The FAST algorithm

Per August Moen
University of Oslo

16 December 2022

FAST

FAST: **FA**st **S**parsity Adap**T**ive Change-point Estimator.

An algorithm for detecting and localizing multiple changes in mean in a sequence of multivariate Gaussian observations.

Joint work with Ingrid Glad (University of Oslo) and Martin Tveten (Norwegian Computing Center)

Setup

Suppose we observe p-dimensional data

$$X_i \overset{\text{ind}}{\sim} \mathsf{N}_p\left(\mu_i, \sigma^2 I\right),$$

where $\mu_i \in \mathbb{R}^p$ for all $i = 1, \ldots, n$.

Suppose there are $J \geq 0$ change-points $\eta_1, \ldots, \eta_J$ such that $\mu_i \neq \mu_{i+1}$ if and only if $i = \eta_j$ for some j.

Goal: estimate J and $\eta_1, \ldots, \eta_J$.

Motivation / background

Liu, Gao, and Samworth (2021) derived the exact minimax testing rate for a single change-point. They also provide a minimax rate optimal test statistic.

For multiple change-points, Pilliat, Carpentier, and Verzelen (2021) derive the detection boundary (in a minimax sense) for consistent estimation of J.

The detection boundary r depends on the sparsity k:

$$r(k) = \begin{cases} \sqrt{p \log n}, & \text{if } k \geq \sqrt{p \log n} \\ k \log\left(\frac{ep \log n}{k^2}\right) \vee \log n, & \text{if } k < \sqrt{p \log n} \end{cases}$$

Our contribution

(1) We propose a sparsity adaptive single change-point estimator with strong theoretical guarantees

(2) Combining the estimator with a search procedure, we propose FAST, a multiple change-point estimation algorithm

Single change-point estimation

Suppose $\sigma = 1$.

Let $C_i(j)$ denote the CUSUM of the j-th time series evaluated in time point i.

For any sparsity level t, define the penalized "score" $S_i^{\lambda}(t)$ as[1]

$$S_i^{\lambda}(t) := \sum_{j=1}^{p} \left(C_i\,(j)^2 - \nu_{a(t)} \right) \mathbb{1}\left\{ |C_i(j)| \geq a(t) \right\} - \lambda r(t),$$

where $\lambda > 0$, $a^2(t) = 4\log\left(\frac{ep\log n}{t^2}\right)\mathbb{1}\left\{t \leq \sqrt{p\log n}\right\}$ and $\nu_a := \mathbb{E}\left(Z^2 \mid |Z| \geq a\right)$ where $Z \sim \mathrm{N}(0,1)$. .

[1] $S_i^{\lambda}(t)$ is derived from a test statistic presented in Liu, Gao, and Samworth (2021)

... continued

Now define

$$\mathcal{T} := \left\{1, 2, 2^2, \ldots, 2^{\lfloor \sqrt{p \log n} \rfloor}\right\} \cup \{p\}.$$

For all candidate change-point positions $i = 1, \ldots, n-1$, define

$$S_i^{\lambda} = \max_{t \in \mathcal{T}} S_i^{\lambda}(t).$$

Then we can estimate the change-point location η by

$$\widehat{\eta}_{\lambda} = \operatorname*{arg\,max}_{i=1,\ldots,n-1} S_i^{\lambda}.$$

Theoretical guarantees

Let $\Delta := \min\{\eta, n-\eta\}$, $\phi = \|\mu_{\eta+1} - \mu_\eta\|_2$, $k = \|\mu_{\eta+1} - \mu_\eta\|_0$, and

$$h(k) = \begin{cases} \sqrt{p(\log n \vee \log\log(ep))}, & \text{if } k \geq \sqrt{p \log n}, \\ k \log\left(\frac{ep \log n}{k^2}\right) \vee \log n, & \text{if } k < \sqrt{p \log n}. \end{cases}$$

Proposition

There exist constants C_0, C_1 and a universal value of λ such that, if

$$\text{SNR} = \frac{\phi^2 \Delta}{\sigma^2} \geq C_0 h(k)$$

then with probability at least $1 - \frac{1}{n}$, we have

$$|\widehat{\eta}_\lambda - \eta| \leq C_1 \frac{\sigma^2}{\phi^2} h(k).$$

Theoretical guarantees

- If $\frac{\phi^2 \Delta}{\sigma^2} / h(k) \to \infty$, then $\widehat{\eta} \xrightarrow{p} \eta$.

In the case where e.g. $k = 1$, the estimation error becomes

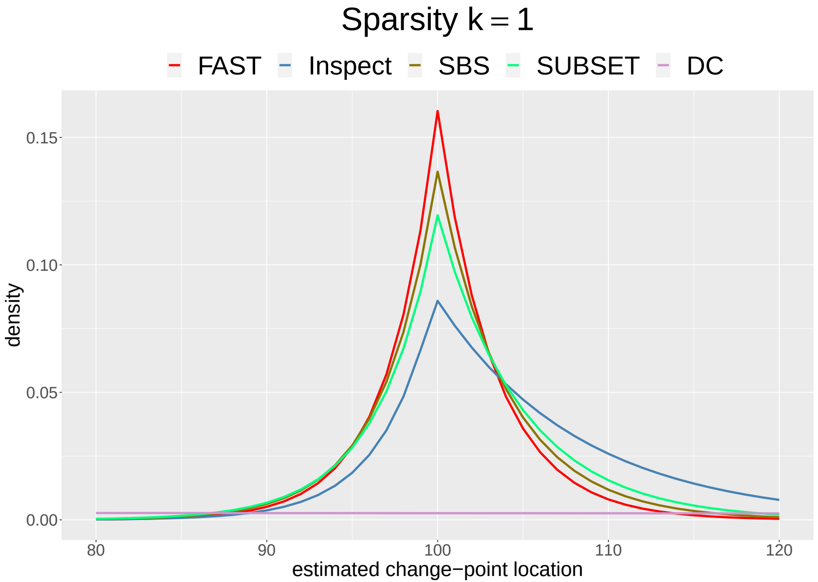
$$|\widehat{\eta} - \eta| \leq C_2 \frac{\sigma^2}{\phi^2} \left(\log n \vee \log p\right).$$

If also $p = \mathcal{O}(n)$, then

$$|\widehat{\eta} - \eta| = \mathcal{O}_p \left(\frac{\sigma^2}{\phi^2} \log n \right).$$

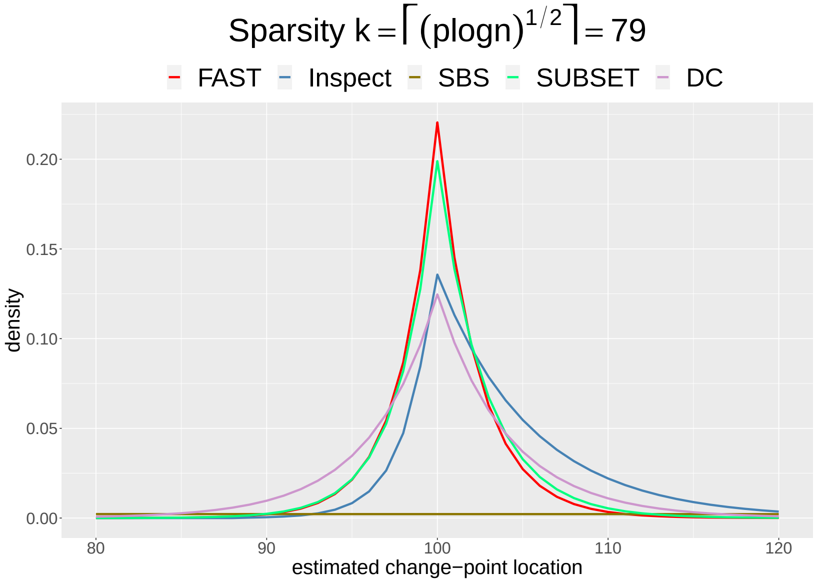
Some simulation results

For $n = 500$, $p = 1000$, $\eta = 100$:



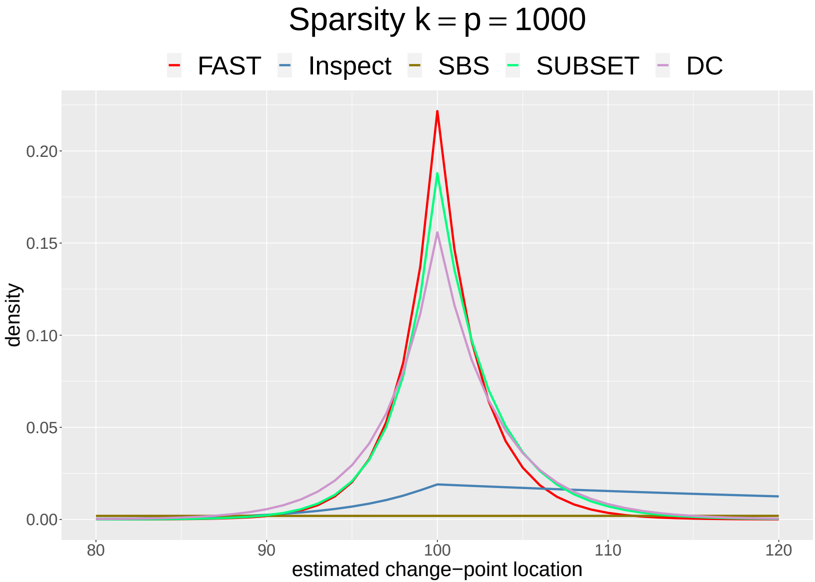
Some simulation results

For $n = 500$, $p = 1000$, $\eta = 100$:



Some simulation results

For $n = 500$, $p = 1000$, $\eta = 100$:



The full FAST algorithm

Building blocks for multiple change-point estimation:

(1) A search procedure

(2) A test statistic for a single change-point

Search procedure

For the search procedure, we use Seeded Binary Segmentation Kovács et al. (2020).

The idea: generate a (non-random) set of integer sub-intervals $[s_m, e_m]$ of $[1, n]$, varying in width and location.

Use a test statistic to iteratively test for a change-point in each of the sub-intervals, in the order of their width.

Once you detect a change-point in an interval, say (s_{m^*}, e_{m^*}), use the change-point estimator to determine the change-point position.

Split the data at the change-point location, and continue the search recursively on each resulting partition.

The test statistic

As test statistic, we use

$$T_\gamma = \max_i \max_{t \in \mathcal{T}} \mathbb{1}_{S_i^\gamma(t) > 0},$$

where

$$S_i^\gamma(t) := \sum_{j=1}^{p} \left(C_i(j)^2 - \nu_{a(t)} \right) \mathbb{1}\left\{ |C_i(j)| \geq a(t) \right\} - \gamma r(t)$$

Theoretical guarantees

For each change-point j, define
$\Delta_j := \min\{\eta_{j+1} - \eta_j, \eta_j - \eta_{j-1}\}$, $\phi_j := \|\mu_{\eta_j+1} - \mu_{\eta_j}\|_2$, and
$k_j := \|\mu_{\eta_j+1} - \mu_{\eta_j}\|_0$.

Theorem
Let $\widehat{J}$ and $\widehat{\eta}_1, \ldots, \widehat{\eta}_{\widehat{J}}$ respectively be estimated number and positions of the change-points from the FAST Algorithm.

There exist constants C_0, C_1 and universal values of λ, γ such that, if

$$\frac{\phi_j (k_j)^2 \Delta_j}{\sigma^2} \geq C_0 r(k_j)$$

holds for all j, then with probability at least $1 - \frac{1}{n}$ we have $\widehat{J} = J$ and

$$|\widehat{\eta}_j - \eta_j| \leq C_2 \frac{\sigma^2}{\phi_j^2} r(k_j)$$

for all j.

Optimality

As $r(t)$ is the detection boundary for multiple change-points, the SNR condition $\frac{\phi_j(k_j)^2\Delta_j}{\sigma^2} \geq C_0 r(k_j)$ is minimal.

The minimax error rate for estimating a single change-point is at least

$$\frac{\sigma^2}{\phi^2}$$

for challenging but feasible problems (see Wang and Samworth (2018)).

Hence FAST has an optimal estimation error up to a factor of $r(k_j)$.

Simulations

Table 1: Multiple changepoints

Parameters				Hausdorff distance					
n	p	Sparsity	J	FAST	SUBSET	Inspect	Pilliat	SBS	DC
500	1000	Dense	2	3.772	**2.384**	28.882	118.728	100.860	73.058
500	1000	Sparse	2	1.962	**1.418**	29.376	97.580	80.287	132.373
500	1000	Mixed	2	2.996	**1.993**	31.149	105.652	90.741	107.468
500	1000	Dense	5	**3.333**	4.148	27.261	89.856	90.271	62.559
500	1000	Sparse	5	**1.775**	3.769	21.906	73.911	71.842	102.847
500	1000	Mixed	5	**2.859**	3.973	23.265	86.692	86.810	81.487

Why the name FAST?

The computational complexity of FAST is
$\mathcal{O}(np \log n (\log p + \log \log n))$

Table 1: Running time for single change-point estimation

Parameters		Time in ms				
n	p	FAST	SUBSET	Inspect	SBS	DC
200	100	**0.297**	2.143	2.171	10.806	10.105
200	1000	**2.138**	13.206	43.256	90.525	91.289
200	5000	**13.038**	81.393	220.202	460.280	469.560
500	100	**0.605**	4.148	5.694	17.584	16.666
500	1000	**6.176**	34.919	265.959	152.294	168.964
500	5000	**39.190**	192.536	1372.398	786.896	939.474

References I

Cho, Haeran. 2016. "Change-Point Detection in Panel Data via Double CUSUM Statistic." *Electronic Journal of Statistics* 10 (2): 2000–2038.

Cho, Haeran, and Piotr Fryzlewicz. 2015. "Multiple-Change-Point Detection for High Dimensional Time Series via Sparsified Binary Segmentation." *Journal of the Royal Statistical Society. Series B* 77 (2): 475–507.

Kovács, Solt, Housen Li, Peter Bühlmann, and Axel Munk. 2020. "Seeded Binary Segmentation: A General Methodology for Fast and Optimal Change Point Detection." https://arxiv.org/abs/2002.06633v1.

Liu, Haoyang, Chao Gao, and Richard J. Samworth. 2021. "Minimax Rates in Sparse, High-Dimensional Change Point Detection." *The Annals of Statistics* 49 (2): 1081–1112.

References II

Pilliat, Emmanuel, Alexandra Carpentier, and Nicolas Verzelen. 2021. "Optimal Multiple Change-Point Detection for High-Dimensional Data." https://hal.archives-ouvertes.fr/hal-03004860.

Tickle, S. O., I. A. Eckley, and P. Fearnhead. 2021. "A Computationally Efficient, High-Dimensional Multiple Changepoint Procedure with Application to Global Terrorism Incidence." *Journal of the Royal Statistical Society: Series A* 184 (4): 1303–25.

Wang, Tengyao, and Richard J. Samworth. 2018. "High Dimensional Change Point Estimation via Sparse Projection." *Journal of the Royal Statistical Society: Series B* 80 (1): 57–83.