Spatial Statistics 14 (2015) 452–471

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

On regression analysis of spatial proportional data with zero/one values

Xiaoping Feng [a], Jun Zhu [a,b,*], Michelle M. Steen-Adams [c]

[a] Department of Statistics, University of Wisconsin, Madison, WI 53706, United States
[b] Department of Entomology, University of Wisconsin, Madison, WI 53706, United States
[c] Department of Environmental Studies, University of New England, Biddeford, ME 04005, United States

ARTICLE INFO

Article history:
Received 12 March 2015
Accepted 15 July 2015
Available online 5 August 2015

Keywords:
Beta-Bernoulli mixture model
Composite likelihood estimation
Gaussian latent variable
Generalized Tobit model
Godambe information
Spatial statistics

ABSTRACT

Proportions including exact zero and/or one values observed at spatial locations in a study area are often encountered in environmental and ecological studies. In this paper, we propose a new spatial beta-Bernoulli mixture model that combines a beta distribution and a Bernoulli distribution. The beta component links the original response on the open unit interval (0, 1) to covariates via regression and is flexible to capture a variety of shapes of the data distributions. Further, the Bernoulli component models the probability of zero/one values with regression. In addition, we propose a novel spatial generalized Tobit model which extends an existing spatial Tobit model by applying an inverse beta cumulative distribution function transformation. A composite likelihood approach is developed for parameter estimation by maximizing a pairwise likelihood function for each model. The standard errors of the parameter estimates are obtained via the inverse of the Godambe information matrix. A simulation study is conducted to evaluate the performance of the proposed models and methods, followed by an ecological data example. Connections among the spatial beta-Bernoulli mixture model, the spatial generalized Tobit model, and the spatial Tobit model are explored using both simulated and real data.

* Correspondence to: Department of Statistics, University of Wisconsin, 1300 University Avenue, Madison, WI 53706, United States.

E-mail addresses: fengx@stat.wisc.edu (X. Feng), jzhu@stat.wisc.edu (J. Zhu), msteenadams@une.edu (M.M. Steen-Adams).

http://dx.doi.org/10.1016/j.spasta.2015.07.007
2211-6753/

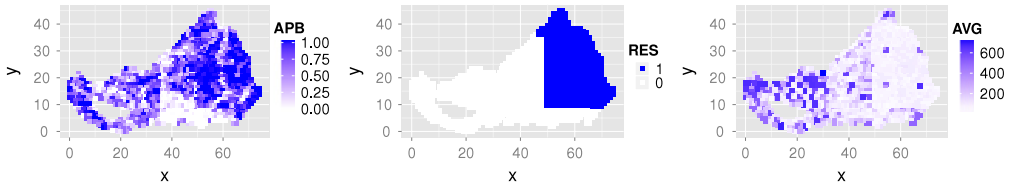

Fig. 1. Data example: maps of the proportions of aspen–paper birch (APB) (left), *reservation* (middle), and *average size of all parcels* (right) in $N = 1429$ quarter sections of a study area in northern Wisconsin.

1. Introduction

Of interest is regression analysis of spatial proportional data with zero and/or one values (henceforth, zero/one values). The motivating example is an integrative landscape ecology and environmental history study that aims to assess the influence of past land-ownership characteristics on landscape structure in northern Wisconsin, USA. The raw data consist of proportions of aspen (*Populus* spp.)–paper birch (*Betula papryfera*) (APB) aerial extent in each quarter section with about 12% of zero/one values, corresponding to no or all APB in those quarter sections (Fig. 1). An objective of the study is to evaluate the influence of various ownership covariates on APB. Standard linear regression analysis may not be appropriate to relate the response to covariates, as the probability model does not restrict the range of the response to the unit interval and the spatial dependence is not accounted for. The purpose of this paper is to develop new statistical models and their inference for the regression analysis of spatial proportional data with zero/one values.

The family of beta distributions characterized by two shape parameters is flexible for modeling proportions on the open unit interval $(0, 1)$, since the density can take on a variety of shapes for different values of the shape parameters. Several recent studies have considered beta regression models. For independent data, Ferrari and Cribari-Neto (2004) proposed a regression model where the response variable follows a beta distribution parameterized by a regression mean and a precision parameter, enabling a more straightforward interpretation of the regression parameters. Simas et al. (2010) extended this model to allow regression for the precision parameter as well. For dependent data, Eskelson et al. (2011) and Guolo and Varin (2014) proposed beta regression models to accommodate spatial dependence and temporal dependence, respectively. However, the aforementioned studies did not consider the possibility of zero/one values in the data.

In practice, the proportional data may contain exact zero/one values and thus are on the closed unit interval $[0, 1]$ (or half-closed intervals $[0, 1)$ and $(0, 1]$). For independent data, Ospina and Ferrari (2010, 2012) described a class of models based on a mixture of continuous and discrete distributions with probability mass at zero/one. For spatial proportional data, Feng et al. (2014) proposed a spatial Tobit model, and the distribution of zero/one values are handled by the thresholding parameter and the regression mean. Although the spatial Tobit model is a natural extension of a spatial probit model for spatial binary data (Heagerty and Lele, 1998; De Oliveira, 2000), the Gaussian latent variable restricts the distribution to be unimodal in the open interval $(0, 1)$, which may not be appropriate for some data distributions in practice (Kieschnick and McCullough, 2003).

Although beta regression models are well established for the open unit interval, it remains a challenge to model spatial proportional data observed on the closed unit interval. We will propose a new spatial beta-Bernoulli mixture model that accommodates exact zero/one values and takes into account spatial dependence in the spatial proportional data using a Gaussian latent process. Compared with the existing spatial Tobit model, the spatial beta-Bernoulli mixture model captures a wider variety of shapes of data distributions on the open interval $(0, 1)$. In addition, we extend the spatial Tobit model by applying an inverse beta cumulative distribution function (CDF) transformation on the Gaussian latent variable, which will be referred as a spatial generalized Tobit model. Similar to the spatial Tobit model, the resulting new model is based on the idea of thresholding a spatial latent process, but overcomes the limitation of the Gaussian latent variable. These two new models assume two distinct data generation mechanisms and can be applied to different situations in practice.