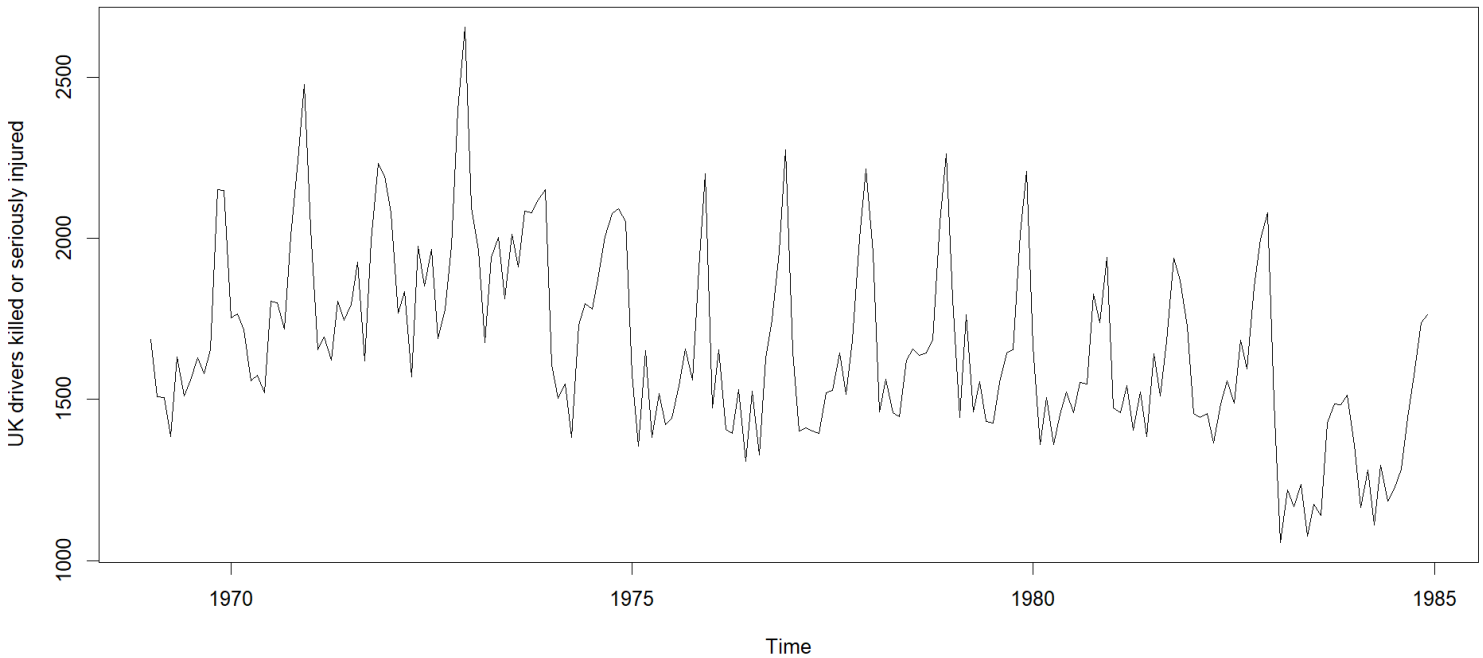
Improving A Bayesian Procedure to Detect Breakpoints in Time Series Data

Dainia Higgins, Nathaniel Wilson, Madison Ell, Sarah Klingbeil
Mentor: Jeff Liebner

Coppin State University, Fullerton College, Southern Adventist University

Friday, June 17, 2022

Introduction



What is a Time Series?

Definition

Time series are observations collected over regular intervals of time.

What are some common types of time series models?

- **Auto-regressive** (AR) model: each output value depends linearly upon previous values and an independent error term

AR(1)

$$x_t = \phi x_{t-1} + \epsilon_t$$

- **Moving average** (MA) model: output value depends linearly upon previous error terms

MA(1)

$$x_t = \epsilon_t + \theta \epsilon_{t-1}$$

What are some common types of time series models?

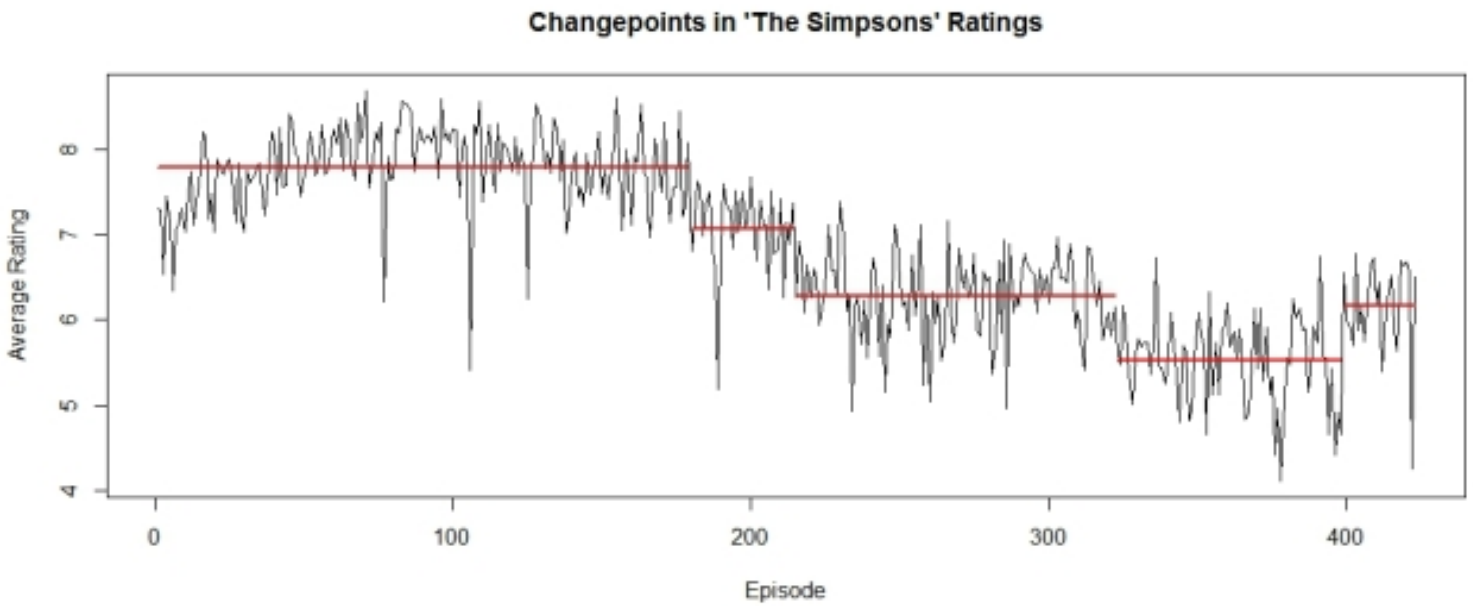
- **Auto-regressive moving average** (ARMA) model: output value depends on a mixed polynomial of previous values and error terms
- **Auto-regressive integrated moving average** (ARIMA) model: output value depends on a mixed polynomial of previous values and error terms, as well as the order of integration (the number of times the original time series is differenced to achieve stationarity)

What is a Breakpoint?

Definition

Breakpoints (also known as Change Points) are points in time where the model changes.

- Breakpoints are significant changes in time series data



How do we find Breakpoints?

1. **Expert Opinion** : breakpoints are approximated by experts in specific field based on historical knowledge.
2. **Reverse Order Cusum** (ROC) : reverses data set and uses historical knowledge to group data points together with boundaries being locations of breakpoints

What is Bai-Perron Test?

Definition

Bai-Perron Test is a general algorithm to find an optimal breakpoint set.

1. a frequentist approach
2. checks almost every single location for a breakpoint and returns the optimal set
3. requires a user to specify the number of breakpoints

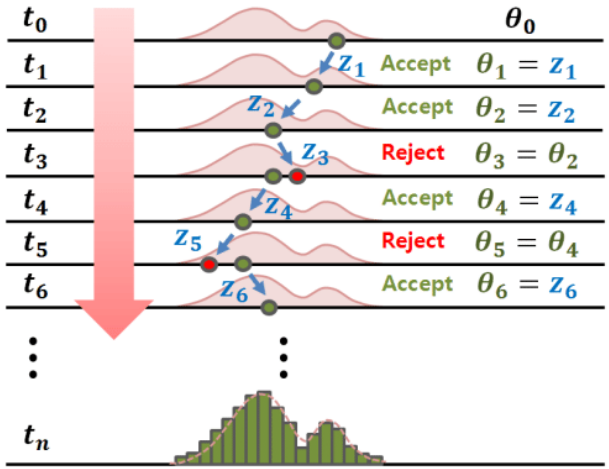
What is the difference between Bayesian and Frequentist approach in Statistics?

- In **Frequentist** Statistics, the parameter is fixed and the data is random. This is the standard type of statistics learned in schools.
- In **Bayesian** Statistics, the parameter is random and the data is fixed. With this method, you apply prior information to the data that is being explored. The posterior is the result.

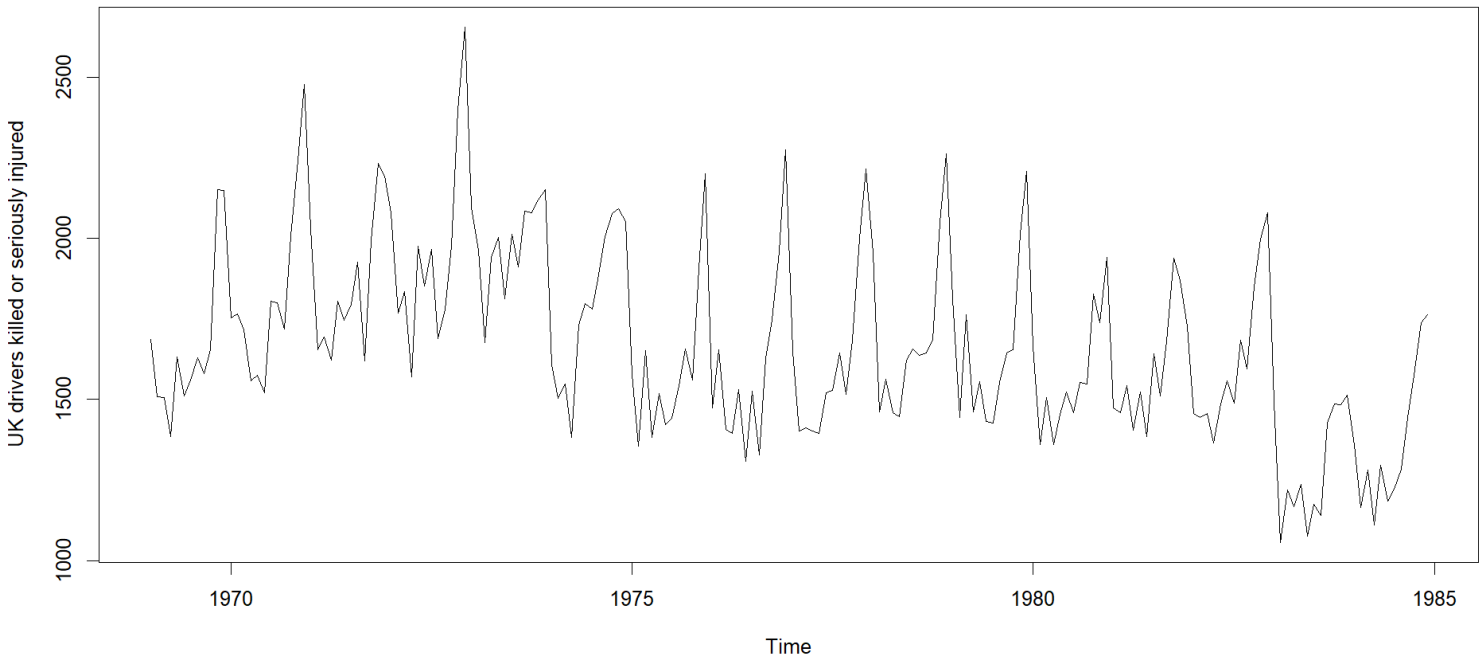
$$f(\theta|x) = \frac{f(\theta|x)\pi(\theta)}{\int f(\theta|x)\pi(\theta)d\theta}$$

What is Markov Chain Monte Carlo (MCMC)?

1. A class of algorithms for sampling from a probability distribution
2. Generates and records a random sample sequence from one sample to another where each proposed sample is accepted or rejected by the algorithm; this process is repeated until a stationary distribution sample is found



Remember...


UK drivers killed or seriously injured
1000
1500
2000
2500
1970
1975
1980
1985
Time

Example

Story time

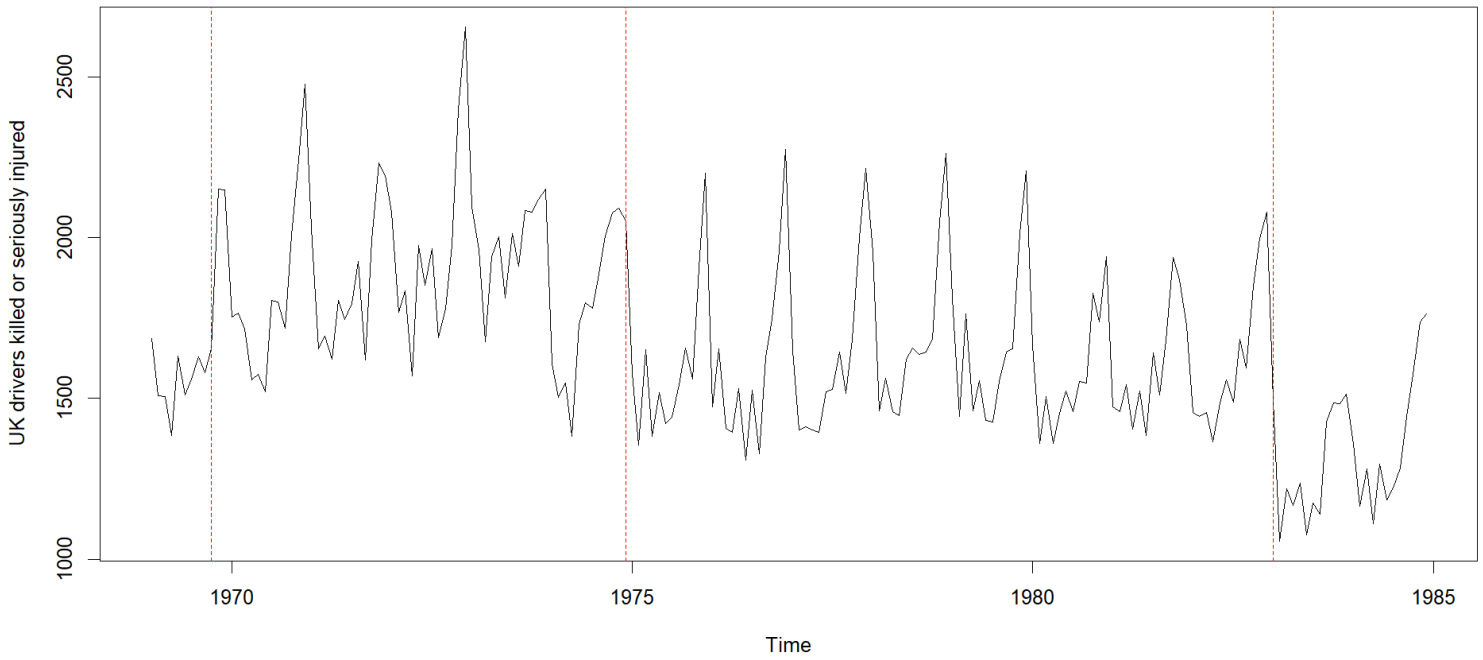


Figure: Breakpoints,represented by the red lines, can be found in Oct. 1969, Dec. 1974, and Jan. 1983.

The Project

2018 REU Goal : Kathryn Haglich, Sarah Neitzel, and Amy Pitts wanted to develop a better quantitative method for locating breakpoints in time series data.

The Project

Bayesian Adaptive Auto-Regession (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

1. needs to have an input stating breakpoint places (Bai-Perron)
2. A new breakpoint set is proposed at each step of the MCMC
3. **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
4. Hence, a distribution of possible breakpoints locations can be obtained

The Project

Bayesian Adaptive Auto-Regession (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

1. needs to have an input stating breakpoint places (Bai-Perron)
2. A new breakpoint set is proposed at each step of the MCMC
3. **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
4. Hence, a distribution of possible breakpoints locations can be obtained

The Project

Bayesian Adaptive Auto-Regession (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

1. needs to have an input stating breakpoint places (Bai-Perron)
2. A new breakpoint set is proposed at each step of the MCMC
3. **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
4. Hence, a distribution of possible breakpoints locations can be obtained

The Project

Bayesian Adaptive Auto-Regession (BAAR)

BAAR is a Bayesian method used to find the location and number of breakpoints in a time series.

1. needs to have an input stating breakpoint places (Bai-Perron)
2. A new breakpoint set is proposed at each step of the MCMC
3. **Metropolis-Hastings** (a MCMC method for receiving a sequence of random samples from the probability distribution when direct sampling is difficult) ratio determines the set's acceptance
4. Hence, a distribution of possible breakpoints locations can be obtained

General Plan

Goal : to improve and expand BAAR (Bayesian Adaptive Auto-Regression) technique developed by previous REU group.

General Plan

Our four goals

1. Find a better starting point
2. Stress Models beyond even-spaced breaks.
3. Evaluate and alter code as needed to enhance the algorithm
4. Explore other models beyond evenly-spaced AR models

General Plan

Our four goals

1. Find a better starting point
2. Stress Models beyond even-spaced breaks.
3. Evaluate and alter code as needed to enhance the algorithm
4. Explore other models beyond evenly-spaced AR models

General Plan

Our four goals

1. Find a better starting point
2. Stress Models beyond even-spaced breaks.
3. Evaluate and alter code as needed to enhance the algorithm
4. Explore other models beyond evenly-spaced AR models

General Plan

Our four goals

1. Find a better starting point
2. Stress Models beyond even-spaced breaks.
3. Evaluate and alter code as needed to enhance the algorithm
4. Explore other models beyond evenly-spaced AR models

General Plan

Our four goals

1. Find a better starting point
2. Stress Models beyond even-spaced breaks.
3. Evaluate and alter code as needed to enhance the algorithm
4. Explore other models beyond evenly-spaced AR models

Moving Forward

1. Familiarize ourselves with the original code and develop more randomized breakpoints.
2. Understand and optimise math equations behind previous REU research
3. Research log likelihood equations for other models, explore math for MA models and start to develop algorithms
4. Explore alternative starting points