UNIVERSITAS INDONESIA

ESTIMASI SKOR PROPENSITAS MENGGUNAKAN METODE VARIASIONAL PADA REGRESI LOGISTIK SPASIAL

MAKALAH SEMINAR

HAKIIM NUR RIZKA
1706047391

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PROGRAM SARJANA STATISTIKA
DEPOK
OKTOBER 2020

HALAMAN PERNYATAAN ORISINALITAS

Makalah ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama : Hakiim Nur Rizka

NPM : 1706047391

Tanda Tangan : ____________________

Tanggal : 27 Oktober 2020

HALAMAN PENGESAHAN

Makalah ini diajukan oleh :

Nama : Hakiim Nur Rizka

NPM : 1706047391

Program Studi : Statistika

Judul Makalah : Estimasi Skor Propensitas Menggunakan Metode Variasional pada Regresi Logistik Spasial

Telah dipresentasikan pada simposium “6th International Symposium on Current Progress in Mathematics and Sciences (ISCPMS)” dan telah diajukan pada prosiding “Proceedings of the 6th International Symposium on Current Progress in Mathematics and Sciences 2020 (ISCPMS 2020)”.

MENGETAHUI

Pembimbing : Dr. Dra. Yekti Widyaningsih, M.Si (...................)

Ditetapkan di : Depok

Tanggal : 27 Oktober 2020

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI MAKALAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama	: Hakiim Nur Rizka
NPM	: 1706047391
Program Studi	: Statistika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis karya	: Makalah Seminar

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul :

Estimasi Skor Propensitas Menggunakan Metode Variasional pada Regresi Logistik Spasial

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan makalah saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 27 Oktober 2020

Yang menyatakan

(Hakiim Nur Rizka)

Propensity Score Estimation Using Variational Method on Spatial Logistic Regression

H. N. Rizka and Y. Widyaningsih[a)]

Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

[a)]Corresponding author: yekti@sci.ui.ac.id

Abstract. Propensity score can be described as a probability of certain treatments conditional to the given observed covariates. Propensity score is one of the known methods to allows an observational study emulating certain characteristics from that of a randomized trial. The most common method used to estimate this score is the logistic regression model. Logistic regression can be used to model the probability of a certain event. With the advancement that is happening to spatial statistics, one can also build a logistic regression model that takes into consideration to that of spatial dependence. Thus, accommodate the spatial effect that is likely happening on observation data that came from different places. Problem arises from this model, that is the estimation of the parameters on the spatial logistic model. EM algorithm which is needed for this problem, still requires another adjustment since the expectation in the E-step is not available in closed form. Variational method modification is then proposed as an alternative for this problem. This paper reviews the propensity score estimation using spatial logistic regression and discusses the variational method as an alternative method to tackle the problem in estimating the parameters on the spatial logistic regression model in a theoretical study.

Keywords: Propensity score, spatial logistic regression, variational method

INTRODUCTION

Propensity score is a conditional probability of a certain treatment given the observed baseline characteristics [1]. In a more specific case, the score estimates the probability of a certain event using the information from the covariates. This score enables one to emulate certain characteristics from a randomized trial on observational study. In other words, it can eliminate the effect that came from confounding factors on observational data.

The method that is commonly used to estimate the propensity score is that of a regression-based model. In particular, this paper focuses on the logistic model. Logistic regression allows predicting the probability of a certain event. Briefly introducing the concept, the probability of "success" given some covariates from the logistic regression is used as the predictor for propensity score. A formal definition of this has been proposed by many and this paper specifically refers to Austin et al. in 2011 and Thavanesvaran et al. in 2009 [1, 2].

The recent study of Hilwin Nisa' et al. in 2019 proposed to take into consideration of the spatial effect on the observational data. This is mostly due to, in practical use, observational data can be taken from different places. By considering the spatial effect, the usual logistic regression does not sufficiently model the data. Thus, spatial logistic regression is proposed as the alternative for accommodating spatial effect [3].

The parameter estimation on spatial logistic regression has a critical intricacy regarding the usage of the maximum likelihood estimate. One of the earliest alternatives on the likelihood-based estimation of the parameter on spatial logistic regression was a numerical method called expectation-maximization (EM) algorithm. EM algorithm uses an iterative method to get the estimation for the parameters in the spatial logistic regression. This iterative procedure uses a form of expectation of the log-likelihood from the model and in the case of the spatial logistic regression model, this expectation is not available in the closed-form. There are some alternatives that have been proposed for this problem such as using Monte Carlo procedures [4] and Laplace approximation to approximate the intractable integral [5].

Rather than conventional EM, a modification of EM method is used in this paper. The name of the method is variational EM [6]. Variational EM uses the lower bound of likelihood function as the new objective function, then optimize this function to get the estimation on the parameters. The variational estimation procedure will always converge to a sufficiently good estimation of the real parameters [7].

This paper will review the propensity score estimation using spatial logistic regression from Hilwin Nisa' et al. in 2019 [3] then modifies the parameter estimation procedures using the variational method from Hardouin et al. in 2019 [6]. As a note, this paper focuses on the theoretical study and the application is not going to be discussed in this paper. For application of the propensity score estimation see for instances Mccafrey et al. in 2004, Beal et al. in 2014 [8, 9].

BERNOULLI DISTRIBUTION

Consider a density function $f(x)$ from a discrete random variable X defined as

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} , x = 0,1,2, \dots, \tag{1}$$

Random variable X is said to be distributed under binomial distribution and it can be written with the notation $X \sim Bin(n,p)$ and p can be defined as "probability of success". A binomial distribution with $n = 1$ is also known as bernoulli distribution. The distribution function for bernoulli distribution can be written as

$$f(x) = p^x (1-p)^{1-x} , x = 0,1 \tag{2}$$

and if a random variable X is distributed under bernoulli distribution, it can be stated as $X \sim Ber(p)$.

LOGISTIC REGRESSION

Logistic regression is one of the methods to model a binary dependent variable with one or more independent variables. Logistic regression builds an equation based on the so-called logistic function. The logistic function can be defined as :

$$f(y) = \frac{e^y}{1+e^y} \tag{3}$$

this function maps $-\infty < y < \infty$ to $0 \leq f(y) \leq 1$. In logistic regression, it is common to use the logarithmic of the logistic function as it leads to an equation of dependent variable with a linear combination of the covariates matrix $\boldsymbol{X}$ with and the regression coefficients vector $\boldsymbol{\beta}$. Suppose there are n observations and j independent variables. For i-th observation its equation can be written as :

$$y_i = \alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + \dots + x_{ji}\beta_j , \;\; i = 1, 2, \dots, n \tag{4}$$

with $\boldsymbol{y}' = [y_1 , y_2, \dots, y_n]$ as vector of value from the dependent variable. If (4) is applied to (3), it is then defined as the probability of i-th observation will be "success" that is :

$$\pi_i = \frac{e^{y_i}}{1+e^{y_i}} = \frac{1}{1+\exp[-(\alpha + x_{1i}\beta_1 + x_{2i}\beta_2 + \cdots + x_{ji}\beta_j)]}, \quad i = 1,2,\ldots,n \tag{5}$$

the form (4) can also be stated as $P(y_i = 1 | X_i)$. As a side note (4) is also strictly increasing function of β_k, $k = 1,2, \ldots,j$.

Regarding the estimation for the parameter in logistic regression, it is common to use the maximum likelihood method. First of all, construct the likelihood function. Consider the parameter that is going to be estimated is $\boldsymbol{\beta}$, then given n observational data the likelihood function for the logistic regression is

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} {\pi_i}^{y_i} \times (1-\pi_i)^{(1-y_i)} \\ L(\boldsymbol{\beta}) &= \prod_{i=1}^{n} \left(\frac{1}{1+e^{-y_i}}\right)^{y_i} \times \left(\frac{1}{1+e^{y_i}}\right)^{(1-y_i)} \end{aligned} \tag{6}$$

then the log-likelihood function for this model can be defined as

$$\begin{aligned} \log L(\boldsymbol{\beta}) &= \log\left[\prod_{i=1}^{n} \left(\frac{1}{1+e^{-y_i}}\right)^{y_i} \times \left(\frac{1}{1+e^{y_i}}\right)^{(1-y_i)}\right] \\ l(\boldsymbol{\beta}) &= \sum_{i=1}^{n} y_i \log\left(\frac{1}{1+e^{-y_i}}\right) + (1-y_i)\log\left(\frac{1}{1+e^{y_i}}\right) \\ &= \sum_{i=1}^{n} y_i \log\left(\frac{1+e^{y_i}}{1+e^{-y_i}}\right) + \log\left(\frac{1}{1+e^{y_i}}\right) \\ &= \sum_{i=1}^{n} y_i \boldsymbol{X_i\beta} - \sum_{i=1}^{n} \log(1+e^{y_i}) \end{aligned} \tag{7}$$

To get the estimation for regression coefficients $\boldsymbol{\beta}$, the next step is to solve the derivative equation based on (7). This procedure calls for a numerical approach. It is caused by the derivative term of $\log(1+e^{y})$ is nonlinear thus makes the derivative equation

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0 \tag{8}$$

has no simple analytical solution.

SPATIAL LOGISTIC REGRESSION

To include spatial dependence in the logistic regression model, this paper follows a similar approach to that of Hardouin in 2019 [6]. First consider a two-dimensional domain $T \equiv \{\boldsymbol{s}_i,\ i = 1,2,\ldots,n\}$, which is a subset of real space, with $\boldsymbol{s}_i = (s_{i1}, s_{i2})$ for $i = 1,2,\ldots,n$. Next for the dependent variable consider $\boldsymbol{Y^*} = (y^*(\boldsymbol{s}_1), y^*(\boldsymbol{s}_2), \ldots, y^*(\boldsymbol{s}_n))$ as a process happening on T. Variable $\boldsymbol{Y^*}$ is of a bernoulli distribution with its mean depend on the process $\boldsymbol{Y} = (y(\boldsymbol{s}_1), y(\boldsymbol{s}_2), \ldots, y(\boldsymbol{s}_n))$. The assumption of independency of $\boldsymbol{Y^*}$ given $\boldsymbol{Y}$ is applied.

Then for each $\boldsymbol{s}$ element of T, the distribution of $\boldsymbol{Y^*}$ given $\boldsymbol{Y}$ can be written as

$$[Y^*(\boldsymbol{s}) | Y(\boldsymbol{s})] \sim Ber(\pi(\boldsymbol{s})) \tag{9}$$

with

$$\pi(\boldsymbol{s}) = \frac{e^{y(\boldsymbol{s})}}{1 + e^{y(\boldsymbol{s})}} \tag{10}$$

and for $Y(\boldsymbol{s})$ it is modeled as

$$Y(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta} + \mathcal{E}(\boldsymbol{s}) \tag{11}$$

with $\boldsymbol{X}(\boldsymbol{s})^T = (X_1(\boldsymbol{s}), X_2(\boldsymbol{s}), \ldots, X_j(\boldsymbol{s}))$ is the j known covariates and $\boldsymbol{\beta}$ is the coefficients regression. The linear combination of these is representing the large-scale spatial variation [6]. The term $\mathcal{E}(\boldsymbol{s})$ is a gaussian spatial process with zero mean and unknown covariance matrix $\sum$, that is

$$\mathcal{E}(\boldsymbol{s}) \sim MN(\boldsymbol{0}, \sum) \tag{12}$$

with little inspection, it can be seen that the condtional distribution in (9) seems to have different parameter π for different locational $\boldsymbol{s}$. This is caused by the notation to give attentionn to unobserved variable which refers to the term $\mathcal{E}(\boldsymbol{s})$ from (11). Thus the conditional probability can be written as

$$p\left(y^*(\boldsymbol{s_i}) = 1 \middle| y(\boldsymbol{s_i})\right) = \frac{e^{y(s_i)}}{1 + e^{y(s_i)}} \tag{13}$$

$$p\left(y^*(\boldsymbol{s_i}) = 0 \middle| y(\boldsymbol{s_i})\right) = \frac{1}{1 + e^{y(s_i)}} \tag{14}$$

Now consider the variable $Y^*(\boldsymbol{s})$ defined under $Y(\boldsymbol{s})$ as :

$$y^*(\boldsymbol{s}_i) \begin{cases} = 0, \text{ for } y(\boldsymbol{s}_i) \leq 0 \\ = 1, \text{ for } y(\boldsymbol{s}_i) > 0 \end{cases} \tag{15}$$

then considering (11) and (12) :

$$\begin{aligned} p(y(\boldsymbol{s}_i) > 0) &= p(X(\boldsymbol{s}_i)^{\boldsymbol{T}}\boldsymbol{\beta} + \mathcal{E}(\boldsymbol{s}_i) > 0) = p(\mathcal{E}(\boldsymbol{s}_i) > - X(\boldsymbol{s}_i)^{\boldsymbol{T}}\boldsymbol{\beta}) \\ &= 1 - p(\mathcal{E}(\boldsymbol{s}_i) < - X(\boldsymbol{s}_i)^{\boldsymbol{T}}\boldsymbol{\beta}) \end{aligned} \tag{16}$$

$$p(y(\boldsymbol{s}_i) \leq 0) = p(X(\boldsymbol{s}_i)^{\boldsymbol{T}}\boldsymbol{\beta} + \mathcal{E}(\boldsymbol{s}_i) \leq 0) = p(\mathcal{E}(\boldsymbol{s}_i) \leq - X(\boldsymbol{s}_i)^{\boldsymbol{T}}\boldsymbol{\beta}) \tag{17}$$

finally the marginal probability of $y^*(\boldsymbol{s}_i)$ can be written with the form [3]

$$p(y^*(\boldsymbol{s}_i) = 1) = \frac{\exp\left(\frac{X(s_i)^T\beta}{\sqrt{\Sigma_{ii}}}\right)}{1+\exp\left(\frac{X(s_i)^T\beta}{\sqrt{\Sigma_{ii}}}\right)} \tag{18}$$

$$p(y^*(\boldsymbol{s}_i) = 0) = \frac{1}{1+\exp\left(\frac{X(s_i)^T\beta}{\sqrt{\Sigma_{ii}}}\right)} \tag{19}$$

where $\sum_{\mathbf{ii}}$ is the i-th diagonal element of $\sum$. The division inside the exponential term is done due to standardization from distribution $\mathcal{E}(\boldsymbol{s}) \sim MN(\boldsymbol{0}, \sum)$.

PARAMETER ESTIMATION OF SPATIAL LOGISTIC REGRESSION

First of all, let us note that the model refers to spatial logistic regression and the parameter needed to be estimated is $\boldsymbol{\beta}$ and $\sum$. Then, this framework considers not a full bayesian approach. That is because the estimation does not assume any prior distribution for the parameter $\sum$. Lastly, notations from here on will follow the notation from Hardouin [6].

To start the procedures, the likelihood function for the model is constructed. It is the complete log-likelihood, l_c, which is going to be the objective function. The term complete refers to the complete data in which involves an observed Y^* and unobserved latent variable $\boldsymbol{\mathcal{E}}$. Using the fact that the following decomposition of distribution is hold

$$[Y^*, \boldsymbol{\mathcal{E}} \mid \boldsymbol{\beta}, \textstyle\sum] = [Y^* \mid \boldsymbol{\beta}, \boldsymbol{\mathcal{E}}][\boldsymbol{\mathcal{E}} \mid \textstyle\sum] \tag{20}$$

as such deriving the complete log-likelihood to be

$$\begin{aligned} l_c[\boldsymbol{y}^*, \boldsymbol{\mathcal{E}} \mid \boldsymbol{\beta}, \textstyle\sum] &= log[\boldsymbol{y}^* \mid \boldsymbol{\beta}, \boldsymbol{\mathcal{E}}] + log[\boldsymbol{\mathcal{E}} \mid \textstyle\sum] \\ &= -\textstyle\sum_{s\in T} \log\left(1 + e^{y(\boldsymbol{s})}\right) + \textstyle\sum_{s\in T} y(\boldsymbol{s}) y^*(\boldsymbol{s}) \\ &\quad - \frac{n}{2}\log 2\pi - \frac{1}{2}[\log(\det \boldsymbol{\Sigma}) + \boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}\boldsymbol{\varepsilon}] \end{aligned} \tag{21}$$

then proceeding with the maximum likelihood method. With the likelihood above, there is a similar problem to parameter estimation on logistic regression that comes from the term $\log(1+e^y)$. It is then the variational EM algorithm [6] can be performed as an alternative method to obtain the estimation of parameters.

In principle, the variational method is an approximation method and it can be applied to estimate the complete log-likelihood from (21). Variational method dictates to approximate the posterior distribution with a simpler distribution which is called variational distribution or variational lower bound. The iteration aims to optimize this lower bound then use it as the estimator for the likelihood function.

The work which most of this paper has cited from Hardouin [6] has explained the details and derivations of the variational lower bound for complete log-likelihood in (21). To simplify, it used the fact that for a logistic function

$$g(y) = \frac{1}{1+e^{-y}} \tag{22}$$

Jaakola et al. in 2000 [10] give an inequality of

$$\log g(y) \geq \log g(\tau) + \frac{y-\tau}{2} - \lambda(\tau)(y^2 - \tau^2) \tag{23}$$

$$\lambda(\tau) = \frac{g(\tau)-1/2}{2\tau} \tag{24}$$

and applying (23) to (21) to get the variational lower bound for the complete log-likelihood. This lower bound is going to be the new objective function. Skipping the detailed derivation from Hardouin [6], the new objective function is

$$\tilde{l}_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon} \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] = T_1(\boldsymbol{\tau}) + T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}^{\boldsymbol{T}}\boldsymbol{M} - \frac{1}{2}[\log(\det \boldsymbol{\Sigma}) + \boldsymbol{\varepsilon}^{\boldsymbol{T}}\mathbf{W}^{-\mathbf{1}}\boldsymbol{\varepsilon}] + const. \tag{25}$$

$$T_1(\boldsymbol{\tau}) = \textstyle\sum_{s\in T} \log g(\boldsymbol{\tau}(\boldsymbol{s})) - \frac{\tau(s)}{2} + \boldsymbol{\tau}(\boldsymbol{s})^2 \lambda(\boldsymbol{\tau}(\boldsymbol{s})) \tag{26}$$

$$T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) = \textstyle\sum_{s\in T} -\lambda\left(\tau(\boldsymbol{s})\right)[\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta}]^2 - [\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta}][y^*(\boldsymbol{s}) - \frac{1}{2}] \tag{27}$$

$$\boldsymbol{W}^{-1} = \boldsymbol{\Sigma}^{-1} + 2\Lambda(\boldsymbol{\tau}) \tag{28}$$

$$\Lambda(\boldsymbol{\tau}) = diag[\lambda(\boldsymbol{\tau})] \tag{29}$$

it introduced new parameters $\boldsymbol{\tau}$ called variational parameters. Also note that $\boldsymbol{\tau} = (\tau(s_1), \tau(s_2), ..., \tau(s_n))$. The constant term represents $(-\frac{n}{2}\log 2\pi)$ and left written without much regard as it does not include any parameter. It also introduced $\boldsymbol{M} = (M(s_1), M(s_2), ..., M(s_n))$ with

$$M(\boldsymbol{s}) = y^*(\boldsymbol{s}) - \frac{1}{2} - 2\lambda(\tau(\boldsymbol{s}))\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}} \tag{30}$$

with inspection of the new objective function (23), it can be seen that there is a proportionality from conditional distribution $[\boldsymbol{\varepsilon}|Y^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ with $[Y^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ for fixed $\boldsymbol{\tau}$. Writing the proportionality of conditional distribution $[\boldsymbol{\varepsilon}|Y^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ as :

$$\begin{aligned} p[\boldsymbol{\varepsilon}|\boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] \ &\propto \exp\left\{T_1(\boldsymbol{\tau}) + \ T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \frac{1}{2}\boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{W}^{-1}\boldsymbol{\mu}\right\} \\ &\times \frac{1}{\sqrt{\log(\det \boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\boldsymbol{\varepsilon} - \boldsymbol{\mu})^{\boldsymbol{T}}\mathbf{W}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{\mu})\right\} \end{aligned} \tag{31}$$

with $\boldsymbol{\mu} = \boldsymbol{WM}$. Considering the right side from (31), it can be seen that the conditional distribution $[\boldsymbol{\varepsilon}|Y^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]$ has the shape of multivariate normal distribution MN($\boldsymbol{\mu}$, $\boldsymbol{W}$). It can also be written as :

$$[\boldsymbol{\varepsilon}|Y^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] = MN(\boldsymbol{\mu}, \boldsymbol{W}) \tag{32}$$

The next step is to proceed with the new objective function (25), which is using the $\tilde{l}_c$ as the approximation of l_c. The inequality from these two functions is

$$l_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]\ \geq \tilde{l}_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] \tag{33}$$

which is directly caused by applying (23) to (21). To add to the inequality (33) suppose the variational method gives the approximation for the parameter as $\boldsymbol{\beta}_{\boldsymbol{max}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{max}}$, also the last update for the variational parameter is given by $\boldsymbol{\tau}_{\boldsymbol{max}}$. The previous sentence means that the objective function has been updated as

$$\tilde{l}_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}_{\boldsymbol{max}}, \boldsymbol{\Sigma}_{\boldsymbol{max}}, \boldsymbol{\tau}_{\boldsymbol{max}}] \tag{34}$$

and it extends the inequality from (33) to

$$l_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]\ \geq\ \tilde{l}_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}_{\boldsymbol{max}}, \boldsymbol{\Sigma}_{\boldsymbol{max}}, \boldsymbol{\tau}_{\boldsymbol{max}}]\ \geq \tilde{l}_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\ \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}] \tag{35}$$

which is a way to say that by optimizing (25), the approximation will be close to that original complete log-likelihood from (21).

The next procedures aim to optimize the new objective function (25). To work with the new objective function (25) the method is going to simply improvise from the conventional EM algorithm. The basic of the EM algorithm for each step can be simplified into two actions: computing the expectation from likelihood function and maximizing the expectation to obtain estimation of the parameter.

With the new objective function (25) also considering (32) Hardouin in 2019 deduced a form for the expectation of the complete log-likelihood as :

$$\begin{aligned} E\ \{l_c[\boldsymbol{y}^*, \boldsymbol{\varepsilon}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\tau}]|\boldsymbol{y}^*, \boldsymbol{\beta}, \boldsymbol{\Sigma}\} &= \ T_1(\boldsymbol{\tau}) + \ T_2(\boldsymbol{\tau}, \boldsymbol{\beta}) + \ \boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{M} \\ &- \frac{1}{2}tr[(\boldsymbol{W} + \boldsymbol{\mu}\boldsymbol{\mu}^{\boldsymbol{T}})\boldsymbol{W}^{-1}] - \frac{1}{2}\log(\det \boldsymbol{\Sigma}) + c \end{aligned} \tag{36}$$

where tr($\boldsymbol{A}$) is trace of matrix $\boldsymbol{A}$. Then, it proceeds with the maximization action to obtain the estimation for the parameter. Since the method also including the new parameter, it finishes the whole step by updating the variational parameter. As the main idea has been given, the next part of this section is going to specify iterative procedures from initialization to the updating procedure for the variational parameter.

To start, initialize the value for $\widehat{\boldsymbol{\beta}}^{(\mathbf{0})}$, $\widehat{\boldsymbol{\Sigma}}^{(\mathbf{0})}$, and $\widehat{\boldsymbol{\tau}}^{(\mathbf{0})}$. For covariance matrix, Hardouin used an exponential covariance matrix with element $\widehat{\boldsymbol{\Sigma}}^{(\mathbf{0})}{}_{\boldsymbol{ij}}$ = C($\mathbf{s_i}$-$\mathbf{s_j}$) where C($\mathbf{h}$) = exp(|$\mathbf{h}$|/θ). It is also recommended to initialize the value for variational parameters which satisfy $\tau(\boldsymbol{s})^2 = y(\boldsymbol{s})^2$ for all $\mathbf{s} \in D$; that is, it can be initialized with $\hat{\tau}^{(0)}(\boldsymbol{s}) = [X(\boldsymbol{s})^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}^{(\mathbf{0})} + \eta(\boldsymbol{s})] \times (2z-1)$, where $\eta(\boldsymbol{s})$ ~ i.i.d of N(0,1). Then for the l-th (l=1,2, ...) iteration follow these procedures [6] :

Compute $\widehat{\boldsymbol{W}}_1^{(l-1)} = \boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l}-\mathbf{1})}\right)$, $\widehat{\boldsymbol{M}}_1^{(l-1)} = \boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l}-\mathbf{1})}\right)$, and $\widehat{\boldsymbol{\mu}}_1^{(l-1)} = \widehat{\boldsymbol{W}}_1^{(l-1)}\widehat{\boldsymbol{M}}_1^{(l-1)}$.

Compute $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\beta}}[T_2\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\beta}\right) + (\widehat{\boldsymbol{W}}_1^{(l-1)}\boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\beta}\right))^{\boldsymbol{T}}\boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\beta}\right)]$. To do this, consider the function $T(\boldsymbol{\beta}) = \sum_{\boldsymbol{s}\in T} -\lambda\left(\tau(\boldsymbol{s})\right)[\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta}]^2 - [\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta}][y^*(\boldsymbol{s}) - \frac{1}{2} - 2\lambda\left(\tau(\boldsymbol{s})\right)\hat{\mu}(\boldsymbol{s})]$ and maximizing this by solving the derivative equation $\frac{\partial T(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$. Then consider the derivative function $G(\boldsymbol{\beta}) = \frac{\partial T(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{\boldsymbol{s}\in T}\left\{-2\lambda\left(\tau(\boldsymbol{s})\right)[\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}\boldsymbol{\beta}] + y^*(\boldsymbol{s}) - \frac{1}{2} - 2\lambda\left(\tau(\boldsymbol{s})\right)\hat{\mu}(\boldsymbol{s})\right\}\boldsymbol{X}(\boldsymbol{s})$. If $\boldsymbol{\beta}$ has one or two dimensions, solving $G(\boldsymbol{\beta}) = 0$ does not need special attention. Otherwise, solve :

$\widehat{\boldsymbol{\beta}}^{(\boldsymbol{k})} = \widehat{\boldsymbol{\beta}}^{(\boldsymbol{k}-\mathbf{1})} - \left[\frac{\partial G(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]^{-\mathbf{1}}_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(\boldsymbol{k}-\mathbf{1})}} \partial G\left(\widehat{\boldsymbol{\beta}}^{(\boldsymbol{k}-\mathbf{1})}\right)$, $\frac{\partial G(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{\boldsymbol{s}\in T} -2\lambda\left(\tau(\boldsymbol{s})\right)\boldsymbol{X}(\boldsymbol{s})\boldsymbol{X}(\boldsymbol{s})^{\boldsymbol{T}}$, until $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{k})} \cong \widehat{\boldsymbol{\beta}}^{(\boldsymbol{k}-\mathbf{1})}$, and use $\widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})} = \widehat{\boldsymbol{\beta}}^{(\boldsymbol{k})}$.

Then update the objective functions $\widehat{\boldsymbol{M}}_2^{(l-1)} = \boldsymbol{M}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right)$ and $\widehat{\boldsymbol{\mu}}_2^{(l-1)} = \widehat{\boldsymbol{W}}_1^{(l-1)}\widehat{\boldsymbol{M}}_2^{(l-1)}$

Compute $\widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\Sigma}}\left\{-\frac{1}{2}\, tr\left[\left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\Sigma}\right) + \left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\Sigma}\right)\widehat{\boldsymbol{M}}_2^{(l-1)}\right)\left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\Sigma}\right)\widehat{\boldsymbol{M}}_2^{(l-1)}\right)^{\boldsymbol{T}}\right)\boldsymbol{W}^{-\mathbf{1}}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \boldsymbol{\Sigma}\right)\right] - \frac{\mathbf{1}}{\mathbf{2}}\log(\det\boldsymbol{\Sigma})\right\}$. To do this, change the term $\boldsymbol{\Sigma}$ function and consider this as $\sigma_\varepsilon^2\boldsymbol{Q}$. And so, minimize

$$f(\boldsymbol{Q}, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon^2}\left[\left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \sigma_\varepsilon^2\boldsymbol{Q}\right) + \left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \sigma_\varepsilon^2\boldsymbol{Q}\right)\widehat{\boldsymbol{M}}_2^{(l-1)}\right)\left(\boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \sigma_\varepsilon^2\boldsymbol{Q}\right)\widehat{\boldsymbol{M}}_2^{(l-1)}\right)^{\boldsymbol{T}}\right)\boldsymbol{W}^{-\mathbf{1}}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \sigma_\varepsilon^2\boldsymbol{Q}\right)\right] + n\log\det\sigma_\varepsilon^2\boldsymbol{Q}$$

respect to σ_ε^2 and $\boldsymbol{Q}$.

Then update the objective functions $\widehat{\boldsymbol{W}}_2^{(l-1)} = \boldsymbol{W}\left(\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l}-\mathbf{1})}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right)$ and $\widehat{\boldsymbol{\mu}}_3^{(l-1)} = \widehat{\boldsymbol{W}}_2^{(l-1)}\widehat{\boldsymbol{M}}_2^{(l-1)}$

Update variational parameters with $\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l})} = \arg\max_{\boldsymbol{\tau}}\left\{T_1(\boldsymbol{\tau}) + T_2\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right) + \left(\boldsymbol{W}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right)\boldsymbol{M}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right)\right)^{T}\boldsymbol{M}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right) - \frac{1}{2}\, tr\left[\left(\boldsymbol{W}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right) + \left(\boldsymbol{W}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right)\boldsymbol{M}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right)\right)\left(\boldsymbol{W}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right)\boldsymbol{M}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\beta}}^{(\boldsymbol{l})}\right)\right)^{\boldsymbol{T}}\right)\boldsymbol{W}^{-\mathbf{1}}\left(\boldsymbol{\tau}, \widehat{\boldsymbol{\Sigma}}^{(\boldsymbol{l})}\right)\right]\right\}$

or use the following closed-form equation : $\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l})}(\boldsymbol{s}) = \sqrt{\widehat{\boldsymbol{\tau}}^{(\boldsymbol{l})}(\boldsymbol{s})^{\mathbf{2}}} \times (2y^*(\boldsymbol{s}) - 1)$; $\hat{\tau}^{(l)}(\boldsymbol{s})^{\mathbf{2}} = [X(\boldsymbol{s})^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}^{(l)}]^2 + 2\left[X(\boldsymbol{s})^{\boldsymbol{T}}\widehat{\boldsymbol{\beta}}^{(l)}\right]\hat{\mu}_3^{(l-1)}(\boldsymbol{s}) + \widehat{\boldsymbol{W}}_{ss}^{(\boldsymbol{l}-\mathbf{1})} + \hat{\mu}_3^{(l-1)}(\boldsymbol{s})^2$

$\widehat{\boldsymbol{W}}_{ss}^{(\boldsymbol{l}-\mathbf{1})}$ is the s-th diagonal element of $\widehat{\boldsymbol{W}}_2^{(\boldsymbol{l}-\mathbf{1})}$.

PROPENSITY SCORE

Following the propensity model from Thavanesvaran et al. [2], the propensity score can be estimated, given the covariates matrix X, as the conditional probability of a certain observation being assigned to a particular treatment. Consider the same notation from previous sections, then for the i-th observation the propensity score is given by :

$$e\left(X(\boldsymbol{s}_i)\right) = p\left(y^*(\boldsymbol{s_i}) = 1 \middle| X(\boldsymbol{s_i})\right) \tag{37}$$

as it already regards the spatial dependence, then proceeds with the spatial logistic regression model, the estimated probability of a particular observation being assigned as "success" (18) to get

$$e\big(X(\boldsymbol{s}_i)\big) = \frac{\exp\left(\frac{X(s_i)^T\beta}{\sqrt{\Sigma_{ii}}}\right)}{1+\exp\left(\frac{X(s_i)^T\beta}{\sqrt{\Sigma_{ii}}}\right)}$$

with the parameters $\boldsymbol{\beta}$ and $\sum$ using variational EM estimator $\widehat{\boldsymbol{\beta}}$and $\widehat{\boldsymbol{\Sigma}}$ specified in the previous section.

CONCLUSION

This paper mentions the problem of estimating parameters on spatial logistic regression that is caused by the term log*(1+exp(-y))* and tackle this problem with the variational EM algorithm. This paper also specified the initialization and *l*-th step from the variational EM algorithm to estimate the parameters in the spatial logistic regression model. Lastly for the propensity score given an observational covariates $X(\boldsymbol{s}_i)$ is

$$e\big(X(\boldsymbol{s}_i)\big) = \frac{\exp\left(\frac{X(s_i)^T\beta}{\Sigma_{ii}}\right)}{1+\exp\left(\frac{X(s_i)^T\beta}{\Sigma_{ii}}\right)}$$

ACKNOWLEDGMENTS

This research was funded by Directorate of Research and Development of Universitas Indonesia (DRPM UI) as a grant of Publikasi Terindeks Internasional (PUTI) Saintekes 2020 No. NKB-2429/UN2.RST/HKP.05.00/2020.

REFERENCES

1. P. C. Austin, Multivariate Behav. Res. **46**, 399-424, (2011).
2. A. Thavaneswaran and L. Lix, *Propensity Score Matching in Observational Studies,* available at http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/protocol/media/propensity_score_matching.pdf. (2008).
3. H. Nisa', M. B. T. Mitakda and S. Astutik, IOP Conf. Ser.: Mater. Sci. Eng. **546,** 052048 (2019).
4. O. Cappé, E. Moulines and T. Rydén, *Inference in Hidden Markov Models*, (New York: Springer-Verlag, 2005).
5. A. Sengupta and N. Cressie, Spat. Stat. **4**, 14-44 (2013).
6. C. Hardouin, Spat. Stat. **31**, 100365 (2019).
7. D. M. Blei, A. Kucukelbir and J. D. McAuliffe, J. Am. Stat. Assoc. **112**, 859-877 (2017).
8. D. F. Mccafrey, G. Ridgeway and A. R. Morral, Psychol. Methods **9**, 403-425 (2004).
9. S. J. Beal and K. A. Kupzyk, J. Early Adolesc. **34**, 66-92 (2014).
10. T. S. Jaakola and M. Jordan, Stat. Comput. **10**, 25-37 (2000).