BAYESIAN INFERENCE ON REGRESSION MODEL WITH AN UNKNOWN CHANGE POINT

Abstract

In this work, we describe a Bayesian procedure for detection of change-point when we have an unknown change point in regression model. Bayesian approach with posterior inference for change points was provided to know the particular change point that is optimal while Gibbs sampler was used to estimate the parameters of the change point model. The simulation experiments show that all the posterior means are quite close to their true parameter values. The performance of this method is recommended for multiple change points.

Keywords: Change point, Gibbs sampler, optimal, Regression, Simulation.

1. Introduction

Detection of change points has recently gained popularity in literature due to its applications in areas of climatology, econometrics, bioinformatics, and network traffic analysis among others (see Aminikhanghahi and Cook (2017), Gold et al. (2018)). Notable work on change point analysis began with the work of Page (1954). Page developed a test for a change in parameter which occurs at an unknown point. This test deals with the identification of subsamples and detection of the changes in the parameter value. Subsequently, development of different detection of changes has emerged for various phenomena.

Change point is referred as location or time point where the observation follows different distributions before and after that point. For a given sample size of n independent observations $\{Y_1, Y_2, \ldots, Y_n\}$, a change point can occurs if and only if there exists a $k \in [1, n-1]$, such that the distributions of $\{Y_1, Y_2, \ldots, Y_k\}$ and $\{Y_1, Y_2, \ldots, Y_n\}$ are different with respect to some criteria (Kang, 2015). Some of the most commonly used criteria are: change in mean, change in regression coefficients, and change in variance. However, change in regression coefficients criterion will be used in this study.

Bayesian approach can provide adequate uncertainty information about location of change point. Literature on Bayesian regression with change point can be seen in the works of

Raftery and Ackman (1986), Baryy and Hartigan (1993), Stephens (1994), Gossl and Kuchenhoff (2001), Moreno, et al. (2005), Giron et al (2007), Pandya et al (2011), Datta et al. (2019) among others.

Bayesian estimation of logistic regression model with unknown threshold limiting value was considered by Gossl and Kuchenhoff (2001). It was assumed that there is no effect of covariate on the response variable under certain unknown threshold limiting value while measurement error in the covariate was also was also accounted in the model. Bayesian method outperformed the likelihood solution with the use of data set.

Pandya et al. (2011) studied a two-phase linear regression with one change point. Bayesian estimators were derived for symmetric loss function and most especially for asymmetric loss functions namely Linex and general entrophy loss functions while effects of correct and wrong prior information on these Bayes estimators were studied. It was observed that the Bayes estimators are robust with correct choice of prior specifications.

Bayesian of high dimensional shrinkage priors in a change point setting to understand segment-specific relationship between the dependent and regressors was examined by Datta et al (2019). Results obtained from both the simulation and real life data reveals that the Bayesian approach can deliver accurate variable selection while inference on the location of change points substantially outperforms the classical LASSO based approach.

This work demonstrates how one can exploit a Bayesian inference on regression model with an unknown change point setup to determine the location of the change point and also identify the true sparse support for each of the linear models.

This paper is organized as follows. In section 2, we introduce a typical regression model with an unknown change. Bayesian procedure for estimation of regression model with unknown change point is presented in section 3. Section 4 gives the numerical study on the model while section 5 presents the results of analyses of data generated from the models using the Bayesian approach. Section 6 concludes this paper.

2. Regression Model with Unknown change Point

Consider a density time series variable, $y_{t,}$, $t = 1,\ldots,T$ conditioned on its lags, the covariates and model parameters, the density of such series can be simply be expressed as:

$$y_t|\psi_1, \psi_2, \beta_1, \beta_2, \sigma^2, \tau^2, \lambda, x_t \sim \begin{cases} N(\psi_1 + \psi_2 x_t,\ \sigma^2), & if \ \ t \le \lambda \\ N(\beta_1 + \beta_2 x_t,\ \tau^2), & if \ \ t > \lambda \end{cases} \quad (1)$$

In model (1), λ is called a change point, it means that for periods until λ, one regression will assumed to generate variable and following λ, another regression will also assumed to generate y.

3. Materials and methods

3.1 Bayesian inference

The likelihood function of the model in (1) is given as:

$$L(\psi_1, \psi_2, \beta_1, \beta_2, \sigma^2, \tau^2, \lambda) = \prod_{t \le \lambda} \phi(y_t; \psi_1 + \psi_2 x_t, \sigma^2) \prod_{t > \lambda} \phi(y_t; \beta_1 + \beta_2 x_t, \tau^2) \quad (2)$$

We utilized priors of the form given as:

For parameter ψ,

$$\psi = (\psi_1, \psi_2)' \sim N(\mu_\psi, Q_\psi) \quad (3)$$

For parameter β,

$$\beta = (\beta_1, \beta_2)' \sim N(\mu_\beta, Q_\beta) \quad (4)$$

while

$$\sigma^2 \sim IG(p_1, p_2) \quad (5)$$

$$\tau_2 \sim IG(q_1, q_2) \quad (6)$$

and

$$\lambda \sim Unif(1, 2, ..., T-1) \quad (7)$$

It is observed λ is known to be parameter of the model, and placing a uniform prior over all the elements, that is, $1, 2, ..., T-1$, it means a change point is assumed to occur Koop et al (2005).

Customarily, the posterior distribution is proportional to both the likelihood times prior and can be written as:

$$\psi \mid \beta, \sigma^2, \tau^2, \lambda, y \sim N(D_\psi d_\psi, D_\psi) \quad (8)$$

where

$$D_\psi = \frac{1}{\left(\frac{x_\psi^{'} x_\psi}{\sigma^2} + Q_\psi^{-1}\right)}$$

and

$$d_\psi = \frac{x_\psi^{'} y_\psi}{\sigma^2} + \frac{\mu_\psi}{Q_\psi}$$

Also,

$$x_\psi = x_\psi(\lambda) = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ . & . \\ 1 & x_\lambda \end{pmatrix}$$

and

$$y_\psi = y_\psi(\lambda) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_\lambda \end{bmatrix}$$

The conditional posterior for β can also be obtained as:

$$\beta \mid \psi, \sigma^2, \tau^2, \lambda, y \sim N(D_\beta d_\beta, D_\beta) \quad (9)$$

where,

$$D_\beta = \frac{1}{\left(\frac{x_\beta^{'} x_\beta}{\tau^2} + \frac{1}{Q_\beta}\right)}$$

and

$$d_\beta = \frac{x_\beta' y_\beta}{\tau^2} + \frac{\mu_\beta}{Q_\beta}$$

Also,

$$x_\beta = x_\beta(\lambda) = \begin{pmatrix} 1 & x_\lambda + 1 \\ 1 & x_2 + 2 \\ \vdots & \vdots \\ . & . \\ 1 & x_T \end{pmatrix}$$

and

$$y_\beta = y_\beta(\lambda) = \begin{bmatrix} y_{\lambda+1} \\ y_{\lambda+2} \\ \vdots \\ y_T \end{bmatrix}$$

while the conditional posteriors for variance parameters are:

$$\sigma^2 \mid \beta, \psi, \tau^2, \lambda, y \sim IG\left[\frac{\lambda}{2} + p_1, \frac{1}{\left(\frac{1}{p_2 + \frac{1}{2}\sum_{t=1}^{\lambda}(y_t - \psi_1 - \psi_2 x_t)^2}\right)}\right] \quad (10)$$

and

$$\tau^2 \mid \beta, \psi, \sigma^2, \lambda, y \sim IG\left[\frac{T-\lambda}{2} + q_1, \frac{1}{\left(\frac{1}{q_2} + \frac{1}{2}\sum_{t=\lambda+1}^{T}(y_t - \beta_1 - \beta_2 x_t)^2\right)}\right] \quad (11)$$

Hence, the posterior for the change point, λ can be given as:

$$P(\lambda \mid \beta, \psi, \sigma^2, \tau^2, y) \propto \prod_{t \le \lambda}^{T} \phi(y_t; \psi_1 + \psi_2 x_t, \sigma^2) \prod_{t > \lambda}^{T} \phi(y_t; \beta_1 + \beta_2 x_t, \tau^2), \lambda = 1, \ldots, T-1 \quad (12)$$

In general, there is no way to simplify the expression in (12), because it does not take the form of any well-known density. Since λ is known to be discrete valued, we can obtain the un-normalized density ordinates when $\lambda = 1, \ldots, T-1$ and also make the distribution proper by simply dividing each of the ordinate by sum of the ordinate values.

4 Numerical study

The lack of analytical results relating to the posterior obtained in section (3.1) suggests a posterior simulator. This posterior simulator will helps to obtain draws from the discrete distribution. In this section, we employ a Gibbs sampler to estimate the parameters of the change point model given in section (3) from the full conditionals posterior distribution.

We describe a simulation study for assessing the performance of the presented method. Thus we have:

$$y_t|\psi_1, \psi_2, \beta_1, \beta_2, \sigma^2, \tau^2, \lambda, x_t \sim \begin{cases} N(3+5.5x_t,\ 1), & if \ \ t \leq \lambda \\ N(5+2x_t,\ 1.5), & if \ \ t > \lambda \end{cases} \tag{13}$$

(i) The values of λ are: 30, 50, and 80 while the sample size are 300 and 500

(ii) The regressor is generated as: $x_t \sim N(0,1)$

(iii)The prior hyper-parameters were set as:

$$\mu_\psi = \mu_\beta = 0$$
$$Q_\psi = Q_\beta = 100I_2$$
$$p_1 = q_1 = 5$$
$$p_2 = q_2 = \frac{1}{2}$$

The Gibbs sampler is run 2000 times while 200 will be used as burn-in period

5 Results and discussion

In Tables 1 and 2, the posterior means and standard error for were given for regression parameters, variance parameters, and change point parameter (β_1, β_2, ψ_1 and ψ_2), (σ^2 and τ^2), and λ respectively. The true values of the parameter were given in the parentheses.

For parameter beta, when the sample sizes are 300 and 500, the optimal change point is 30 having the smallest standard error while the posterior means are closest to the true parameters. For parameter tau, the optimal change point is 80 having the smallest standard error while the posterior means are closest to the true parameters. It is apparent that the optimal change points for variance parameters σ^2 and τ^2 are 80 and 30, respectively having least standard error for sample sizes of 300 and 500.

It is important to note that for all models, standard error of change point tends when λ = 80 for sample size of 300 is the smallest while standard error of change point when λ = 80 for sample size of 500 is tends to be the smallest. This behaviour is expected as for λ =80, both the sample sizes (300 and 500) for estimating beta and tau are effectively 80.0011 and 80, respectively.

Table 1: Results from simulation when the sample size, n=300 for change points at 30, 50, and 80.

	Mean			Standard error		
	Change point = λ			Change point= λ		
Parameters	30	50	80	30	50	80
β_1 =(5)	4.9780	5.0609	5.0549	0.0402	0.0454	0.0425
β_2 =(2)	2.0354	1.9321	1.9675	0.0398	0.0469	0.0457
ψ_1 = (3)	2.9740	2.9461	3.0414	0.1058	0.0685	0.0522
ψ_2 = (5.5)	5.4605	5.5967	5.4687	0.0976	0.0780	0.0508
σ^2 = (0.2)	0.3236	0.2228	0.2126	0.0812	0.0466	0.0335
τ^2 = (0.5)	0.4489	0.5319	0.4050	0.0378	0.0477	0.0391
λ	29.2850	49.9472	80.0011	0.4515	0.2237	0.0577

Table 2: Results from simulation when the sample size, n=500 for change points at 30, 50, and 80.

	Mean			Standard error		
	λ			λ		
Parameters	30	50	80	30	50	80
β_1 =(5)	4.9628	4.9730	5.0385	0.0311	0.0339	0.0330
β_2 =(2)	1.9811	1.9984	1.9597	0.0323	0.0372	0.0353
ψ_1 = (3)	3.0362	2.9793	2.8103	0.1086	0.0746	0.0546
ψ_2 = (5.5)	5.4797	5.5278	5.5339	0.1047	0.0776	0.0598
σ^2 = (0.2)	0.3432	0.2740	0.2175	0.0849	0.0549	0.0335
τ^2 = (0.5)	0.4634	0.5445	0.4507	0.0305	0.0352	0.0310
λ	30.0478	49.8539	80	0.2134	0.3806	0

6 Conclusion

In Bayesian modelling framework, computationally efficient technique for applying Bayesian in the analysis of regression model with unknown change point has been applied. The applied Bayesian technique was used to know the particular change point that is optimal. It was observed that all the posterior means are quite close to their true parameter values. The optimal value of the change point is 80 in the estimation of regression model while β and ψ have the same support. It is recommended that the performance of this method can also be used for multiple change points.

References

Aminikhanghahi, S. And Cook, D. J. 2017. A Survey of Methods for Time series change point detection. *Knowl Inf Syst.* 51(2), 339-367.

Barry, D. and Hartigan, J. A. 1993. A Bayesian analysis for change point problems. *Journal of American statistical association*, 88, 309-319.

Datta, A., Zou, H. and Banerjee, S. 2019. Bayesian high-dimensional regression for change point analysis. *Stat Interface* 12(2), 253-264.

Giron, F. J., Moreno. E. and Casella, G. 2007. Objective Bayesian analysis of multiple change points for linear models. *Bayesian Statistics* 8, 1-27.

Gold, N., Frasch, M. G., Herry, C. L., Richardson, B. S. and Wang, X. 2018. A Doubly Stochatic change point Detection Algorithm for Noisy Biological signals. *Front. Physiol.* 8:1112. doi: 10.3389/fphys.2017.01112

Gossl, C. and Kuchenhoff, H. 2001. Bayesian analysis of logistic regression with an unknown change point and covariate measurement error. *Statistics in med*icine 23(20).

Kang, S. 2015. Bayesian change-point analysis in linear regression model with scale mixtures of normal distributions. Masters dissertation, Michigan technological University.

Kook, G., Poirier, D. J. and Tobias, J. L. 2005. Bayesian econometric methods, Cambridge press University.

Moreno, E. Casella, G. and Garcia-Ferrer, A. 2005. Objective Bayesian analysis of the change point problem. Stochastic Environ. *Res. Risk assessment* 19, 191-204.

Pandya, M., Bhatt, K., and Andharia, M. 2011. Bayes estimation of two-phase linear regression model. *International Journal of Quality, Statistics and Reliability* 2011, 1-9.

Page, E. S. 1954. Continuous inspection scheme. *Biometrika* 41,100-115.

Raftery, A. E. And Akman, V. E. 1986. Bayesian analysis of a Poisson process with a change point. *Biometrika* 73, 85-89.

Stephens, D. A. 1994. Bayesian retrospective multiple-change point identification. *Applied Statist*, 43, 159-168