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

Marshall Honaker & John Schwenck

10 December 2020

Abstract

We propose a spatial adaptation of the score-based working likelihood function for quantile regression by deriving a non-parametric spline-based weight matrix capable of performing inference for multiple conditional quantiles simultaneously. We outline the methodology for Bayesian quantile regression for a single quantile and then generalize to multiple, making effective use of the Importance Sampling algorithm to compute posterior summaries. We then demonstrate our method's effectiveness through both a simulation study and a real-world example of NYC crime data. Whereas existing methods for quantile regression in the spatial domain lack . . . , our method . . . . (punchline) (no citations)

we provide 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.

Introduction

Quantile regression, introduced by Koenker & Bassett (1978), has become a widely studied topic for its usefulness in characterizing an entire distribution, particularly when tail risk and extreme behavior is of interest and/or the error terms exhibit heteroskedasticity. Whereas the more common conditional mean regression tends to be sensitive to outliers, quantile regression by contrast allows for a much more general relationship between Y and X as the regression coefficients change with different levels of τ.

More formally, for a given continuous response, Y, and p-dimensional design matrix, $\boldsymbol{X}$, let $\tau \in (0,1)$ represent a given quantile, and let $Q_Y(\tau|\boldsymbol{X}=\boldsymbol{x}) = \boldsymbol{x}^T\ \boldsymbol{\beta}(\tau)$ denote the τth conditional quantile of Y given $\boldsymbol{X}=\boldsymbol{x}$ where $\boldsymbol{\beta}(\tau) \in \mathbb{R}^p$. The standard quantile regression estimator is therefore given by

$$\hat{\beta}(\tau) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\boldsymbol{\beta})$$

where $\rho_\tau(u) = u(\tau - I(u<0))$ is the asymmetric loss function, whose asymptotically normal results from Koenker & Bassett (1978) lead to valid inference on the quantile regression parameters. From a frequentist perspective, inference for multiple quantiles follows directly. In hypothesis testing scenarios, we would test whether the effect of a given covariate has a different effect on the response at multiple separate quantiles, i.e. for two quantile levels: at τ_1 and τ_2.

There has been considerable work in this area since its inception, but it has been slower to gain traction within the Bayesian community due to the need for a specified parametric likelihood. The common approach for the Bayesian framework is to base the likelihood model on the Asymmetric Laplace Distribution (ALD) proposed by Yu & Moyeed (2001)(is this the correct source?). Despite the consistency of the resulting posterior distribution demonstrated by Sriram et al. (2013), and the fairly straightforward implementation of standard MCMC procedures, Sriram (2015) and Yang et al. (2016) have shown that the variance is incorrect, leading to invalid posterior inference.

While these studies propose corrections to account for the bias in the variance (???)induced by the ALD based likelihood, Wu and Narisetty (2020) recently proposed a score-based working likelihood approach that yields valid posterior results, thus circumventing the the need for any such corrections. Their proposed methodology revolves around Bayesian multiple quantile regression for linear models, where the parameters of the resulting posterior distribution were approximated by a variant of the Importance Sampling algorithm.

Our proposed method extends their approach to a spatial setting by utilizing an augmented spline-based feature matrix to incorporate spatial dependencies within observed data and smooth the spatial variability, building off of (Sang's paper). There have been previous studies that explored the use of Bayesian quantile regression for spatially-dependent data, but to the authors' knowledge, there have been no attempts to implement the score-based framework at the time of writing. Reich et al. (2011) proposed a semi-parametic approach using Bernstein basis polynomials to model the distribution of the covariates at a particular quantile level which was later expanded by Smith et al. (2015), who used cubic splines to jointly model multiple quantiles simultaneously. (mention drawback of Reich/Smith?)Reich et al. (2011)'s approach is typically the most popular approach in practice and has been implemented in several applicationss, such as Ramsey (2019) who used this approach to model the effects of adverse weather conditions and technological advances on crop yields. However, these proposals fail to ensure valid frequentist coverage.

Methodology

outline reminders for tomorrow: (1) wu likelihood –> our likelihood with spatial (2) overview of nonparam in spatial and why it can be applied here (3) ada-is 1 quantile (4) ada-is mult quantiles (5)

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

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

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 $\boldsymbol{\beta}$. It can be shown that using the above working likelihood $L(Y|X,\boldsymbol{\beta})$ 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.

This section needs to be revised to incorporate/describe the changes made to our approach. Describe augmented feature matrix and discuss how it draws from Dr. Sang's paper. Give overview of Wu & Narisetty and how our feature matrix will give us a novel Bayesian spatial quantile regression.

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:

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

Simulation (Marshall)

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.)

Real-World Application: NYC Crime

(until code is finalized, there is not much to say here…)

Discussion

Limitations

(we will need to revisit this later with less (or no) code, but for final report purposes this is fine)

With regard to the weight matrix… (something about super small values)

The problems of the weight matrix naturally led to difficulties with the correct likelihood. Because of the extremely large negative value in our exponent, we nearly always returned a likelihood value of zero explicitly. By nature of importance sampling, if the likelihood is zero or very close to it, as is the prior, then the resulting importance weights will be zero as well. Once we diagnose the problem within the likelihood function, the rest of the method should work as intended.

Granted, our simulation relied on a drastically different sample size than in the original Wu & Narisetty paper, but our results are still essentially zero.

Future Directions

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. It is worth noting that this extension need not be specific to a spatial context, but the regularization literature is much less established in such a domain.

From a reproducibility and optimization standpoint, we will alleviate bottlenecks within the necessary algorithms by translating R code to C++ code, and will restructure it to incorporate compatibility checks and other fail-safe measures.

Supplementary Material

All code is hosted on GitHub through the repository: https://github.com/johnschwenck/crimeBQR

References

1. Banerjee, S., Gelfand, A., Finley, A., & Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *J R Stat Soc Series B Stat Methodol*, 70(4), 825–848. doi:10.1111/j.1467-9868.2008.00663.x.

2. Koenker, R., & Bassett, G. (1978). Regression Quantiles. *Econometrica*, 46(1), 33-50. doi:10.2307/1913643

3. Ramsey, A.F. (2020), Probability Distributions of Crop Yields: A Bayesian Spatial Quantile Regression Approach. *Amer. J. Agr. Econ.*, 102(1), 220-239. doi:10.1093/ajae/aaz029

4. Reich, B. J., Fuentes, M., & Dunson, D. B. (2011). Bayesian Spatial Quantile Regression. *Journal of the American Statistical Association*, 106(493), 6–20. doi:10.1198/jasa.2010.ap09237

5. Smith, L. B., Reich, B. J., Herring, A. H., Langlois, P. H., & Fuentes, M. (2015). Multilevel quantile function modeling with application to birth outcomes. *Biometrics*, 71(2), 508–519. doi:10.1111/biom.12294

6. Sriram K. (2015). A sandwich likelihood correction for Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Statistics & Probability Letters*, 107, 18-26. doi:10.1016/j.spl.2015.07.035

7. Sriram, K., Ramamoorthi, R.V., Ghosh, P. (2013). Posterior Consistency of Bayesian Quantile Regression Based on the Misspecified Asymmetric Laplace Density. *Bayesian Analysis.* 8(2), 479-504. doi:10.1214/13-BA817.

8. Wu, T., & Narisetty, N. (2020). Bayesian Multiple Quantile Regression for Linear Models Using a Score Likelihood. *Bayesian Analysis*, adv issue. doi:10.1214/20-BA1217

9. Yang, Y., Wang, H., & He, X. (2016). Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. *International Statistical Review*, 84(3), 327-344. doi:10.1111/insr.12114

10. Yu, K., & Moyeed, R. (2001). Bayesian Quantile Regression. *Statistics & Probability Letters*, 54(4), 437-447. doi:10.1016/S0167-7152(01)00124-9