

Extension of Bayesian MDL for Changepoint Detection

Yingbo Li*

1 Current Bayesian MDL Formula

1.1 Linear model

Given a changepoint model $\boldsymbol{\eta}$, the sampling distribution has the regression representation

$$\mathbf{X}_{1:N} = \mathbf{A}_{1:n}\mathbf{s} + \mathbf{D}_{1:N}\boldsymbol{\mu} + \boldsymbol{\epsilon}_{1:n}, \quad (1)$$

with $\mathbf{A}_{1:n} \in \mathbb{R}^{n \times T}$ and $\mathbf{D}_{1:n} \in \mathbb{R}^{n \times m}$ as seasonal and regime indicator matrices, respectively:

$$\begin{aligned} [\mathbf{A}_{1:n}]_{t,v} &= \mathbf{1}(\text{time } t \text{ is in season } v), \quad v = 1, \dots, T, \\ [\mathbf{D}_{1:n}]_{t,r-1} &= \mathbf{1}(\text{time } t \text{ is in regime } r), \quad r = 2, \dots, m+1, \end{aligned}$$

where $\mathbf{1}(A)$ denotes the indicator of the event A .

2 Permitting Changes in Linear Trends

2.1 Harmonic regression

Order k

*e-mail: carolli13@gmail.com.

2.2 Linear trends

The maximum number of changepoints

$$n - p > 2k + 1 + 2m + 1 \rightarrow m \leq \left\lceil \frac{n - p - 2k - 2}{2} \right\rceil - 1$$

3 Handling Outliers

Total number of outliers

$$l = \sum_{t=1}^n \xi_t$$