STAT 647 Final Presentation

John Schwenck & Marshall Honaker

11/24/2020

Spatial Bayesian Multiple Quantile Regression Using Adaptive Importance Sampling

Motivation

We propose a spatial adaptation of the score-based likelihood approach to quantiel regression proposed in *Bayesian Multiple Quantile Regression for Linear Models Using a Score Likelihood* by Wu & Narisetty (2020) and use the "Ada-IS" importance sampling algorithm to approximate posterior summaries

Example Application

Draw inference about various crimes in NYC based on geographic region. Data from *NYC Open Data*.

Quantile Regression - Overview

Oftentimes, conditional mean regression is sensitive to extreme events/outliers and fails to adequately describe these characteristics of the underlying population. Quantile regression (QR) can be used to investigate specific parts of a distribution

More formally, QR models the relationship between a p-dimensional vector of covariates $\mathbf{X} = \mathbf{x}$ and $Q_\tau(Y|\mathbf{x})$ where

$$Q_\tau(Y|\mathbf{x}) \equiv Q_\tau(Y|\mathbf{X} = \mathbf{x}) = \inf\{y : P(Y \leq y|X = \mathbf{x}) \geq \tau\} = \mathbf{x}^\top \beta$$

where $\tau \in (0, 1)$

The objective function for QR is therefore:

$$\widehat{\beta} = \arg\min_\beta \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \beta)$$

where the check loss function is given by $\rho_\tau(u) = u(\tau - I(u < 0))$

Quantile Regression - Overview

Remark

- QR does not make any assumptions about a parametric likelihood which is problematic for the Bayesian framework for the posterior estimatio.
- Previous attempts have used the Asymmetric Laplace Distribution. However, inference procedures based on the posteriors from these procedures are not valid. This paper proposes a *working* likelihood that incorporates the check loss function (this allows for valid frequentist inference procedures based on the posterior distribution)

Spatial Bayesian Multiple Quantile Regression - Interpretation

In the Bayesian framework, testing & inference about multiple quantile levels re-parameterize the QR model with $\theta = \beta_1(\tau_1) - \beta_2(\tau_2)$ and test for differences in the slope (i.e. whether $\theta = 0$). The hypotheses for two different quantiles ($\tau_1 > \tau_2$) would be stated as follows:

$$H_0 : \beta_1(\tau_1) = \beta_2(\tau_2) \text{ vs } H_1 : \beta_1(\tau_1) \neq \beta_2(\tau_2)$$

Interpretation

- Test whether the effects of the first covariate X_1 on Y are different for quantiles τ_1 and τ_2 based on the location
- If 0 is not within the credible interval of the posterior, reject H_0

Spatial Bayesian Multiple Quantile Regression - Method

The authors define the ***score function*** as a p-dimensional vector of the following form:

$$s_\tau(\beta) = \sum_{i=1}^{n} x_i \psi_i(y_i - x_i^\top \beta)$$

where $\psi_i(u) = \tau - I(u < 0)$

Note:

- since $\hat{\beta}(\tau)$ minimizes the quantile loss function, $s_\tau(\hat{\beta}(\tau)) \approx 0$

Spatial Bayesian Multiple Quantile Regression - Method

The authors also propose the following as the ***working likelihood***:

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

where $\mathbf{W}$ is a $p \times p$ positive definite matrix given by

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

and C is a normalizing constant free of β

Spatial Bayesian Multiple Quantile Regression - Method

To account for the spatial dependencies in the data, we augment the weight matrix $\mathbf{W}$ with spatial basis functions to smooth and account for the spatial variability

Importance Sampling

Importance sampling gives us a way to approximate quantities of interest for a given distribution, even if we can't directly sample from that particular distribution

$$\mathbb{E}_{\color{red}g}[X] = \sum_x x \frac{\color{red}g(x)}{\color{blue}f(x)} {\color{blue}f(x)} = \mathbb{E}_{\color{blue}f}\left[X \frac{g(x)}{f(x)}\right]$$

So,

$$\mathbb{E}_g(X) \approx \frac{1}{n} \sum_{i=1}^{n} x_i \frac{g(x_i)}{f(x_i)}$$

where $w_i = \frac{g(x_i)}{f(x_i)}$ are the importance sampling weights

Algorithm - Single Quantile

1. Initialize the proposal distribution with a linear mean (or median) regression model where $q(\boldsymbol{b}) = N(\boldsymbol{a}, \hat{\Sigma})$ with

$$\hat{\Sigma} = c\tau(1-\tau)\left(\sum_{i=1}^{n} x_i x_i^\top / n\right)^{-1}$$

2. Generate samples $b^{(1)}, b^{(2)}, ..., b^{(m)}$ from $q(b)$ for some large M and calculate the importance weights

$$w^{(r)} = \frac{L(b^{(r)})\pi(b^{(r)})}{q(b^{(r)})}$$

and Effective Sample Size $ESS = \frac{M}{1+cv^2}$ where $cv^2 = \frac{(M-1)^{-1}\sum_{r=1}^{M}(w^{(r)}-\bar{w})^2}{\bar{w}^2}$ and $\bar{w} = \frac{1}{n}\sum_{r=1}^{M} w^{(r)}$

Algorithm - Single Quantile

3. Estimate posterior mean and covariance:

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, ..., \hat{\mu}_m) = \frac{\sum_{i=1}^{M} w^{(r)}\mathbf{b}^{(r)}}{\sum_{i=1}^{M} w^{(r)}}$$

   and

$$\hat{\sigma}_{j,k} = \frac{\sum_{r=1}^{M} w^{(r)}\mathbf{b}_j^{(r)}\mathbf{b}_k^{(r)}}{\sum_{r=1}^{M} w^{(r)}} - \hat{\mu}_j\hat{\mu}_k$$

   for the jk^{th} entry of the posterior covariance matrix where $b_j^{(r)}$ is the j^{th} coordinate of $b^{(r)}$

4. Use the mean and covariance matrix from (3) to form a new normal proposal distribution and repeat steps 2-4 a pre-specified number of times

5. Estimate posterior mean and covariance using importance sampling from the final proposal distribution from (4)

Algorithm - Multiple Quantiles

1. Implement the Ada-IS algorithm for a single quantile separately for each $\tau \in \{\tau_1, ..., \tau_m\}$ which will result in $(\hat{\boldsymbol{\mu}}_\mathbf{1}, ..., \hat{\boldsymbol{\mu}}_\boldsymbol{m})$ and $(\hat{\boldsymbol{\Sigma}}_\mathbf{1}, ..., \hat{\boldsymbol{\Sigma}}_\boldsymbol{m})$
2. Use an mp-dimensional multivariate normal proposal with the results from (1) as the mean and block diagonal joint covariance, respectively. The off diagonal covariances for quantile levels τ_i, τ_j are imputed using the following:

$$\hat{\Sigma}_{i,j} = \gamma(\min(\tau_i, \tau_j) - \tau_i\tau_j)\left(\frac{(\tau_i(1-\tau_i)\hat{\Sigma}_i^{-1} + \tau_j(1-\tau_j)\hat{\Sigma}_j^{-1})}{2}\right)^{-1}$$

3. Generate samples from the proposal distribution and estimate the mean and covariance matrix and perform (3) from the single quantile QR described previously
4. Repeat a pre-specified number of times and estimate the posterior using the importance samples from (3) above

Simulation

To assess the performance of the model and make sure that it works properly, we conducted a simulation study

10,000 values of 5 covariates (each from a distinct distribution) taken from random locations on a 10x10 grid were used to generate simulated response values

For the augmented feature matrix approach, the simulation will allow us to test other aspects of the model, such as:

1. Effects of different numbers of bases (though this could be chosen using cross validation)
2. Predictions and approximations to the posterior predictive distribution (this will also allow us to compute training and test errors)
3. Assess validity of frequentist inference procedures based on the posterior (are we able to retain this property from Bayesian QR model proposed in the original paper?)

Real-World Application - NYC Crime Data

To apply our method, we obtained data from the NYC Open Data API via Python. Due to size limits, we had to constrain our data to 5 GB worth which includes all historical and YTD data. All variables included a corresponding latitude and longitude.

Response: Location-based crime severity index

Covariates:

- Shootings by location and demographics
- Arrests by location and demographics
- Motor vehicle crashes / injuries by location
- Crime complaints by location and crime type
- Court summons by location
- Income by district

Given different levels of the crime index, the effects of the covariates change according to the quantile level

Future Directions

- Convert coding bottlenecks to C++ / Rcpp for improved speed
- Incorporate Kernel Basis Functions to non-parametrically estimate weight matrix
- Apply our method to other cities that offer open source crime data
- Addition of regularization term(s) to the objective function (what can we learn if certain variables are included/larger at certain quantile levels but not at others?)