Locating Changes in Highly Dependent Data

A. Khaleghi D. Ryabko

INRIA, Lille

November 16, 2012

Outline for section 1

The Change-Point Problem

We are given a sequence

$$\mathbf{x} := X_1, \dots, X_n \in \mathbb{R}^n, \;\; n \in \mathbb{N}$$

which is the concatenation of some $k+1$ sequences

$$\mathbf{X_1, \dots, X_{\pi_1}}, \mathbf{X_{\pi_1+1}, \dots, X_{\pi_2}}, \dots, X_{\pi_k+1}, \dots, X_n$$

The Change-Point Problem

We are given a sequence

$$\mathbf{x} := X_1, \dots, X_n \in \mathbb{R}^n,\ \ n \in \mathbb{N}$$

which is the concatenation of some $k+1$ sequences

$$\underbrace{X_1, \dots, X_{\pi_1}}_{\sim\rho}, \underbrace{X_{\pi_1+1}, \dots, X_{\pi_2}}_{\sim\rho'}, \underbrace{X_{\pi_2} + 1, \dots, X_{\pi_3}}_{\sim\rho}, \dots, \underbrace{X_{\pi_k+1}, \dots, X_n}_{\sim\rho''}$$

The consecutive sequences separated by $\pi_i,\ i = 1..k$ are generated by **different unknown process distributions**.
$\Rightarrow$ *The indices* $\pi_i,\ i = 1..k$ *are called* **change-points**.

The Change-Point Problem

Objective

We want to find an **asymptotically consistent estimate** $\hat{\pi}_i$ for every change-point $\pi_i,\ i = 1..k$ so that with probability one we have

$$\lim_{n\to\infty} \frac{1}{n}|\hat{\pi}_i - \pi_i| = 0$$

Related Work

- Change point analysis is a classical problem.
- In a typical formulation samples are assumed i.i.d. and the change is in the mean.
- More general formulations usually assume that the samples belong to specific model classes (ex. HMM)
- In typical nonparametric settings the form of the change and/or the nature of dependence are usually restricted; ex. the time-series are assumed to be strongly mixing
- It is almost exclusively assumed that the finite-dimensional marginals before and after the change are *different*.

See Khaleghi and Ryabko, arxiv 2012 and NIPS 2012 & references therein.

Related Work

What distinguishes our work is that ...

1. We do **not** require the finite-dimensional marginals of any fixed size before and after the change to be different.
   - $\rightarrow$ We consider the general case: the process distributions are different.
2. We make **no** such assumptions as **independence**,**memory**,**mixing**...
   - $\rightarrow$ We allow the samples to be **dependent**.
   - $\rightarrow$ The dependence can be **arbitrary**.
3. Our **only assumption** is that the process distributions are **stationary ergodic**.

Our framework is similar to that of Ryabko and Ryabko, IEEE Trans. 2010 where the case of **one change-point** was considered.

$\rightarrow$ It turns out that extensions to multiple change-points is *not* trivial!

Assumptions on Data

The process distributions generating the data are **stationary ergodic**.

Stationarity (time index does not matter)

A measure μ is stationary if for every $m, t \in \mathbb{N}$ and any measurable set A we have

$$\mu(X_1, \dots, X_m \in A) = \mu(X_{1+t}, \cdots, X_{m+t} \in A)$$

Ergodicity (it makes sense to measure frequencies)

Ergodicity means that the frequency of occurrence of any event converges to its probability almost surely.

Remark

The assumption that the distributions are stationary-ergodic is one of the weakest assumptions in statistics.

Outline for section 2

Problem Formulation

Let

$$x := X_1, \dots, X_{\pi_1}, X_{\pi_1+1}, \dots, X_{\pi_2}, \dots, X_{\pi_k+1}, \dots, X_n.$$

be a sequence with k change-points where,

- Each of the sequences $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ is generated by an **unknown stationary ergodic** process distribution.
- The lengths of $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ are linear in n

  → The change-points are linear in n:

  $$\pi_i := n\theta_i,\ i = 1..k,\ \theta_i \in (0,1)$$

  → The change-points are at least $n\lambda_{\min}$ apart for

  $$\lambda_{\min} := \min_{i=1..k+1} \theta_i - \theta_{i-1} > 0$$

  where $\theta_0 := 0$ and $\theta_{k+1} := 1$.

Problem Formulation

Let

$$x := X_1, \dots, X_{\pi_1}, X_{\pi_1+1}, \dots, X_{\pi_2}, \dots, X_{\pi_k+1}, \dots, X_n.$$

be a sequence with k change-points where,

- Each of the sequences $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ is generated by an **unknown stationary ergodic** process distribution.
- The lengths of $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ are linear in n
  - $\rightarrow$ The change-points are linear in n:

$$\pi_i := n\theta_i,\ i = 1..k,\ \theta_i \in (0,1)$$

  - $\rightarrow$ The change-points are at least $n\lambda_{\min}$ apart for

$$\lambda_{\min} := \min_{i=1..k+1} \theta_i - \theta_{i-1} > 0$$

where $\theta_0 := 0$ and $\theta_{k+1} := 1$.

Problem Formulation

Let

$$x := X_1, \dots, X_{\pi_1}, X_{\pi_1+1}, \dots, X_{\pi_2}, \dots, X_{\pi_k+1}, \dots, X_n.$$

be a sequence with k change-points where,

- Each of the sequences $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ is generated by an **unknown stationary ergodic** process distribution.
- The lengths of $X_{1..\pi_1},\ X_{\pi_1+1..\pi_2}, \dots,\ X_{\pi_k+1..n}$ are linear in n
  - $\rightarrow$ The change-points are linear in n:

$$\pi_i := n\theta_i,\ i = 1..k,\ \theta_i \in (0,1)$$

  - $\rightarrow$ The change-points are at least $n\lambda_{\min}$ apart for

$$\lambda_{\min} := \min_{i=1..k+1} \theta_i - \theta_{i-1} > 0$$

where $\theta_0 := 0$ and $\theta_{k+1} := 1$.

An Impossibility Theorem

Theorem (Ryabko, 2010)

Assume that the sequences $\mathbf{x}$ and $\mathbf{y}$ are generated by stationary ergodic distributions. It is impossible to distinguish between the case where they are generated by *the same* process distribution or by *two different* ones.

$\rightarrow$ In this general framework it is impossible to estimate k.

Number k of Change-Points

1. If k is **known**
   → we can have a consistent algorithm;
   see (Khaleghi, Ryabko, arxiv 12).
2. If k is **unknown** we can either
   → make **stronger assumptions** or
   → have **an alternative goal**
   - Produce **a sorted list of change points**
   - The **first k elements** of the produced list are the **true change points**
   - An algorithm that achieves this goal is called **consistent**

Number k of Change-Points

1. If k is **known**
   - $\rightarrow$ we can have a consistent algorithm; see (Khaleghi, Ryabko, arxiv 12).
2. If k is **unknown** we can either
   - $\rightarrow$ ~~make stronger assumptions~~
   - $\rightarrow$ have **an alternative goal**
     - Produce **a sorted list of change points**
     - The **first k elements** of the produced list are the **true change points**
     - An algorithm that achieves this goal is called **consistent**

Analogy to Hierarchical Clustering

In clustering with an unknown # k of clusters

- A hierarchical clustering algorithm produces a tree
- Some pruning of this tree is the ground-truth

In change-point estimation with an unknown # k of change-points

- Produce **a sorted list of change points**
- The **first k elements** of the produced list are the **true change points**
- An algorithm that achieves this goal is called **consistent**

Theorem

Let $\mathbf{x} := X_1, \dots, X_n,\ n \in \mathbb{N}$ be a sequence with an *unknown* number of change-points. If the distributions generating $\mathbf{x}$ are stationary ergodic and some $\lambda \in (0, \lambda_{\min}]$ is provided, then there exists a multiple change-point estimation algorithm producing a *s*orted list of estimates

$$\hat{\pi}_1, \dots, \hat{\pi}_{1/\lambda}$$

whose first k elements consistently estimate the change-points $\pi_1, \dots, \pi_k$, i.e.

$$\lim_{n\to\infty} \sup_{i=1..k} \frac{1}{n} |\hat{\pi}_{[i]} - \pi_i| = 0.$$

Proposed Algorithm

Distance Measure

We need a notion of distance between the sequences

How **'close'** are the two sequences $\mathbf{x} := X_1, X_2, \ldots, X_n$ and $\mathbf{y} := Y_1, Y_2, \ldots, Y_{n'}$?

→ For binary sequences

- let $B^m,\ m \in \mathbb{N}$ be all possible words of length m

$$0, 1, 00, 01, 10, 11, 000, 001, \ldots$$

- let $\nu(\mathbf{x}, B^m)$ be the frequency of the word B^m in $\mathbf{x} := X_1, X_2, \ldots, X_n$
- define the distance

$$\hat{d}(\mathbf{x}, \mathbf{y}) := \sum_{m=1}^{\infty} 2^{-m} |\nu(\mathbf{x}, B^m) - \nu(\mathbf{y}, B^m)|$$

Distance Measure

We need a notion of distance between the sequences

How **'close'** are the two sequences $\mathbf{x} := X_1, X_2, \ldots, X_n$ and $\mathbf{y} := Y_1, Y_2, \ldots, Y_{n'}$?

$\rightarrow$ For binary sequences

- let $B^m,\ m \in \mathbb{N}$ be all possible words of length m

$$0, 1, 00, 01, 10, 11, 000, 001, \ldots$$

- let $\nu(\mathbf{x}, B^m)$ be the frequency of the word B^m in $\mathbf{x} := X_1, X_2, \ldots, X_n$
- define the distance

$$\hat{d}(\mathbf{x}, \mathbf{y}) := \sum_{m=1}^{\infty} 2^{-m} |\nu(\mathbf{x}, B^m) - \nu(\mathbf{y}, B^m)|$$

Distance Measure

We need a notion of distance between the sequences

How **'close'** are the two sequences $\mathbf{x} := X_1, X_2, \dots, X_n$ and $\mathbf{y} := Y_1, Y_2, \dots, Y_{n'}$?

$\rightarrow$ For real-valued sequences

- let $\mathcal{B}^{m,l}$ $m, l \in \mathbb{N}$ be the set of all hypercubes of dimension m and edge-length 2^{-l}
- let $\nu(\mathbf{x}, B)$ be the frequency with which $\mathbf{x}$ **crosses** $B \in \mathcal{B}^{m,l}$
- define the distance

$$\hat{d}(\mathbf{x}, \mathbf{y}) := \sum_{m,l=1}^{\infty} \sum_{B \in \mathcal{B}^{m,l}} 2^{-(m+l)} |\nu(\mathbf{x}, B) - \nu(\mathbf{y}, B)|$$

Distance Measure

We need a notion of distance between the sequences

How **'close'** are the two sequences $\mathbf{x} := X_1, X_2, \ldots, X_n$ and $\mathbf{y} := Y_1, Y_2, \ldots, Y_{n'}$?

$\rightarrow$ For real-valued sequences

- let $\mathcal{B}^{m,l}$ $m, l \in \mathbb{N}$ be the set of all hypercubes of dimension m and edge-length 2^{-l}
- let $\nu(\mathbf{x}, B)$ be the frequency with which $\mathbf{x}$ **crosses** $B \in \mathcal{B}^{m,l}$
- define the distance

$$\hat{d}(\mathbf{x}, \mathbf{y}) := \sum_{m,l=1}^{\infty} \sum_{B \in \mathcal{B}^{m,l}} 2^{-(m+l)} |\nu(\mathbf{x}, B) - \nu(\mathbf{y}, B)|$$

Distance Measure

Assume that $\mathbf{x}_1$ is generated by ρ_1 and $\mathbf{x}_2$ by ρ_2.
If ρ_1 and ρ_2 are **stationary ergodic**, $\hat{d}(\cdot,\cdot)$ converges to the so-called **distributional-distance**:

$$d(\rho_1, \rho_2) := \sum_{m,l=1}^{\infty} 2^{-(m+l)} \sum_{B \in \mathcal{B}^{m,l}} |\rho_1(B) - \rho_2(B)|$$

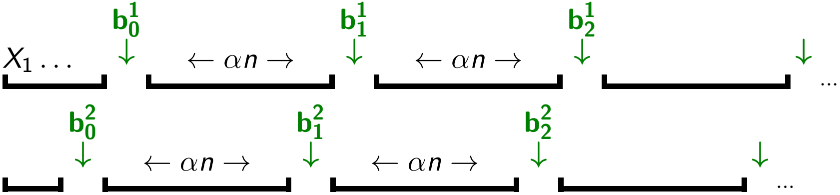
See Gray, 1988 for more on $d(\cdot,\cdot)$.

Algorithm

input: $\mathbf{x} := X_1, \dots, X_n,\ \lambda \in (0, \lambda_{\min}]$

1. Set interval size $\alpha \leftarrow \frac{\lambda}{3}$
2. Generate 2 sets of Separators

$$b_i^t \leftarrow n\alpha(i + \frac{1}{t+1}),\ i = 0..\frac{1}{\alpha},\ t = 1, 2$$



- $b_0^1 := \frac{n\alpha}{2}$
- $b_0^2 := \frac{n\alpha}{3}$

Algorithm

3. Estimate a change-point $\hat{\pi}_i^t$ in every segment

$$\ldots \quad \overset{\mathbf{b_i^t}}{\downarrow} \quad \underbrace{\leftarrow \alpha n \rightarrow} \quad \overset{\mathbf{b_{i+1}^t}}{\downarrow} \quad \ldots$$

4. Calculate a performance score for every estimate $\hat{\pi}_i^t$
5. Remove duplicate estimates

   Start from the set of all estimates

   Do

   i. **Add** *an available estimate* $\hat{\pi}$ of **highest score** to the output list

   ii. **Remove** all **estimates within a radius of** $\lambda n/2$ from $\hat{\pi}$

   While estimates are still available

output: A list of change-point estimates

Algorithm

3. Estimate a change-point $\hat{\pi}_i^t$ in every segment

$$\ldots \quad \overset{\mathbf{b_i^t}}{\downarrow} \quad \underbrace{\leftarrow \alpha n \rightarrow} \quad \overset{\mathbf{b_{i+1}^t}}{\downarrow} \quad \ldots$$

4. Calculate a performance score for every estimate $\hat{\pi}_i^t$
5. Remove duplicate estimates

   Start from the set of all estimates

   Do

   i. **Add** *an available estimate* $\hat{\pi}$ of **highest score** to the output list

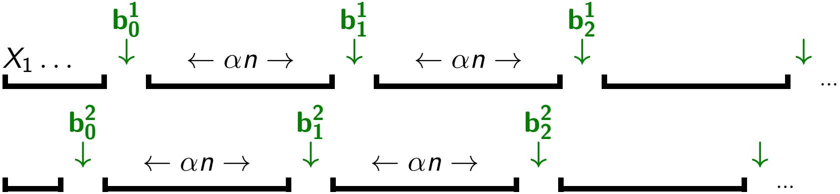
   ii. **Remove** all **estimates within a radius of $\lambda n/2$** from $\hat{\pi}$

   While estimates are still available

output: A list of change-point estimates

Estimating the Change-Points

First, we generate two sets of segments



with the property that if $\pi_j \in [b_i^t, b_{i+1}^t]$ for some $j \in 1..k$ then

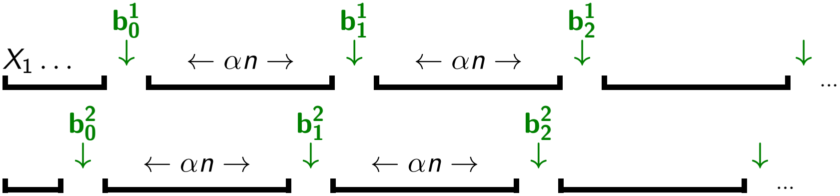
$$[\pi_{j-1}, \pi_{j+1}] \subseteq [b_{i-1}^t, b_{i+2}^t]$$

$$\dots \underbrace{X_{b_{i-1}^t+1} \dots X_{b_i^t} \downarrow X_{b_i^t+1} \dots X_{\pi_j}}_{\sim\rho} \underbrace{X_{\pi_j+1} \dots X_{b_{i+1}^t} \downarrow X_{b_{i+1}^t+1} \dots X_{b_{i+2}^t}}_{\sim\rho'} \dots$$

$$\hat{\pi}_j := \operatorname*{argmax}_{t' \in b_i^t..b_{i+1}^t} \hat{d}(X_{b_{i-1}^t+1..t'}, X_{t'+1..b_{i+2}^t})$$

Estimating the Change-Points

First, we generate two sets of segments



with the property that if $\pi_j \in [b_i^t, b_{i+1}^t]$ for some $j \in 1..k$ then

$$[\pi_{j-1}, \pi_{j+1}] \subseteq [b_{i-1}^t, b_{i+2}^t]$$

$$\ldots \underbrace{X_{b_{i-1}^t+1} \ldots X_{b_i^t} \downarrow X_{b_i^t+1} \ldots X_{\pi_j}}_{\sim\rho} \underbrace{X_{\pi_j+1} \ldots X_{b_{i+1}^t} \downarrow X_{b_{i+1}^t+1} \ldots X_{b_{i+2}^t}}_{\sim\rho'} \ldots$$

$$\hat{\pi}_j := \operatorname*{argmax}_{t' \in b_i^t..b_{i+1}^t} \hat{d}(X_{b_{i-1}^t+1..t'}, X_{t'+1..b_{i+2}^t})$$

Estimating the Change-Points

If $\pi_j \in (b_i^t, b_{i+1}^t)$ then we estimate it as the index in $b_i^t..b_{i+1}^t$ that breaks $X_{b_{i-1}^t...b_{i+2}^t}$ into two sequences of maximum distance, i.e.

$$\hat{\pi}_j := \operatorname*{argmax}_{t' \in b_i^t..b_{i+1}^t} \hat{d}(X_{b_{i-1}^t+1..t'}, X_{t'+1..b_{i+2}^t})$$

Lemma

For all $\alpha \in (0, \lambda_{\text{min}}/3]$ and sequences of separators $b_i^t := n\alpha(i + \frac{1}{t+1})$ $i = 0..\alpha^{-1}$, $t = 1, 2$ we have

$$\lim_{n\to\infty} \sup_{\substack{i \in 1..\alpha^{-1}-1 \\ \exists j \in 1..k \text{ s.t. } \pi_j \in (b_i^t, b_{i+1}^t)}} \frac{1}{n}|\hat{\pi}_j - \pi_j| = 0$$

Estimating the Change-Points

If $\pi_j \in (b_i^t, b_{i+1}^t)$ then we estimate it as the index in $b_i^t..b_{i+1}^t$ that breaks $X_{b_{i-1}^t \dots b_{i+2}^t}$ into two sequences of maximum distance, i.e.

$$\hat{\pi}_j := \operatorname*{argmax}_{t' \in b_i^t..b_{i+1}^t} \hat{d}(X_{b_{i-1}^t+1..t'}, X_{t'+1..b_{i+2}^t})$$

Lemma

For all $\alpha \in (0, \lambda_{\min}/3]$ and sequences of separators $b_i^t := n\alpha(i + \frac{1}{t+1})$ $i = 0..\alpha^{-1}$, $t = 1, 2$ we have

$$\lim_{n\to\infty} \sup_{\substack{i \in 1..\alpha^{-1}-1 \\ \exists j \in 1..k \text{ s.t. } \pi_j \in (b_i^t, b_{i+1}^t)}} \frac{1}{n} |\hat{\pi}_j - \pi_j| = 0$$

A Performance Indicator

For every segment $X_{b_i^t..b_{i+1}^t}$ we have two cases:

It does not contain any change-points

$$\underbrace{\dots X_{b_i} \downarrow X_{b_i+1} \dots X_{b_{i+1}} \downarrow X_{b_{i+1}+1} \dots}_{\sim\rho}$$

or it contains exactly one change-point

$$\underbrace{\dots X_{b_i} \downarrow X_{b_i+1} \dots X_{\pi_j}}_{\sim\rho} \underbrace{X_{\pi_j+1} \dots X_{b_{i+1}} \downarrow X_{b_{i+1}+1} \dots}_{\sim\rho'}$$

Question

Which segments contain true change-points?

A Performance Indicator

Definition

For a segment $X_{b\dots b'}$, $b, b' \in 1..n$ define

$$\Delta_{\mathbf{x}}(b, b') := \hat{d}(X_{b..\frac{b+b'}{2}}, X_{\frac{b+b'}{2}..b'})$$

$$X_1 X_2 \dots X_{b-1} \overset{\mathbf{b}}{\downarrow} \overbrace{X_b \dots X_{\frac{b+b'}{2}}} \underbrace{X_{\frac{b+b'}{2}+1} \dots X_{b'}} \overset{\mathbf{b'}}{\downarrow} X_{b'+1} \dots X_n$$

A Performance Indicator

We can show that,
if $X_{b_i^t..b_{i+1}^t}$ does not contain any change-points

$$\underbrace{\dots X_{b_i} \downarrow X_{b_i+1} \dots X_{b_{i+1}} \downarrow X_{b_{i+1}+1} \dots}_{\sim\rho}$$

then $\Delta_{\mathbf{x}}(b_i^t, b_i^t) \to 0$ and
if it contains exactly one change-point

$$\underbrace{\dots X_{b_i} \downarrow X_{b_i+1} \dots X_{\pi_j}}_{\sim\rho}\underbrace{X_{\pi_j+1} \dots X_{b_{i+1}} \downarrow X_{b_{i+1}+1} \dots}_{\sim\rho'}$$

then $\Delta_{\mathbf{x}}(b_i^t, b_{i+1}^t) \to$ **non-zero constant**.

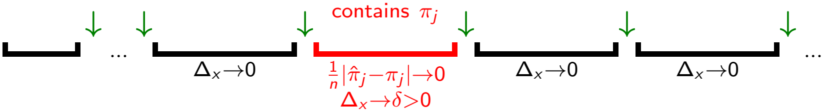
Lemma

For all $\alpha \in (0, \lambda_{\min}/3]$ and sequences of separators $b_i^t := n\alpha(i + \frac{1}{t+1})$ $i = 0..\alpha^{-1}, \ t = 1, 2$ we have

1. $$\lim_{n\to\infty} \sup_{\substack{i\in 1..\alpha^{-1}, t=1,2 \\ \nexists j\in 1..k \text{ s.t. } \pi_j \in (b_{i-1}^t, b_i^t)}} \frac{1}{n}\Delta_{\mathbf{x}}(b_{i-1}^t, b_i^t) = 0$$
2. There exists some $\zeta \in (0, 1)$ such that
$$\lim_{n\to\infty} \inf_{\substack{i\in 1..\alpha^{-1}, t=1,2 \\ \exists j\in 1..k \text{ s.t. } \pi_j \in (b_{i-1}^t, b_i^t)}} \frac{1}{n}\Delta_{\mathbf{x}}(b_{i-1}^t, b_i^t) \geq \zeta\delta_{\min}$$
where $\delta_{\min} \in (0, 1)$ is the minimum distance between the distributions.

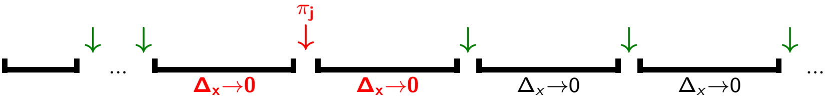
Recap

- Every change-point can be estimated consistently
- The performance scores of the **false estimates** converge to 0
- The performance scores of the **true estimates** converge to some non-zero constant



A Technical Problem

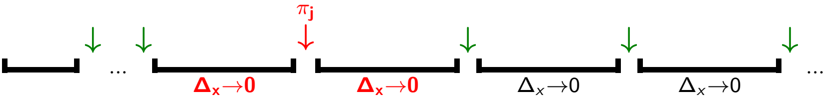
A separator may *'hit'* a change-point



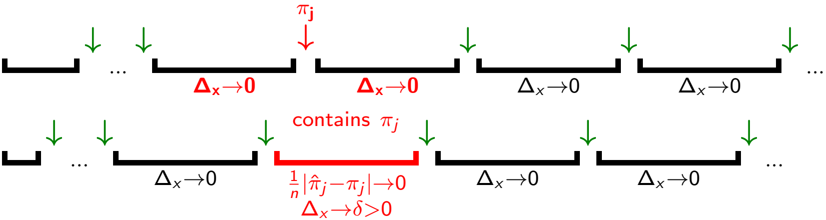
We need to ensure that every change-point is estimated at least once.

A Technical Problem

A separator may *'hit'* a change-point



We need to ensure that every change-point is estimated at least once.



Consistency

- The algorithm provides a list of change-point estimates.
- The estimates are sorted according to their performance scores.
- The estimates are at least $n\lambda$ apart.
- The first k estimates converge to some permutation of the true change-points.

Consistency

Theorem

Let $\mathbf{x} := X_1, \ldots, X_n,\ n \in \mathbb{N}$ be a sequence with an *unknown* number of change-points. If the distributions generating $\mathbf{x}$ are stationary ergodic and some $\lambda \in (0, \lambda_{\text{min}}]$ is provided, then there exists a multiple change-point estimation algorithm producing a *s*orted list of estimates

$$\hat{\pi}_1, \ldots, \hat{\pi}_{1/\lambda}$$

whose first k elements consistently estimate the change-points $\pi_1, \ldots, \pi_k$, i.e.

$$\lim_{n \to \infty} \sup_{i=1..k} \frac{1}{n} |\hat{\pi}_{[i]} - \pi_i| = 0.$$

Outline for section 3

Time-Series Generation

To generate a binary sequence $\mathbf{x} = X_{1..n}$ we,

1. Fix some parameter $\alpha \in (0, 1)$.
2. Select $r_0 \in [0, 1]$ at random.
3. For each $i = 1..n$ obtain $r_i := r_{i-1} + \alpha - \lfloor r_{i-1} + \alpha \rfloor$.
4. Generate $\mathbf{x} = (X_1, \cdots, X_n)$ by thresholding each element at 0.5, i.e.

$$X_i := \mathbb{I}\{r_i > 0.5\}$$

- If α is irrational then $\mathbf{x}$ forms a **stationary ergodic time-series** which does **not** belong to any "simpler" class
  - $\rightarrow$ cannot be modeled by a hidden Markov process with a finite state-space; see (Shields, 1996).
- We simulate α by a `longdouble` with a long mantissa.

Time-Series Generation

To generate an input sequence $\mathbf{x} = X_{1..n}$ we proceed as follows

- Fix $\lambda_{\min} = 0.23$
- Randomly generate $k = 3$ change-points at a minimum distance $n\lambda_{\min}$
- Set $\alpha_1 := 0.30...,\ \alpha_2 := 0.35..., \alpha_3 := 0.40..., \alpha_4 := 0.45...$
- Use $\alpha_i,\ i = 1..4$ to generate the four segments.

Consistency

Algorithm with $\lambda = 0.18$

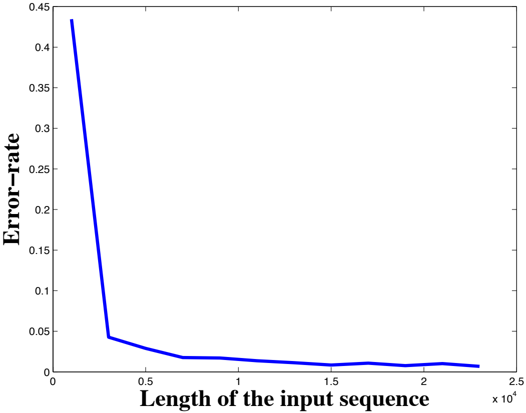


Figure: Avg (over 20 runs) error as a function of the length of the input sequence

Dependence on λ

Algorithm with $n = 20000$

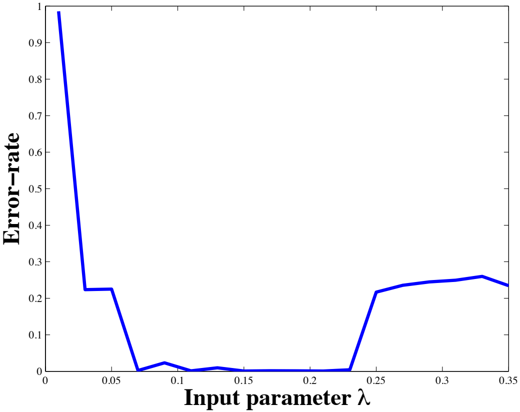


Figure: Avg (over 25 runs) error as a function of λ; recall $\lambda_{\text{min}} = 0.23$

Outline for section 4

Concluding Remarks

- We proposed a consistency framework for the change-point problem.
- If the number k of change-points is **known**
  - $\rightarrow$ We can consistently estimate every change-point; see (Khaleghi, Ryabko, arxiv).
- If k is **unknown** then it cannot be estimated (in this general setting).
  - $\rightarrow$ We produce a list of estimates, the first k of which converge to some permutation of the true change-points (Khaleghi, Ryabko, NIPS12).
  - $\rightarrow$ Other *intermediate* formulations may also be interesting
    - e.x. The total number of distributions is known, but k is unknown.

Concluding Remarks

- We proposed a consistency framework for the change-point problem.
- If the number k of change-points is **known**
  - → We can consistently estimate every change-point; see (Khaleghi, Ryabko, arxiv).
- If k is **unknown** then it cannot be estimated (in this general setting).
  - → We produce a list of estimates, the first k of which converge to some permutation of the true change-points (Khaleghi, Ryabko, NIPS12).
  - → Other *intermediate* formulations may also be interesting
    - e.x. The total number of distributions is known, but k is unknown.

I would appreciate any pointers for my spotlight presentation.

Thanks