Computational Statistics and Data Analysis 73 (2014) 87–102

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Efficient estimation of the link function parameter in a robust Bayesian binary regression model

Vivekananda Roy *

Department of Statistics, Iowa State University, Ames, IA 50011, United States

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 17 October 2013
Accepted 20 November 2013
Available online 4 December 2013

Keywords:
Data augmentation
Empirical Bayes
Importance sampling
Markov chain
Robit regression
Robust regression

ABSTRACT

It is known that the robit regression model for binary data is a robust alternative to the more popular probit and logistic models. The robit model is obtained by replacing the normal distribution in the probit regression model with the Student's t distribution. Unlike the probit and logistic models, the robit link has an extra degrees of freedom (df) parameter. It is shown that in practice it is important to estimate (rather than use a prespecified fixed value) the df parameter. A method for effectively selecting the df parameter of the robit model is described. The proposed method becomes computationally more effective if efficient MCMC algorithms are available for exploring the posterior distribution associated with a Bayesian robit model. Fast mixing parameter expanded DA (PX–DA) type algorithms based on an appropriate Haar measure are developed for significantly improving the convergence of DA algorithms for the robit model. The algorithms built for sampling from the Bayesian robit model shed new light on the construction of efficient PX–DA type algorithms in general. In spite of the fact that Haar PX–DA algorithms are known to be asymptotically "optimal", through an empirical study it is shown that it may take millions of iterations before they provide improvement over the DA algorithms. Contrary to the popular belief, it is demonstrated that a partially reparameterized DA algorithm can outperform a fully reparameterized DA algorithm. The proposed methodology of selecting the df parameter is illustrated through two detailed examples.

1. Introduction

Suppose $y = (y_1, y_2, \ldots, y_n)$ are n independent observations where y_i is either 0 or 1. Binary regression models using Generalized Linear Models (GLMs) assume that

$$Y_i \sim \text{Ber}(p_i),$$

where $F^{-1}(p_i) = x_i^T \beta$ for some link function $F^{-1}(\cdot)$, the x_i's, $i = 1, 2, \ldots, n$ are $p \times 1$ covariate vectors and β is the $p \times 1$ vector of regression coefficients. The two most popular choices for the link function are $F(\cdot) = \Phi(\cdot)$, the standard normal cdf which leads to the probit model and $F(\eta) = e^{\eta}/(1 + e^{\eta})$, the cdf of the standard logistic distribution which leads to the logistic model. It is well known that the estimates of regression coefficients for logistic and probit models are not robust to outliers (Pregibon, 1982). A robust alternative to logistic and probit models is obtained by assuming $F(\cdot) = F_{\nu}(\cdot)$, where F_{ν} is the cdf of the standard Student's t distribution with degrees of freedom ν (Liu, 2004). Following Liu (2004), this model is called the robit regression model. Both logit and probit link functions are well approximated by a robit link function with an appropriate degrees of freedom parameter ν (Mudholkar and George, 1978; Albert and Chib, 1993). In fact, a robit link with about seven degrees of freedom provides an excellent approximation to the logit link, and the probit link can be well

* Tel.: +1 515 294 8701; fax: +1 515 294 4040.
E-mail address: vroy@iastate.edu.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.11.013

approximated by a robit model with large degrees of freedom. Gelman and Hill (2007, Chapter 6) showed that in the presence of outliers, the robit model, unlike logistic and probit models, can effectively downweight the discordant data points, for a better model fitting. Therefore, if the degrees of freedom parameter is chosen appropriately, the robit model will replicate the logistic or probit models if the data follows one of those models, but will provide a robust alternative when outliers are present. Here we consider a Bayesian robit model for analyzing binary data.

For a Bayesian analysis we need a prior distribution for the vector of regression coefficients β. We consider the multivariate t prior on β given by

$$\pi(\beta) = \frac{\Gamma((\nu_0+p)/2)}{\Gamma(\nu_0/2)\nu_0^{p/2}\pi^{p/2}}|\Sigma_0|^{1/2}\left[1+\nu_0^{-1}\beta^T\Sigma_0\beta\right]^{-\frac{p+\nu_0}{2}}. \tag{1.1}$$

Here, the prior for β is $t_p(0, \Sigma_0^{-1}, \nu_0)$, the p dimensional multivariate Student's t distribution with a known $p \times p$ positive definite scatter matrix Σ_0, and known degrees of freedom ν_0, centered at 0. As mentioned in Gelman et al. (2008), the t family of prior distributions allow for robust inference. Let X be the $n \times p$ design matrix whose ith row is x_i^T. Note that if $\Sigma_0 = cX^TX$ for some constant c, then $\pi(\beta)$ is the marginal prior for β under Zellner's g-prior for the normal linear model (Zellner, 1983). The posterior density is

$$\pi_\nu(\beta|y) = \frac{1}{m_\nu(y)}\ell_\nu(\beta|y) \times \pi(\beta), \tag{1.2}$$

where $\ell_\nu(\beta|y)$ is the likelihood function given by

$$\ell_\nu(\beta|y) = \prod_{i=1}^{n}\left(F_\nu(x_i'\beta)\right)^{y_i}\left(1-F_\nu(x_i'\beta)\right)^{1-y_i} \tag{1.3}$$

and $m_\nu(y) := \int_{\mathbb{R}^p}\ell_\nu(\beta|y) \times \pi(\beta)d\beta$ is the normalizing constant.

As we will see in Section 4 that in practice it is important to estimate (rather than use a prespecified fixed value) the degrees of freedom parameter ν for the robit model to provide a robust alternative to more popular logistic and probit models. Here we consider an empirical Bayes approach for estimating ν, that is, we select that value of ν which maximizes the marginal likelihood of the data $m_\nu(y)$. Henceforth we simply write m_ν instead of $m_\nu(y)$. If we are interested in a family of robit models indexed by $\nu \in \mathcal{N}$ for some set $\mathcal{N} \subset (0, \infty)$, we can calculate m_ν for all $\nu \in \mathcal{N}$ and then select that value of ν which maximizes m_ν. Note that the value of ν that maximizes m_ν, is same as the value of ν that maximizes am_ν for $\nu \in \mathcal{N}$ where a is a constant. It is often much easier to calculate am_ν than m_ν for all $\nu \in \mathcal{N}$ if a is properly chosen. For selecting models that are better than other models when ν varies across $\mathcal{N}$, we can calculate and subsequently compare the values of m_ν/m_{ν_1}, where ν_1 is a suitably chosen fixed value of the degrees of freedom parameter. (Note that, in this case $a = 1/m_{\nu_1}$.) We denote m_ν/m_{ν_1} by B_{ν,ν_1}. Ideally we would like to calculate and compare B_{ν,ν_1} for a large number of values of ν. Recently, Doss (2010) described a method based on importance sampling for selecting prior hyperparameters by estimating a large family of Bayes factors. Following Doss (2010) we consider a method that is based on importance sampling and control variates to efficiently estimate B_{ν,ν_1} for a large set of possible values of ν. The method proposed here is not specific to the robit model and can be used for any model especially when it may be difficult to effectively specify a prior on certain parameters.

Availability of fast mixing MCMC algorithms with stationary density $\pi_\nu(\beta|y)$ (for fixed ν) makes the above method of estimating ν more computationally efficient. Using the fact that the t distribution can be represented as a scale mixture of normal distributions a data augmentation (DA) algorithm can be constructed for $\pi_\nu(\beta|y)$ (see Liang et al., 2010, Section 2.4.2). It is well known that DA algorithms often converge to their stationary distributions very slowly. On the other hand, over the last decade several authors have shown that convergence of DA algorithms can be significantly improved by introducing an efficient parameter expansion step into the DA algorithm (see, e.g., Meng and van Dyk, 1999, Liu and Wu, 1999, van Dyk and Meng, 2001, Hobert and Marchev, 2008). Following these works we construct three parameter expanded DA (PX–DA) type algorithms, which are similar to the DA algorithm in terms of computational complexity. We show that *two* of these algorithms significantly improve the convergence of the DA algorithm, while the third algorithm does not show improvement over the DA algorithm. The PX–DA type algorithms that we build in the context of our Bayesian robit model shed new light on the construction of efficient PX–DA type algorithms in general. Firstly, it has been shown in the literature that the Haar PX–DA algorithms, which are PX–DA type algorithms based on appropriate Haar measures, are always asymptotically at least as efficient as their underlying DA algorithms (Roy, 2012b; Hobert and Marchev, 2008). Also it is known that Haar PX–DA algorithms are "optimal" PX–DA algorithms in the sense that the Haar PX–DA algorithm is at least as good as any other PX–DA algorithm in terms of asymptotic convergence rate and asymptotic efficiency (Hobert and Marchev, 2008) (see Section 2.2.1 for details). We construct a Haar PX–DA algorithm and empirically show that it fails to provide improvement over the DA algorithm even when the algorithm is run for several thousands of iterations. This example shows that care must be taken while constructing effective PX–DA type algorithms by introducing a parameter expansion step into the DA algorithms. Secondly, we show that a Haar PX–DA algorithm with a partial reparameterization of the augmented space can outperform a Haar PX–DA algorithm where all the augmented variables are reparameterized, which is commonly done in practice.