Bayesian Density Regression for Count Data

Charalampos Chanialidis[1], Ludger Evers[2], Tereza Neocleous[2]

Second Bayesian Young Statisticians Meeting (BAYSM 2014)
Vienna, September 18–19, 2014

[1] University of Glasgow, Department of Mathematics and Statistics, Glasgow, United Kingdom
c.chanialidis.1@research.gla.ac.uk

[2] University of Glasgow, Department of Mathematics and Statistics, Glasgow, United Kingdom
<ludger.evers,tereza.neocleous>@glasgow.ac.uk

Abstract

Quantile regression models for count data have so far received little attention. The main quantile regression technique for count data involves adding uniform random noise, thus overcoming the problem that the conditional quantile function is not a continuous function of the parameters of interest. This method has the drawback that, for small values of the response variable Y, the added noise can have a large influence on the estimated quantiles. In addition, quantile regression can lead to "crossing" quantiles. We propose a Bayesian Dirichlet process (DP)-based approach that is based on an adaptive DP mixture of COM-Poisson regression models and determines the quantiles by estimating the density of the data, thus eliminating all the aforementioned problems.

Keywords: Dirichlet process; quantile regression; Bayesian nonparametrics.

1 Introduction

Quantile regression was introduced as a nonparametric method for modelling a variable of interest as a function of covariates [3]. By estimating the conditional quantiles rather than the mean, it gives a more complete description of the conditional distribution of the response variable than least squares regression.

The problem with applying quantile regression to count data is that the cumulative distribution function of the response variable is not continuous, resulting in quantiles that are not continuous, which can not be expressed as a continuous function of the covariates. One way to overcome this problem is by adding uniform random noise ("jittering") to the counts [4].

We propose an adaptive Dirichlet process mixture approach which estimates the conditional density of the data. The approach is based on an adaptive Dirichlet Process mixture (DPM) of COM-Poisson regression models.

2 COM-Poisson distribution

The COM-Poisson distribution [1, 8] is a two-parameter generalisation of the Poisson distribution that allows for different levels of dispersion. Its probability mass function is

$$P(Y = y|\mu, \nu) = \left(\frac{\mu^y}{y!}\right)^\nu \frac{1}{Z(\mu, \nu)} \quad \text{where } Z(\mu, \nu) = \sum_{j=0}^{\infty} \left(\frac{\mu^j}{j!}\right)^\nu \quad \text{and} \quad y = 0, 1, 2, \dots \tag{1}$$

where

$$\mathbb{E}[Y] \approx \mu, \quad \mathbb{V}[Y] \approx \frac{\mu}{\nu} \tag{2}$$

One can obtain a point mass by letting the variance parameter ν tend to infinity. Thus one can show that mixtures of COM-Poisson distributions can provide an arbitrarily precise approximation to any discrete distribution with support $\mathbb{N}_0$, which is why COM-Poisson distributions are used by our method.

A regression model can be defined based on (1),

$$\log \boldsymbol{\mu_i} = \boldsymbol{x_i}^\intercal \boldsymbol{\beta} \tag{3}$$

$$\log \boldsymbol{\nu_i} = \boldsymbol{x_i}^\intercal \boldsymbol{c} \tag{4}$$

It can be seen, in the next subsection, that the calculation of the normalisation constant of this distribution is redundant.

2.1 Exchange algorithm

Any probability density function $p(y|\theta)$ can be written as

$$p(y|\theta) = \frac{q_\theta(y)}{Z(\theta)} \tag{5}$$

where $q_\theta(y)$ is the unnormalised density and the normalising constant $Z(\theta) = \int p(y, \theta)\,\mathrm{d}y$ is unknown. In this case the Metropolis-Hastings acceptance ratio is

$$\alpha = \min\left(1, \frac{q_{\theta^*}(y)\pi(\theta^*)Z(\theta)h(\theta|\theta^*)}{q_\theta(y)\pi(\theta)Z(\theta^*)h(\theta^*|\theta)}\right) \tag{6}$$

where $\pi(\theta)$ is the prior of θ. The ratio in (6) involves computing unknown normalising constants. Introducing auxiliary variables θ^*, y^* and sampling from an augmented distribution

$$\pi(\theta^*, y^*, \theta|y) \propto p(y|\theta)\pi(\theta)p(y^*|\theta^*)h(\theta^*|\theta) \tag{7}$$

results in

$$\begin{aligned}
\alpha &= \min\left(1, \frac{p(y|\theta^*)\pi(\theta^*)p(y^*|\theta)h(\theta|\theta^*)}{p(y|\theta)\pi(\theta)p(y^*|\theta^*)h(\theta^*|\theta)}\right) && (8)\\
&= \min\left(1, \frac{q_\theta(y^*)\pi(\theta^*)h(\theta|\theta^*)q_{\theta^*}(y)Z(\theta)Z(\theta^*)}{q_\theta(y)\pi(\theta)h(\theta^*|\theta)q_{\theta^*}(y^*)Z(\theta^*)Z(\theta)}\right) && (9)\\
&= \min\left(1, \frac{q_\theta(y^*)\pi(\theta^*)q_{\theta^*}(y)}{q_\theta(y)\pi(\theta)q_{\theta^*}(y^*)}\right) && (10)
\end{aligned}$$

where the normalising constants cancel out and $h()$ is a symmetric distribution [6, 5].

3 Bayesian density regression

Density regression allows flexible modelling of the response variable Y given the covariates $\boldsymbol{x} = (x_1, \ldots, x_p)'$. Features of the conditional distribution of the response variable, vary with $\boldsymbol{x}$, so, depending on the predictor values, these features can change in a different way than the population mean. Bayesian methods for density regression are considered in [2] where the conditional distribution of the response variable is expressed as a mixture of regression models where the mixing weights vary with covariates.

This paper focuses on the following mixture:

$$f(y_i|\boldsymbol{x}_i) = \int f(y_i|\boldsymbol{x}_i, \phi_i) G_{\boldsymbol{x}_i}(\mathrm{d}\phi_i) \text{ where } f(y_i|\boldsymbol{x}_i, \phi_i) = \text{COM-P}(y_i; \exp(\boldsymbol{x}_i'\boldsymbol{b}_i), \exp(\boldsymbol{x}_i'\boldsymbol{c}_i)) \tag{11}$$

the conditional density of the response variable is expressed as a mixture of COM-Poisson regression models with $\phi_i = (\boldsymbol{b}_i, \boldsymbol{c}_i)$ and $G_{\boldsymbol{x}_i}$ is an unknown mixture distribution that depends on the location of $\boldsymbol{x}_i$.

3.1 MCMC algorithm

The MCMC algorithm alternates between:

- updating the allocation parameter, assigning each observation either to a new mixture component or to an already existing one
- and updating the parameters (μ, ν) for each mixture component.

Unlike the model in [2], there is no closed form expression for the posterior distribution and approximation of the probability of allocating observations to a new cluster is difficult.

We overcome this problem by bridging: i) an MCMC algorithm for sampling from the posterior distribution of a Dirichlet process model, with a non-conjugate prior, found in [7]; ii) the MCMC algorithm found in [2]; and iii) a variation of the MCMC exchange algorithm.

4 Simulations

We consider two simulated datasets to compare the proposed method to the "jittering" method. These are

$$Y_i|X_i = x_i \sim \text{Binomial}(10, 0.3x_i) \tag{12}$$

$$Y_i|X_i = x_i \sim 0.4\text{Pois}(\exp(1 + x_i)) + 0.2\text{Binomial}(10, 1 - x_i) + 0.4\text{Geom}(0.2) \tag{13}$$

where $x_i \sim \text{Unif}(0, 1)$. Table (1) shows the absolute mean errors, $\mathbb{E}[|q_p - \hat{q_p}|]$, where q_p is the true conditional quantile when $x = p$ and $\hat{q_p}$ is the estimated conditional quantile. The new method outperforms the "jittering" method and in almost all cases the "jittering" method leads to crossing quantiles (except when $n = 500$).

References

[1] Conway, Richard W. and Maxwell, William L. (1962) "A Queuing Model with State Dependent Service Rate." *Journal of Industrial Engineering*, **12**, 132–136.

Method	Number of Observations					
	Binomial			Mixture		
	20	100	500	20	100	500
Density regression	0.5576	**0.2820**	**0.2421**	**0.7435**	**0.5833**	**0.3589**
Jittering (linear)	**0.5256**	0.8461	0.4765	1.1923	0.6666	0.4294
Jittering (splines)	0.7820	0.5128	0.3020	1.9487	0.8269	0.3910

Table 1: Mean absolute error obtained using the different density/quantile regression methods.

[2] Dunson, David B. and Pillai, Natesh and Park, Ju-Hyun. (2007) "Bayesian Density Regression." *Journal of the Royal Statistical Society: Series B*, **69**, 163–183.

[3] Koenker, Roger and Bassett, Gilbert (1978) "Regression Quantiles." *Econometrica*, **46**, 33–50.

[4] Machado, Josè António Ferreira and Santos Silva, João M.C.. (2005) "Quantiles for Counts." *Journal of the American Statistical Association*, **100**, 1226–1237.

[5] Møller, J. and Pettitt, A. N. and Reeves, R. and Berthelsen, K. K. (2006) "An Efficient Markov Chain Monte Carlo Method for Distributions with Intractable Normalising Constants." *Biometrika*, 93**2**, 451–458.

[6] Murray, Ian and Ghahramani, Zoubin and MacKay, David J. C. (2006) "MCMC for Doubly-intractable Distributions." *Proceedings of the 22nd Annual UAI Conference*, 359–366.

[7] Neal, Radford M. (2000) "Markov Chain Sampling Methods for Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics*, **9**, 249–265.

[8] Shmueli, Galit and Minka, Thomas P. and Kadane, Joseph B. and Borle, Sharad and Boatwright, Peter (2008) "A Useful Distribution for Fitting Discrete Data: Revival of the Conway-Maxwell-Poisson Distribution." *Journal of the Royal Statistical Society: Series C*, **54**, 127–142.