A Spatially Adapted Score-Based Likelihood Approach to Bayesian Quantile Regression

John Schwenck & Marshall Honaker

10 December 2020

Abstract

Since its introduction by Koenker & Bassett in 1978, quantile regression has become a widely studied topic within statistics. In the Bayesian paradigm, a common approach to quantile regression is to assume likelihood model based on the Asymmetric Laplace Distribution. While this approach is fairly easy to implement using standard MCMC procedures, it has been shown that the variance of the resulting posterior distribution is incorrect, leading to invalid posterior inference. (Sriram (2015) and Yang et al. (2016)) Wu and Narisetty (2020) propose a score-based working likelihood function that yields valid posterior inference. We expand this approach to incorporate the spatial dependencies within observed data by augmenting the standard feature matrix. Using this augmented feature matrix, we attempt to implement the framework outlined by Wu and Narisetty to obtain a spatially adapted Bayesian quantile regression. Unfortunately, several issues were encountered during this process that prohibited the successful development of such a model. We outline these issues and discuss what steps will be taken to address them as well as potential avenues for future research.

Introduction

Wu & Narisetty (2020) propose a working likelihood function of the form:

$$L(Y|X, \vec{\beta}) = C \exp\left(-\frac{1}{2n} s_\tau(\vec{\beta})^T W s_\tau(\vec{\beta})\right)$$

where $s_\tau(\vec{\beta}) = \sum_{i=1}^{n} x_i \psi_\tau(y_i - x_i^T \vec{\beta})$ is the score function (in which $\psi_\tau(u) = \tau - I_{\{u<0\}}(u)$ is the check loss function for the τth quantile), W is a p x p positive definite weight matrix given by:

$$W = \frac{n}{\tau(1-\tau)} \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1}$$

and C is a constant that does not depend on $\vec{\beta}$. It can be shown that using the above working likelihood $L(Y|X, \vec{\beta})$ with weight matrix W as our sampling model leads to a posterior distribution that is, asymptotically, very nearly normal. Using this working likelihood as the sampling model for the model parameters of a Bayesian regression, Wu & Narisetty propose an adaptive Importance Sampling algorithm to approximate the mean vector and covariance matrix of the resulting posterior distribution.

When working with spatial data, it is imperative that the model consider and account for the spatial variability and dependencies within the data. Failure to do so can lead to inaccurate predictions, loss of power, and invalid inferential procedures. To introduce these spatial dependencies to the score-based working likelihood approach outlined above, we will generate an augmented feature matrix by appending a matrix of spatial basis covariate functions to the standard n x p feature matrix X. Then, implementing the framework outlined by Wu & Narisetty will yield a Bayesian quantile regression model that takes into account the locations at which various observations were made as part of its input.

Review of the Literature

Bayesian quantile regression has been studied and implemented in a wide variety of contexts. However, one encountered issue when working with such models is the specification of a parametric likelihood. Yu & Moyeed (2001) propose that an appropriate working likelihood could be based on the Asymmetric Laplace Distribution (ALD) and the consistency of the resulting posterior distribution was demonstrated by Sriram et al. (2013). However, Sriram (2015) and Yang et al. (2016) show that the variance of the posterior distribution based on such a likelihood is incorrect, indicating that any inference procedures based on the posterior distribution will be invalid. While these studies propose corrections to account for the bias in the variance induced by the ALD based likelihood, the score based likelihood approach proposed by Wu & Narisetty (2020) yields a posterior distribution with correct variance, circumventing the the need for any such corrections. Furthermore, the authors propose a novel Importance Sampling algorithm to approximate the parameters of the resulting posterior distribution.

Previous studies have explored the use of Bayesian quantile regression to model spatially dependent data. Reich et al. (2011) proposes a semiparametric approach using Bernstein basis polynomials to model the distribution of the covariates at a particular quantile level to examine tropospheric ozone levels. This approach was expanded by Smith et al. (2015), who uses cubic splines to jointly model multiple quantiles simultaneously. The approach proposed by Reich et al. (2019) is currently the most popular and has been implemented in several situations, such as Ramsey (2019) who used this approach to model the effects of adverse weather conditions and technological advances on crop yields.

To the author's knowledge, there has been no attempt to implement the framework proposed by Wu & Narisetty (2020) in a spatial context as of the writing of this paper. The approach outlined in this paper provides does just this, providing an alternative to methods outlined in Reich et al. (2011) and Smith et al. (2015) and circumventing the bias in the posterior variance induced by ALD based likelihoods.

Methodology

To account for the spatial dependence of observed data, we introduce an augmented feature matrix $\tilde{X}$. $\tilde{X}$ is obtained by appending an n x b matrix of spatial basis covariate functions to the standard n x p feature matrix.

where x_{ij} is the j^{th} value of the i^{th} covariate and s_{ij} is the j^{th} value of the i^{th} spatial basis function. Implementing the framework outlined above using the augmented feature matrix $\tilde{X}$ yields a Bayesian quantile regression model that takes into consideration the spatial dependencies within the observed data and includes the locations at which observations were made as part of the model's input.

The matrix of spatial basis functions was automatically generated by applying the `auto_basis()` function from `R`'s `FRK` package to the locations. **WE SHOULD FIND ANOTHER, MORE METHODOLOGICAL WAY TO COME UP WITH OUR BASIS FUNCTIONS**

In many cases, a relatively large number of spatial basis covariates (say, 20) are used to model the spatial variability of the data. Because the effects of each of these spatial basis functions enters into the model additively, we will scale each spatial basis covariate by its L^2 norm to ensure that they do not exert a disproportionate influence on the model. Taken all together, the model will be of the form:

$$\tilde{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} & \frac{s_{11}}{||s_1||_2} & \frac{s_{12}}{||s_2||_2} & \dots & \frac{s_{1b}}{||s_b||_2} \\ x_{21} & x_{22} & \dots & x_{2p} & \frac{s_{21}}{||s_1||_2} & \frac{s_{22}}{||s_2||_2} & \dots & \frac{s_{2b}}{||s_b||_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} & \frac{s_{n1}}{||s_1||_2} & \frac{s_{n2}}{||s_2||_2} & \dots & \frac{s_{nb}}{||s_b||_2} \end{bmatrix}$$

Simulation

To assess the performance of this model based on the proposed augmented feature matrix, we generated a set of 10,000 observations from spatial Gaussian process from 10,000 random locations on a 10 x 10 grid. We then split these 10,000 simulated observations into training and test data sets. (75% training, 25% test)

We weren't able to get things up and running, but it might be worth talking about how the imulation would have worked if it had. (For reference when we do get it to work and try to move forward/publish it.)

Issues and Attempted Solutions

Discussion of issues involving weight matrix, likelihood, and importance weights. It would also be worth mentioning the size of the simulation conducted by Wu & Narisetty and the size of our simulation.

Suggestions for Further Research

While previous studies have discussed and even implemented Bayesian spatial quantile regression models, this paper presents a framework for a general model that provides valid posterior inference. However, this framework could be extended to include a regularization term that would allow for model selection at various quantile levels. By comparing the variables included in the model at various quantile levels, one would be able to examine which variables play an active role in modeling the response at various points in the distribution.

Another direction might explore the possibility of applying a kernel function to the inner products of the observations in the weight matrix W. Doing so would allow the model to encode or account for non-linearities in the observed data at no additional cost to the model. However, to do so, it must be shown that the likelihood depends on the observations only through their inner products. Doing so may not be trivial, since the observations also appear as arguments in the check loss function within the score function.