

Bayesian Online Changepoint Detection

COVID Data and National Policy Analysis

Yuhan Fu

Denison University

fu_y1@denison.edu

January 21, 2023



Table of Contents

1 Introduction

- Time Series Analysis
- Changepoint Detection
- Bayesian Online Changepoint Detection

2 Methods

- Partitions and Run Length
- Recursive run length posterior estimation
- Sequential inference algorithm
- Changepoint detection methods

3 Real-World Applications

- Sierra Leone
- Egypt

4 Conclusion

Table of Contents

1 Introduction

- Time Series Analysis
- Changepoint Detection
- Bayesian Online Changepoint Detection

2 Methods

- Partitions and Run Length
- Recursive run length posterior estimation
- Sequential inference algorithm
- Changepoint detection methods

3 Real-World Applications

- Sierra Leone
- Egypt

4 Conclusion

Introduction

Time Series Analysis

Crude Oil Imports from the Crude Oil Imports Database

USA 01/1981 - 06/2022

Data Planet™: A SAGE Publishing Resource, Source: Energy Information Administration

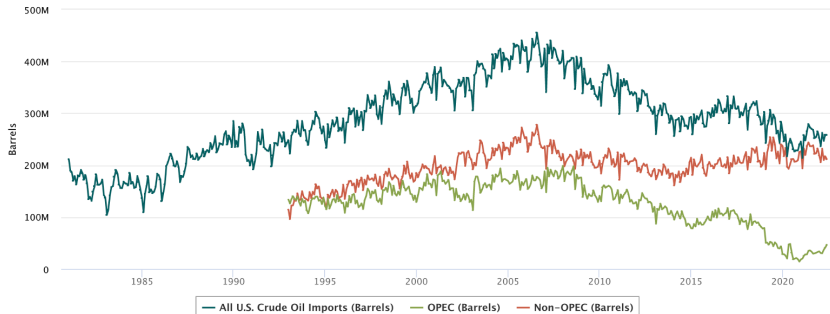


Figure 1: Energy Information Administration (2022-09-01). Crude Oil Imports: Crude Oil Imports | All U.S. Crude Oil Imports, 01/1981 - 06/2022. Data Planet™ Statistical Datasets: A SAGE Publishing Resource.

Changepoint Detection

- According to Adams and Mackay, "changepoints are abrupt variations in the generative parameters of a data sequence."

Introduction

Time Series Analysis

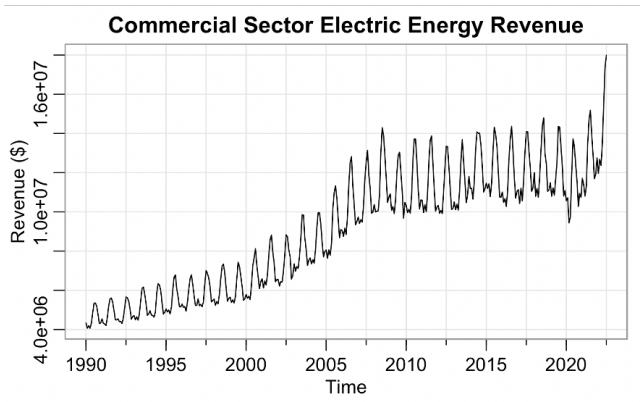


Figure 2: Monthly data of commercial sector electric energy revenue from January 1990 to August 2022. Data source: Energy Information Administration.

Introduction

Time Series Analysis

- Stationarity: No trend or periodic fluctuations; mean and variance don't vary in time.

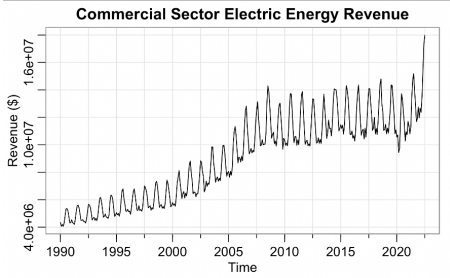


Figure 3: The time series of commercial sector electric energy revenue.

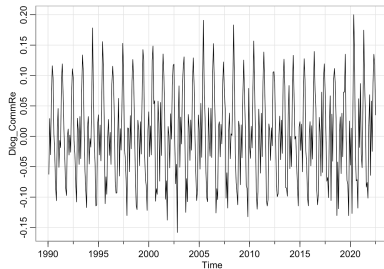


Figure 4: The first-order differencing of the log of the time series.

Bayesian Online Changepoint Detection

The online detection algorithm:

- Generates a new distribution for each upcoming datum in a sequence based on already observed data
- Differs from offline and retrospective changepoint detection (offline signal processing):
 - ▶ EEG analysis [1]
 - ▶ DNA segmentation [2]
 - ▶ NMR (nuclear magnetic resonance) analysis [3]

Table of Contents

1 Introduction

- Time Series Analysis
- Changepoint Detection
- Bayesian Online Changepoint Detection

2 Methods

- Partitions and Run Length
- Recursive run length posterior estimation
- Sequential inference algorithm
- Changepoint detection methods

3 Real-World Applications

- Sierra Leone
- Egypt

4 Conclusion

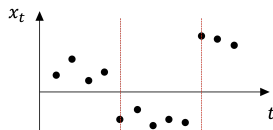
Partitions and Run Length

- Let $\mathbf{x}_{s:t}$ denote a sequence of observations $x_s, x_{s+1}, \dots, x_{t-1}, x_t$ for $s \leq t$. Assume the T data points $\mathbf{x}_{1:T}$ can be divided into non-overlapping product partitions such that the data within each partition are i.i.d. from some probability distribution [4].
- Changepoints occur between partitions.
- Run length: The time since the last changepoint occurs, denoted r_t at time t .

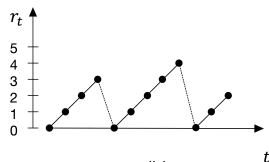
$$r_t = \begin{cases} 0 & \text{if changepoint at time } t \\ r_{t-1} + 1 & \text{else.} \end{cases}$$

Methods

Partitions and Run Length



(a)



(b)

Figure 5: Conceptual diagrams of (a) data partitioned by two changepoints, and (b) the run length with respect to time.

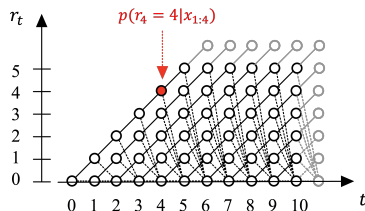


Figure 6: The trellis of message-passing algorithm. Each node has associated mass. For example, the red node has probability $p(x_4 = 4 \mid x_{1:4})$.

Recursive run length posterior estimation

- The changepoint prior
- The posterior predictive
- Message-passing parameter

Recursive run length posterior estimation (Methods)

The changepoint prior

The run length has only two outcomes with nonzero mass: either $r_t = r_{t-1} + 1$ or $r_t = 0$.

The changepoint prior has the following cases:

$$p(r_t \mid r_{t-1}) = \begin{cases} H(r_{t-1} + 1) & \text{if } r_t = 0 \\ 1 - H(r_{t-1} + 1) & \text{if } r_t = r_{t-1} + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Recursive run length posterior estimation (Methods)

The changepoint prior (cont.)

The function $H(\tau)$ is the *hazard function* (Forbes. et al., 2011).

$$H(\tau) = \frac{f(\tau)}{S(\tau)}. \quad (2)$$

Let T be a discrete, nonnegative random variable for the current run length, then $f(\tau)$ is the probability that $T = \tau$ (the current run length is τ). And $S(\tau)$ is the *survival function* at τ , which is the probability that $T \geq \tau$ or

$$S(\tau) = P(T \geq \tau) = \sum_{\tau'=\tau}^{\infty} f(\tau'). \quad (3)$$

Recursive run length posterior estimation (Methods)

The changepoint prior (cont.)

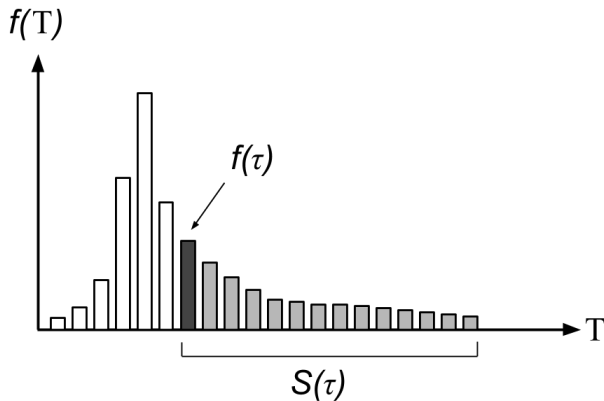


Figure 7: Figure from Gregory

Gundersen(<http://gregorygundersen.com/blog/2019/08/13/bocd/>). The probability $H(\tau)$ for a prior distribution $f(\tau)$.

Recursive run length posterior estimation (Methods)

The posterior predictive

- Conjugate-exponential models
- The general form of an exponential family likelihoods:

$$p(\mathbf{x} \mid \boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta})\exp\{\boldsymbol{\eta}^\top u(\mathbf{x})\} \quad (4)$$

- ▶ $\boldsymbol{\eta}$: natural parameter;
 $h(\mathbf{x})$: underlying measure;
 $u(\mathbf{x})$: sufficient statistic of the data;
 $g(\boldsymbol{\eta})$: normalizer.

Recursive run length posterior estimation (Methods)

The posterior predictive

- Use ν and χ as hyperparameters
- The conjugate prior:

$$p(\eta \mid \chi, \nu) = f(\chi, \nu) g(\eta)^\nu \exp\{\eta^\top \chi\} \quad (5)$$

- Express the posterior by multiplying the likelihood and the prior:

$$\begin{aligned} p(\eta \mid X, \chi, \nu) &= \left(\left(\prod_{i=1}^N h(\mathbf{x}_n) \right) g(\eta)^n \exp\{\eta^\top \sum_{n=1}^N u(\mathbf{x}_n)\} \right) \\ &\quad \left(f(\chi, \nu) g(\eta)^\nu \exp\{\eta^\top \chi\} \right) \\ &= \left(\prod_{i=1}^N h(\mathbf{x}_n) \right) f(\chi, \nu) g(\eta)^{N+\nu} \exp\{\eta^\top \sum_{n=1}^N u(\mathbf{x}_n) + \eta^\top \chi\}. \end{aligned} \quad (6)$$

Recursive run length posterior estimation (Methods)

The posterior predictive

- Since the first $N + 1$ terms are constant with respect to η ,

$$p(\eta \mid X, \chi, \nu) \propto g(\eta)^{N+\nu} \exp\{\eta^\top (\sum_{n=1}^N u(\mathbf{x}_n) + \chi)\}. \quad (7)$$

- Express the updated hyperparameters ν' and χ' :

$$\begin{aligned} \nu' &= \nu_{\text{prior}} + N \\ \chi' &= \chi_{\text{prior}} + \sum_{n=1}^N u(\mathbf{x}_n). \end{aligned} \quad (8)$$

- Since the sufficient statistics are additive, they allow us to compute the posterior predictive efficiently and sequentially.

Recursive run length posterior estimation (Methods)

Message-passing parameter

$$\begin{aligned}\nu_t^{(0)} &= \nu_{prior} \\ \chi_t^{(0)} &= \chi_{prior} \\ \nu_t^{(l)} &= \nu_{t-1}^{(l-1)} + 1 \\ \chi_t^{(l)} &= \chi_{t-1}^{(l-1)} + u(x_t).\end{aligned}\tag{9}$$

- $\nu_{t-1}^{(l)}$ and $\chi_{t-1}^{(l)}$ denote the exponential family parameters at time t with $r_{t-1} = l$.

Sequential inference algorithm

- 1 Set the initial conditions and priors

$$p(r_0) = \begin{cases} 1 & \text{if changepoint at time } t = 0 \\ p(r_0 = \tau) & \text{else} \end{cases}$$

$$\mu_0^{(l)} = \mu_{\text{prior}}^{(l)}$$

$$\lambda_0^{(l)} = \lambda_{\text{prior}}^{(l)}$$

(here we use the precision $\lambda = \sigma^{-2}$).

- 2 Observe new datum x_t .
- 3 Compute predictive probability

$$\pi_{t-1}^{(l)} = p(x \mid \mathbf{x}_{1:n}).$$

- 4 Compute growth probabilities

$$p(r_t = l, \mathbf{x}_{1:t}) = p(r_{t-1}, \mathbf{x}_{1:t-1}) \pi_{t-1}^{(l)} (1 - H(r_{t-1})).$$

- 5 Compute changepoint probabilities

$$p(r_t = 0, \mathbf{x}_{1:t}) = \sum_{r_{t-1}} p(r_{t-1}, \mathbf{x}_{1:t-1}) \pi_{t-1}^{(l)} H(r_t - 1).$$

- 6 Compute the evidence

$$p(\mathbf{x}_{1:t}) = \sum_{r_t} p(r_t, \mathbf{x}_{1:t}).$$

7 Compute the run length posterior

$$p(r_t \mid \mathbf{x}_{1:t}) = \frac{p(r_t, \mathbf{x}_{1:t})}{p(\mathbf{x}_{1:t})}.$$

8 Update sufficient statistics

$$\mu_0^{(l)} = \mu_{\text{prior}}^{(l)}$$

$$\lambda_0^{(l)} = \lambda_{\text{prior}}^{(l)}$$

$$\mu_t^{(l)} = \frac{1}{\lambda_t^{(l)}} \left(\lambda_0^{(l)} \mu_0^{(l)} + \lambda_t^{(l)} \sum_{i=1}^t x_i \right)$$

$$\lambda_t^{(l)} = \lambda_0^{(l)} + t \lambda^{(l)}$$

- 9 Perform prediction

$$p(x_{t+1} \mid \mathbf{x}_{1:t}) = \sum_{r_t} p(x_{t+1} \mid r_t, \mathbf{x}^{(l)}) p(r_t \mid \mathbf{x}_{1:t}).$$

- 10 Set $t = t + 1$. Return to Step 2.

Changepoint detection methods

- plot the probability that a changepoint occurs at each time point t
- find all local maxima that meet a certain threshold probability i.e. the lowest probability we accept to suggest a changepoint

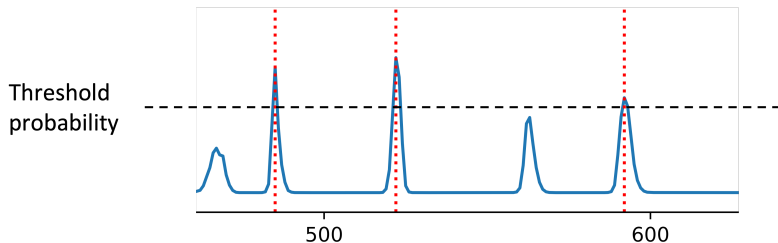


Figure 8: Blue line: The probability that a changepoint occurs at t . Black dashed line: Threshold probability. Red dashed line: Indication of changepoints.

Table of Contents

1 Introduction

- Time Series Analysis
- Changepoint Detection
- Bayesian Online Changepoint Detection

2 Methods

- Partitions and Run Length
- Recursive run length posterior estimation
- Sequential inference algorithm
- Changepoint detection methods

3 Real-World Applications

- Sierra Leone
- Egypt

4 Conclusion

Real-World Applications

Sierra Leone

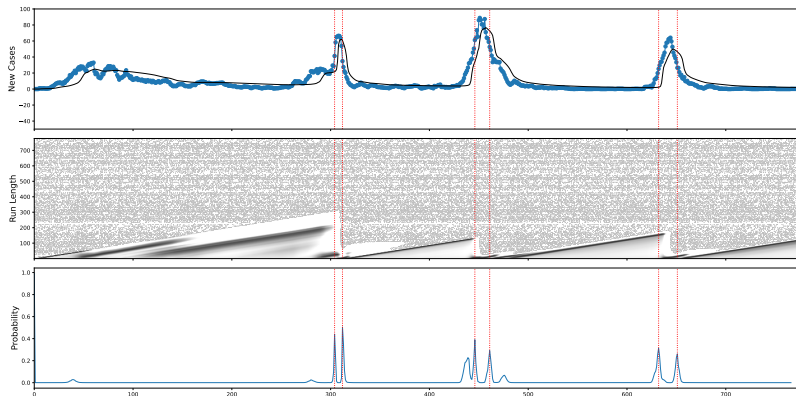


Figure 9: (Top) The smoothed daily new cases of COVID-19 (blue line) in Sierra Leone and the predictive mean (black line). (Middle) The run length posterior at each time step using a logarithmic color scale. Darker pixels indicate higher probability. (Bottom) The probability that a changepoint occurs at each time step. Red dashed lines denote changepoints.

Real-World Applications

Egypt

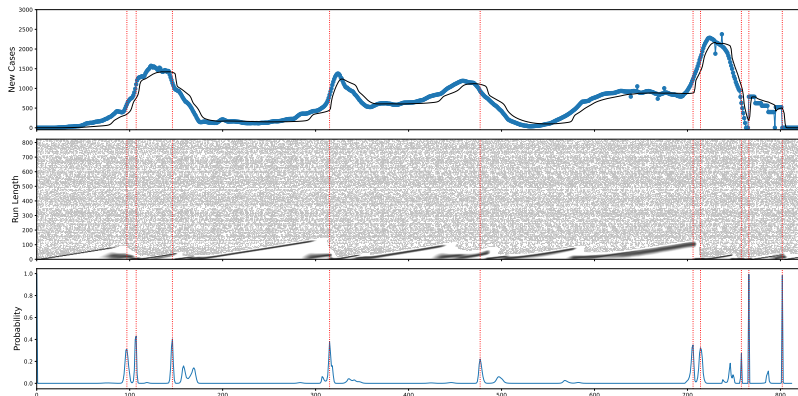


Figure 10: (Top) The smoothed daily new cases of COVID-19 (blue line) in Egypt and the predictive mean (black line). (Middle) The run length posterior at each time step using a logarithmic color scale. Darker pixels indicate higher probability. (Bottom) The probability that a changepoint occurs at each time step. Red dashed lines denote changepoints.

Table of Contents

1 Introduction

- Time Series Analysis
- Changepoint Detection
- Bayesian Online Changepoint Detection

2 Methods

- Partitions and Run Length
- Recursive run length posterior estimation
- Sequential inference algorithm
- Changepoint detection methods

3 Real-World Applications

- Sierra Leone
- Egypt

4 Conclusion

Conclusion

- This Bayesian Online Changepoint Detection method is based on a Gaussian univariate model. Depending on the type and our prior knowledge of the data, we can also parameterize various distributions differently, such as using an inverse gamma or an inverse chi-squared [5].
- Higher probability often results from a larger number of changepoints. We sometimes sacrifice the probability to achieve a decent amount of changepoints. The threshold probability also varies for different data.

Acknowledgement

This research project is supported by Denison University Research Foundation. Thanks Dr. Zhe Wang for giving me useful instructions during the research and paper writing process. I also appreciate the work done by Mark Raney and Riley Coburn, who contribute to other parts of the entire project, including the data mining, the cluster analysis and the dynamic linear modeling.

References

- ❶ JS Barlow, OD Creutzfeldt, D Michael, J Houchin, and H Epelbaum. Automatic adaptive segmentation of clinical eegs. *Electroencephalography and Clinical Neurophysiology*, 51(5):512–525, 1981.
- ❷ Jerome V Braun, RK Braun, and Hans-Georg Müller. Multiple changepoint fitting via quasilielihood, with application to dna sequence segmentation. *Biometrika*, 87(2):301–314, 2000.
- ❸ Paul Fearnhead. Exact and efficient bayesian inference for multiple changepoint problems. *Statistics and computing*, 16(2):203–213, 2006.
- ❹ Daniel Barry and John A Hartigan. Product partition models for change point problems. *The Annals of Statistics*, pages 260–279, 1992.
- ❺ Kevin P Murphy. Conjugate bayesian analysis of the gaussian distribution. *def*, 1($2\sigma^2$):16, 2007.