ut} + \frac{1}{2}K_{rtu}))$

For autologistic model, employ the pseudo-likelihood function as mentioned in (6) ,

$$l_{PL}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log \frac{\exp(Y_i(\boldsymbol{X}_i^T\boldsymbol{\beta} + \eta\sum_{j\in\mathcal{N}_i} Y_j))}{1+\exp(\boldsymbol{X}_i^T\boldsymbol{\beta} + \eta\sum_{j\in\mathcal{N}_i} Y_j)}$$

If we write $\boldsymbol{\theta} = (\beta_1, \beta_2, ...\beta_p, \eta)$ and $\boldsymbol{X}_i^{*T} = (\boldsymbol{X}_i^T, \sum_{j\in\mathcal{N}_i} Y_j)^T$, $i = 1, ..., n$, then the corresponding U_r, V_{rs}, W_{rst} are:

$$U_r = \sum_{i=1}^{n}\left[Y_i X_{ir}^* - \frac{\exp \boldsymbol{X}_i^{*T}\boldsymbol{\theta}}{1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta})}X_{ir}^*\right],$$

$$V_{rs} = \sum_{i=1}^{n} -\frac{\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta})}{(1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta}))^2}X_{ir}^* X_{is}^*,$$

$$W_{rst} = \sum_{i=1}^{n}\left[\frac{1}{(1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta}))^2} - \frac{2}{(1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta}))^3}\right]X_{ir}^* X_{is}^* X_{it}^*.$$

Then

$$I_{rs} = E(-V_{rs}) = -V_{rs}, \quad K_{rst} = E(W_{rst}) = W_{rst}, \tag{14}$$

$$J_{r,st} = E(U_r V_{rs}) = \sum_{i=1}^{n} E(U_r^i V_{st}^i) \tag{15}$$

Since V_{st}^i is constant and $EU_r^i = 0$, we have $J_{r,st} = 0$

As a result the corresponding bias is:

$$E(\hat{\beta}_r - \beta_r) = \frac{1}{2}(I^{-1})_r \cdot \boldsymbol{R} \tag{16}$$

where $\boldsymbol{R} = (R_1, R_2, ...R_{p+1})^T$ and $R_r = \sum_{t,u=1}^{p}((I^{-1})_{tu} \cdot K_{rtu})$. Subtracting this bias from the original maximum pseudo-likelihood estimate leads to bias reduced estimates.

ii. Firth Method

In regular problems the maximum likelihood estimate is derived as a solution to the score equation

$$U(\theta) = \frac{\partial}{\partial\theta}l(\theta) \equiv 0$$

where $l(\theta) = \log L(\theta)$ is the log likelihood function.

Firth (1993)[8] provided a bias reduction technique by introducing a penalty function into log likelihood function. For exponential families, if θ is the canonical parameter, the new log likelihood function can be written as

$$l^*(\theta) = l(\theta) + \frac{1}{2}\log|I(\theta)|,$$

where $I(\theta)$ is the fisher information matrix. This new log likelihood function is the original log likelihood function plus a penalty term. This penalty function $|I(\theta)|^{\frac{1}{2}}$ happens to be the Jeffreys (1946)[15] invariant prior.

We use pseudo-likelihood function (6) for autologistic model. To perform the Newton-Raphson algorithm, we first need to derive score function and fisher information for our new pseudo-likelihood function. Then,

$$\begin{aligned} l^*(\boldsymbol{\theta}) &= l_{PL}(\boldsymbol{\theta}) + \frac{1}{2}\log|I(\boldsymbol{\theta})| \\ &= \sum_{i=1}^{n}\left(Y_i\big(\boldsymbol{X}_i^T\boldsymbol{\beta} + \eta\sum_{j\in\boldsymbol{\mathcal{N}}_i}Y_j\big) - \log\left(1+\exp\big(\boldsymbol{X}_i^T\boldsymbol{\beta}+\eta\sum_{j\in\boldsymbol{\mathcal{N}}_i}Y_j\big)\right)\right) + \frac{1}{2}\log|I(\boldsymbol{\theta})| \\ &= \sum_{i=1}^{n}\left(Y_i\boldsymbol{X}_i^{*T}\boldsymbol{\theta} - \log\left(1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta}\right)\right) + \frac{1}{2}\log|I(\boldsymbol{\theta})|, \end{aligned}$$

where the fisher information $I(\theta)$ is derived from pseudo-likelihood function $l_{PL}(\boldsymbol{\theta})$. For our new score function

$$\begin{aligned} \frac{\partial}{\partial\theta_r}l^*(\boldsymbol{\theta}) &= \frac{\partial}{\partial\theta_r}l_{PL}(\boldsymbol{\theta}) + \frac{1}{2}\frac{\partial}{\partial\theta_r}\log|I(\boldsymbol{\theta})| \\ \Rightarrow U_r^*(\boldsymbol{\theta}) &= U_r(\boldsymbol{\theta}) + \frac{1}{2}tr\left[I(\boldsymbol{\theta})^{-1}\left(\frac{\partial}{\partial\theta_r}I(\boldsymbol{\theta})\right)\right] \end{aligned} \tag{17}$$

Here,

$$U_r(\boldsymbol{\theta}) = \sum_{i=1}^{n}((Y_i - p_i)X_{ir}^*), \quad \left(\text{here } p_i = \frac{\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta})}{1+\exp(\boldsymbol{X}_i^{*T}\boldsymbol{\theta})}\right)$$

then,

$$U(\boldsymbol{\theta}) = \boldsymbol{X}^{*T}(\boldsymbol{Y} - \boldsymbol{P}) \tag{18}$$

$$V_{rs}(\boldsymbol{\theta}) = \sum_{i=1}^{n}(-p_i(1-p_i)X_{ir}^*X_{is}^*) \tag{19}$$

$$\begin{aligned} V(\boldsymbol{\theta}) &= -\boldsymbol{X}^{*T}\boldsymbol{A}\boldsymbol{X}^* \\ \Rightarrow I(\boldsymbol{\theta}) &= E(-V(\boldsymbol{\theta})) = \boldsymbol{X}^{*T}\boldsymbol{A}\boldsymbol{X}^*, \end{aligned} \tag{20}$$

where $\boldsymbol{A} = diag(p_1(1-p_1), \ldots, p_n(1-p_n))$ a diagonal matrix.

$$W_{rst}(\boldsymbol{\theta}) = \sum_{i=1}^{n}(p_i(1-p_i)(2p_i-1)X_{ir}^*X_{is}^*X_{it}^*) \tag{21}$$

Let

$$\boldsymbol{B}_r = diag(p_1(1-p_1)(2p_1-1)X_{1r}^*, \ldots, p_n(1-p_n)(2p_n-1)X_{nr}^*)$$

then,

$$\begin{aligned} &W_r(\boldsymbol{\theta}) = \boldsymbol{X}^{*T}\boldsymbol{B}_r\boldsymbol{X}^* \\ \Rightarrow\quad &\frac{\partial}{\partial\theta_r}I(\boldsymbol{\theta}) = -W_r(\boldsymbol{\theta}) \end{aligned} \tag{22}$$

combine equation (18), (20) and (22) into (17), then we have

$$U_r^*(\boldsymbol{\theta}) = (\boldsymbol{X}^{*T}(\boldsymbol{Y} - \boldsymbol{P}))_r - \frac{1}{2} tr \left[(\boldsymbol{X}^{*T} \boldsymbol{A} \boldsymbol{X}^*)^{-1} (\boldsymbol{X}^{*T} \boldsymbol{B}_r \boldsymbol{X}^*) \right].$$

We can apply Newton-Raphson algorithm to solve the new pseudo-likelihood function to get a bias reduced estimate.

IV. Simulation

In order to examine the performance of the two bias reduction methods, we ran simulations for $\beta = 1$ and $\eta = 0.25, 0.5, 0.75$ respectively. Under each scenario, we generated a lattice for 15×15, 20×20, 25×25, 30×30, 40×40, and 50×50. The results for the uncorrected, Cox & Snell, and Firth bias reduction methods for each sample size can be found in table 1. Simulations were all replicated 5000 times. Figures 1, 2, and 3 are included to better summarize the simulation results.

Table 1: *Bias and standard error (in parentheses) for autologistic model simulations*

Method	$n = 15 \times 15$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00890 (0.00299)	0.00204 (0.00262)
Cox & Snell	0.00351 (0.00296)	-0.01105 (0.00255)
Firth	0.00355 (0.00296)	-0.01076 (0.00255)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.00388 (0.00299)	0.01904 (0.00290)
Cox & Snell	-0.00119 (0.00295)	-0.00385 (0.00281)
Firth	-0.00106 (0.00295)	-0.00335 (0.00281)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	0.00749 (0.00307)	0.02875 (0.00630)
Cox & Snell	0.00198 (0.00304)	-0.00745 (0.00340)
Firth	0.00221 (0.00304)	-0.00644 (0.00341)

Method	$n = 20 \times 20$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00810 (0.00290)	0.00087 (0.00157)
Cox & Snell	0.00381 (0.00288)	-0.00296 (0.00155)
Firth	0.00385 (0.00288)	-0.00292 (0.00155)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.01192 (0.00306)	0.00382 (0.00181)
Cox & Snell	0.00748 (0.00303)	-0.00434 (0.00179)
Firth	0.00755 (0.00303)	-0.00427 (0.00179)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	0.00955 (0.00318)	0.00962 (0.00221)
Cox & Snell	0.00434 (0.00315)	-0.00498 (0.00216)
Firth	0.00450 (0.00315)	-0.00481 (0.00216)

Method	$n = 25 \times 25$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00391 (0.00143)	0.00135 (0.00278)
Cox & Snell	0.00230 (0.00142)	-0.01281 (0.00272)
Firth	0.00230 (0.00142)	-0.01254 (0.00272)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.00207 (0.00141)	0.01423 (0.00295)
Cox & Snell	0.00057 (0.00140)	-0.00659 (0.00287)
Firth	0.00058 (0.00140)	-0.00624 (0.00284)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	0.00169 (0.00143)	0.02936 (0.00342)
Cox & Snell	0.00019 (0.00142)	0.00030 (0.00331)
Firth	0.00022 (0.00142)	0.00084 (0.00331)

Method	$n = 30 \times 30$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00137 (0.00123)	0.00210 (0.00167)
Cox & Snell	0.00021 (0.00123)	-0.00409 (0.00165)
Firth	0.00022 (0.00123)	-0.00403 (0.00165)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.00144 (0.00124)	0.00505 (0.00180)
Cox & Snell	0.00033 (0.00124)	-0.00461 (0.00178)
Firth	0.00033 (0.00124)	-0.00461 (0.00178)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	0.00195 (0.00125)	0.00753 (0.00208)
Cox & Snell	0.00080 (0.00125)	-0.00674 (0.00204)
Firth	0.00081 (0.00125)	-0.00657 (0.00204)

Method	$n = 40 \times 40$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00278 (0.00086)	0.00340 (0.00214)
Cox & Snell	0.00217 (0.00086)	-0.00531 (0.00211)
Firth	0.00217 (0.00086)	-0.00522 (0.00211)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.00072 (0.00086)	0.00947 (0.00226)
Cox & Snell	0.00015 (0.00086)	-0.00296 (0.00222)
Firth	0.00015 (0.00085)	-0.00283 (0.00222)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	0.00118 (0.00087)	0.01574 (0.00254)
Cox & Snell	0.00061 (0.00087)	-0.00125 (0.00249)
Firth	0.00062 (0.00087)	-0.00107 (0.00249)

Method	$n = 50 \times 50$	
	$\beta = 1$	$\eta = 0.25$
Uncorrected	0.00012 (0.00068)	0.00295 (0.00188)
Cox & Snell	-0.00027 (0.00068)	-0.00395 (0.00186)
Firth	-0.00028 (0.00068)	-0.00389 (0.00186)
	$\beta = 1$	$\eta = 0.5$
Uncorrected	0.00025 (0.00067)	0.00636 (0.00203)
Cox & Snell	-0.00011 (0.00066)	-0.00342 (0.00201)
Firth	-0.00013 (0.00066)	-0.00332 (0.00201)
	$\beta = 1$	$\eta = 0.75$
Uncorrected	-0.00006 (0.00068)	0.00921 (0.00227)
Cox & Snell	-0.00042 (0.00068)	-0.00406 (0.00224)
Firth	-0.00042 (0.00068)	-0.00394 (0.00224)

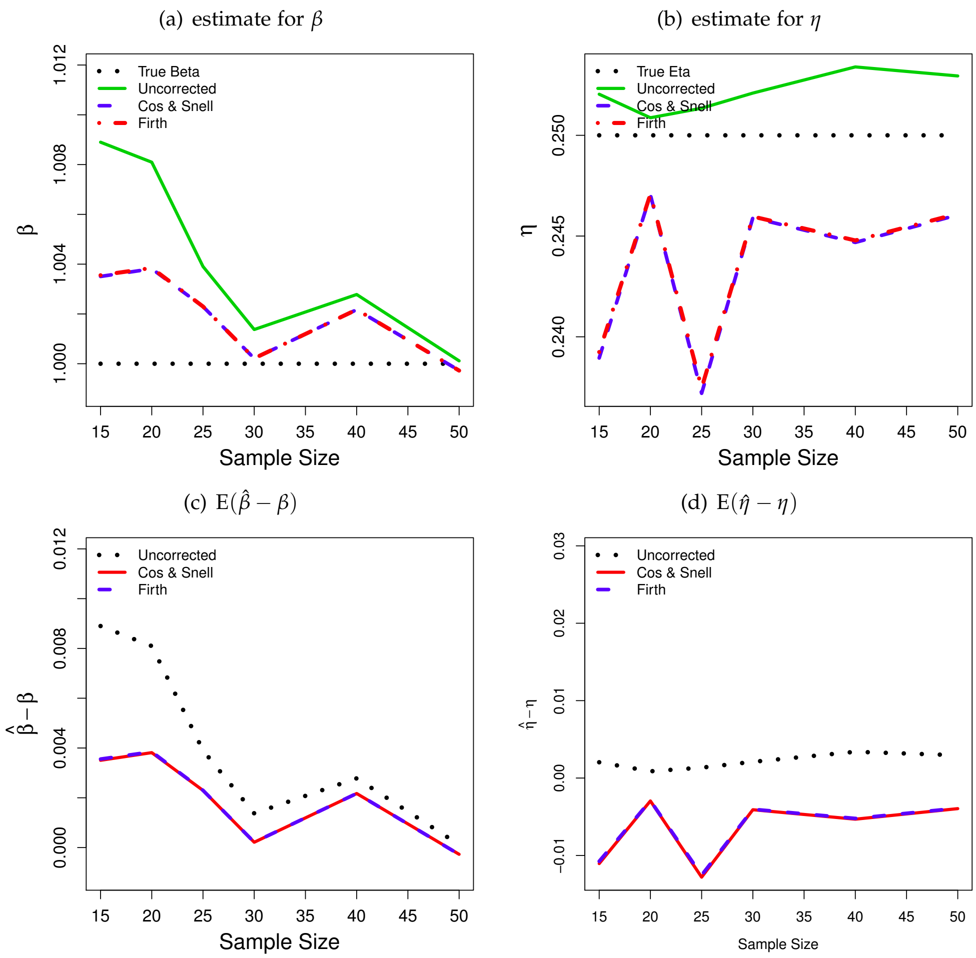

Figure 1: $\beta = 1$, $\eta = 0.25$

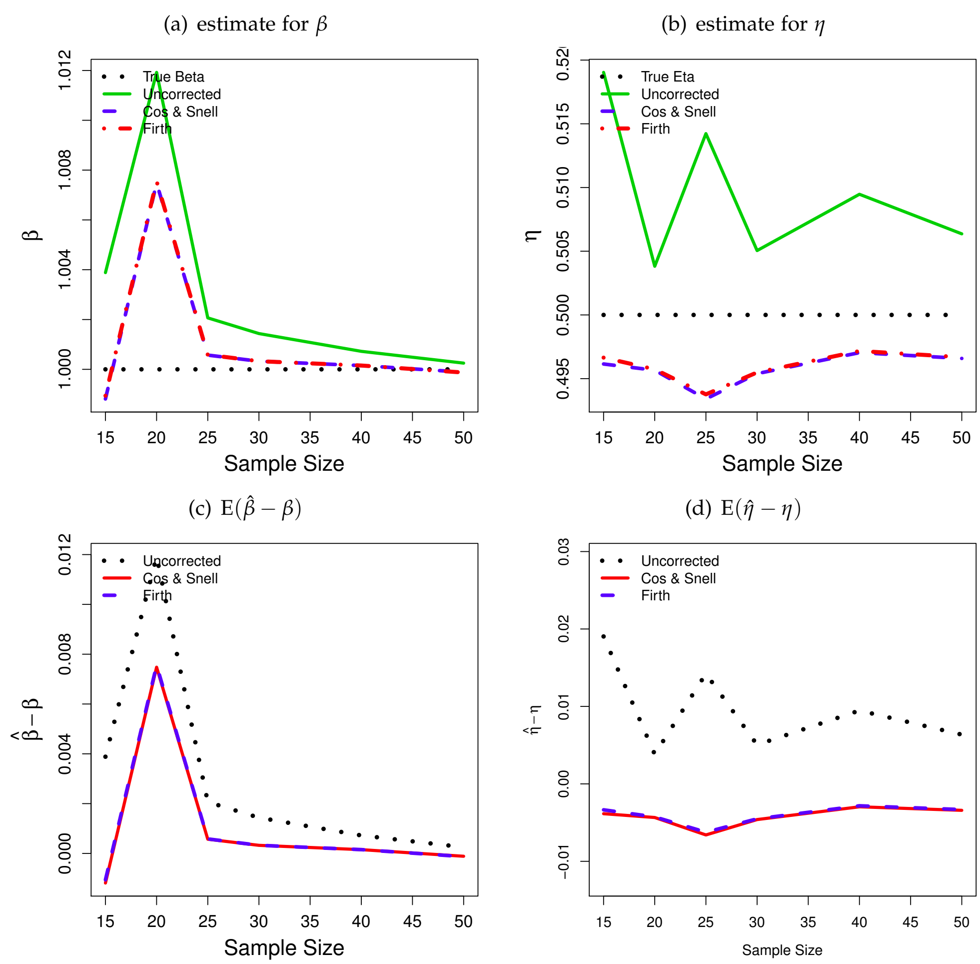

Figure 2: $\beta = 1$, $\eta = 0.5$

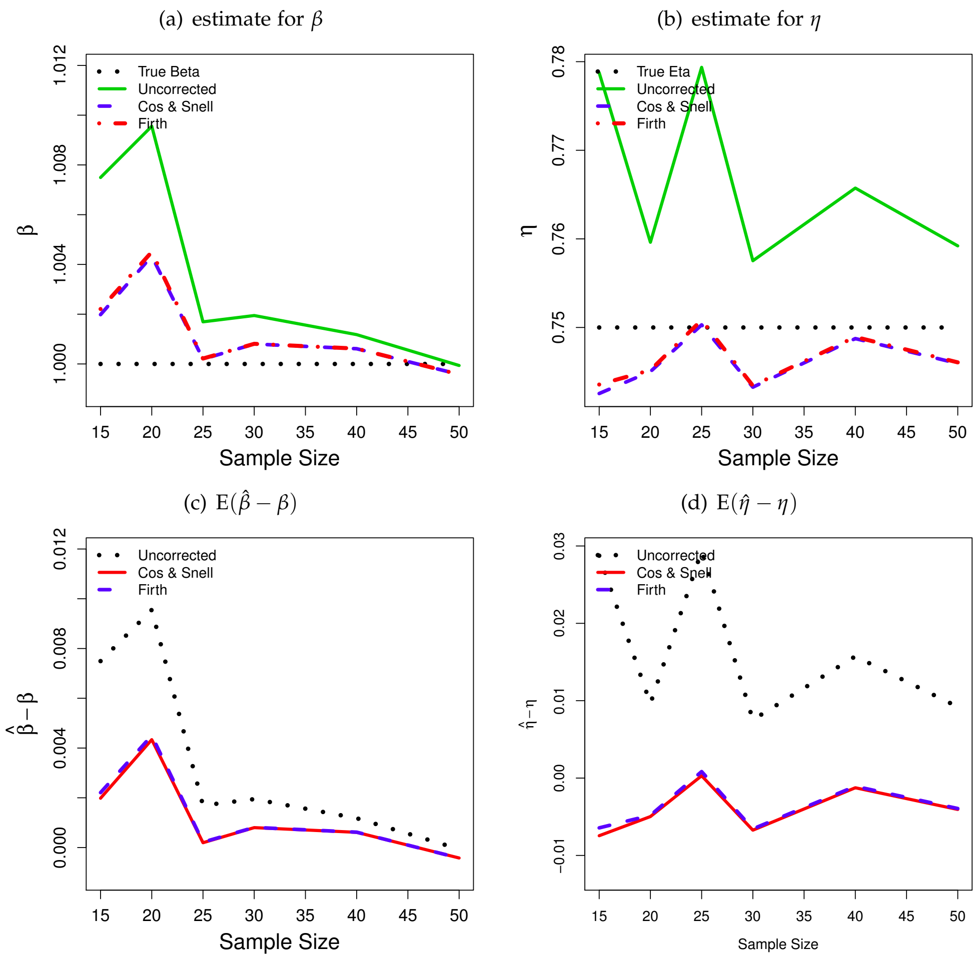

Figure 3: $\beta = 1$, $\eta = 0.75$

V. Discussion

We employed Cox & Snell's and Firth's method to reduce the bias of the maximum pseudo-likelihood estimate for an autologistic model. By using the pseudo-likelihood, the model becomes an exponential family model with canonical parameter, in which Cox & Snell's method and Firth's method will be equivalent. The numerical result confirms this assertion because both methods perform similar concerning bias reduction (table 1).

Both methods perform well when estimating β. Figures 1, 2, and 3 show that the bias corrected estimates for β always exhibit less bias than the uncorrected estimates except in the largest sample size ($n = 50 \times 50$) where the uncorrected bias was extremely low anyway. The estimate for β also becomes less bias as sample sizes increase. This is coincident with the intuition that larger sample sizes provide more information for inference which leads to better parameter estimation. Further, we did not specifically account for any edge effect issues that occur when the lattice is small or when the lattice has dimension of $n \times m$ where n is much smaller than m. Both of these situations result in a large proportion of observations on the edge of the lattice. Since edge observations have fewer neighbors than interior cells, this may lead to poorer parameter estimation. Edge effects may also be contributing to the larger bias at smaller lattice sizes.

For η, the bias reducing techniques work well when $\eta = 0.5$ and $\eta = 0.75$ (figures 2 and 3). They fail when $\eta = 0.25$ (figure 1). It seems that when the strength of the spatial dependence for the lattice is low ,i.e. $\eta = 0.25$, the grid cell observations are nearly independent thus, the effective sample size is large enough to get good parameter estimations before bias reduction (which are almost of order n^{-1}). Since the bias of η is already very small, and the amount reduced by the two methods are not exactly of order n^{-1}(may be larger), the bias reduction methods fail when $\eta = 0.25$.

Further exploration concerning the effect of the magnitude of the spatial dependence parameter needs to be done. Here, we only looked at three different η values. In future research we would also like to examine how the Cox & Snell and Firth bias reduction procedures behave when additional covariates are introduced to the model.

References

[1] G Arbia, R Benedetti, and G Espa. Contextual classification in image analysis: an assessment of accuracy of ICM. *Computational Statistics & Data Analysis*, 30(4):443–455, Jun 1999.

[2] NH Augustin, MA Mugglestone, and ST Buckland. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33(2):339–347, 1996.

[3] J Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):pp. 192–236, 1974.

[4] J Besag. Statistical-Analysis of Non-Lattice Data. *Statistician*, 24(3):179–195, 1975.

[5] J. E. Besag. Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(1):pp. 75–83, 1972.

[6] D. R. Cox and E. J. Snell. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(2):pp. 248–275, 1968.

[7] Noel A. C. Cressie. *Statistics for Spatial Data*. Wiley, revised edition, January 1993.

[8] D Firth. Bias Reduction of Maximum-Likelihood-Estimates. *Biometrika*, 80(1):27–38, sep 1993.

[9] R Fu, AL Thurman, T Chu, MM Steen-Adams, and J Zhu. On Estimation and Selection of Autologistic Regression Models via Penalized Pseudolikelihood. *Journal of Agricultural Biological and Environmental Statistics*, 18(3):429–449, Sep 2013.

[10] Marcia L. Gumpertz, Jonathan M. Graham, and Jean B. Ristaino. Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2):pp. 131–156, Jun 1997.

[11] FL He, J Zhou, and HT Zhu. Autologistic regression model for the distribution of vegetation. *Journal of Agricultural Biological and Environmental Statistics*, 8(2):205–222, Jun 2003.

[12] JA Hoeting, M Leecaster, and D Bowden. An improved model for spatially correlated binary responses. *Journal of Agricultural Biological and Environmental Statistics*, 5(1):102–114, 2000.

[13] John Hughes, Murali Haran, and Petruta C. Caragea. Autologistic models for binary data on a lattice. *Environmentrics*, 22(7):857–871, 2011.

[14] Kathryn M. Irvine, Jamie Thornton, Vickie M. Backus, Matthew G. Hohmann, Erik A. Lehnhoff, Bruce D. Maxwell, Kurt Michels, and Lisa J. Rew. A comparison of adaptive sampling designs and binary spatial models : a simulation study using a census of Bromus inermis. *Environmentrics*, 24(6):407–417, 2013.

[15] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186(1007):pp. 453–461, 1946.

[16] MD Piorecky and DRC Prescott. Multiple spatial scale logistic and autologistic habitat selection models for northern pygmy owls, along the eastern slopes of Alberta's Rocky Mountains. *Biological Conservation*, 129(3):360–371, May 2006.

[17] M Sherman, TV Apanasovich, and RJ Carroll. On estimation in binary autologistic spatial models. *Journal of Statistical Computation and Simulation*, 76(2):167–179, Feb 2006.