Computational Statistics and Data Analysis 71 (2014) 464–476

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Bayesian binary regression with exponential power link

L. Naranjo[a,*], J. Martín[a], C.J. Pérez[b]

[a] *Departamento de Matemáticas, Facultad de Ciencias, Universidad de Extremadura, Avda. de Elvas s/n 06006 Badajoz, Spain*

[b] *Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, Avda. de la Universidad s/n 10003 Cáceres, Spain*

ARTICLE INFO

Article history:
Received 10 November 2011
Received in revised form 20 July 2012
Accepted 23 July 2012
Available online 1 August 2012

Keywords:
Bayesian methods
Binary data
Data augmentation
Exponential power distribution
Generalized linear models
Gibbs sampling

ABSTRACT

A flexible Bayesian approach to a generalized linear model is proposed to describe the dependence of binary data on explanatory variables. The inverse of the exponential power cumulative distribution function is used as the link to the binary regression model. The exponential power family provides distributions with both lighter and heavier tails compared to the normal distribution, and includes the normal and an approximation to the logistic distribution as particular cases. The idea of using a data augmentation framework and a mixture representation of the exponential power distribution is exploited to derive efficient Gibbs sampling algorithms for both informative and noninformative settings. Some examples are given to illustrate the performance of the proposed approach when compared with other competing models.

1. Introduction

Suppose that n independent binary random variables $Y_1, \ldots, Y_n$ are observed, where Y_i is Bernoulli distributed with success probability $\mathrm{p}(Y_i = 1|\boldsymbol{\beta}, \mathbf{x}_i) = \Psi(\mathbf{x}_i^T\boldsymbol{\beta})$. $\boldsymbol{\beta}$ is a k vector of unknown parameters, $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{ik})$ is a vector of known covariates, and Ψ is a known nonnegative function ranging between 0 and 1. The standard approach to modeling the dependence of binary data on explanatory variables under the generalized linear model setting is performed through a cumulative density function (cdf) Ψ. For instance, the probit model is obtained when Ψ is the standard normal cdf and the logit model when Ψ is the logistic cdf. See, for example, Cox (1971), McCullagh and Nelder (1989), and Collett (1991).

In Bayesian learning, the approach starts with a prior probability distribution for the unknown model parameters. Specifically, one denotes by $\pi(\boldsymbol{\beta})$ the (proper or improper) prior density function for the unknown parameter vector $\boldsymbol{\beta}$. Then the posterior density of $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\beta}|\mathbf{y}, \mathbf{x}) = \frac{\pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \Psi(\mathbf{x}_i^T\boldsymbol{\beta})^{y_i}(1 - \Psi(\mathbf{x}_i^T\boldsymbol{\beta}))^{1-y_i}}{\int \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \Psi(\mathbf{x}_i^T\boldsymbol{\beta})^{y_i}(1 - \Psi(\mathbf{x}_i^T\boldsymbol{\beta}))^{1-y_i} d\boldsymbol{\beta}},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ is the vector of binary data and $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is the matrix of explanatory variables. In general, this distribution is largely intractable.

* Corresponding author. Tel.: +34 924289300x86108; fax: +34 924272911.
E-mail addresses: lizbeth@unex.es (L. Naranjo), jrmartin@unex.es (J. Martín), carper@unex.es (C.J. Pérez).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.022

Bayesian approaches to binary regression models have been extensively developed in the literature. Zellner and Rossi (1984) examined the generalized linear models considering link functions such as the probit or logit. They commented on the inaccuracy of the normal approximation for small samples. For a small number of parameters (k small), they summarized the posterior distribution by using numerical integration. For large models (k large), they computed posterior moments by Monte Carlo integration (importance sampling with a multivariate Student's t as importance function).

With the upsurge of MCMC methods in the nineties (Gelfand and Smith, 1990), the new simulation tools allowed one to efficiently apply Bayesian regression methods. For example, Dellaportas and Smith (1993) used the Gibbs sampling procedure to compute the posterior distribution of the regression parameters for log-concave densities by using the adaptive rejection algorithm. A key paper on this topic was written by Albert and Chib (1993). It proposed a data augmentation framework for the Bayesian probit model by using Gibbs sampling. It also proposed extensions of the probit link by using mixtures of normal or Student's t distributions. The logit regression is studied as a particular case. Holmes and Held (2006) highlight a technique to improve performance in probit regression simulation by jointly updating the regression coefficients and the auxiliary variables. They show that the proposed approach is also possible for logistic regression by using a scale mixture of normal distributions for the noise process. Meza et al. (2009) proposed a stochastic approximation of the EM algorithm to analyze binary data under a two stage probit normal model with random effects. Haro-López et al. (2000) proposed a binary regression model that chooses an arbitrary link function from a set of different inverse platykurtic cdfs. The Bayesian analysis (focused on robustness) is implemented using cdfs of scale mixtures of normal distributions. Another remarkable result is that of Eyheramendy and Madigan (2007). They presented a class of sparse generalized linear models that include probit and logit regressions as special cases and offer some extra flexibility. They provided an EM algorithm for learning the regression parameter. Nott and Leng (2010) proposed a Bayesian approach to variable selection in generalized linear models. This approach has been implemented and validated for a logistic binary regression model.

A different kind of approach to modeling binary data is based on nonparametric models. Diaconis and Freedman (1993) studied a nonparametric Bayesian binary regression model from a theoretical viewpoint. They reviewed the relationship between consistency of Bayes estimates and rules for model selection, as well as sieves and orthogonal series estimation. Newton et al. (1996) proposed a semiparametric regression model for binary response data that places no structural restrictions on the link function other than monotonicity and known location and scale. By modifying the Dirichlet process they obtained a prior measure over the semiparametric model, and used Polya sequence theory to formulate the proposed measure in terms of a finite number of unobservables variables. When there is only one predictor, the method in Newton et al. (1996) is fully nonparametric. Qian et al. (2000) introduced a nonparametric Bayesian binary regression model with a single predictor variable that is usually more flexible than the commonly used logistic or probit models. The model presented by Qian et al. (2000) is a special case of the semiparametric binary regression model of Newton et al. (1996), in which a simpler computing algorithm is provided. However, the extension to more than one predictor variable is not obvious. Wood and Kohn (1998) proposed a Bayesian approach to binary nonparametric regression which assumes that the argument of the link is an additive function of the explanatory variables, and uses splines for the calculations. Although the nonparametric binary regression models offer robustness because of the flexibility of the link, the methods (except in special cases) are very difficult to implement and the results are not interpretable in terms of parameters.

A Bayesian approach to a binary regression model is proposed here. The inverse of the exponential power cdf is used as the link function. The exponential power (EP) family (Box and Tiao, 1973) includes the normal distribution and incorporates additional shapes, including platykurtic and leptokurtic ones. This means that distributions with both lighter and heavier tails compared to the normal case can be achieved, which is always an advantage when analyzing robustness. These distributions allow the modeling of kurtosis, providing, in general, more flexible fits to experimental data than the normal distribution. The proposed approach contains, among others, the probit model and an approximation to the logistic model as special cases.

The implementation of the approach is based on two main ideas. The first is to develop a data augmentation framework by introducing latent variables in a similar way to Albert and Chib (1993). The second is to use the mixture representation for the EP distribution suggested by Walker and Gutiérrez-Peña (1999). These two ideas are exploited to derive efficient Gibbs sampling algorithms for both informative and noninformative settings. All the full conditional distributions can be easily generated. The applicability of the approach is illustrated through some examples that show its good performance when compared with other competing models.

The outline of the paper is as follows. In Section 2, the EP distribution and some properties are presented. Section 3 presents the proposed binary regression model, including the derivation of the full conditional distributions that are necessary to apply Gibbs sampling. Section 4 illustrates the performance of the proposed approach through some examples. Finally, Section 5 presents the conclusions.

2. Exponential power distribution

The EP family (see, for example, Box and Tiao, 1973, and Gómez et al., 1998) includes the normal distribution and incorporates additional shapes, including platykurtic (lighter tails compared to the normal) and leptokurtic ones (heavier tails compared to the normal). These distributions allow the modeling of kurtosis, providing, in general, more flexible fits to experimental data than the normal distribution.