Parallelisation of a Common Changepoint Detection Method

Supplementary Materials

S. O. Tickle, I. A. Eckley, P. Fearnhead, K. Haynes

July 4, 2019

1 Yao's Results and Extension

The following two lemmas are due to Yao (1988).

Lemma 1.1. *Suppose* $Z_1, ..., Z_n \sim^{i.i.d.} N(0, \sigma^2)$. *Then for any* $\epsilon > 0$ *as* $n \to \infty$*:*

$$\mathbb{P}\left(\max_{0 \leq i < j \leq n} \frac{(Z_{i+1} + ... + Z_j)^2}{(j-i)} > 2(1+\epsilon)\sigma^2 \log n\right) \to 0. \tag{1}$$

Lemma 1.2. *Let* m_U *be an upper bound on the number of changes, and let* $(\hat{\tau}_1, ..., \hat{\tau}_{\hat{m}})$ *be the set of estimated changes generated (by Yao's procedure). For every* $\hat{m}$ *s.t.* $m < \hat{m} \leq m_U$ *and* $1 \leq r \leq m$*,*

$$\mathbb{P}((\hat{\tau}_1, ..., \hat{\tau}_{\hat{m}}) \in B_i^2(n)) \to 0$$

as $n \to \infty$*, where:*

$$B_i^{\delta}(n) = \{(\xi_1, ..., \xi_t) : 0 < \xi_1 < ... < \xi_t < n \text{ and } |\xi_s - \tau_r| \geq \left\lceil (\log n)^{\delta} \right\rceil \text{ for } 1 \leq s \leq \hat{m}\}.$$

Corollary 1.2.1. *Lemma 1.2 can be extended to $B_i^{1+\alpha}(n)$, for any $\alpha > 0$.*

Proof of Corollary 1.2.1: The argument for the location accuracy being $(\log n)^2$ in Yao (1988) comes from showing that the residual sum of squares for a segmentation that misses a change by more than this amount can be reduced by an amount that is greater than $3(2+\epsilon)\log n$ with probability tending to 1 as n increases, by adding three changes at the changepoint plus or minus $(\log n)^2$. Thus such a segmentation cannot be optimal as the penalised cost for the latter segmentation will be less than the original one. We therefore need only show that this argument holds if we replace an accuracy of $(\log n)^2$ with $(\log n)^{1+\alpha}$ for any $\alpha > 0$.

To do this it suffices to show that a segmentation $\hat{\tau}_1, ..., \hat{\tau}_{\hat{m}}$ which misses a particular change τ_i by at least $\left\lceil (\log n)^{1+\alpha} \right\rceil$ has a residual sum of squares between the points $\tau_i - \left\lceil (\log n)^{1+\alpha} \right\rceil$ and $\tau_i + \left\lceil (\log n)^{1+\alpha} \right\rceil$ which when normalised by the true fit has term of leading order $\left\lceil (\log n)^{1+\alpha} \right\rceil$.

For a segmentation $\hat{\tau}_{1:\hat{m}}$ define $\text{RSS}(y_{s:t}; \hat{\tau}_{1:\hat{m}})$ to be the residual sum of squares obtained if we fit the changepoints to the subset of data $y_{s:t}$. Note that this will only depend on the changepoints, if any, that lie between timepoints s and t. Then for any $\hat{\tau}_{1:\hat{m}} \in B_i^{1+\alpha}(n)$:

$$\text{RSS}(y_{1:n}; \hat{\tau}_{1:\hat{m}}) \geq \text{RSS}\left(y_{1:n}; \hat{\tau}_{1:\hat{m}}, \tau_1, \ldots, \tau_{i-1}, \tau_i - \left\lceil (\log n)^{1+\alpha} \right\rceil, \tau_i + \left\lceil (\log n)^{1+\alpha} \right\rceil, \tau_{i+1}, \ldots, \tau_m\right). \tag{2}$$

As Yao (1988) remarks, RHS of (2) can be decomposed as:

$$\text{RSS}(y_{1:\tau_1}; \mathcal{T}_1) + \ldots + \text{RSS}\left(y_{\tau_{i-1}+1:\tau_i - \left\lceil (\log n)^{1+\alpha} \right\rceil}; \mathcal{T}_i\right) + \text{RSS}\left(y_{\tau_i - \left\lceil (\log n)^{1+\alpha} \right\rceil + 1:\tau_i + \left\lceil (\log n)^{1+\alpha} \right\rceil}; \emptyset\right)$$
$$+\text{RSS}\left(y_{\tau_i + \left\lceil (\log n)^{1+\alpha} \right\rceil + 1:\tau_{i+1}}; \mathcal{T}_{i+1}\right) + \ldots + \text{RSS}\left(y_{\tau_m+1:n}; \mathcal{T}_{m+1}\right),$$

where $\mathcal{T}_h$ is the subset of $\hat{\tau}_{1:\hat{m}}$ which falls inside the corresponding segment of the univariate time series. By Lemma 1.1, each term in this decomposition involving $\mathcal{T}_h$ is such that:

$$\mathrm{RSS}(y_{a+1:b}; \mathcal{T}_h) = \sum_{j=a+1}^{b} Z_j^2 + O_p(\log n),$$

while, if without loss of generality we assume that the mean at the changepoint τ_i changes from 0 to μ, then letting $c_n^{(\alpha)} = \lceil (\log n)^{1+\alpha} \rceil$:

$$\begin{aligned}
\mathrm{RSS}\left(y_{\tau_i - c_n^{(\alpha)}+1:\tau_i+c_n^{(\alpha)}}; \emptyset\right) &= \sum_{j=\tau_i-c_n^{(\alpha)}+1}^{\tau_i+c_n^{(\alpha)}} \left(Y_i - \bar{Y}\left(\tau_i - c_n^{(\alpha)} + 1, \tau_i + c_n^{(\alpha)}\right)\right)^2 \\
&= \sum_{j=\tau_i-c_n^{(\alpha)}+1}^{\tau_i+c_n^{(\alpha)}} Z_j^2 + \frac{\mu^2}{2} c_n^{(\alpha)} - \frac{1}{2c_n^{(\alpha)}} \left(\sum_{j=\tau_i-c_n^{(\alpha)}+1}^{\tau_i+c_n^{(\alpha)}} Z_j \right)^2 + D,
\end{aligned}$$

where $D \sim N\left(0, 2\sigma^2 c_n^{(\alpha)} \mu^2\right)$. Therefore:

$$\begin{aligned}
\left\{ \sum_{j=\tau_i-c_n^{(\alpha)}+1}^{\tau_i+c_n^{(\alpha)}} Z_j^2 - \mathrm{RSS}\left(y_{\tau_i-c_n^{(\alpha)}+1:\tau_i+c_n^{(\alpha)}}; \emptyset\right) \right\} / c_n^{(\alpha)} &= \frac{\mu^2}{2} - \frac{1}{2\left(c_n^{(\alpha)}\right)^2} \left(\sum_{j=\tau_i-c_n^{(\alpha)}+1}^{\tau_i+c_n^{(\alpha)}} Z_j \right)^2 + D/c_n^{(\alpha)} \\
&\to \frac{\mu^2}{2} \text{ by Lemma 1.1.}
\end{aligned}$$

In particular, $\forall \hat{\tau}_{1:\hat{m}} \in B_i^{1+\alpha}(n)$:

$$\left\{ \mathrm{RSS}\left(x_{1:n}; \hat{\tau}_{1:\hat{m}}, \tau_{1:n}^{-i}, \tau_i - c_n^{(\alpha)}, \tau_i + c_n^{(\alpha)}\right) - \sum_{j=1}^{n} Z_j^2 \right\} / c_n^{(\alpha)} \to \frac{\mu^2}{2}.$$

Thus, as any segmentation from $B_i^{1+\alpha}(n)$ is strictly worse than a corresponding segmentation, which in turn is worse (in probability) than fitting the truth under a penalty of $\beta = 2(1+\epsilon)\log n$, uniformly in $B_i^{1+\alpha}(n)$, $\mathbb{P}(\hat{\tau}_{1:\hat{m}} \in B_i^{1+\alpha}(n)) \to 0$. $\square$

2 Unparallelised Consistency Results

Proof of Proposition 3.1: Let $\hat{m}$ be the number of changes estimated by the procedure. The aim is firstly to show that:

(a): $\mathbb{P}(\hat{m} > m) \to 0$,

(b): $\mathbb{P}(\hat{m} < m) \to 0$.

Proof of (a): Under Corollary 1.2.1, for $\hat{m} > m$, with probability 1 as $n \to \infty$ it must be the case that m of the estimated changes are within $(\log n)^{1+\alpha}$, some $\alpha > 0$, of the true changes. We will now show that with probability tending to 1 these segmentations cannot be optimal.

To do this we will compare the penalised cost of any such segmentation with the penalised cost of the true segmentation. The latter cost can be bounded above by $\sum_{t=1}^{n} Z_t^2 + m\,(2+\epsilon)\log n$. Our approach is to split the comparison of the residual sum of squares of a segmentation $\hat{\tau}_{1:\hat{m}}$ with $\sum_{t=1}^{n} Z_t^2$ into comparisons for a fixed number of regions of data. To do this, define $c_n^{(\alpha)} = \lceil (\log n)^{1+\alpha} \rceil$, $u_0 = 0$, $l_{m+1} = n$, and for $i = 1, \ldots, m$, $l_i = \tau_i - c_n^{(\alpha)}$ and $u_i = \tau_i + c_n^{(\alpha)}$. We can partition the time points $1, \ldots, n$ into regions $\mathcal{B}_i = \{u_{i-1}+1, \ldots, l_i\}$, for $i = 1, \ldots, m+1$ and regions $\mathcal{L}_i = \{l_i+1, \ldots, \tau_i\}$ and $\mathcal{R}_i = \{\tau_i+1, \ldots, u_i\}$ for $i = 1, \ldots, m$. These can be viewed as regions more than $c_n^{(\alpha)}$ from a changepoint, and regions of length $c_n^{(\alpha)}$ that are respectively left and right of a changepoint.

It is straightforward to show that for any segmentation:

$$\mathrm{RSS}(y_{1:n}; \hat{\tau}_{1:\hat{m}}) \geq \sum_{i=1}^{m+1} \mathrm{RSS}(y_{\mathcal{B}_i}; \hat{\tau}_{1:\hat{m}}) + \sum_{i=1}^{m+1} \mathrm{RSS}(y_{\mathcal{L}_i}; \hat{\tau}_{1:\hat{m}}) + \sum_{i=1}^{m+1} \mathrm{RSS}(y_{\mathcal{R}_i}; \hat{\tau}_{1:\hat{m}}).$$

The proof proceeds by showing that on each region $\mathcal{B}_i$ if we have $k = k(\hat{\tau}_{1:\hat{m}})$ changepoints that lie within this region then with probability tending to 1:

$$\max_{\hat{\tau}_{1:\hat{m}}} \left\{ \mathrm{RSS}(y_{\mathcal{B}_i}; \hat{\tau}_{1:\hat{m}}) + 2\,(1+\epsilon/2)\,k \log n - \sum_{t=u_i+1}^{l_i} Z_t^2 \right\} > -4 \log\log n.$$

Then we show that on each region $\mathcal{L}_i$ (and similarly each region $\mathcal{R}_i$) that if there are $k = k(\hat{\tau}_{1:\hat{m}})$ changepoints, then with probability tending to 1:

$$\max_{\hat{\tau}_{1:\hat{m}}} \left\{ \mathrm{RSS}(y_{\mathcal{L}_i}; \hat{\tau}_{1:\hat{m}}) - \sum_{t=u_i+1}^{l_i} Z_t^2 \right\} > -4\,(k+1) \log\log n.$$

Taken together we have, with probability tending to 1, a uniform bound on the difference in cost between any segmentation with more than m changepoints, that has one change within $c_n^{(\alpha)}$ of each true change, and the true segmentation. As such a segmentation can only have, at most, $\hat{m} - m$ changes in regions $\mathcal{B}_i$ this difference is bounded by:

$$(\hat{m} - m)\epsilon \log n - 4(2m+3) \log\log n > \epsilon \log n - 4(2m+3) \log\log n,$$

which is positive for large enough n.

Note that on each region $\mathcal{B}_i, \mathcal{L}_i, \mathcal{R}_i$ there are no true changes so any estimated changes we do fit inside these regions will involve fitting changes to the noise. Take a generic region of length $\tilde{n}$ which contains no true changes. We examine the reduction in the residual sum of squares when we add 0 and $k > 0$ estimated changes. Note that in the former case it is true that:

$$-\mathrm{RSS}(y_{\mathcal{A}_i}; \hat{\tau}_{1:\hat{m}}) + \sum_{t=a_i+1}^{a_i+\tilde{n}} Z_t^2 = \frac{1}{\tilde{n}} \left(\sum_{t=a_i+1}^{a_i+\tilde{n}} Z_t \right)^2,$$

where $\mathcal{A}$ is used as a placeholder to refer to any of the three types of region such that $\mathcal{A}_i = \{a_i + 1, \ldots, a_i + \tilde{n}\}$. Thus, the negative of the expression of interest is distributed according to χ_1^2. Therefore, for sufficiently large n, the probability that this quantity is greater than $4 \log\log n$ tends to 0.

So we need focus only on the case where $k > 0$. Label, without loss of generality, the estimated changes which lie in the region $\mathcal{A}_i$ as $\hat{\tau}_1, \ldots, \hat{\tau}_k$, and let:

$$\text{Diff} = \text{RSS}(y_{\mathcal{A}_i}; \hat{\tau}_{1:\hat{m}}) - \sum_{t=a_i+1}^{a_i+\tilde{n}} Z_t^2.$$

Then:

$$\text{Diff} = \frac{1}{\hat{\tau}_1 - a_i} \left(\sum_{t=a_i+1}^{\hat{\tau}_1} Z_t \right)^2 + \ldots + \frac{1}{a_i + \tilde{n} - \hat{\tau}_k} \left(\sum_{t=\hat{\tau}_k+1}^{a_i+\tilde{n}} Z_t \right)^2$$

We demonstrate that this difference is less than $2k(1+\epsilon)\log \tilde{n}$, for any $\epsilon > 0$. Note that, collectively, the positive terms in the expression follow a χ^2_{k+1} distribution. By Laurent and Massart (2000), for any quantity U which follows a chi-squared distribution with D degrees of freedom, then for any $x > 0$:

$$\mathbb{P}\left(U - D \geq 2\sqrt{Dx} + 2x\right) \leq \exp(-x). \tag{3}$$

Letting $D = k+1$ and $x = \frac{d \log \tilde{n} - \sqrt{(2d \log \tilde{n} - (k+1))(k+1)}}{2}$, for some $d > 0$ such that $\tilde{n} \geq e^{\frac{k+1}{2d}}$. In practice $d > k$ (see below) so almost all positive integer values of $\tilde{n}$ will be sufficient. With this choice of x, the LHS of (3) corresponds to $\mathbb{P}(U > d \log \tilde{n})$, and for large enough $\tilde{n}$ (3) becomes:

$$\mathbb{P}(U \geq d \log n) \leq \tilde{n}^{-\frac{d}{2}+\delta}, \text{ for any } \delta > 0 \tag{4}$$

There are then $\binom{\tilde{n}}{k}$ possible segmentations of these (incorrectly) fitted changes in this region. Given that $\binom{\tilde{n}}{k} < \frac{\tilde{n}^k}{k!}$ then by employing a Bonferroni correction, for the best segmentation involving k changes in the region:

$$\mathbb{P}(\text{Diff} \geq d \log \tilde{n}) \leq \tilde{n}^{-\frac{d}{2}+\delta} \tilde{n}^{k}$$
$$= \tilde{n}^{k+\delta-\frac{d}{2}} \to 0 \text{ for } d = 2k(1+\epsilon), \text{ if we set, for example, } \delta = \epsilon/2.$$

(For $d = 2k(1+\epsilon)$, if $\delta = \epsilon/2$ - as (4) permits any strictly positive value of δ - then $k + \delta - \frac{d}{2} = -(2k-1)\epsilon/2 < 0$.)

Note that this establishes the appropriate bound only in the case where k is fixed and positive. To obtain the uniform bound over all k, we must sum over all $k = 1, ..., \tilde{n}$.

So for a given $\tilde{n}$ and ϵ:

$$\sum_{k=1}^{\tilde{n}} \mathbb{P}(\text{Diff} \geq 2k(1+\epsilon) \log \tilde{n}) \leq \sum_{k=1}^{\tilde{n}} \tilde{n}^{-(2k-1)\epsilon/2}$$
$$= \frac{\tilde{n}^{-\epsilon/2}\left(1 - \tilde{n}^{-\tilde{n}\epsilon}\right)}{1 - \tilde{n}^{-\epsilon}} \to 0, \ \forall \epsilon > 0.$$

This establishes the required results for both regions of type $\mathcal{B}_i$ and $\mathcal{L}_i$ ($\mathcal{R}_i$) by substituting $\tilde{n} = \lambda n$, $\lambda \leq 1$ and $\tilde{n} = \lceil (\log n)^{1+\alpha} \rceil$ (for $\alpha < 1$ to obtain the constant 4 in the two initial statements) respectively.

Hence $\mathbb{P}(\hat{m} > m) \to 0$.

Proof of (b): Now have that $\hat{m} < m$. For n sufficiently large, it is guaranteed that there is at least one true change (which shall be labelled τ) such that the closest estimated change is at least $\lceil (\log n)^{1+\alpha} \rceil$ time points away. Thus, by the proof of Corollary 1.2.1, given that a change has been missed by this error, adding in estimated changes to the model at the points $\tau - \lceil (\log n)^{1+\alpha} \rceil$, τ, $\tau + \lceil (\log n)^{1+\alpha} \rceil$ gives that the reduction in the RSS is greater than the incurred penalty for adding 3 changes. Thus, the original segmentation was not optimal.

Hence $\mathbb{P}(\hat{m} < m) \to 0$.

Lastly, we need to establish that when $\hat{m} = m$, the event that each of the estimated changes is within $\lceil (\log n)^{1+\alpha} \rceil$ of a true change tends to 1. Suppose we have a segmentation with $\hat{m} = m$ which contains a true change, τ_i, with no estimated changes within $\lceil (\log n)^{1+\alpha} \rceil$. Then by comparing this segmentation to an equivalent segmentation which also fits estimated changes at $\tau_i - \lceil (\log n)^{1+\alpha} \rceil, \tau_i, \tau_i + \lceil (\log n)^{1+\alpha} \rceil$, we again obtain a saving of greater than the cost of adding 3 changes by Yao (1988) and Corollary 1.2.1. $\square$

Note that this result extends naturally to a multivariate analogue:

Lemma 2.1. *Take a procedure which exactly minimises the squared error loss for the multivariate problem:*

$$\boldsymbol{Y}_i = \boldsymbol{\epsilon}_i + \boldsymbol{\mu}_k, \text{ for } \tau_{k-1} + 1 \leq i \leq \tau_k, \text{ and } k \in \{1, ..., m+1\}, \tag{5}$$

where $\boldsymbol{Y}_i = \left(Y_i^{(1)}, ..., Y_i^{(d)}\right)^T$, $\forall i \in \{1, ..., n\}$; $\boldsymbol{\mu}_k \neq \boldsymbol{\mu}_{k+1}$, $\forall k \in \{1, ..., m\}$; $\boldsymbol{\epsilon}_i \sim^{i.i.d.} N_d(\boldsymbol{0}, \sigma^2 I)$, *some* d. *In addition, take the penalty for fitting a change to be* $(d+1)(1+\epsilon)\log n$, *for any* $\epsilon > 0$. *Then defining* $\mathcal{E}_n^\alpha$ *as for Proposition 3.1 for any* $\alpha > 0$ *again gives that* $\mathbb{P}(\mathcal{E}_n^\alpha) \to 1$ *as* $n \to \infty$.

Proof of Lemma 2.1: We define the natural extension of the residual sum of squares in the multivariate case as:

$$\begin{aligned}
\text{RSS}(\mathbf{y}_{1:n}; \hat{\tau}_{1:\hat{m}}) &= \sum_{i=1}^{\hat{\tau}_1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_1)^T (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_1) + ... + \sum_{i=\hat{\tau}_{\hat{m}}+1}^{n} (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{\hat{m}+1})^T (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_{\hat{m}+1}) \\
&= \sum_{j=1}^{\hat{m}+1} \sum_{i=\hat{\tau}_{j-1}+1}^{\hat{\tau}_j} \sum_{k=1}^{d} (y_{i,k} - \hat{\mu}_{j,k})^2, \text{ with } \hat{\mu}_{j,k} = \frac{1}{\hat{\tau}_j - \hat{\tau}_{j-1}} \sum_{i=\hat{\tau}_{j-1}+1}^{\hat{\tau}_j} y_{i,k} = \bar{y}_{j,k}.
\end{aligned}$$

Using this, we proceed along the same trajectory as for the previous proof. Suppose that $\hat{m}$ changes are detected by the procedure. Then we first show that:

(a): $\mathbb{P}(\hat{m} > m) \to 0$,

(b): $\mathbb{P}(\hat{m} < m) \to 0$.

Proof of (a): Again let $c_n^{(\alpha)} = \left\lceil (\log n)^{1+\alpha} \right\rceil$. Note first that an equivalent result to Corollary 1.2.1 holds in the multivariate case as the residual sum of squares between the points $\tau_i - c_n^{(\alpha)}$ and $\tau_i + c_n^{(\alpha)}$ (where τ_i is some true change missed by the procedure as before) satisfies:

$$\begin{aligned}
\frac{\text{RSS}\left(\mathbf{y}_{\tau_i - c_n^{(\alpha)}+1:\tau_i + c_n^{(\alpha)}}; \emptyset\right) - \sum_{k=1}^{d} \sum_{j=\tau_i - c_n^{(\alpha)}+1}^{\tau_i + c_n^{(\alpha)}} Z_{j,k}^2}{c_n^{(\alpha)}} &= \sum_{k=1}^{d} \frac{\left(\mu_k^{(i)} - \mu_k^{(i+1)}\right)^2}{2} - \frac{1}{2c_n^{(\alpha)}} \sum_{k=1}^{d} \left(\sum_j Z_{j,k}\right)^2 \\
&\quad + \frac{D_k}{c_n^{(\alpha)}} \\
&\to \sum_{k=1}^{d} \frac{\left(\mu_k^{(i)} - \mu_k^{(i+1)}\right)^2}{2} \text{ as } n \to \infty,
\end{aligned}$$

where D_k is normally distributed with a variance equivalent to the deterministic term scaled by $4\sigma^2$.

Hence, as per the previous proof, we can compare the residual sum of squares of the fit of a set of estimated changes with $\hat{m} > m$ across (equivalent) regions $\mathcal{B}_i, \mathcal{L}_i, \mathcal{R}_i$ to the null fit. Across a region bounded by the points (a, b) containing estimated changes $\hat{\tau}_1, ..., \hat{\tau}_p$, the relevant difference term is:

$$\text{Diff} = \sum_{k=1}^{d} \left[\frac{1}{\hat{\tau}_1 - a} \left(\sum_{j=a+1}^{\hat{\tau}_1} Z_{j,k} \right)^2 + ... + \frac{1}{b - \hat{\tau}_p} \left(\sum_{j=\hat{\tau}_p+1}^{b} Z_{j,k} \right)^2 \right],$$

giving that $\text{Diff} \sim \chi^2_{d(p+1)}$. A similar argument to before then gives that:

$$\mathbb{P}(\text{Diff} \geq p\,(d+1)\,(1+\epsilon)\log n) \to 0,$$

and in particular:

$$\sum_{p=1}^{n} \mathbb{P}(\text{Diff} \geq p\,(d+1)\,(1+\epsilon)\log n) \to 0.$$

Hence $\mathbb{P}(\hat{m} > m) \to 0$.

Proof of (b): This follows immediately from considering the multivariate equivalent to Corollary 1.2.1 shown above, inferring the presence of a missed change, τ_i, and fitting three estimated changes at $\tau_i - \left\lceil (\log n)^{1+\alpha} \right\rceil, \tau_i, \tau_i + \left\lceil (\log n)^{1+\alpha} \right\rceil$. This segmentation will produce a lower residual sum of squares than the original with probability approaching 1.

Hence $\mathbb{P}(\hat{m} < m) \to 0$.

All that remains is to show that this correct number of changes falls within $\left\lceil (\log n)^{1+\alpha} \right\rceil$. However, this again follows the same line of reason as for the univariate case by the result established above. $\square$

References

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302 – 1338.

Yao, Y.-C. (1988). Estimating the number of change-points via schwarz' criterion. *Statistics & Probability Letters*, 6(3):181–189.