Mahesh Pandurang Mohite
IIT Kanpur

Self Project
Parameter and State estimation

Bayesian Online Changepoint Detection (BOCD)

Please note: Refer to problem statement pdf file for complete question statement.

Critical Review of paper

Bayesian model for online detection of changepoints in time-series data.
There could be abrupt variations in a data sequence that could be characterized by any combination of parameters such as mean or variance that may abruptly change after a particular time. Change point is referred to as finding the earliest time at which this variation occurs. Bayesian online change point detection is a Bayesian framework for online estimation of change points. Online Change point detection has many applications in the areas such as finance, biometrics and robotics.

Main Goals

- Most Bayesian Approaches to change point detection is offline and are using retrospective Segmentation algorithms. So In this paper, the main goal of the author is to present the Bayesian Change point Detection for Online Inference.
- So, given only the data already observed, the author focuses on generating accurate distributions of the next unseen datum in the sequence using Bayesian approach.

BOCD Algorithm

- Recursive Bayesian estimation is the main algorithm in this paper which is used to detect changes in streaming data in an online fashion. The model parameters before and after the change point are independent is the main assumption. The time since the last change point happens is referred to run length. The main idea is to compute the posterior distribution of the Run length. Run length is a discrete random variable which can takes positive integers
- At each time, we will observe the new data and for each time, the run length can take t+1 values. We have to find out the posterior probability for each of the t+1 run length values. The run length value may grow, that is it will be added with 1 (growth probabilities) in the next time instant or it will fall to zero (change point probabilities). We have to find growth and change point probabilities for all possible values of run length at each time instant. If change point probability is higher than all the growth probabilities, then the change point happened at that time instant.
- Given the data, in order to calculate the posterior distribution of run length we need three probability values. The predictive distribution which depends only on the recent data, The hazard function and the probability distribution of run length of previous time which we can get it by message passing (recursive). Here in order to calculate the predictive distribution which is called UPM, The author is using the properties of Exponential family curves and the advantage of using conjugate priors. So we can calculate the UPM distribution just by updating the parameters for the next iteration in the current time. It makes the algorithm very simple and computationally more efficient. The initial conditions and the priors will be taken accordingly

Usefulness and contribution

1. The algorithm functions at real time on streaming data. (Online nature)
2. The algorithm will detect any change in time series data — either mean or variance. Proper Conjugate priors, initial conditions and hazard functions should be used.
3. It has a provision for inputting all the domain knowledge information (priors) since it is using Bayesian approach.
4. The algorithm is robust to noise, occasional false positives and outliers.
5. The algorithm is elegant and we can able to understand the inner workings of the algorithm and the algorithm can be extended to suit our needs
6. The algorithm is highly modular as it can be applied to variety of types of data.

Limitations

1. The author didn't explain how to update the parameters and the priors after the first change point is detected. It makes the implementation tougher once after the first change point is detected.
2. We need to have proper prior information and proper initial conditions. For that we need to know the distribution of the data sequence.
3. The author also follows some assumptions like exponential family and to use conjugate priors. If these assumptions fail, predicting the UPM probabilities is highly difficult.

Implementation

The data sequence given is the well-drilling nuclear magnetic response data. Given that this process can be assumed to be piece-wise Gaussian and has multiple change points in mean. The standard deviation is assumed to be constant across change points but unknown. They have suggested that the change point occurrence in this process is memoryless and can thus be modelled as a geometric distribution with the timescale of 250. Hence for the Hazard function H0, 1/250 used for calculating change point probability where (1 – H0) will be used for finding growth probability.

The normal-gamma prior is conjugate for a Gaussian likelihood and it is used as a prior and can be constructed as

$$
\begin{aligned}
p(\mu, \lambda, \mu_0, k, , \beta) &= p(\mu|\lambda; \mu_0, k) \times p(\lambda; \alpha, \beta) \\
\text{where}, p(\mu|\lambda; \mu_0, k) &= \mathcal{N}(\mu_0, \frac{1}{k\lambda}) \\
p(\lambda; \alpha, \beta) &= \text{Gamma}(\alpha, \beta)
\end{aligned}
$$

Hence the posterior distribution for mean and variance will have the same family as the prior and the parameters are will be updated as:

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_0 + n}$$
$$\kappa_n = \kappa_0 + n$$
$$\alpha_n = \alpha_0 + \frac{n}{2}$$
$$\beta_n = \beta_0 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\kappa_0 n(\bar{x} - \mu_0)^2}{2(\kappa_0 + n)}$$

The initial parameters are μ_0= 1:15, k = 0:01, κ_0 = 20, and β_0 = 2. The posterior predictive distribution (UPM predictive) will be (since the mean is varying),

$$p(\mu|D) = T_{2\alpha_n}(\mu|\mu_n, \beta_n/(\alpha_n\kappa_n))$$

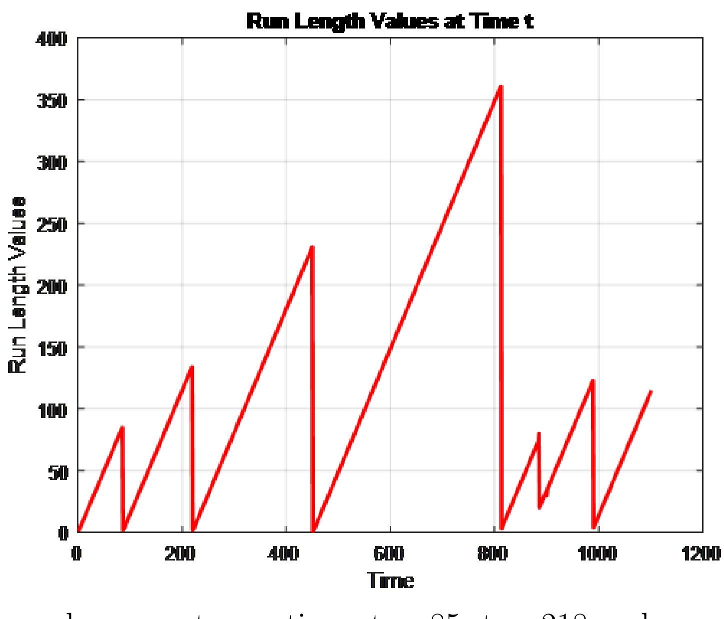
Which is the Students T distribution and its pdf is given by

$$t_v(x|\mu,\sigma^2) = c\left[1 + \frac{1}{\nu}(\frac{x-\mu}{\sigma})^2\right]^{\frac{\nu+1}{2})}$$
$$c = \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)} \times \frac{1}{\sigma\sqrt{\nu\pi}}$$

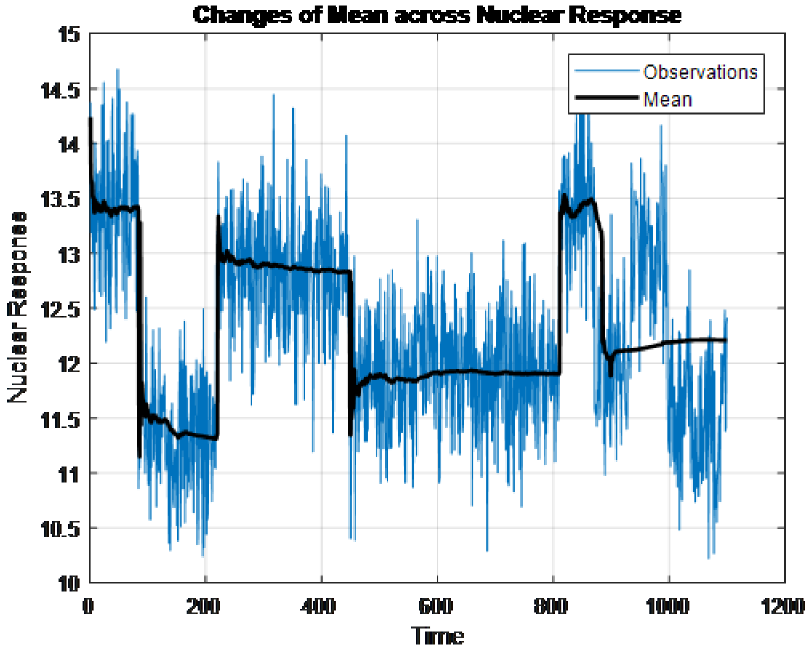
for the given observation x, degrees of freedom v. The probability of run length before the data comes is 1. R (0,0) = 1. With all the initial conditions and priors, the matrix R(r,t) (posterior distribution of Run length) is constructed by message passing algorithm at each time instant. Run length is taken in rows and time is taken in columns. At each column the row value with the maximum probability will be the corresponding run length value at that time instant. If row value 1 have the maximum probability then change point occurs. After the change point the initial parameters are updated accordingly.

Results

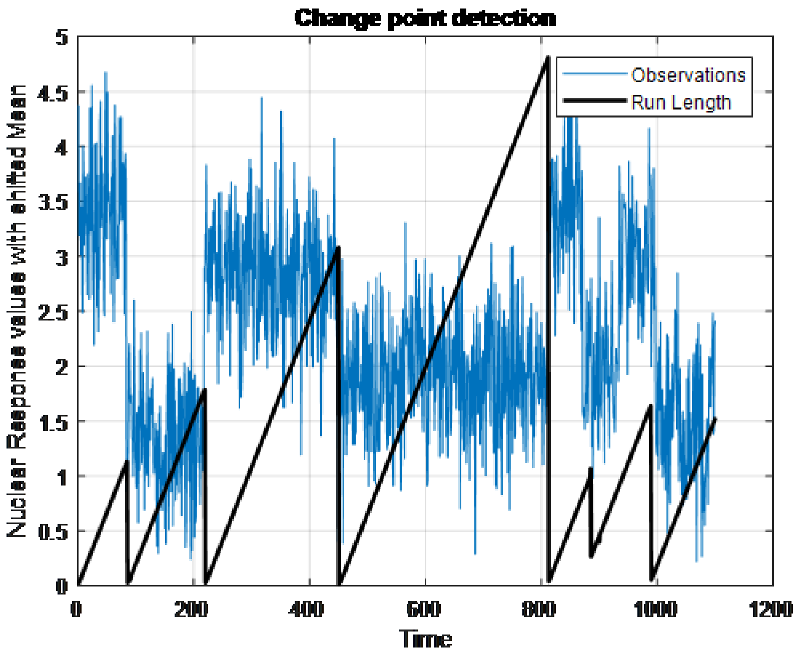
- **Run Length values with maximum Probability at time t** :



It can be seen at sometimes t = 85, t = 218 and so on the run length drops to zero. These time points are the change points

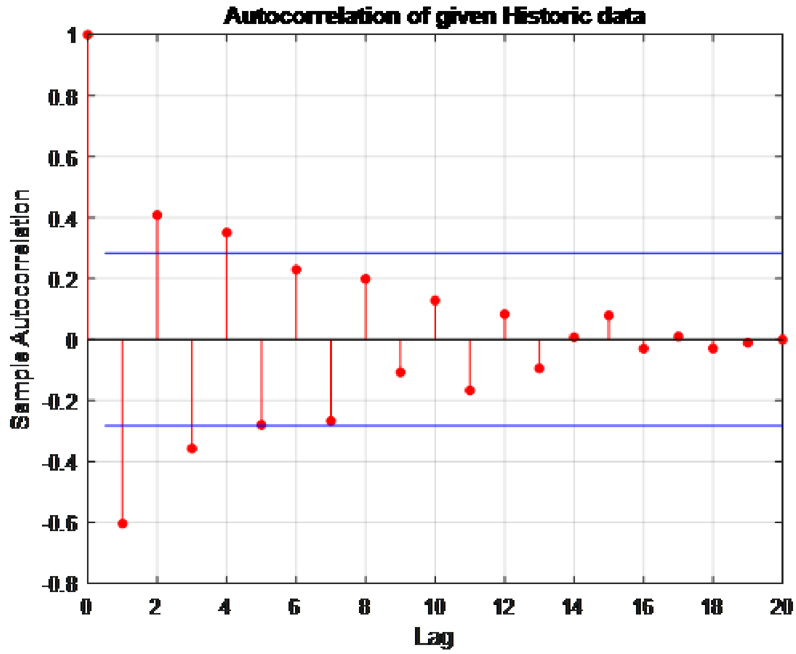


In the above plot we can see the parameter mean is changing across different change points. Initial mean 1.15 is not included in the plot for visualization purposes.

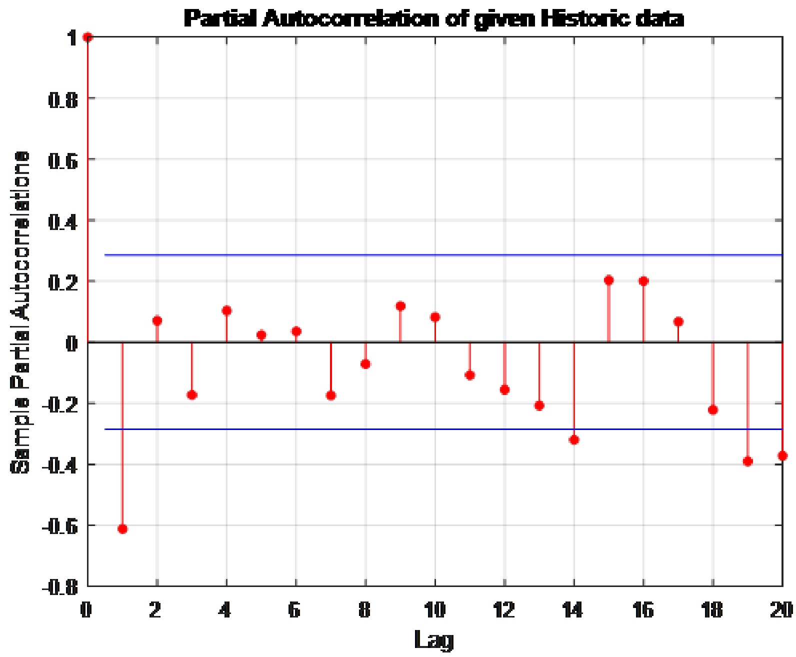


In the above plot change points are detected wherever the run length drops to zero. The nuclear response values are shifted with mean and run length values are scaled accordingly for visual comparison.

- **Extension of BOCD by integrating with RLS**
  Identifying AR model order and its parameters estimates:
  The autocorrelation and partial autocorrelation is taken for historic sample of a mystery time series signal to find the order of the AR model.



The autocorrelation function of given data is decaying exponentially with alternative positive and negative values. From this we can say the parameters will be having negative values but we cant say order of the model.



From the PACF plot it can be seen that only the lag 1 has the value greater than the 95 percentage confidence interval. Hence the order of the AR model is 1 and the initial parameter to estimate is 1.

- **Initial Parameter Estimate using LS :**

```
theta_LS =


Linear regression model:
    y ~ 1 + x1

Estimated Coefficients:
                    Estimate        SE        tStat        pValue
                    _________    ________    ________    __________

    (Intercept)     -0.018793    0.043134    -0.43569        0.6651
    x1               -0.61278     0.11738     -5.2203    4.1825e-06


Number of observations: 48, Error degrees of freedom: 46
Root Mean Squared Error: 0.299
R-squared: 0.372,  Adjusted R-Squared 0.358
F-statistic vs. constant model: 27.3, p-value = 4.18e-06
```

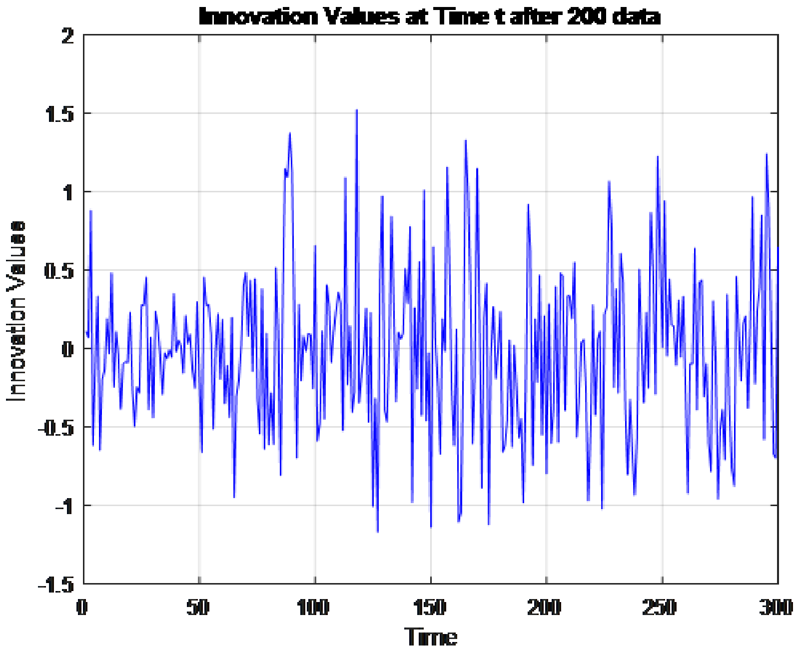
The intercept term won't be considered since it fails to pass the significant test. p value is greater than 0.5.

```
theta_LS =


Linear regression model:
    y ~ x1

Estimated Coefficients:
          Estimate      SE        tStat       pValue
          ________    _______    _______    __________

    x1    -0.61035    0.11624    -5.2509    3.5842e-06


Number of observations: 48, Error degrees of freedom: 47
Root Mean Squared Error: 0.296
```

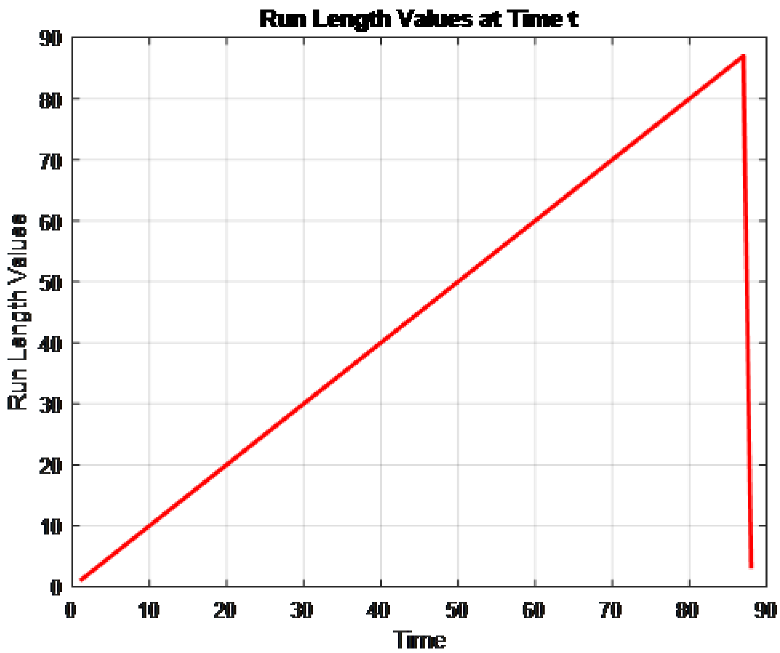
The Initial Parameter Estimate for RLS will be -0.61035 and the number of parameters to be estimated will be equal to 1.

Recursive LS is modeled with 1 parameter and initial estimate and its run for next 200 new data streaming online. The parameter estimate and the predicted output for each data is given RLS model and the difference between predicted data and actual data is called innovations. The innovations e follows Gaussian distribution and it is undergoing change in its distribution (as its variance undergoes abrupt change at a change point) after 200 data.

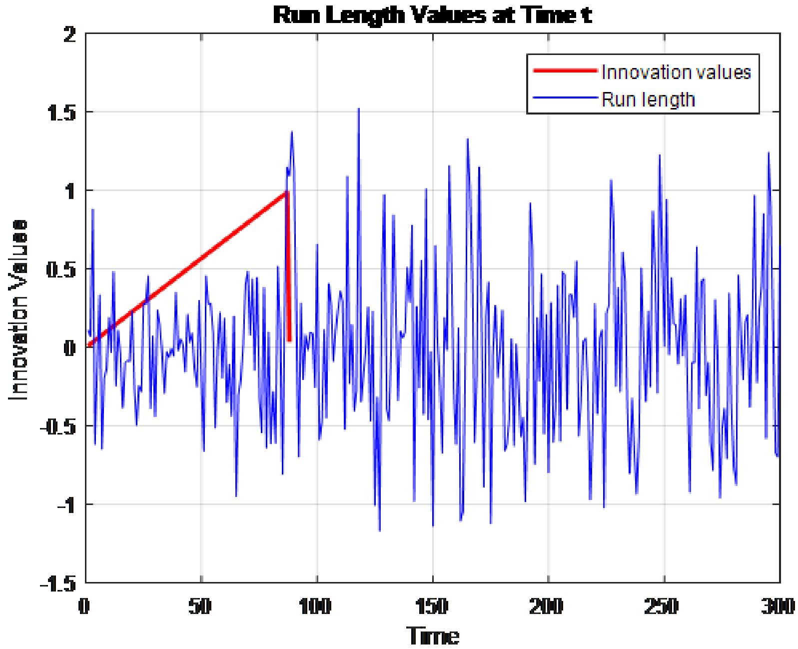
After 200th data streamed online, in order to find the change point in which the variance of innovations undergoes an abrupt change, the Bayesian Online Change Point detection is applied for innovations to find the change in its distributions. Since Innovations follow Gaussian distribution the same Gaussian Normal prior is used with slight changes in initial parameters and the T distribution can be used to find the UPM predictive probability.



After 200 data, the change point in the innovations obtained at the time 87 and the BOCD algorithm is stopped.



- **Innovation Values with Run Length Comparison:**



That's the change point is detected at the time t = 88 after 200 innovations in which at this time the variance is changing abruptly.
