Journal of the Korean Statistical Society 45 (2016) 466–476

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Bayesian variable selection in quantile regression using the Savage–Dickey density ratio

Man-Suk Oh [a,*], Jungsoon Choi [b,1], Eun Sug Park [c,2]

[a] Department of Statistics, Ewha Womans University, Seo-dae-moon Gu, Seoul 120-750, Republic of Korea
[b] Department of Mathematics, Hanyang University, Seong-dong Gu, Seoul 133-791, Republic of Korea
[c] Texas A&M Transportation Institute, 3135 TAMU, College Station, TX 77843-3135, USA

ARTICLE INFO

Article history:
Received 3 October 2015
Accepted 19 January 2016
Available online 2 March 2016

AMS 2000 subject classifications:
primary 62F15
secondary 62F03

Keywords:
Markov chain Monte Carlo
Asymmetric Laplace distribution
Bayesian model selection
Bayes factor

ABSTRACT

In this paper we propose a Bayesian variable selection method in quantile regression based on the Savage–Dickey density ratio of Dickey (1976). The Bayes factor of a model containing a subset of variables against an encompassing model is given as the ratio of the marginal posterior and the marginal prior density of the corresponding subset of regression coefficients under the encompassing model. Posterior samples are generated from the encompassing model via a Gibbs sampling algorithm and the Bayes factors of all candidate models are computed simultaneously using one set of posterior samples from the encompassing model. The performance of the proposed method is investigated via simulation examples and real data sets.

1. Introduction

Quantile regression, introduced by Koenker and Bassett (1978), has been rapidly expanding over recent years, particularly in econometrics, finance, survival analysis, social sciences, and microarray research (Hendrick & Koenker, 1992; Koenker & Geling, 2001; Koenker & Hallock, 2001; Wang & He, 2007; Yang, 1999). Quantile regression explores the relationships between a set of covariates and different parts of the conditional distribution of a response variable, and hence provides more comprehensive information on the relationship between the response variable and the covariates than classical mean regression. See Koenker (2005) and Yu, Lu, and Stander (2003) for an overview.

Let y_i denote the response variable and $\mathbf{x}_i$ the $q \times 1$ vector of covariates for the ith observation. For $0 < p < 1$, let $Q_{y_i}(p|\mathbf{x}_i)$ be the pth quantile of y_i given $\mathbf{x}_i$. Suppose that the relationship between $Q_{y_i}(p|\mathbf{x}_i)$ and $\mathbf{x}_i$ can be modeled as $Q_{y_i}(p|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}_p$, for $i = 1, \ldots, n$, where $\boldsymbol{\beta}_p$ is an unknown vector of coefficients that depends on p. Then, a linear quantile regression model can be expressed as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_p + \varepsilon_i, \tag{1}$$

* Corresponding author. Tel.: +82 2 3277 2374; fax: +82 2 3277 3607.
E-mail addresses: msoh@ewha.ac.kr (M.-S. Oh), jungsoonchoi@hanyang.ac.kr (J. Choi), e-park@tamu.edu (E.S. Park).
[1] Tel.: +82 2 2220 2621; fax: +82 2 2281 0019.
[2] Tel.: +1 979 845 9942; fax: +1 979 845 6008.

http://dx.doi.org/10.1016/j.jkss.2016.01.006
1226-3192/

where ε_i is an error term whose distribution is restricted such that the pth quantile is equal to zero. A consistent estimate of the quantile regression coefficient $\boldsymbol{\beta}_p$ can be obtained by minimizing

$$\sum_{i=1}^{n} \rho_p(y_i - \mathbf{x}_i'\boldsymbol{\beta}_p), \tag{2}$$

where ρ_p is the check loss function, given by

$$\rho_p(u) = u(p - I(u < 0)), \tag{3}$$

and I is the indicator function. Since explicit solutions to the minimization cannot be obtained, linear programming methods are commonly used to obtain estimates of $\boldsymbol{\beta}_p$ (Koenker & Park, 1996; Portnoy & Koenker, 1997).

Koenker and Machado (1999) noted that minimizing (2) is equivalent to maximizing a likelihood function under asymmetric Laplace (AL) error distribution. Yu and Moyeed (2001) used this finding to propose a Bayesian modeling approach to quantile regression, and obtained samples of $\boldsymbol{\beta}_p$ from its posterior distribution using Markov chain Monte Carlo (MCMC) methods. Kozumi and Kobayashi (2011), Reed and Yu (2009) and Tsionas (2003) proposed Gibbs sampling algorithms (Gelfand & Smith, 1990) for posterior estimates of $\boldsymbol{\beta}_p$ by using a location-scale mixture representation of the asymmetric Laplace distribution.

As in mean regression, variable selection in quantile regression is a crucial aspect of improving the precision of model fit. Recently, a Bayesian approach for variable selection in quantile regression has received attention from researchers, since Bayesian methods are often more competitive for small or moderate data sets with a low signal-to-noise ratio (Antoniadis, Bigot, & Sachs, 2009; Ji, Lin, & Zhang, 2012). Li, Xi, and Lin (2010) studied regularization, e.g. lasso, in quantile regression from a Bayesian perspective. Alhamzawi and Yu (2012), Ji et al. (2012), and Yu, Chen, Reed, and Dunson (2013) proposed a variant of stochastic search variable selection (SSVS George & McCulloch, 1993) method in quantile regression using the asymmetric Laplace error distribution.

In this paper we introduce an alternative Bayesian variable selection method in quantile regression based on the Savage–Dickey (SD) density ratio, referred to as the SD method. When there is a set of candidate models having subsets of $\boldsymbol{\beta}$ equal to zero, each candidate model can be compared with an encompassing model which encompasses all the candidate models. Under suitable priors the Bayes factor of each candidate model against the encompassing model can be represented as the ratio of the posterior and the prior marginal density of the corresponding subset of $\boldsymbol{\beta}$ at zero, where the marginal densities are derived from the joint density of $\boldsymbol{\beta}$ under the encompassing model. This is known as the Savage–Dickey density ratio (Dickey, 1971, 1976; Dickey & Lieutz, 1970).

Since the prior densities are known in most cases, the variable selection problem now becomes a problem of estimating the marginal posterior densities of subsets of $\boldsymbol{\beta}$ at zero. Estimation of the marginal posterior densities can be achieved easily using the known conditional posterior densities of subsets of $\boldsymbol{\beta}$ and the posterior samples of $\boldsymbol{\beta}$ from the encompassing model. Key features of the SD method are that it fits *only* the encompassing model and that it estimates the Bayes factors of *all* candidate models *simultaneously*. Also, the method can be easily implemented for constrained variable selection in which variables are selected in groups or in hierarchy or in anti-hierarchy. In anti-hierarchy constraints, the inclusion of one variable forces another variable to be excluded from the model (Farcomeni, 2010). We illustrated a simple hierarchy constrained case in Section 5.2.

The rest of paper is organized as follows. Section 2 presents the posterior inference on parameters in quantile regression using the Gibbs sampling algorithm. Section 3 derives the Bayes factors of models using the SD density ratio and proposes an efficient estimation of the Bayes factors using the posterior samples. Section 4 presents the results from a series of simulation studies to investigate the performance of the SD method. Section 5 contains an application of the method to real data sets. The paper is concluded in Section 6.

2. Posterior inference

We consider model (1) and assume that the error term ε_i follows the AL distribution with density

$$f_p(\varepsilon_i) = p(1-p)\exp\left[-\rho_p(\varepsilon_i)\right],$$

where ρ_p is defined in (3). Then the likelihood function of $\boldsymbol{\beta}_p$ is proportional to

$$\exp\left[-\sum_{i=1}^{n} \rho_p(y_i - \mathbf{x}_i'\boldsymbol{\beta}_p)\right],$$

and maximizing the likelihood function is equivalent to minimizing (2). From here on we suppress the subscript p on $\boldsymbol{\beta}$ to simplify notation.

The AL distribution has various mixture representations (Kotz, Kozubowski, & Podgorski, 2001). Among those, mixture representations based on normal and exponential distributions are useful for implementation of the Gibbs sampler (Ji et al., 2012; Kozumi & Kobayashi, 2011; Yu et al., 2013). We adopt the mixture representation given in Yu et al. (2013). Let w_i follow an exponential distribution with rate $p(1-p)$ and assume that, given w_i, ε_i follows $N((1-2p)w_i, 2w_i)$ distribution,