Statistical Methodology 13 (2013) 25–41

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

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

Posterior propriety for Bayesian binomial regression models with a parametric family of link functions

Vivekananda Roy *, Mark S. Kaiser

Department of Statistics, Iowa State University, United States

A R T I C L E I N F O

Article history:
Received 1 August 2012
Received in revised form
29 November 2012
Accepted 3 December 2012

Keywords:
Bayesian regression
Binary
Binomial
Link function parameter
Posterior propriety
Tolerance distribution
Toxicity tests

A B S T R A C T

We consider a Bayesian analysis of Binomial response data with covariates. To describe the problem under investigation, suppose we have n independent binomial observations $Y_1, \ldots, Y_n$ where $Y_i \sim \text{Bin}(m_i, \theta_i)$ and let $\mathbf{x}_i$ be p-dimensional covariate vector associated with Y_i for $i = 1, \ldots, n$. Binomial observations can be analyzed through a generalized linear model (GLM) where we assume $\theta_i = F(\mathbf{x}_i^T \boldsymbol{\beta})$ for some known distribution function $F(\cdot)$ and $\boldsymbol{\beta}$ is the vector of unknown regression coefficients. In this paper, we state necessary and sufficient conditions for propriety of the posterior distribution of $\boldsymbol{\beta}$ if an improper uniform prior is used on $\boldsymbol{\beta}$. We also consider situations where the link function is not pre-specified but belongs to a parametric family and the link function parameters are estimated along with the regression coefficients. In this case, we investigate the propriety of the joint posterior distributions of $\boldsymbol{\beta}$ and the link function parameters. There are a number of parametric families of link functions available in the literature. As a specific example, we consider Pregibon's (1980) [17] link function and show that our general posterior propriety results can be used to establish propriety of the posterior distributions corresponding to the Pregibon's (1980) [17] link. We show that Pregibon's (1980) [17] simple one parameter family of link function can be used to fit both positively and negatively skewed response curves. Moreover, the conditions for posterior propriety corresponding to the Pregibon's (1980) [17] link can be easily checked and are milder than those required by the flexible GEV link of Wang and Dey (2010) [24]. As an illustration, we analyze a data set from Ramsey and Schafer (2002) [18] regarding

* Corresponding author. Tel.: +1 515 294 8701; fax: +1 515 294 4040.
E-mail address: vroy@iastate.edu (V. Roy).

1572-3127/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2012.12.003

the relationship between dose of Aflatoxicol and odds of liver tumor in rainbow trouts. In this example, the symmetric logit link fails to fit the data, whereas Pregibon's (1980) [17] skewed link yields a slightly better fit than the GEV link.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Binomial and binary response data with one or more covariates can be analyzed through a basic generalized linear model with a specified link function. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be n independent binomial or binary random variables with probability mass functions

$$f(y_i|\theta_i) = \binom{m_i}{y_i} \theta_i^{y_i}(1-\theta_i)^{m_i-y_i}; \quad y_i = 0, 1, \ldots, m_i, \tag{1}$$

for $m_i \geq 1,\ i = 1, 2, \ldots, n$, with

$$\theta_i = F(\mathbf{x}_i^T \boldsymbol{\beta}), \tag{2}$$

where F is a cumulative distribution function (cdf) (F^{-1} is typically called the link function), $\mathbf{x}_i$'s are p-dimensional covariates and $\boldsymbol{\beta}$ is the vector of unknown regression coefficients. Some frequently used links are logit, probit and complimentary log–log links. The logit link is obtained by setting $F^{-1}(\theta_i) = \log\{\theta_i/(1-\theta_i)\}$, whereas the function $F(\theta_i) = \Phi(\theta_i)$ yields the probit link, where Φ is the cdf of $N(0, 1)$. The complementary log–log link is specified as $F^{-1}(\theta_i) = -\log(-\log(\theta_i))$. Among these three links, both logit and probit are symmetric link functions whereas the complimentary log–log is an asymmetric link. However these popular link functions do not always provide good fit to the given data set and this can yield a substantial bias in the mean response estimates [12]. In this case one might wish to extend the above formulation to one in which the link belongs to a parameterized family of functions, and the link function parameter is estimated along with the regression parameters.

Suppose the link function F is embedded into a parametric family $\{F_\lambda(\eta), \lambda \in \Lambda\}$ where $\eta = x^T\boldsymbol{\beta}$ is the linear predictor and $\Lambda \subset \mathbb{R}^l$ for some $l \geq 1$. So the joint probability mass function of $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ becomes

$$f(\mathbf{y}|\boldsymbol{\beta}, \lambda) = \prod_{i=1}^{n} \binom{m_i}{y_i} \{F_\lambda(\eta_i)\}^{y_i}\{1 - F_\lambda(\eta_i)\}^{m_i-y_i}, \tag{3}$$

where $\eta_i = \mathbf{x}_i^T\boldsymbol{\beta}$. While choosing an appropriate family of link functions $\{F^{-1}(\eta|\lambda)\}$, we need to consider the flexibility of the link functions in fitting a variety of response curves, and also the complexity of the link functions, for example, the dimension of Λ. To the best of our knowledge, consideration of parameterized link function began with Pregibon [17], who showed that a score test for adequacy of a hypothesized link function could be obtained by embedding the hypothesized link in a parametric family of link functions. We will describe what we will call Pregibon's [17] link function in detail in Section 3. Pregibon's [17] link function has only one parameter and we show that it can fit both positively and negatively skewed response curves. Stukel [23] proposed a class of generalized logistic models. Stukel's [23] two-parameter models are general and several commonly used link functions, such as the probit and complimentary log–log link models can be approximated by members of this family. Czado [9] studied a standardized link family called the generalized probit regression model. Chen, Dey and Shao [5] considered a class of skewed link functions, where an underlying latent variable has a mixed-effects model structure. Kim, Chen and Dey [15] introduced a class of generalized skewed t-link models. Some other link functions and related works include [25,10,11]. More recently, Wang and Dey [24] considered a flexible skewed link function based on the generalized extreme value (GEV) distribution. The superiority of the GEV link over the logit, probit and the complimentary log–log link is due to the fact that the shape parameter of the GEV distribution purely controls its tail behavior. In fact, Wang and Dey [24] also demonstrated that the GEV link even often dominates the Kim et al.'s [15] skewed-t link, due to more flexibility of the GEV distribution.

Download English Version:

https://daneshyari.com/en/article/1151078

Download Persian Version:

https://daneshyari.com/article/1151078

Daneshyari.com