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 a working likelihood 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 on this approach by introducing an augmented feature matrix that takes into account the spatial dependencies in a given data set and show that this approach DOES SOMETHING. *Insert blurb about simulation/simulation results here.* **Insert a sentence about conclusions here.**

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 weight 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 $L(Y|X, \vec{\beta})$ as our sampling model leads to a posterior distribution that is very nearly normal, especially asymptotically. 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, is 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 and include them in the resulting model, 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. Doing so will result in a Bayesian quantile regression model that considers the location at which observations were made as part of its input.

Review of the Literature

Previous research has lead to the development and implementation of Bayesian spatial quantile regression models, most notably Reich et al. (2011) but also Ramsey (2019) and King & Song (2016). However, as of the writing of this paper, there has been no attempt to develop or implement a Bayesian quantile regression model using the framework outlined above that also accounts for the spatially dependent structure of the observations to the author's knowledge.

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 jth value of the ith covariate and s_{ij} is the jth value of the ith 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:

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

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're meeting with Dr. Sang tomorrow and she mentioned she wanted to examine the simulation to make sure it was alright. I'll avoid writing more about the simulation now so that if we need to change the simulation at all tomorrow we will have the most recent information here and won't have to update anything.

Results

Once the simulation is actually running, there will be two (possibly three) things we will really want to look at: 1). training and test error (to get these, we will need to derive the posterior predicitive distribution, but since the posterior for the covartiates is normal that shouldn't be too bad), 2). validity of posterior inference procedures (run a bunch of tests using the simulated data and check to see if the type 1 error rate is actually close to α), and (possibly) 3). comparison to the resutls given in Wu & Narisetty and the other Bayesian spatial quantile regression papers.

Conclusions

This section will depend on what we observe for the previous section.

Suggestions for Furher 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.