Bayesian Minimal Description Lengths for Multiple Changepoint Detection

Yingbo Li

Dept of Mathematical Sciences, Clemson University

Co-authors: Robert Lund (Clemson University),
Anuradha Hewaarachchi (University of Kelaniya)
Monthly maximum temperature series in Tuscaloosa, AL

Yingbo Li (Clemson)
Monthly maximum temperature series in Tuscaloosa, AL

Observed data — sample seasonal mean

Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
Instrumentation changes: 1956 Nov, 1987 May

Yingbo Li (Clemson)
Bayesian MDL Changepoint Detection
QPRC 2017
Monthly maximum temperature series in Tuscaloosa, AL

Observed data – sample seasonal mean

Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
Instrumentation changes: 1956 Nov, 1987 May
Monthly maximum temperature series in Tuscaloosa, AL

Observed data — sample seasonal mean

Metadata (station history logs)
- Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
- Instrumentation changes: 1956 Nov, 1987 May
A Motivating Example

Monthly maximum temperature series in Tuscaloosa, AL

Observed data – sample seasonal mean

Metadata (station history logs): more likely to induce mean shifts
- Station relocations: 1921 Nov, 1939 Mar, 1956 Jun, 1987 May
- Instrumentation changes: 1956 Nov, 1987 May
Monthly maximum and minimum temperature series

Observed data — sample seasonal mean
### Monthly maximum and minimum temperature series

**Observed data — sample seasonal mean**

![Graph showing monthly maximum and minimum temperature series with changepoints marked in 1918 Feb, 1957 Jul, and 1990 Jan.]

- **Tmax (°F)**
  - 1900 to 2010
  - Values range from -10 to 10

- **Tmin (°F)**
  - 1900 to 2010
  - Values range from -10 to 15

- **Changepoints:**
  - 1918 Feb
  - 1957 Jul
  - 1990 Jan

---

Yingbo Li (Clemson)  
Bayesian MDL Changepoint Detection  
QPRC 2017
A Motivating Example

Monthly maximum and minimum temperature series

Observed data — sample seasonal mean

Tmax and Tmin are likely to shift at the same time.
MDL for Changepoint Detection in Piecewise AR Series

- Automatic MDL by Davis et al. [2006]: a penalized likelihood, with penalty being the code length (CL) of parameters

\[
\log(m) + (m + 1) \log(N) + \sum_{r=1}^{m+1} \log p_r + \sum_{r=1}^{m+1} \frac{p_r+2}{2} \log N_r
\]

- Automatic MDL rules:
  - CL of an unbounded positive integer \( I \): \( \log(I) \)
  - CL of a positive integer bounded above by \( U \): \( \log(U) \)
  - CL of the MLE of a real-valued parameter estimated by \( N \) observations: \( \frac{1}{2} \log N \)
Model Selection Using MDL Principle

- Description length [Rissanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
Model Selection Using MDL Principle

- Description length [Rissanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

\[
\mathcal{L}(\mathbf{X}, \theta) = \mathcal{L}(\mathbf{X} \mid \theta) + \mathcal{L}(\theta)
\]

transmit \( \mathbf{X} \) transmit \( \theta \)
Model Selection Using MDL Principle

- Description length [Rissanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

\[
L(X, \theta) = - \log f(X | \theta) - \log \pi(\theta)
\]

transmit \( X \)  transmit \( \theta \)
Model Selection Using MDL Principle

- Description length [Rissanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

\[
\mathcal{L}(X, \theta) = -\log f(X | \theta) - \log \pi(\theta)
\]

transmit \(X\) transmit \(\theta\)

- Mixture MDL

\[
\mathcal{L}(X) = -\log \int f(X | \theta)\pi(\theta)d\theta
\]

marginal likelihood
Model Selection Using MDL Principle

- Description length [Rissanen, 1989, Hansen and Yu, 2001]: the number of storage units to transmit a random dataset.
- In model selection, the true model has the smallest MDL.
- Two-part MDL

\[ \mathcal{L}(\mathbf{X}, \theta) = -\log f(\mathbf{X} | \theta) - \log \pi(\theta) \]

transmit $\mathbf{X}$ transmit $\theta$

- Mixture MDL

\[ \mathcal{L}(\mathbf{X}) = -\log \int f(\mathbf{X} | \theta)\pi(\theta)d\theta \]

marginal likelihood

- If the prior $\pi(\theta | \tau)$, combine with two-part MDL:

\[ \mathcal{L}(\mathbf{X}, \hat{\tau}) = -\log \int f(\mathbf{X} | \theta)\pi(\theta | \hat{\tau})d\theta - \log \pi(\hat{\tau}) \]

Closely related with empirical Bayes marginal likelihood.
Bayesian MDL

- Observed time series: $X_{1:N} = (X_1, X_2, \ldots, X_N)'$
- Suppose $m$ changepoints $1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N$ partition the timeline $m + 1$ distinct regimes (segments).

  Time $t$ is in regime $r \iff \tau_{r-1} \leq t < \tau_r$
Bayesian MDL

- Observed time series: $X_{1:N} = (X_1, X_2, \ldots, X_N)'$
- Suppose $m$ changepoints $1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N$ partition the timeline $m + 1$ distinct regimes (segments).
  
  Time $t$ is in regime $r \iff \tau_{r-1} \leq t < \tau_r$
- Any time in $t = \{p + 1, p + 2, \ldots, N\}$ can be a changepoint

**Definition**

Denote a *multiple changepoint configuration* as an indicator vector $\eta = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)'$, such that

$$\eta_t = \begin{cases} 1, & \text{if time } t \text{ is a changepoint} \\ 0, & \text{if time } t \text{ is not a changepoint} \end{cases}$$
Bayesian MDL

- Observed time series: \( X_{1:N} = (X_1, X_2, \ldots, X_N)' \)
- Suppose \( m \) changepoints \( 1 \leq \tau_1 < \tau_2 < \cdots < \tau_m \leq N \) partition the timeline \( m + 1 \) distinct regimes (segments).
  
  Time \( t \) is in regime \( r \) \( \iff \tau_{r-1} \leq t < \tau_r \)

- Any time in \( t = \{p + 1, p + 2, \ldots, N\} \) can be a changepoint

**Definition**

Denote a *multiple changepoint configuration* as an indicator vector

\[
\eta = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)', \quad \text{such that}
\]

\[
\eta_t = \begin{cases} 
1, & \text{if time } t \text{ is a changepoint} \\
0, & \text{if time } t \text{ is not a changepoint}
\end{cases}
\]

- Number of changepoints in \( \eta \): \( m = \sum_{t=p+1}^{N} \eta_t \)
- Total number of models: \( 2^{N-p} \)
Prior distribution $\pi(\eta)$: Beta-Binomial

- If time $t$ is not in metadata
  $$\eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(1)} \right), \quad \rho^{(1)} \sim \text{Beta} \left( a, b^{(1)} \right)$$

- If time $t$ is in metadata
  $$\eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(2)} \right), \quad \rho^{(2)} \sim \text{Beta} \left( a, b^{(2)} \right)$$
Prior distribution $\pi(\eta)$: Beta-Binomial

- If time $t$ is not in metadata
  \[
  \eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(1)} \right), \quad \rho^{(1)} \sim \text{Beta} \left( a, b^{(1)} \right)
  \]

- If time $t$ is in metadata
  \[
  \eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(2)} \right), \quad \rho^{(2)} \sim \text{Beta} \left( a, b^{(2)} \right)
  \]

- $\pi(\eta)$ has a closed form:
  \[
  \pi(\eta) = \prod_{k=1}^{2} \int_{0}^{1} \left[ \prod_{t^{(k)}} \pi \left( \eta_{t^{(k)}} \mid \rho^{(k)} \right) \right] \pi \left( \rho^{(k)} \right) d\rho^{(k)}
  \]
Prior distribution $\pi(\eta)$: Beta-Binomial

- If time $t$ is not in metadata
  \[ \eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(1)} \right), \quad \rho^{(1)} \sim \text{Beta} \left( a, b^{(1)} \right) \]

- If time $t$ is in metadata
  \[ \eta_t \overset{iid}{\sim} \text{Bernoulli} \left( \rho^{(2)} \right), \quad \rho^{(2)} \sim \text{Beta} \left( a, b^{(2)} \right) \]

- $\pi(\eta)$ has a closed form:

  \[
  \pi(\eta) \propto \prod_{k=1}^{2} \Gamma \left( a + m^{(k)} \right) \Gamma \left( b^{(k)} + N^{(k)} - m^{(k)} \right)
  \]

Number of:
- undocumented times $N^{(1)}$, undocumented changepoints $m^{(1)}$
- documented times $N^{(2)}$, documented changepoints $m^{(2)}$
Parameter elicitation

- Metadata times are more likely to be changepoints

\[
E \left( \rho^{(1)} \right) = \frac{a}{a + b^{(1)}} < \frac{a}{a + b^{(2)}} = E \left( \rho^{(2)} \right)
\]
Parameter elicitation

- Metadata times are more likely to be changepoints

\[ E(\rho^{(1)}) = \frac{a}{a + b^{(1)}} < \frac{a}{a + b^{(2)}} = E(\rho^{(2)}) \]

- US temperature: 6 changepoints per century [Mitchell, 1953],

0.005 changepoint / month

Default parameters:

\[ a = 1, \quad b^{(1)} = 239 \implies E(\rho^{(1)}) = 0.004 \]
\[ b^{(2)} = 47 \implies E(\rho^{(2)}) = 0.021 \]
**Likelihood: multivariate normal**

Under a specific model $\eta$,

$$X_t = s_{\nu(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$ 

- $\nu(t)$ time $t$ is in season $\nu(t)$
- $r(t)$ time $t$ is in regime $r(t)$
- $s_{\nu}$ seasonal mean, $\nu = 1, 2, \ldots, 12.$
- $\mu_{r}$ regime mean, $r = 1, 2, \ldots, m + 1$
- $\{\epsilon_t\}$ Gaussian AR($p$) errors, with white noise variance $\sigma^2$ and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
Likelihood: multivariate normal

Under a specific model \( \eta \),

\[
X_t = s_{v(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.
\]

- \( v(t) \) time \( t \) is in season \( v(t) \)
- \( r(t) \) time \( t \) is in regime \( r(t) \)
- \( s_v \) seasonal mean, \( v = 1, 2, \ldots, 12. \)
- \( \mu_r \) regime mean, \( r = 1, 2, \ldots, m + 1 \)
- \( \{\epsilon_t\} \) Gaussian AR(\( p \)) errors, with white noise variance \( \sigma^2 \) and autoregression coefficients \( \phi_1, \ldots, \phi_p \)

- For identifiability, \( \mu_1 = 0 \)
- Likelihood function: \( f (X_{(p+1):N} \mid \mu, s, \sigma^2, \phi, \eta, X_{1:p}) \)
Likelihood: multivariate normal

Under a specific model $\eta$,

$$X_t = s_{\nu(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$  

$\nu(t)$ time $t$ is in season $\nu(t)$
$r(t)$ time $t$ is in regime $r(t)$
$s_{\nu}$ seasonal mean, $\nu = 1, 2, \ldots, 12.$
$\mu_{r}$ regime mean, $r = 1, 2, \ldots, m + 1$
$\{\epsilon_t\}$ Gaussian AR($p$) errors, with white noise variance $\sigma^2$ and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
- Likelihood function: $f\left( X_{(p+1):N} \mid \mu, s, \sigma^2, \phi, \eta, X_{1:p} \right)$
Likelihood: multivariate normal

Under a specific model $\eta$,

$$X_t = s_{\nu(t)} + \mu_{r(t)} + \epsilon_t, \quad t = 1, 2, \ldots, N.$$  

- $\nu(t)$ time $t$ is in season $\nu(t)$
- $r(t)$ time $t$ is in regime $r(t)$
- $s_{\nu}$ seasonal mean, $\nu = 1, 2, \ldots, 12$
- $\mu_{r}$ regime mean, $r = 1, 2, \ldots, m + 1$
- $\{\epsilon_t\}$ Gaussian AR($p$) errors, with white noise variance $\sigma^2$ and autoregression coefficients $\phi_1, \ldots, \phi_p$

- For identifiability, $\mu_1 = 0$
- Likelihood function: $f(\mathbf{X}_{(p+1):N} | \mu, s, \sigma^2, \phi, \eta, \mathbf{X}_{1:p})$
- Using mixture MDL on $\mu$, and two-part MDL on the rest.
BMDL: has a closed form

\[
\text{BMDL}(\eta) = \mathcal{L}(X | \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \mathcal{L}(\hat{s}, \hat{\sigma}^2, \hat{\phi} | \eta) + \mathcal{L}(\eta)
\]

- mixture MDL
- two-part MDL
- two-part MDL
BMDL: has a closed form

\[
\text{mixture MDL} \quad \text{two-part MDL} \quad \text{two-part MDL}
\]
\[
\text{BMDL}(\eta) = \mathcal{L}(X \mid \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \mathcal{L}(\hat{s}, \hat{\sigma}^2, \hat{\phi} \mid \eta) + \mathcal{L}(\eta)
\]

- Prior distribution on \( \mu = (\mu_2, \mu_3, \ldots, \mu_{m+1})' \)

\[
\mu \mid \sigma^2, \eta \sim \mathcal{N}(0, \nu \sigma^2 I_m).
\]

- \( \nu \) is pre-specified; default \( \nu = 5 \).
- Normal-normal conjugacy: marginal likelihood has a closed form

\[
f(X \mid s, \sigma^2, \phi, \eta) = \int f(X \mid \mu, s, \sigma^2, \phi, \eta) \pi(\mu \mid \sigma^2, \eta) d\mu
\]
BMDL: has a closed form

\[ \text{BMDL}(\eta) = \mathcal{L}(\mathbf{X} \mid \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \mathcal{L}(\hat{s}, \hat{\sigma}^2, \hat{\phi} \mid \eta) + \mathcal{L}(\eta) \]

- Prior distribution on \( \mu = (\mu_2, \mu_3, \ldots, \mu_{m+1})' \)

  \[ \mu \mid \sigma^2, \eta \sim \mathcal{N}(0, \nu \sigma^2 \mathbf{I}_m). \]

  - \( \nu \) is pre-specified; default \( \nu = 5 \).
  - Normal-normal conjugacy: marginal likelihood has a closed form

  \[ f(\mathbf{X} \mid s, \sigma^2, \phi, \eta) = \int f(\mathbf{X} \mid \mu, s, \sigma^2, \phi, \eta) \pi(\mu \mid \sigma^2, \eta) d\mu \]

- \( \hat{s}, \hat{\sigma}^2 = \arg \max f(\mathbf{X} \mid s, \sigma^2, \phi, \eta) \) have closed forms.

- \( \hat{\phi} : \) Yule-Walker estimator.
BMDL: has a closed form

\[
\text{BMDL}(\eta) = \mathcal{L}(\mathbf{X} | \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \mathcal{L}(\hat{s}, \hat{\sigma}^2, \hat{\phi} | \eta) + \mathcal{L}(\eta) - \log f(\mathbf{X} | \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \frac{p+13}{2} \log(N - p) - \log \pi(\eta)
\]

- Prior distribution on \( \mu = (\mu_2, \mu_3, \ldots, \mu_{m+1})' \)
  \[
  \mu | \sigma^2, \eta \sim \mathcal{N}(0, \nu \sigma^2 I_m).
  \]
  - \( \nu \) is pre-specified; default \( \nu = 5 \).
  - Normal-normal conjugacy: marginal likelihood has a closed form

\[
f(\mathbf{X} | s, \sigma^2, \phi, \eta) = \int f(\mathbf{X} | \mu, s, \sigma^2, \phi, \eta) \pi(\mu | \sigma^2, \eta) d\mu
\]

- \( \hat{s}, \hat{\sigma}^2 = \text{arg max } f(\mathbf{X} | s, \sigma^2, \phi, \eta) \) have closed forms.
- \( \hat{\phi} \): Yule-Walker estimator.
## BMDL: has a closed form

<table>
<thead>
<tr>
<th>Mixture MDL</th>
<th>Two-part MDL</th>
<th>Two-part MDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{BMDL}(\eta) = \mathcal{L}(\mathbf{X}</td>
<td>\hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) + \mathcal{L}(\hat{s}, \hat{\sigma}^2, \hat{\phi}</td>
<td>\eta) + \mathcal{L}(\eta)$</td>
</tr>
</tbody>
</table>

- Prior distribution on $\mu = (\mu_2, \mu_3, \ldots, \mu_{m+1})'$

$$\mu | \sigma^2, \eta \sim \mathcal{N}(0, \nu \sigma^2 \mathbf{I}_m).$$

- $\nu$ is pre-specified; default $\nu = 5.$
- Normal-normal conjugacy: marginal likelihood has a closed form

$$f(\mathbf{X} | s, \sigma^2, \phi, \eta) = \int f(\mathbf{X} | \mu, s, \sigma^2, \phi, \eta) \pi(\mu | \sigma^2, \eta) d\mu$$

- $\hat{s}, \hat{\sigma}^2 = \text{arg max } f(\mathbf{X} | s, \sigma^2, \phi, \eta)$ have closed forms.
- $\hat{\phi}$: Yule-Walker estimator.
Bayesian MDL formula

\[
BMDL(\eta) = \frac{N - p}{2} \log (\hat{\sigma}^2) + \frac{m}{2} \log(\nu) + \frac{1}{2} \log \left( \left| \hat{D}'\hat{D} + \frac{I_m}{\nu} \right| \right) \\
- \sum_{k=1}^{2} \log \left[ \Gamma \left( a + m^{(k)} \right) \Gamma \left( b^{(k)} + N^{(k)} - m^{(k)} \right) \right].
\]
Computations: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\eta$
  \[
p_{EB}(\eta \mid X) \propto \int f\left(X \mid \mu, \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta\right) \pi(\mu \mid \hat{\sigma}^2, \eta) d\mu \cdot \pi(\eta)
  \]

- BMDL is closely related to EB.
  \[
  BMDL(\eta) = -\log \int f\left(X \mid \mu, \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta\right) \pi(\mu \mid \hat{\sigma}^2, \eta) d\mu - \log \pi(\eta)
  \]
Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\eta$
  \[
p_{EB}(\eta | X) \propto \int f(X | \mu, \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) \pi(\mu | \hat{\sigma}^2, \eta) d\mu \cdot \pi(\eta)
  \]

- BMDL is closely related to EB.
  \[
  \text{BMDL}(\eta) = - \log p_{EB}(\eta | X)
  \]
Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\eta$

$$p_{EB}(\eta \mid X) \propto \int f \left( X \mid \mu, \hat{s}, \hat{\sigma}^2, \phi, \eta \right) \pi(\mu \mid \hat{s}^2, \eta) d\mu \cdot \pi(\eta)$$

- BMDL is closely related to EB.

$$BMDL(\eta) = - \log p_{EB}(\eta \mid X)$$

- Borrow stochastic model search algorithms from Bayesian model selection literature: Metropolis-Hastings (MCMC).

$$\eta^{[0]} \rightarrow \eta^{[1]} \rightarrow \eta^{[2]} \rightarrow \ldots \rightarrow \eta^{[t]} \rightarrow \ldots$$
Computation: stochastic model search using MCMC

- Empirical Bayes posterior probability of model $\eta$
  
  \[ p_{\text{EB}}(\eta \mid X) \propto \int f(X \mid \mu, \hat{s}, \hat{\sigma}^2, \hat{\phi}, \eta) \pi(\mu \mid \hat{\sigma}^2, \eta) d\mu \cdot \pi(\eta) \]

- BMDL is closely related to EB.
  
  \[ \text{BMDL}(\eta) = - \log p_{\text{EB}}(\eta \mid X) \]

- Borrow stochastic model search algorithms from Bayesian model selection literature: Metropolis-Hastings (MCMC).
  
  \[ \eta[0] \rightarrow \eta[1] \rightarrow \eta[2] \rightarrow \ldots \rightarrow \eta[t] \rightarrow \ldots \]

- R package BayesMDL.
Scenario 1: monotonic shifts $\mu = (0, \Delta, 2\Delta, 3\Delta)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

A sample simulated series
Scenario 1: monotonic shifts $\mu = (0, \Delta, 2\Delta, 3\Delta)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$
Scenario 1: monotonic shifts $\mu = (0, \Delta, 2\Delta, 3\Delta)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

A sample simulated series (minus seasonal mean)

Detection percentage: without metadata

36.1, 41.4, 37.7
Scenario 1: monotonic shifts $\mu = (0, \Delta, 2\Delta, 3\Delta)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$
Scenario 2: non-monotonic shifts $\mu = (0, -\Delta, \Delta, 0)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

A sample simulated series
Scenario 2: non-monotonic shifts $\mu = (0, -\Delta, \Delta, 0)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$
Scenario 2: non-monotonic shifts $\mu = (0, -\Delta, \Delta, 0)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

A sample simulated series (minus seasonal mean)

Detection percentage: without metadata

36.7
84.3
39.2
Scenario 2: non-monotonic shifts $\mu = (0, -\Delta, \Delta, 0)'$

$p = 3, \phi = (0.2, 0.1, 0.05)', \Delta/\sigma = 1.5.$

A sample simulated series (minus seasonal mean)

Detection percentage: with metadata

77.3 84.6 38.2
Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$

Scale time to $[0, 1]$ by mapping time $t$ to $t/N$.

\[\begin{align*}
\lambda : & \quad 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\
\eta : & \quad 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \\
\end{align*}\]

\[\eta_r = \lfloor \lambda_r N \rfloor \]
Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)' \)
  
  Scale time to \([0, 1]\) by mapping time \(t\) to \(t/N\).

\[
\begin{align*}
\lambda & : \quad 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\
\eta & : \quad 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \end{align*}
\]

\[\eta_r = \lfloor \lambda_r N \rfloor\]

- True model \(\lambda^0\) has \(m^0\) changepoints.
  The true number of changepoints \(m^0\) is unknown.
Infill asymptotics

- Literature: Davis and Yau [2013], Du et al. [2016]
- Relative changepoint configuration: \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)' \)
  Scale time to \([0, 1]\) by mapping time \(t\) to \(t/N\).

\[
\begin{align*}
\lambda : & \quad 0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_m < 1 \\
\eta : & \quad 0 < \eta_1 < \cdots < \eta_r < \cdots < \eta_m < N \\
\end{align*}
\]
\[\iff \eta_r = \lfloor \lambda_r N \rfloor\]

- True model \(\lambda^0\) has \(m^0\) changepoints.
  The true number of changepoints \(m^0\) is unknown.
- Consider all relative changepoint configurations in

\[
\Lambda = \{ \lambda : 0 \leq m \leq M, \min_{r=1,2,\ldots,m+1} \lambda_r - \lambda_{r-1} \geq d \}
\]

- \(M\): a large integer, fixed, \(M > m^0\)
- \(d\): a very small positive constant
Asymptotic selection consistency

- The estimated relative changepoint model:

\[
\hat{\lambda}_N = \arg \min_{\lambda \in \Lambda} \text{BMDL}(\lambda),
\]

- The estimated number of changepoints is \( \hat{m}_N = |\hat{\lambda}_N| \).

**Theorem (Consistency for changepoint model estimation)**

As \( N \to \infty \), we have

\[
\hat{m}_N \xrightarrow{P} m^0, \quad \text{and} \quad \hat{\lambda}_N \xrightarrow{P} \lambda^0.
\]

Furthermore, for each \( r = 1, \ldots, m^0 \),

\[
|\hat{\lambda}_r - \lambda^0_r| = O_P \left( \frac{1}{N} \right).
\]
Parameter estimation

Under the estimated changepoint model \( \hat{\lambda} \),

- Yule-Walker estimator for \( \phi \).
- Optimizers of BMDL for \( \sigma^2 \) and \( s \).
- Conditional posterior mean for \( \mu \).

**Theorem (Consistency for parameter estimation)**

As \( N \to \infty \), all parameter estimators converge to the true values,

\[
\begin{align*}
\hat{\mu}_N & \xrightarrow{P} \mu^0, \\
\hat{s}_N & \xrightarrow{P} s^0, \\
\hat{\sigma}_N^2 & \xrightarrow{P} (\sigma^2)^0, \\
\hat{\phi}_N & \xrightarrow{P} \phi^0.
\end{align*}
\]
Outline

1. A Motivating Example

2. A Brief Review of MDL

3. Bayesian MDL
   - Method
   - Simulation examples
   - Asymptotic consistency

4. Generalization of BMDL: bivariate case
   - Method
   - Simulation examples

5. Tuscaloosa data
Joint detection of $T_{\text{max}}$ and $T_{\text{min}}$

- Any time in $t = \{p + 1, p + 2, \ldots, N\}$ can be a changepoint, for either $T_{\text{max}}$ or $T_{\text{min}}$, or both.
- If both, can shift in the same or opposite directions.
Joint detection of $T_{\text{Max}}$ and $T_{\text{Min}}$

- Any time in $t = \{p + 1, p + 2, \ldots, N\}$ can be a changepoint, for either $T_{\text{Max}}$ or $T_{\text{Min}}$, or both.
- If both, can shift in the same or opposite directions.

**Definition (bivariate changepoint model $\eta$)**

$$\eta = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)' \in \mathbb{R}^{(N-p) \times 2},$$

$$\eta_t = \begin{cases} 
(1, 1)', & \text{if time } t \text{ is a concurrent changepoint} \\
(1, 0)', & \text{if time } t \text{ is only a changepoint for } T_{\text{Max}} \\
(0, 1)', & \text{if time } t \text{ is only a changepoint for } T_{\text{Min}} \\
(0, 0)', & \text{if time } t \text{ is not a changepoint} 
\end{cases}$$

- Number of changepoints in $\eta$: $m_1$ ($T_{\text{Max}}$), $m_2$ ($T_{\text{Min}}$)
Joint detection of $T_{\text{Max}}$ and $T_{\text{Min}}$

- Any time in $t = \{p + 1, p + 2, \ldots, N\}$ can be a changepoint, for either $T_{\text{Max}}$ or $T_{\text{Min}}$, or both.
- If both, can shift in the same or opposite directions.

**Definition (bivariate changepoint model $\eta$)**

$$
\eta = (\eta_{p+1}, \eta_{p+2}, \ldots, \eta_N)' \in \mathbb{R}^{(N-p) \times 2},
$$

$$
\eta_t = \begin{cases} 
(1, 1)', & \text{if time } t \text{ is a concurrent changepoint} \\
(1, 0)', & \text{if time } t \text{ is only a changepoint for } T_{\text{Max}} \\
(0, 1)', & \text{if time } t \text{ is only a changepoint for } T_{\text{Min}} \\
(0, 0)', & \text{if time } t \text{ is not a changepoint}
\end{cases}
$$

- Number of changepoints in $\eta$: $m_1$ ($T_{\text{Max}}$), $m_2$ ($T_{\text{Min}}$)
- Dirichlet-Multinomial prior: closed form $\pi(\eta)$.
- Hyper parameter choices: encourage concurrent shifts.
Bivariate model

Under a specific model $\eta$,

$$(X_{t,1}, X_{t,2}) = (s_{\nu,1}, s_{\nu,2}) + (\mu_{r_1,1}, \mu_{r_2,2}) + (\epsilon_{t,1}, \epsilon_{t,2}), \quad t = 1, 2, \ldots, N.$$  

- $\nu$ time $t$ is in season $\nu$
- $r_i$ time $t$ is in regime $r_i$, where $i = 1(T_{\text{max}}), 2(T_{\text{min}})$
- $s_{\nu,i}$ seasonal mean; $s_1 \in \mathbb{R}^{12}, s_2 \in \mathbb{R}^{12}$
- $\mu_{r_i,i}$ regime mean; $\mu_1 \in \mathbb{R}^{m_1}, \mu_2 \in \mathbb{R}^{m_2}$
- $\{\epsilon_t\}$ Gaussian VAR($p$) errors, with white noise covariance $\Sigma \in \mathbb{R}^{2 \times 2}$ and autoregression coefficients $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^{2 \times 2}$.

- Likelihood: normal
Bivariate model

Under a specific model $\eta$,

\[
\begin{pmatrix}
X_{t,1} \\
X_{t,2}
\end{pmatrix} = \begin{pmatrix}
sv,1 \\
sv,2
\end{pmatrix} + \begin{pmatrix}
\mu_{r1,1} \\
\mu_{r2,2}
\end{pmatrix} + \begin{pmatrix}
\epsilon_{t,1} \\
\epsilon_{t,2}
\end{pmatrix}, \quad t = 1, 2, \ldots, N.
\]

- $\nu$ time $t$ is in season $\nu$
- $r_i$ time $t$ is in regime $r_i$, where $i = 1(T_{\text{max}}), 2(T_{\text{min}})$
- $s_{\nu,i}$ seasonal mean; $s_1 \in \mathbb{R}^{12}, s_2 \in \mathbb{R}^{12}$
- $\mu_{r_i,i}$ regime mean; $\mu_1 \in \mathbb{R}^{m_1}, \mu_2 \in \mathbb{R}^{m_2}$
- $\{\epsilon_t\}$ Gaussian VAR($p$) errors, with white noise covariance $\Sigma \in \mathbb{R}^{2 \times 2}$ and autoregression coefficients $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^{2 \times 2}$.

- Likelihood: normal
- Prior $\pi(\mu_1, \mu_2)$: normal
- $\hat{s}_1, \hat{s}_2$: closed forms
- $\hat{\Sigma}, \hat{\Phi}_1, \ldots, \hat{\Phi}_p$: Yule-Walker estimators
Bivariate model

Under a specific model $\eta$,

$$
\begin{pmatrix}
X_{t,1} \\
X_{t,2}
\end{pmatrix} =
\begin{pmatrix}
s_{v,1} \\
s_{v,2}
\end{pmatrix} +
\begin{pmatrix}
\mu_{r_1,1} \\
\mu_{r_2,2}
\end{pmatrix} +
\begin{pmatrix}
\epsilon_{t,1} \\
\epsilon_{t,2}
\end{pmatrix}, \quad t = 1, 2, \ldots, N.
$$

$v$ time $t$ is in season $v$

$r_i$ time $t$ is in regime $r_i$, where $i = 1(T_{\text{max}}), 2(T_{\text{min}})$

$s_{v,i}$ seasonal mean; $s_1 \in \mathbb{R}^{12}, s_2 \in \mathbb{R}^{12}$

$\mu_{r_i,i}$ regime mean; $\mu_1 \in \mathbb{R}^{m_1}, \mu_2 \in \mathbb{R}^{m_2}$

$\{\epsilon_t\}$ Gaussian VAR($p$) errors, with white noise covariance $\Sigma \in \mathbb{R}^{2 \times 2}$ and autoregression coefficients $\Phi_1, \ldots, \Phi_p \in \mathbb{R}^{2 \times 2}$.

- Likelihood: normal
- Prior $\pi(\mu_1, \mu_2)$: normal \quad \text{BMDL has a closed form.}
- $\hat{s}_1, \hat{s}_2$: closed forms
- $\hat{\Sigma}, \hat{\Phi}_1, \ldots, \hat{\Phi}_p$: Yule-Walker estimators
Combining simulated data Scenario 1 and 2

Two concurrent shifts: 150 (↑↓), 300 (↑↑).
- Tmax (Series 1): shifts at 150, 300, 450. \( \mu_1 = (0, \Delta, 2\Delta, 3\Delta)' \)
- Tmin (Series 2): shifts at 150, 300, 375. \( \mu_2 = (0, -\Delta, \Delta, 0)' \)
Combining simulated data Scenario 1 and 2

Two concurrent shifts: 150 (↑↓), 300 (↑↑).
- Tmax (Series 1): shifts at 150, 300, 450. $\mu_1 = (0, \Delta, 2\Delta, 3\Delta)'$
- Tmin (Series 2): shifts at 150, 300, 375. $\mu_2 = (0, -\Delta, \Delta, 0)'$

VAR parameters
- $\Sigma = \begin{pmatrix} 9 & 2 \\ 2 & 9 \end{pmatrix}$, $\Delta/3 = 1.5$
- $p = 3$, $\Phi_1 = \begin{pmatrix} 0.2 & 0.02 \\ 0.02 & 0.2 \end{pmatrix}$, $\Phi_2 = \begin{pmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{pmatrix}$, $\Phi_3 = \begin{pmatrix} 0.05 & 0.005 \\ 0.005 & 0.05 \end{pmatrix}$
Combining simulated data Scenario 1 and 2

A sample simulated series, Tmax

A sample simulated series, Tmin
Combining simulated data Scenario 1 and 2

A sample simulated series (minus seasonal mean), $T_{\text{max}}$

A sample simulated series (minus seasonal mean), $T_{\text{min}}$
Combining simulated data Scenario 1 and 2

Detection percentage using univariate BMDL: Tmax

Detection percentage using univariate BMDL: Tmin

36.1 41.4 37.7
36.7 84.3 39.2
Combining simulated data Scenario 1 and 2

Detection percentage using bivariate BMDL: Tmax

Detection percentage using bivariate BMDL: Tmin
Outline

1. A Motivating Example

2. A Brief Review of MDL

3. Bayesian MDL
   - Method
   - Simulation examples
   - Asymptotic consistency

4. Generalization of BMDL: bivariate case
   - Method
   - Simulation examples

5. Tuscaloosa data
Univariate BMDL: no metadata

![Temperature graphs showing potential changepoints]

1957Mar 1990Jan
1918Feb 1957Jul 1990Jan
Univariate BMDL: with metadata

Yingbo Li (Clemson)  Bayesian MDL Changepoint Detection  QPRC 2017  25 / 26
Bivariate BMDL: with metadata
Reference I


