An Analysis of a Circular Changepoint Model
– A Covid-19 Case Study –

Matthew Scholes with Owen Li

University of Lancaster - STOR-i

September 17, 2022
1 Introducing Changepoint Models
   - Linear Models
   - Circular Model

2 Circular Model Performance

3 Covid-19 Case Study
   - Data Suitability
   - Segment Neighbourhood vs Circular
   - Potential Error
Introducing Changepoint Models

1. Introducing Changepoint Models
   - Linear Models
   - Circular Model

2. Circular Model Performance

3. Covid-19 Case Study
   - Data Suitability
   - Segment Neighbourhood vs Circular
   - Potential Error
Introducing Changepoint Models

The aim of changepoints regression is to reduce the cost function, a common example is using the negative log likelihood (for normalised data).

\[ L(M_\kappa) = -2\log\max L(\theta_\kappa) + p_\kappa \phi(n) \] (2).

**Figure:** Graphs show examples of how parameters change given at changepoint locations: LHS - mean; RHS - trend (1).
Linear Models

- Binary Segmentation (BinSeg):
  - Computationally fast $O(n)$ (3).
  - Only able to find local minimum.
Introducing Changepoint Models

Linear Models

- Binary Segmentation (BinSeg):
  - Computationally fast $O(n)$ (3).
  - Only able to find local minimum.

- Segment Neighbourhood (SegN):
  - For max $m$ changepoints at time $\tau$.
  - Minimize $\sum_{i=0}^{m} C(y(\tau_{i+1} : \tau_{i+1})$.
  - Computationally slow $O(n^2)$ (4).
  - Exact, will find global minimum.
Introducing Changepoint Models

Linear Models

- **Binary Segmentation (BinSeg):**
  - Computationally fast $O(n) \ (3)$.
  - Only able to find local minimum.

- **Segment Neighbourhood (SegN):**
  - For max $m$ changepoints at time $\tau$.
  - Minimize $\sum_{i=0}^{m} C(y_{(\tau_i+1):\tau_{i+1}})$.
  - Computationally slow $O(n^2) \ (4)$.
  - Exact, will find global minimum.

- **Pruned Exact Linear Time (PELT):**
  - Removes values of $\tau$ that can never be considered a changepoint.
  - Computationally fast $O(n \log(n)) \ (5)$.
  - Exact, will find global minimum.
Introducing Changepoint Models

Circular Model

**Linear Data**

![Linear Data](image)

**Circular Data**

![Circular Data](image)

**Figure**: Normally distributed data; LHS initial linear form, RHS data wrapped around a period of 20.

- Instead of representing the data linearly, wrap the time axis on itself so every time point has multiple data entries.
- Then minimize \( \sum_{i=0}^{m} C(x_{(\tau_i+1):\tau_{i+1}}) \) where \( x \) is a vector of points.
Contents

1 Introducing Changepoint Models
   • Linear Models
   • Circular Model

2 Circular Model Performance

3 Covid-19 Case Study
   • Data Suitability
   • Segment Neighbourhood vs Circular
   • Potential Error
Circular Model Performance

- generated normalised, periodic, data with a constant variance of $\sigma = 1$ and a varying mean ($\mu$).

- Success is defined only if both changepoints are located at the correct place.

- Circular method can detect changepoints that are more 'subtle'.

- Finds both changepoints with a success of 99% the up to and including $\Delta \mu = 2$.

**Figure:** A comparison of linear and circular methods for detecting a change in mean for a time series.
Contents

1 Introducing Changepoint Models
   • Linear Models
   • Circular Model

2 Circular Model Performance

3 Covid-19 Case Study
   • Data Suitability
   • Segment Neighbourhood vs Circular
   • Potential Error
Background

- Data of positive Covid-19 cases recorded in England from 30/01/20.
- Due to multiple social factors, less tests are processed at weekends.
- This causes a dip in positive cases over the weekend which leads to periodicity (6).
- Aim: Identify this periodicity and see if circular method can identify changepoints hidden to traditional Segment Neighbourhood search.

**Figure:** Number of positive cases in England from 30/01/20.
Checking Periodicity

- Circular Method requires period of data as an input.
- Trends are determined using PELT.
- Conducting a FFT on the cleaned data a peak frequency of $0.143$ was found.
- This corresponds to a period of $6.98$ days.
Segment Neighbourhood vs Circular

Figure: LHS, Results of SegN and RHS Results of Circular Method.

- Segment Neighbourhood picks up similar changepoints to PELT, failing to register the periodicity.
- 43% of cpts are found on Friday and Sunday, but why is this inconsistent?
Areas of Error

- The times where the method does not locate Friday or Sunday as a changepoint are boxed in red.
- The majority of these are areas of high variance, which correspond to times of numerous positive cases in raw data.
- Could the increased variance account for erroneous changepoint location?
- What about the first box? This is low variance.

**Figure:** Boxes areas are locations where neither Friday or Saturday are located as changepoints.
Areas of Error

- Requirement of the method is that the data is normalised.
- The QQ plot shows that the data is too heavy at the tails to be completely normalised.
  - Areas of low variance are more normalised than the boxed areas.
  - This does not mean that the high variance areas cannot be analysed, but a different cost function would be needed.
- Could political reasoning explain the error from low variance?

Figure: LHS: QQ plot of cleaned data, RHS headlines from BBC regarding 100k testing target.
Conclusion

- Circular method was developed to locate changepoint locations, normally hidden from linear methods.
- It achieves this as it can detect smaller changes.
- When used with a Covid-19 case study, weekly points of change were located as expected.
- Revealed further information about the time series, prompting areas for further research.
- A very powerful tool, when used in isolation and in conjunction with other linear changepoint models.
Thank you for listening
Are there any questions?
References


