

# A Proposal of Two-step Autoregressive Model

Loc Nguyen (✉ [ng\\_phloc@yahoo.com](mailto:ng_phloc@yahoo.com))

Loc Nguyen's Academic Network <https://orcid.org/0000-0001-5192-8106>

---

## Method Article

**Keywords:** autoregressive (AR) model, conditional autoregressive (CAR) model, logistic function, spatial-temporal data analysis

**Posted Date:** February 4th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1322456/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# A Proposal of Two-step Autoregressive Model

Loc Nguyen

Loc Nguyen's Academic Network, Vietnam

Email: ng\_phloc@yahoo.com

Homepage: www.locnguyen.net

## Abstract

Autoregressive (AR) model and conditional autoregressive (CAR) model are specific regressive models in which independent variables and dependent variable imply the same object. They are powerful statistical tools to predict values based on correlation of time domain and space domain, which are useful in epidemiology analysis. In this research, I combine them by the simple way in which AR and CAR is estimated in two separate steps so as to cover time domain and space domain in spatial-temporal data analysis. Moreover, I integrate logistic model into CAR model, which aims to improve competence of autoregressive models.

**Keywords:** Autoregressive (AR) model, conditional autoregressive (CAR) model, logistic function, spatial-temporal data analysis.

## 1. Two-step autoregressive model

Autoregressive (AR) model evaluates new value of a variable based on its old values; in other words, independent variables called regressors and dependent variable called responsor refer to the same concerned object (CO). AR model is often used to predict new values in time series, which is called *temporal autoregressive model*. Conditional autoregressive (CAR) model also follows ideology of AR model where regressors and responsor imply the same CO but CAR model added more values of neighbors of CO. Therefore, CAR model is often used to predict new values in space domain, which is called *spatial autoregressive model*. There are some researches like (Mariella & Tarantino, 2010) to integrate AR model and CAR model but here I combine them consequently by the simple way in which AR and CAR is estimated in two separate processes. In other words, the research uses mainly autoregressive model through two steps:

1. *Step 1:* CAR model known as spatial autoregressive model is estimated to predict responsor based on its neighbors and its environmental features like humidity, precipitation, etc. As a convention, responsor is called case like disease case in epidemiology.
2. *Step 2:* AR model known as temporal autoregressive model is estimated to forecast next case based on current cases.

Let  $S = \{1, 2, \dots, n\}$  be spatial domain and so each site  $s$  can correspond to an area. Let  $nb(s)$  be set of neighbor sites of site  $s$ . Let  $X_s = (x_{s1}, x_{s2}, \dots, x_{sk})$  be the vector of environmental features at site  $s$  and so  $x_{sj}$  is the  $j^{\text{th}}$  environmental feature at site  $s$ . Let  $Y_s$  be the outcome number of cases at site  $s$ . If time is concerned,  $X_s$  and  $Y_s$  are replaced by  $X_{st}$  and  $Y_{st}$ , respectively. Suppose the time interval is  $T = 2$ , this research firstly uses spatial autoregressive model to estimate the number of cases  $Y_{st} = Y_s$  at current timepoint  $t$ . Later on, the research uses temporal autoregressive model to forecast the future number of cases  $Y_{s(t+1)}$  at timepoint  $t+1$  based on current  $Y_{st}$ . Therefore, the computational model of the research is called the two-step autoregressive model.

In the first step, spatial autoregressive model is estimated. Given current timepoint, we have  $Y_s = Y_{st}$  and  $X_s = X_{st}$ . When building up spatial autoregressive model, we ignore timepoints in dataset. Because the number of cases is moved around an average number, Poisson distribution is more suitable to build spatial autoregressive model. However, for simplicity, I use normal

distribution for spatial autoregressive model. Let  $W$  be the symmetric weighted adjacency matrix (Mariella & Tarantino, 2010, p. 227):

$$W = (w_{ss^*}) \text{ where } w_{ss^*} = \begin{cases} 0 & \text{if } s^* = s \\ \varphi(s, s^*) & \text{if } s^* \in \text{nb}(s) \\ 0 & \text{otherwise} \end{cases} \quad (1.1)$$

Where  $\varphi(s, s^*)$  measures the approximation of site  $s$  and site  $s^*$  with note that  $\text{nb}(s)$  denotes the set of neighbors of site  $s$ . Moreover, the set  $\text{nb}(s)$  does not contain  $s$ . Because  $W$  is symmetric, we have (Mariella & Tarantino, 2010, p. 227):

$$\forall s, s^* \in S, w_{ss^*} = w_{s^*s}$$

Note,  $W$  is constant. Let  $W_D$  be the diagonal matrix of normalization (Mariella & Tarantino, 2010, p. 227):

$$W_D = \text{diag}(w_{1+}, w_{2+}, \dots, w_{n+})$$

$$w_{s+} = \sum_{s^* \in \text{nb}(s)} w_{ss^*} \quad (1.2)$$

Let  $A$  be the matrix of interaction (Mariella & Tarantino, 2010, p. 227):

$$A = (a_{ss^*}) = W_D^{-1}W$$

$$a_{ss^*} = \frac{w_{ss^*}}{w_{s+}} \quad (1.3)$$

Note,  $A$  is constant. Let  $\Sigma_D$  be the diagonal matrix of variances (Mariella & Tarantino, 2010, p. 227):

$$\Sigma_D = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2) = \sigma^2 W_D^{-1}$$

$$\sigma_s^2 = \frac{\sigma^2}{w_{s+}} \quad (1.4)$$

Note,  $\sigma^2$  is called global variance. The probability density function (PDF) of the number of cases  $Y_s$  at  $X_s$  given its neighbors  $X_{s^*}$  is defined according to normal distribution as follows (Mariella & Tarantino, 2010, p. 227):

$$f(Y_s | X_s, X_{s^*}, \alpha_s, \sigma^2)$$

$$= \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{\left(Y_s - (\alpha_s^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (Y_{s^*} - \alpha_{s^*}^T X_{s^*}))\right)^2}{2\sigma_s^2}\right) \quad (1.5)$$

Where  $\alpha_s = (\alpha_{s1}, \alpha_{s2}, \dots, \alpha_{sk})$  is site regressive coefficient vector and  $\sigma_s^2$  is site variance. The constant  $\rho$  is called effective neighborhood constant,  $|\rho| < 1$ . Note, the superscript “ $T$ ” denotes vector/matrix transposition operator. Each site variance (local variance)  $\sigma_s^2$  was defined.

$$\sigma_s^2 = \frac{\sigma^2}{w_{s+}}$$

Therefore, only parameters  $\alpha_s$  and  $\sigma^2$  are parameters in the PDF of  $Y_s$  at  $X_s$  given its neighbors. The mean and variance of  $Y_s$  at  $X_s$  given its neighbors are:

$$\text{mean}(Y_s | X_s, X_{s^*}, \alpha_s, \sigma^2) = \alpha_s^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (Y_{s^*} - \alpha_{s^*}^T X_{s^*}) \quad (1.6)$$

$$\text{var}(Y_s | X_s, X_{s^*}, \alpha_s, \sigma^2) = \sigma_s^2$$

From the mean of  $Y_s$  at  $X_s$  given its neighbors, the spatial autoregressive model of the number of cases is defined as follows:

$$Y_s = \alpha_s^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (Y_{s^*} - \alpha_{s^*}^T X_{s^*}) \quad (1.7)$$

The essence of the first step in this research is to estimate all regressive coefficients  $\alpha_s$  when the effective neighborhood constant  $\rho$  is pre-defined. Let  $x_s^{(i)}$  be the  $i^{\text{th}}$  instance of the vector of

environmental feature  $X_s$  at site  $s$ . Let  $y_s^{(i)}$  be the  $i^{\text{th}}$  instance of  $Y_s$  at site  $s$ . Let  $\mathbf{X}_s$  be the set of  $N_s$  instances  $\mathbf{x}_s^{(i)}$  and let  $\mathbf{Y}_s$  be the set of  $N_s$  instances  $y_s^{(i)}$ . Each site  $s$  owns one dataset  $(\mathbf{X}_s, \mathbf{Y}_s)$ .

$$\mathbf{X}_s = \begin{pmatrix} (\mathbf{x}_s^{(1)})^T \\ (\mathbf{x}_s^{(2)})^T \\ \vdots \\ (\mathbf{x}_s^{(N_s)})^T \end{pmatrix} = \begin{pmatrix} 1 & x_{s1}^{(1)} & x_{s2}^{(1)} & \cdots & x_{sn}^{(1)} \\ 1 & x_{s1}^{(2)} & x_{s2}^{(2)} & \cdots & x_{sn}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{s1}^{(N_s)} & x_{s2}^{(N_s)} & \cdots & x_{sn}^{(N_s)} \end{pmatrix} \quad (1.8)$$

$$\mathbf{x}_s^{(i)} = \begin{pmatrix} 1 \\ x_{s1}^{(i)} \\ x_{s2}^{(i)} \\ \vdots \\ x_{sn}^{(i)} \end{pmatrix}, \mathbf{x}_s^{(j)} = \begin{pmatrix} 1 \\ x_{s1}^{(j)} \\ x_{s2}^{(j)} \\ \vdots \\ x_{sn}^{(j)} \end{pmatrix}, \mathbf{y}_s = \begin{pmatrix} y_s^{(1)} \\ y_s^{(2)} \\ \vdots \\ y_s^{(N_s)} \end{pmatrix}$$

Let  $\mathbf{X}$  and  $\mathbf{Y}$  be the sets of all site dataset  $\mathbf{X}_s$  and  $\mathbf{Y}_s$ , respectively. Let  $\alpha$  and  $\sigma^2$  be the sets of all  $\alpha_s$  and all  $\sigma_s^2$ , respectively. The joint probability of the global dataset  $(\mathbf{X}, \mathbf{Y})$  is:

$$f(\mathbf{X}, \mathbf{Y} | \alpha, \sigma^2) = \prod_{s=1}^{|\mathcal{S}|} f(Y_s | X_s, X_{s^*}, \alpha_s, \sigma_s^2)$$

$$= \left( \prod_{s=1}^{|\mathcal{S}|} (2\pi\sigma_s^2)^{-\frac{N_s}{2}} \right) * \prod_{s=1}^{|\mathcal{S}|} \exp \left( -\frac{1}{2\sigma_s^2} \sum_{i=1}^{N_s} \left( y_s^{(i)} - \left( \alpha_s^T \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (y_{s^*}^{(i)} - \alpha_{s^*}^T \mathbf{x}_{s^*}^{(i)}) \right) \right)^2 \right)$$

The log-likelihood function is logarithm function of such joint probability:

$$l(\alpha, \sigma^2) = -\frac{1}{2} \sum_{s=1}^{|\mathcal{S}|} N_s \log(2\pi\sigma_s^2)$$

$$- \frac{1}{2} \sum_{s=1}^{|\mathcal{S}|} \frac{1}{\sigma_s^2} \sum_{i=1}^{N_s} \left( y_s^{(i)} - \left( \alpha_s^T \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (y_{s^*}^{(i)} - \alpha_{s^*}^T \mathbf{x}_{s^*}^{(i)}) \right) \right)^2$$

Let  $\alpha_s^* = (\alpha_{s1}^*, \alpha_{s2}^*, \dots, \alpha_{sk}^*)$  be the optimal estimate of regressive coefficients and thus,  $\alpha_s^*$  is a maximizer of  $l(\alpha, \sigma^2)$ .

$$\alpha_s^* = \underset{\alpha_1, \alpha_2, \dots, \alpha_s, \dots, \alpha_{|\mathcal{S}|}}{\text{argmax}} l(\alpha, \sigma^2)$$

By taking first-order partial derivatives of  $l(\alpha, \sigma^2)$  with regard to  $\alpha_s$ , we obtain:

$$\frac{\partial l(\alpha, \sigma^2)}{\partial \alpha_s} = \frac{1}{\sigma_s^2} \sum_{i=1}^{N_s} \left( y_s^{(i)} - \left( \alpha_s^T \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (y_{s^*}^{(i)} - \alpha_{s^*}^T \mathbf{x}_{s^*}^{(i)}) \right) \right) (\mathbf{x}_s^{(i)})^T$$

When the first-order partial derivatives of  $l(\alpha, \sigma^2)$  with regard to  $\alpha_s$  are equal to zero, it gets locally maximal. In other words,  $\alpha_s^*$  is solution of the following equation resulted from setting such derivatives to be zero.

$$\sum_{i=1}^{N_s} \left( \mathbf{y}_s^{(i)} - \left( \alpha_s^T \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\mathbf{y}_{s^*}^{(i)} - \alpha_{s^*}^T \mathbf{x}_{s^*}^{(i)}) \right) \right) (\mathbf{x}_s^{(i)})^T = \mathbf{0}^T$$

As a convention,  $a_{ss^*} = 0$  if  $y_{s^*}^{(i)}$  or  $x_{s^*}^{(i)}$  does not exist. The equation above is re-written as follows:

$$\begin{aligned} \mathbf{y}_s^T \mathbf{X}_s - \alpha_s^T \mathbf{X}_s^T \mathbf{X}_s - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\mathbf{y}_{s^*}^T \mathbf{X}_s - \alpha_{s^*}^T \mathbf{X}_s^T \mathbf{X}_s) \\ = \left( \mathbf{y}_s^T - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \mathbf{y}_{s^*}^T \right) \mathbf{X}_s - \left( \alpha_s^T - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \alpha_{s^*}^T \right) \mathbf{X}_s^T \mathbf{X}_s = \mathbf{0}^T \end{aligned}$$

This implies

$$\alpha_s - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \alpha_{s^*} = (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \left( \mathbf{y}_s - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \mathbf{y}_{s^*} \right)$$

In general, the estimates  $\alpha_s^*$  for all regressive coefficient vectors  $\alpha_s$  where  $s$  from 1 to  $|S|$  is solution of the linear equation system consisting of  $|S|$  equations as follows:

$$\alpha_s - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \alpha_{s^*} = (\mathbf{X}_s^T \mathbf{X}_s)^{-1} \mathbf{X}_s^T \left( \mathbf{y}_s - \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \mathbf{y}_{s^*} \right) \Big|_{s=1, |S|} \quad (1.9)$$

It is not difficult to solve the linear equation system above. As a result, the spatial autoregressive model is re-written as follows:

$$Y_s = (\alpha_s^*)^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (Y_{s^*} - (\alpha_{s^*}^*)^T X_{s^*}) \quad (1.10)$$

In the second step, temporal autoregressive model is estimated. For simplicity, we follow the Markov condition in which, future cases only depend on current cases. When building up spatial autoregressive model, we ignore environmental features  $X_s$  in dataset or we focus on a given site  $s$ . Suppose the current number of cases  $Y_{st}$  at site  $s$  was determined previously by spatial autoregressive model. As a convention, the “ $s$ ” index is removed from the response variable and so we have  $Y_{t+1} = Y_{s(t+1)}$  and  $Y_t = Y_{st}$ . We will forecast next (unknown)  $Y_{t+1}$  based on current (known)  $Y_t$ . The temporal autoregressive model is defined as follows (Eshel, 2003, p. 1):

$$Y_{t+1} = \beta Y_t + \varepsilon \quad (1.11)$$

Where  $\varepsilon$  is the error random variable and  $\beta$  is temporal coefficient with note that mean of  $\varepsilon$  is 0 such that

$$E(Y_{t+1}) = \beta Y_t$$

Let  $y_t$  be the value of  $Y_t$  at timepoint  $t$ . Suppose there are  $T$  timepoints and we have  $t = 1, 2, \dots, T$ . The sum of squared errors is defined as follows:

$$S(\beta) = \sum_{t=1}^{T-1} (y_{t+1} - \beta y_t)^2$$

Let  $\beta^*$  be the estimate of  $\beta$ . The least squares method sets  $\beta^*$  as the minimizer of  $S(\beta)$ .

$$\beta^* = \underset{\beta}{\text{argmin}} S(\beta)$$

By taking first-order derivative of  $S(\beta)$  with regard to  $\beta$ , we obtain (Eshel, 2003, p. 1):

$$\frac{dS(\beta)}{d\beta} = -2 \sum_{t=1}^{T-1} (y_{t+1} - \beta y_t) y_t$$

The temporal coefficient  $\beta^*$  is solution of the following equation resulted from setting such derivative to be zero.

$$-2 \sum_{t=1}^{T-1} (y_{t+1} - \beta y_t) y_t = 0$$

It is easy to determine the formula for calculating  $\beta^*$  (Eshel, 2003, p. 1).

$$\beta^* = \frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T-1} y_t^2} \quad (1.12)$$

Please pay attention that values  $y_t$  are calculated by CAR model estimated in the first step. The temporal autoregressive model is re-written as follows:

$$Y_{s(t+1)} = \beta^* Y_{st} \quad (1.13)$$

Equation 1.12 to calculate  $\beta^*$  is simplest form of the well-known Yule-Walker equation.

## 2. Integrated with logistic function

In normal CAR model, the resporor  $Y_s$  which is the outcome number of cases at site  $s$  is real variable but in the situation that  $Y_s$  is category variable whose  $G$  ordered values are  $v_1, v_2, \dots, v_G$ , new modification of CAR model is required. The multinomial logistic function of  $Y_s$  with regard to value  $v_g$  at site  $s$ , which is associated spatial autoregression, is proposed as follows:

$$f_g(Y_s | X_s, X_{s^*}, \alpha_{sg}) = \frac{1}{1 + \exp\left(-(\alpha_{sg}^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\alpha_{sg}^T X_{s^*} - \alpha_{s^*g}^T X_{s^*}))\right)} \quad (2.1)$$

Note that  $\alpha_{sg} = (\alpha_{sg1}, \alpha_{sg2}, \dots, \alpha_{sgk})$  which is site regressive coefficient vector with regard to value  $y_g$  is the main parameter which is estimated here. Let  $p_g$  be the probability that  $Y_s$  receives value  $y_g$ , where  $g$  is from 1 to  $G$ . If  $G$  values  $v_1, v_2, \dots, v_G$  are not ordered,  $p_g$  is the same to the proposed multinomial logistic function of  $Y_s$ .

$$p(X_s | \alpha_{sg}) = P(Y_s = v_g | X_s, X_{s^*}, \alpha_{sg}) = f_g(Y_s | X_s, X_{s^*}, \alpha_{sg}) \\ = \frac{1}{1 + \exp\left(-(\alpha_{sg}^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\alpha_{sg}^T X_{s^*} - \alpha_{s^*g}^T X_{s^*}))\right)} \quad (2.2)$$

If  $G$  values  $v_1, v_2, \dots, v_G$  are ordered such that  $v_1 \leq v_2 \leq \dots \leq v_G$  then, according to Bender and Grouven (Bender & Grouven, 1997, p. 547),  $p_g$  becomes the cumulative probability such that  $Y_s \leq v_g$  as follows:

$$p(X_s | \alpha_{sg}) = P(Y_s \leq v_g | X_s, X_{s^*}, \alpha_{sg}) = \sum_{i=1}^g f_i(Y_s | X_s, X_{s^*}, \alpha_{si}) \quad (2.3)$$

Obviously,  $p(X_s | \alpha_{sg})$  is the CAR model integrated with logistic function, which is called LCAR model. Essentially, LCAR is determined by estimating parameters  $\alpha_{sg}$ . As a convention, let

$$p_{sg} = p(X_s | \alpha_{sg}) \\ \sum_{g=1}^G p_{sg} = 1 \quad (2.4)$$

The logarithm of  $p_{sg}$  which is called logit of  $p_{sg}$  is defined as follows:

$$\text{logit}(p_{sg}) = \log\left(\frac{p_{sg}}{1 - p_{sg}}\right) = \alpha_{sg}^T X_s + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\alpha_{sg}^T X_{s^*} - \alpha_{s^*g}^T X_{s^*}) \quad (2.5)$$

Suppose  $Y_s$  obeys multinomial distribution so that the parameters  $\alpha_{sg}$  at site  $s$  are determined by maximum likelihood estimation (MLE) method (Czepiel, 2002). The joint probability of the global dataset  $(\mathbf{X}, \mathbf{Y})$  in reduced form is:

$$f(\mathbf{X}, \mathbf{Y} | \alpha) = \prod_{s=1}^{|\mathcal{S}|} \prod_{g=1}^G \prod_{i=1, y_s^{(i)}=v_g}^{N_s} p_{sg}^{(i)}$$

Where  $p_{sg}^{(i)}$  is the evaluation of  $p_{sg}$  with the instance  $X_s = \mathbf{x}_s^{(i)}$ . The log-likelihood function is logarithm function of such joint probability:

$$l(\alpha) = - \sum_{s=1}^{|S|} \sum_{g=1}^G \sum_{i=1, y_s^{(i)}=v_g}^{N_s} \log \left( 1 + \exp \left( - \left( \alpha_{sg}^T \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \left( \alpha_{sg}^T \mathbf{x}_{s^*}^{(i)} - \alpha_{s^*g}^T \mathbf{x}_{s^*}^{(i)} \right) \right) \right) \right)$$

Let  $\alpha_{sg}^* = (\alpha_{sg1}^*, \alpha_{sg2}^*, \dots, \alpha_{sgk}^*)$  be the optimal estimate of regressive coefficients which is a maximizer of  $l(\alpha)$  with the constraint  $\sum_{g=1}^G p_{sg} = 1$  which is equivalent to:

$$\log \left( \sum_{g=1}^G p_{sg} \right) = 0$$

Let  $L_0(\alpha, \lambda)$  be Lagrange function defined as follows:

$$L_0(\alpha, \lambda) = l(\alpha) + \sum_{s=1}^{|S|} \sum_{g=1}^G \sum_{i=1, y_s^{(i)}=v_g}^{N_s} \lambda_i^{(sg)} \log(p_{sg}^{(i)})$$

Where  $\lambda_i$  are Lagrange multipliers such that  $\lambda_i \geq 0$  with note that each  $\lambda_i^{(sg)}$  implies  $\lambda_i$  at site  $s$  given value  $v_g$ . Because logarithm is concave function, we apply Jessen inequation to approximate the Lagrange function as follows:

$$L_0(\alpha, \lambda) \leq L(\alpha, \lambda) = l(\alpha) + \sum_{s=1}^{|S|} \sum_{g=1}^G \sum_{i=1, y_s^{(i)}=v_g}^{N_s} \lambda_i^{(sg)} \log(p_{sg}^{(i)})$$

We now focus on maximizing  $L(\alpha, \lambda)$ . The first-order partial derivative of  $L(\alpha, \lambda)$  with regard to  $\alpha_{sg}$  is:

$$\frac{\partial L(\alpha, \lambda)}{\partial \alpha_{sg}} = \sum_{i=1, y_s^{(i)}=v_g}^{N_s} (\lambda_i + 1) (1 - p_{sg}^{(i)}) \left( (\mathbf{x}_s^{(i)})^T + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} (\mathbf{x}_{s^*}^{(i)})^T \right)$$

Note,  $a_{ss^*} = 0$  if  $y_{s^*}^{(i)}$  or  $\mathbf{x}_{s^*}^{(i)}$  does not exist. Let

$$\mathbf{Z}_s^{(i)} = \mathbf{x}_s^{(i)} + \rho \sum_{s^* \in \text{nb}(s)} a_{ss^*} \mathbf{x}_{s^*}^{(i)} \quad (2.6)$$

We have:

$$\frac{\partial L(\alpha, \lambda)}{\partial \alpha_{sg}} = \sum_{i=1, y_s^{(i)}=v_g}^{N_s} (\lambda_i + 1) (1 - p_{sg}^{(i)}) (\mathbf{Z}_s^{(i)})^T$$

Setting the first-order partial derivative of  $L(\alpha, \lambda)$  with regard to  $\lambda_i$  to be 0, we obtain:

$$\frac{\partial L(\alpha, \lambda)}{\partial \lambda_i} = \lambda_i \sum_{g=1}^G \log(p_{sg}^{(i)}) = 0 \Rightarrow \lambda_i = 0$$

Substituting all  $\lambda_i = 0$  into the derivative  $\frac{\partial L(\alpha, \lambda)}{\partial \alpha_{sg}}$ , we have:

$$\frac{\partial L(\alpha, \lambda)}{\partial \alpha_{sg}} = \sum_{i=1, y_s^{(i)}=v_g}^{N_s} (1 - p_{sg}^{(i)}) (\mathbf{Z}_s^{(i)})^T$$

By setting such derivative  $\frac{\partial l(\alpha)}{\partial \alpha_{sg}}$  to be zero, we obtain:

$$\sum_{i=1, y_s^{(i)}=v_g}^{N_s} p_{sg}^{(i)} Z_s^{(i)} = \sum_{i=1, y_s^{(i)}=v_g}^{N_s} Z_s^{(i)}$$

As a result, the estimates  $\alpha_{sg}^*$  for all regressive coefficient vectors  $\alpha_{sg}$  over all sites and all values  $v_1, v_2, \dots, v_G$  is solution of the linear equation system consisting of  $|S|G$  equations as follows:

$$\sum_{i=1, y_s^{(i)}=v_g}^{N_s} p_{sg}^{(i)} Z_s^{(i)} = \sum_{i=1, y_s^{(i)}=v_g}^{N_s} Z_s^{(i)} \quad \left|_{s=1, |S|, g=1, \overline{G}} \right. \quad (2.7)$$

In general, the equation 2.7 is used to estimate LCAR model, which can be solved by numerical methods like Newton-Raphson method for nonlinear equation system (Burden & Faires, 2011, pp. 638-642).

### 3. Conclusions

In this research, I proposed a simple integration model for spatial-temporal autoregression which consists of two consequent and separated steps. Moreover, I also proposed the integration of logistic model and autoregressive model called LCAR which is slightly different from traditional CAR model. However, the traditional AR model known as temporal autoregressive model is still simple with assumption of first-order Markov property. In the future trend, I will improve AR model with autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA).

### References

- Bender, R., & Grouven, U. (1997, September 5). Ordinal logistic regression in medical research. *Journal of the Royal College of Physicians of London*, 31(5), 546-551. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/9429194/>
- Burden, R. L., & Faires, D. J. (2011). *Numerical Analysis* (9th Edition ed.). (M. Julet, Ed.) Brooks/Cole Cengage Learning.
- Czepiel, S. A. (2002). *Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation*. Czepiel's website <http://czep.net>.
- Eshel, G. (2003). *The Yule Walker Equations for the AR Coefficients*. University of Wharton, Statistics Department. Michael Steele Homepage. Retrieved 11 24, 2018, from <http://www-stat.wharton.upenn.edu/~steele/Courses/956/ResourceDetails/YWSourceFiles/YW-Eshel.pdf>
- Mariella, L., & Tarantino, M. (2010). Spatial Temporal Conditional Auto-Regressive Model: A New Autoregressive Matrix. *Austrian Journal of Statistics*, 39(3), 223-244. doi:10.17713/ajs.v39i3.246