GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Statistical Blind Source Separation (with Applications in Cancer Genetics)

Merle Behr [1], Chris Holmes [3], and Axel Munk [1 2]

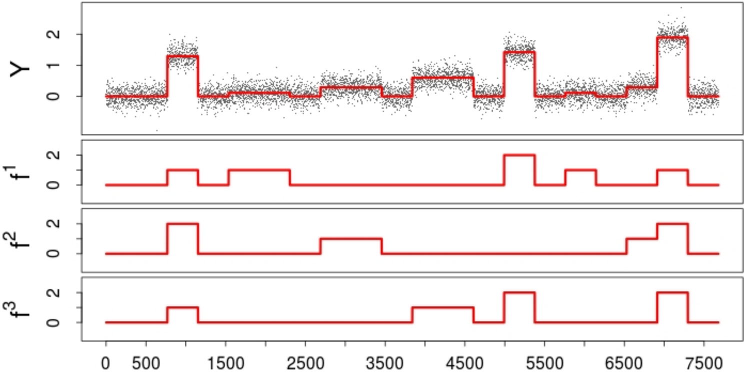
[1] Institute for Mathematical Stochastics, Göttingen,

[2] Max Planck Institute for Biophysical Chemistry, Göttingen,

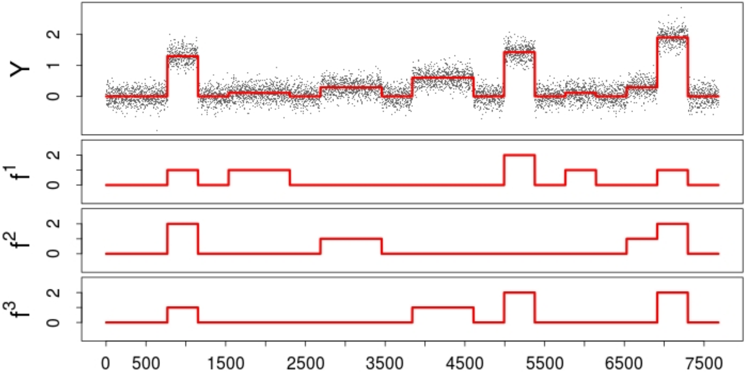
[3] University of Oxford, Department of Statistics

Mathematical Statistics Inverse Problems Week
C.I.R.M., February 8th, 2016

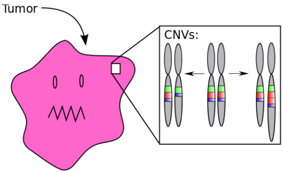
Blind Source Separation (BSS): Information transmission, cancer genetics, ...



Blind Source Separation (BSS): Information transmission, cancer genetics, ...



1. Introduction of the SBSSR Model.
2. Identifiability conditions.
3. Estimator, which SEparateS finite Alphabet MixturEs (SESAME) and yields confidence statements.
4. Applications and Simulations.



Linear mixtures of finite alphabet step functions

Blind source separation:

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$

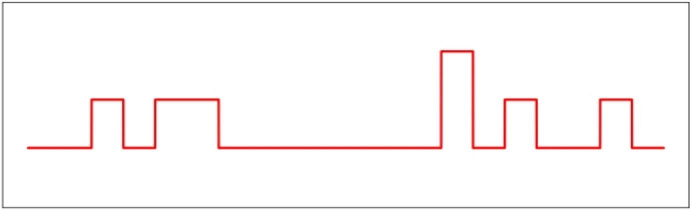
Linear mixtures of finite alphabet step functions

Blind source separation:

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$

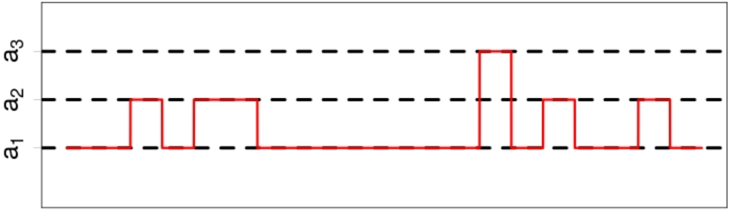
Linear mixtures of finite alphabet **step functions**

$$\underbrace{f^1, \dots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$



Linear mixtures of **finite alphabet** step functions

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$



$\mathfrak{A} := \{a_1, \ldots, a_k\}$ finite alphabet, known

Linear mixtures of finite alphabet step functions

$$\underbrace{f^1, \dots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$

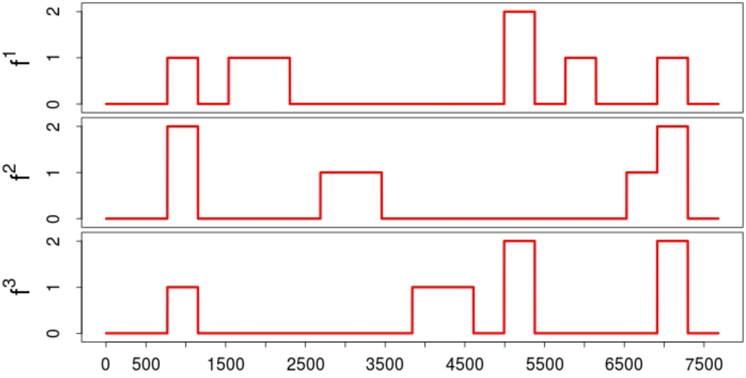
$$g = \sum_{i=1}^{m} \omega_i f^i, \quad \text{with} \quad \sum_{i=1}^{m} \omega_i = 1 \text{ and } 0 < \omega_i < 1.$$

Remark: Extensions to general ω possible (not shown).

Linear mixtures of finite alphabet step functions

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$

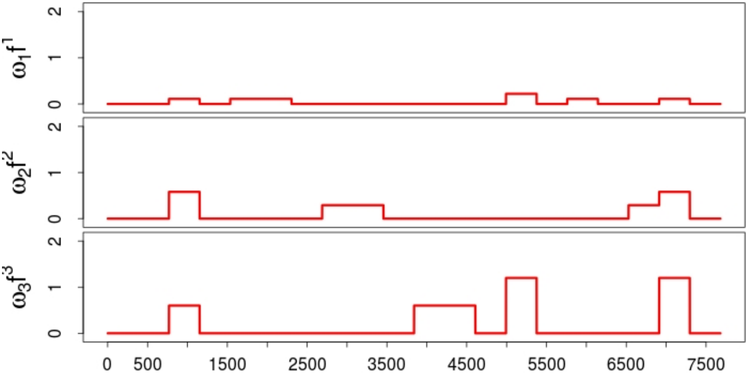
Example: $m = 3$, $\mathfrak{A} = \{0, 1, 2\}$



Linear mixtures of finite alphabet step functions

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \underbrace{g}_{\text{mixture}}$$

Example: $m = 3$, $\mathfrak{A} = \{0, 1, 2\}$, $\omega = (0.11, 0.29, 0.6)$

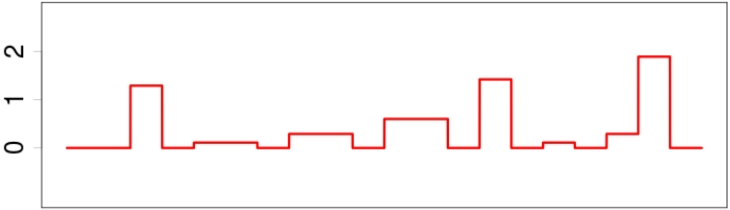


Linear mixtures of finite alphabet step functions

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}}$$

Example: $m = 3$, $\mathfrak{A} = \{0, 1, 2\}$, $\omega = (0.11, 0.29, 0.6)$

$$g = \omega_1 f^1 + \omega_2 f^2 + \omega_3 f^3$$

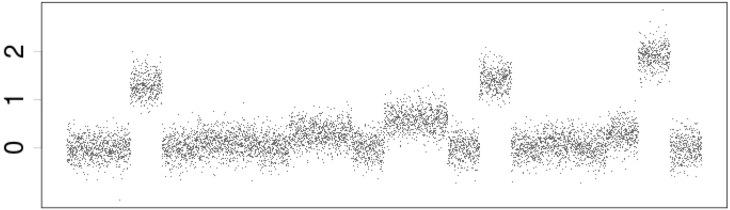


Linear mixtures of finite alphabet step functions

$$\underbrace{f^1, \ldots, f^m}_{\text{source functions}} \quad \underbrace{\xrightarrow{\omega}}_{\text{linear mixing}} \quad \underbrace{g}_{\text{mixture}} + \epsilon$$

Example: $m = 3$, $\mathfrak{A} = \{0, 1, 2\}$, $\omega = (0.11, 0.29, 0.6)$

$$Y = \omega_1 f^1 + \omega_2 f^2 + \omega_3 f^3 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

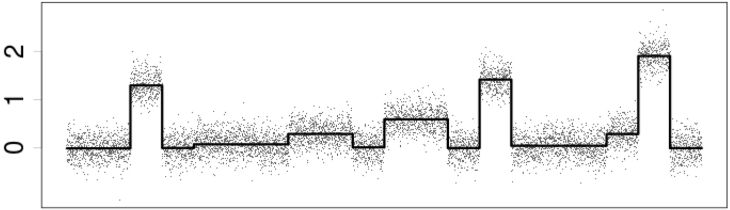


First Attempts

Idea 1. : Estimate the mixture and decompose afterwards.

- Small signal differences will be hard to recover
- Not every step function can be decomposed: alphabet-specific restrictions on function values! ↯

$$Y = \omega_1 f^1 + \omega_2 f^2 + \omega_3 f^3 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

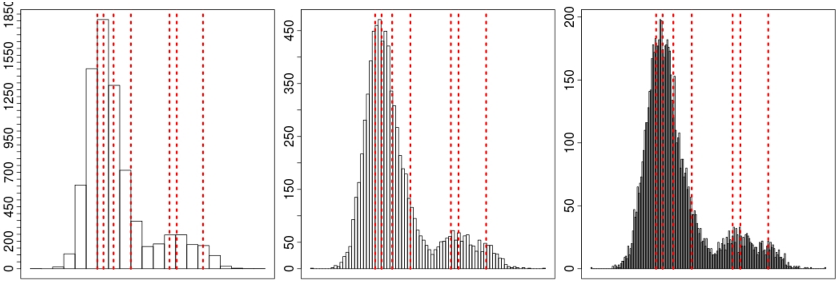


Naive clustering approach (has been advocated in SP literature...)

Idea 2. : Pre-estimate the mixture function values.

Clustering of (at most) k^m modes is known to be a hard problem!

$$Y = \omega_1 f^1 + \omega_2 f^2 + \omega_3 f^3 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

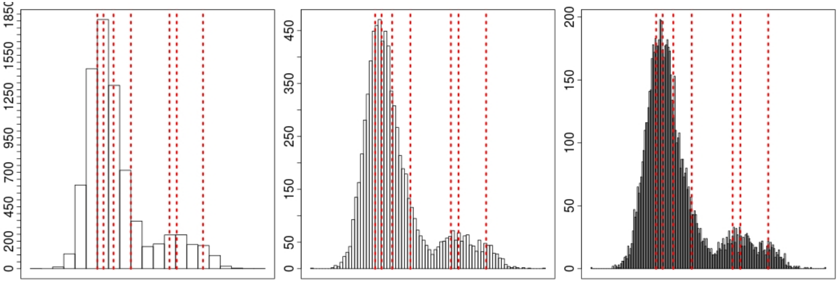


Naive clustering approach (has been advocated in SP literature...)

Idea 2. : Pre-estimate the mixture function values.

Clustering of (at most) k^m modes is known to be a hard problem!

$Y = \omega_1 f^1 + \omega_2 f^2 + \omega_3 f^3 + \epsilon, \qquad \epsilon \sim \mathcal{N}(0, \sigma^2 I)$



SESAME avoids preclustering
– recovers all quantities based on mixture model & finite alphabet
– simultaneous multiscale inference

Statistical Blind Source Separation Regression (SBSSR)

$$Y_j = \sum_{i=1}^{m} \omega_i f^i(x_j) + \epsilon_j, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I), \quad j = 1, \dots, n$$

$$\text{with } f^1, \dots, f^m \in \mathcal{S}(\mathfrak{A}) \text{ and } \omega \in \Omega(m), \quad x_j = j/n.$$

Mixing weights:

$\Omega(m) := \left\{\omega \in \mathbb{R}^m : 0 < \omega_1 < \dots < \omega_m \text{ and } \sum_{i=1}^m \omega_i = 1\right\}$

Finite alphabet step functions:

$\mathcal{S}(\mathfrak{A}) := \left\{ \sum_{i=0}^K \theta_i \mathbb{1}_{[\tau_i, \tau_{i+1})} : \theta_i \in \mathfrak{A}, 0 = \tau_0 < \dots < \tau_{K+1} = 1, K \in \mathbb{N} \right\}$

Known are

1. the alphabet $\mathfrak{A} = \{a_1, \dots, a_k\}$,
2. the number of source functions $m \in \mathbb{N}$, and
3. the (pre-estimated) standard deviation σ.

Statistical Blind Source Separation Regression (SBSSR)

$$Y_j = \sum_{i=1}^{m} \omega_i f^i(x_j) + \epsilon_j, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I), \quad j = 1, \ldots, n$$

$$\text{with } f^1, \ldots, f^m \in \mathcal{S}(\mathfrak{A}) \text{ and } \omega \in \Omega(m), \quad x_j = j/n.$$

Mixing weights:

$\Omega(m) := \left\{\omega \in \mathbb{R}^m : 0 < \omega_1 < \ldots < \omega_m \text{ and } \sum_{i=1}^m \omega_i = 1\right\}$

Finite alphabet step functions:

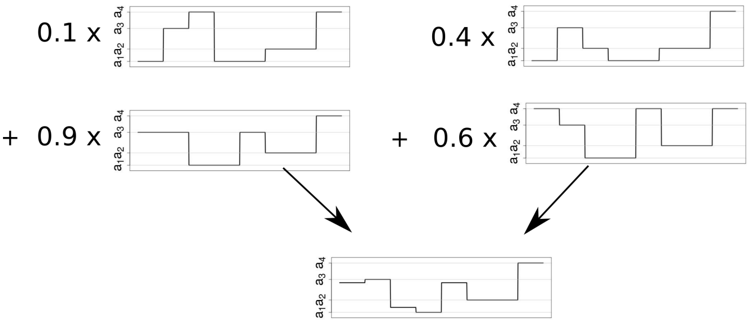
$\mathcal{S}(\mathfrak{A}) := \left\{ \sum_{i=0}^{K} \theta_i \mathbb{1}_{[\tau_i, \tau_{i+1})} : \theta_i \in \mathfrak{A}, 0 = \tau_0 < \ldots < \tau_{K+1} = 1, K \in \mathbb{N} \right\}$

Unknowns are

1. the mixing weights $\omega = (\omega_1, \ldots, \omega_m)$ and
2. the source functions $f^1, \ldots, f^m$, i.e.
   - 2.1 their number of change-points K^i,
   - 2.2 their change-point locations $\tau_1, \ldots, \tau_K$, and
   - 2.3 their function values ($\theta \in \mathfrak{A}$).

A non-identifiable mixture

For $m = 2$ and $\mathfrak{A} = \{a_1, a_2, a_3, a_4\} = \{10, 13.75, 20, 25\}$:



$\Rightarrow$ Identifiability is a necessary assumption for valid signal recovery in the SBSSR model!

Identifiability (for given $\mathfrak{A}$ and $m \in \mathbb{N}$)

g is identifiable $\Leftrightarrow$ $\exists!\ (\omega, f) \in \Omega(m) \times \mathcal{S}(\mathfrak{A})^m$ s.t. $g = \omega^\top f$.

The following two conditions ensure identifiability: [1]

1. **Alphabet separation boundary** for ω **(ASB)**

&

2. **Variability of sources** f **(VS)**

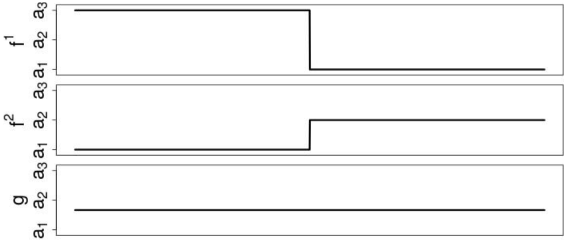
$\Rightarrow$ **Identifiability**

[1] [Diamantaras, 2006, Behr and Munk, 2015]

Necessary identifiability condition

1. (Necessary) condition on the weights ω for the source functions $f^1, \ldots, f^m$ to be identifiable:

Example: $m = 2$, $\mathfrak{A} = \{10, 20, 30\}$, *and* $(\omega_1, \omega_2) = (\frac{1}{3}, \frac{2}{3})$:



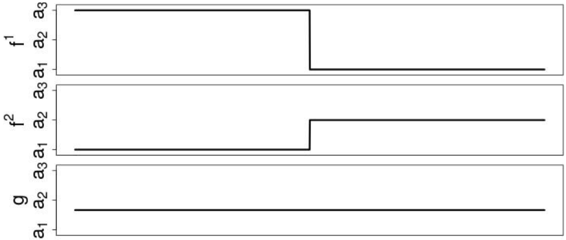
Necessary identifiability condition

1. (Necessary) condition on the weights ω for the source functions $f^1, \ldots, f^m$ to be identifiable:

Finite alphabet separation boundary:

$$0 < \delta := \min_{a \neq a' \in \mathfrak{A}^m} \left| \omega^\top a - \omega^\top a' \right| \qquad \textbf{(ASB)}$$

Example: $m = 2$, $\mathfrak{A} = \{10, 20, 30\}$, *and* $(\omega_1, \omega_2) = (\frac{1}{3}, \frac{2}{3})$:



Sufficient identifiability condition

2. (Sufficient) condition on the source functions $f^1, \ldots, f^m$ for the weights ω to be identifiable:

Example: *If* $f^1 = \ldots = f^m$*, then* $g = \sum_{i=1}^m \omega_i f^i = f^1$ *irrespective of* $\omega \Rightarrow f^1, \ldots, f^m$ ***must differ sufficiently much.***

Sufficient identifiability condition

2. (Sufficient) condition on the source functions $f^1, \ldots, f^m$ for the weights ω to be identifiable:

Variability of source function:
For all $r = 1, \ldots, m$ exists a sampling point x_r such that

$$(f^1(x_r), \ldots, f^m(x_r)) = (a_1, \ldots, a_1, \underbrace{a_2}_{r\text{th position}}, a_1, \ldots, a_1) \qquad \textbf{(VS)}$$

Example: *If* $f^1 = \ldots = f^m$, *then* $g = \sum_{i=1}^m \omega_i f^i = f^1$ *irrespective of* $\omega \Rightarrow f^1, \ldots, f^m$ ***must differ sufficiently much.***

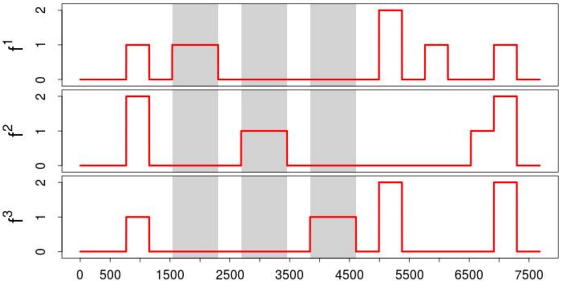
Sufficient identifiability condition

2. (Sufficient) condition on the source functions $f^1, \ldots, f^m$ for the weights ω to be identifiable:

Variability of source function:

For all $r = 1, \ldots, m$ exists a sampling point x_r such that

$$(f^1(x_r), \ldots, f^m(x_r)) = (a_1, \ldots, a_1, \underbrace{a_2}_{r\text{th position}}, a_1, \ldots, a_1) \qquad \textbf{(VS)}$$



Stable recovery of weights and sources

Let $g = \omega^\top f, \tilde{g} = \tilde{\omega}^\top \tilde{f}$ be two mixtures, both satisfying the identifiability conditions 1. and 2. (for the same ASB δ). Let ϵ be such that $0 < \epsilon < \delta(a_2 - a_1)/(2m(a_k - a_1))$. If

$$\sup_{x \in [0,1)} |g(x) - \tilde{g}(x)| < \epsilon,$$

1. then the weights satisfy the **stable approximate recovery (SAR)** property $\max_{i=1,\ldots,m} |\omega_i - \tilde{\omega}_i| < \epsilon/(a_2 - a_1)$ and
2. the sources satisfy the **stable exact recovery (SER)** property $f = \tilde{f}$.

SESAME (SEparateS finite Alphabet MixturEs)

1. Construct a confidence region $\mathcal{C}_{1-\alpha}$ for the mixing weights ω *(characterized by acceptance region of a multiscale test)*, $\rightarrow$ with diameter $\ln(n)/\sqrt{n}$.
2. Estimate $\hat{\omega} \in \mathcal{C}_{1-\alpha}(Y)$.
3. Estimate $\hat{f}^1, \ldots, \hat{f}^m$ as a constrained maximum likelihood estimator *(With the same multiscale constraint as for $\mathcal{C}_{1-\alpha}$ but with a possibly different level β)*
4. This yields asymptotically uniform multivariate (honest) confidence bands $\mathcal{H}(\beta)$ for $f^1, \ldots, f^m$.

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$

$$\omega = (0.11, 0.29, 0.6) \quad \Rightarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \rightarrow \omega_1 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \rightarrow \omega_1 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

Example: $m = 3$, $\mathfrak{A} = \{0, 1\}$:

$$\omega = (0.11, 0.29, 0.6) \quad \Leftarrow \quad \begin{aligned} \omega_1 0 + \omega_2 0 + \omega_3 0 &= 0 \\ \omega_1 1 + \omega_2 0 + \omega_3 0 &= 0.11 \rightarrow \omega_1 \\ \omega_1 0 + \omega_2 1 + \omega_3 0 &= 0.29 \rightarrow \omega_2 \\ \omega_1 1 + \omega_2 1 + \omega_3 0 &= 0.4 \\ \omega_1 0 + \omega_2 0 + \omega_3 1 &= 0.6 \\ \omega_1 1 + \omega_2 0 + \omega_3 1 &= 0.71 \\ \omega_1 0 + \omega_2 1 + \omega_3 1 &= 0.89 \\ \omega_1 1 + \omega_2 1 + \omega_3 1 &= 1 \end{aligned}$$

Inferring the mixing weights ω

For ω and f satisfying the identifiability conditions (ASB) and (VS):

> The mixing weights are in one-to-one correspondence to the mixture function values.

$\Longrightarrow$ exact recovery algorithm for ω, $O(k^m)$.

- Recovery of the weights relies on a good estimate of the function values of $g = \omega^\top f$ taking into account the specific structure of underlying step functions $f^i \in \mathcal{S}(\mathfrak{A})$.
- Estimation of f is not required.
- From this one can construct a honest confidence set for ω

Multiscale statistic

As the jump locations may occur at any place, a natural way for inferring the function values of g is to use local log-likelihood ratio test statistics in a multiscale fashion[2]. For the local test problem

$$H_0 : g|_{[x_i,x_j]} \equiv g_{ij} \quad \text{vs.} \quad H_1 : g|_{[x_i,x_j]} \not\equiv g_{ij}$$

we employ the test statistic

$$T_i^j(Y_i,\ldots,Y_j,g_{ij}) = \frac{(\sum_{l=i}^{j} Y_l - g_{ij})^2}{\sigma^2(j-i+1)},$$

in a multiscale fashion

$$T_n(Y,\tilde{g}) := \max_{\substack{1\le i\le j\le n \\ \tilde{g}|_{[i/n,j/n]}\equiv\tilde{g}_{ij}}} \frac{|\sum_{l=i}^{j} Y_l - \tilde{g}_{ij}|}{\sigma\sqrt{j-i+1}} - \sqrt{2\ln\left(\frac{en}{j-i+1}\right)}.$$

---

[2] [Siegmund and Yakir, 2000, Dümbgen and Spokoiny, 2001, Davies and Kovac, 2001, Dümbgen and Walther, 2008]

Geometric interpretation of the statistic T_n

$$T_n(Y,\tilde{g}) \leq q \;\Leftrightarrow \tilde{g}_{ij} \in B(i,j)\; \forall 1 \leq i \leq j \leq n \text{ with } \tilde{g}|_{[i/n,j/n]} \equiv \tilde{g}_{ij},$$

for $q \in \mathbb{R}$, with intervals

$$B(i,j) := \left[\overline{Y}_i^j - \frac{q + pen(j-i+1)}{\sqrt{j-i+1}/\sigma}, \overline{Y}_i^j + \frac{q + pen(j-i+1)}{\sqrt{j-i+1}/\sigma}\right].$$

From simulations one obtains $q_n(\alpha)$, $\alpha \in (0,1)$, the $1-\alpha$ quantile of $T_n = T_n(Y,0)$, i.e.,

$$\inf_g \mathbf{P}(T_n(Y,g) \leq q_n(\alpha)) \geq 1-\alpha.$$

Hence, for $B(i,j)$ with $q = q_n(\alpha)$,

$$\inf_g \mathbf{P}(g_{ij} \in B(i,j)\; \forall 1 \leq i \leq j \leq n \text{ with } g|_{[i/n,j/n]} \equiv g_{ij}) \geq 1-\alpha.$$

Confidence boxes

Let $\mathfrak{B} = \{B(i,j) : 1 \le i \le j \le n\}$ with $q = q_n(\alpha)$ and assume $B^\star := B(i_1^\star, j_1^\star) \times \ldots \times B(i_m^\star, j_m^\star) \in \mathfrak{B}^m$ has been constructed, such that

$$f|_{[i_r^\star, j_r^\star]} \equiv [A]_r, \tag{1}$$

with A as in **(VS)**. Then

$$\{\omega \in A^{-1}B^\star\} \supset \bigcap_{1 \le r \le m} \{g|_{[i_r^\star j_r^\star]} \equiv \omega^\top [A]_r \in B(i_r^\star, j_r^\star)\}$$

and

$$\{T_n(Y,g) \le q_n(\alpha)\} = \bigcap_{\substack{1 \le i \le j \le n \\ g|_{[i/n, j/n]} \equiv g_{ij}}} \{g_{ij} \in B(i,j)\}$$

which implies

$$\{\omega \in A^{-1}B^\star\} \supset \{T_n(Y,g) \le q_n(\alpha)\}$$

Confidence boxes

One cannot obtain $B^\star$ directly as $f^1, \dots, f^m$ are unknown. ↯

$\Rightarrow$ Construct $\mathfrak{B}^\star \subset \mathfrak{B}^m$, with $\mathbf{P}\left(B^\star \in \mathfrak{B}^\star \middle| T_n \leq q_n(\alpha)\right) = 1$ and define

$$\boxed{\mathcal{C}_{1-\alpha} := \bigcup_{B \in \mathfrak{B}^\star} A^{-1} B.}$$

$$\begin{aligned}
&\mathbf{P}\left(\omega \in \mathcal{C}_{1-\alpha}\right) \\
=&\mathbf{P}\left(\omega \in \mathcal{C}_{1-\alpha} \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
=&\mathbf{P}\left(\omega \in \bigcup_{B \in \mathfrak{B}^\star} A^{-1} B \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
\geq&\mathbf{P}\left(\omega \in A^{-1} B^\star \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
\geq&1 - \alpha.
\end{aligned}$$

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 1. Delete $B \in \mathfrak{B}^m$ if there exists an $r \in \{1, \ldots, m\}$, s.t. $\text{proj}_r(B) \in$

$$\{B(i,j) \in \mathfrak{B} : \exists [s,t],[u,v] \subset [i,j] \text{ with } B(s,t) \cap B(u,v) = \emptyset\}.$$

All boxes, s.t. $\tilde{g} \in \mathcal{M}$ satisfies MS constraint, cannot be constant on $[i,j]$

$\rightarrow$ Exploring the fact that $f = (f_1, \ldots, f^m)^\top$ is constant on $[i_r^\star, j_r^\star]$, with $B^\star := B(i_1^\star, j_1^\star) \times \ldots \times B(i_m^\star, j_m^\star)$, conditioned on $\{T_n \leq q_n(\alpha)\}$.

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 2. Delete $B \in \mathfrak{B}^m$, with $[\underline{b}_r, \overline{b}_r] := \text{proj}_r(B)$,

1. for any $2 \leq r \leq m$

$$\frac{a_2 + (m-1)a_1 - \sum_{k=1}^{r-1} \underline{b}_k}{m - r + 1} \leq \underline{b}_r \quad \text{or} \quad \underline{b}_{r-1} \geq \overline{b}_r, or$$

2. ...

→ Exploring the structure of $\Omega(m)$, e.g., $\omega_{i-1} < \omega_i < (1 - \sum_{j=1}^{i-1} \omega_j)/(m - i + 1)$, ..., together with the specific choice of the matrix A in **(VS)**.

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 3. Delete $B \in \mathfrak{B}^m$, if there exists a $k \in \{1, \dots, n\}$ such that for all $[i,j] \in \{[i,j] : k \in [i,j] \text{ and } B(i,j) \notin \mathfrak{B}_{\text{nc}}\}$

$$\left[\max_{i \le u \le v \le j} \underline{b}_{uv}, \min_{i \le u \le v \le j} \overline{b}_{uv}\right] \cap \left\{\tilde{\omega}^\top a : a \in \mathfrak{A}^m \text{ and } \tilde{\omega} \in A^{-1}B\right\}$$

is empty, with $B(u,v) = [\underline{b}_{uv}, \overline{b}_{uv}] \in \mathfrak{B}$.

$\rightarrow$ Exploring the fact that $g = \omega^\top f$ maps to $\{\tilde{\omega}^\top a : a \in \mathfrak{A}^m \text{ and } \tilde{\omega} \in A^{-1}B^\star\}$ conditioned on $\{T_n \le q_n(\alpha)\}$.

Diameter of $\mathcal{C}_{1-\alpha}$

Assume a minimal scale for jumps $\lambda > 0$ and the identifiability conditions (ASB) and (VS), then

$$\mathbf{P}\left(\operatorname{dist}(\omega, C_{1-\alpha_n}(Y)) < \frac{c_2}{a_2 - a_1}\frac{\ln(n)}{\sqrt{n}}\right) \geq 1 - \exp(-c_1 \ln^2(n))$$

for all $n \geq N^\star$ and $\alpha_n := \exp(-c_1 \ln^2(n))$, with some constants $c_1 = c_1(\lambda, \delta)$, $c_2 = c_2(\lambda, \delta)$ and some explicit $N^\star \in \mathbb{N}$, with $\ln(N^\star) > c(\mathfrak{A}, m)\sigma^2/(\lambda\delta^2)$, where $\operatorname{dist}(d, D) := \sup_{\tilde{d}\in D} \|d - \tilde{d}\|_\infty$.

Estimating the mixing weights

SESAME estimates ω by

$$\hat{\omega} := \frac{1}{\sum_{i=1}^{m}(\underline{\omega}_i + \overline{\omega}_i)}(\underline{\omega}_1 + \overline{\omega}_1, \ldots, \underline{\omega}_m + \overline{\omega}_m),$$

with $\mathcal{C}_{1-\alpha} =: [\underline{\omega}_1, \overline{\omega}_1] \times \ldots \times [\underline{\omega}_m, \overline{\omega}_m]$.

$\rightarrow \alpha$ can be seen as tuning parameter for $\hat{\omega}$.

$\Rightarrow$ Data driven selection method (MVT- and SST-method [3]).

[3] [Behr et al., 2015]

Inferring the source functions $f^1, \ldots, f^m$

For $\hat{\omega} \in \mathcal{C}_{1-\alpha}(Y)$ we estimate $f^1, \ldots, f^m$ with a **constrained maximum likelihood estimator**[4]:

$$(\hat{f}^1, \ldots, \hat{f}^m) := \text{argmax}_{f \in \mathcal{H}(\beta)} L_Y(f),$$

with L being the likelihood function and

$$\mathcal{H}(\beta) := \{f \in \mathcal{S}(\mathfrak{A})^m : T_n\left(Y, \hat{\omega}^\top f\right) \leq q_n(\beta) \text{ and } K\left(\hat{\omega}^\top f\right) = \hat{K}\}.$$

and

$$\hat{K} := \inf_{f \in \mathcal{S}(\mathfrak{A})^m} K(\hat{\omega}^\top f) \quad \text{s.t.} \quad T_n(Y, \hat{\omega}^\top f) \leq q_n(\beta).$$

($K(g)$ denotes the number of change-points of g)

[4] [Frick et al., 2014]

Exact recovery of the source functions $f^1, \ldots, f^m$

Assuming a minimal scale for jumps $\lambda > 0$, the identifiability condition (ASB) and (VS) and choosing

$$\alpha_n, \beta_n = \exp(-C_* \ln^2(n)),$$

then for $n \geq N^\star$ with probability at least $1 - \alpha_n$ the estimator $\hat{f}^1, \ldots, \hat{f}^m$

1. estimates the number of change-points of f^i correctly for $i = 1, \ldots, m$,
2. estimates the change-point locations with rate $\frac{\ln^2(n)}{n}$, and
3. estimates the function values of $f^1, \ldots, f^m$ exactly (up to the uncertainty in the change point location).

Confidence bands

Let $\tilde{T}_n$ be as T_n, but with penalty term increased by $\left(\frac{(a_2 - a_1)\ln(n)}{m} + \sqrt{\frac{8\sigma^2 \ln(e/\lambda)}{\lambda}}\right)\sqrt{\frac{j-i+1}{n}}$, and let $\tilde{\mathcal{H}}$ be as $\mathcal{H}$ but with T_n replaced by $\tilde{T}_n$.

Assume the identifiability conditions (ASB) and (VS), then for $\hat{\omega} = \hat{\omega}(\alpha_n)$ in $\tilde{\mathcal{H}}(\beta)$

$$\lim_{n \longrightarrow \infty} \inf_{g} \mathbf{P}((f^1, ..., f^m) \in \tilde{\mathcal{H}}(\beta)) \geq 1 - \beta.$$

SESAME's rates of convergence

1. SESAME recovers the **change point locations** of f^i in probability with rate $\ln^2(n)/n$.
   $\rightarrow$ Estimation rate is bounded from below by the sampling rate $1/n \Rightarrow$ optimal rate up to a $\ln^2(n)$ factor.

[5] [Dümbgen and Walther, 2008, Frick et al., 2014]

SESAME's rates of convergence

1. SESAME recovers the **change point locations** of f^i in probability with rate $\ln^2(n)/n$.
   $\rightarrow$ Estimation rate is bounded from below by the sampling rate $1/n \Rightarrow$ optimal rate up to a $\ln^2(n)$ factor.
2. The **minimal scale** λ may depend on n. If $\lambda_n^{-1} \in o(\ln(n))$ SESAME's estimates remain consistent.
   $\rightarrow$ No method can recover finer details of the mixture g below its detection boundary which is of the same order[5].

[5] [Dümbgen and Walther, 2008, Frick et al., 2014]

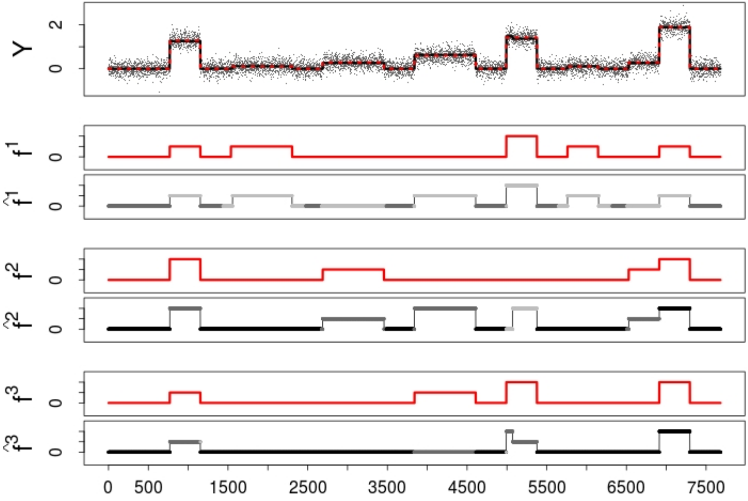
SESAME's rates of convergence

1. SESAME recovers the **change point locations** of f^i in probability with rate $\ln^2(n)/n$.
   $\rightarrow$ Estimation rate is bounded from below by the sampling rate $1/n \Rightarrow$ optimal rate up to a $\ln^2(n)$ factor.
2. The **minimal scale** λ may depend on n. If $\lambda_n^{-1} \in o(\ln(n))$ SESAME's estimates remain consistent.
   $\rightarrow$ No method can recover finer details of the mixture g below its detection boundary which is of the same order[5].
3. The **weights' estimation rate** $\ln(n)/\sqrt{n}$, arises from the box height with $q_n(\alpha_n) \in \mathcal{O}(\ln(n))$ and attains the optimal rate $\mathcal{O}(1/\sqrt{n})$ up to a $\ln(n)$ term.

[5] [Dümbgen and Walther, 2008, Frick et al., 2014]

Example

$m = 3$, $\mathfrak{A} = \{0,1,2\}$, $\sigma = 0.22$, and $n = 7680$, with $\omega = (0.11, 0.29, 0.6)$. We estimated $\hat{\omega} = (0.11, 0.26, 0.63)$ (with $\mathcal{C}_{0.9} = [0.00, 0.33] \times [0.07, 0.41] \times [0.39, 0.71]$).



Color code: deviation with confidence$\{\mathbf{0}, 1, 2\}$

Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.

Violation of identifiability

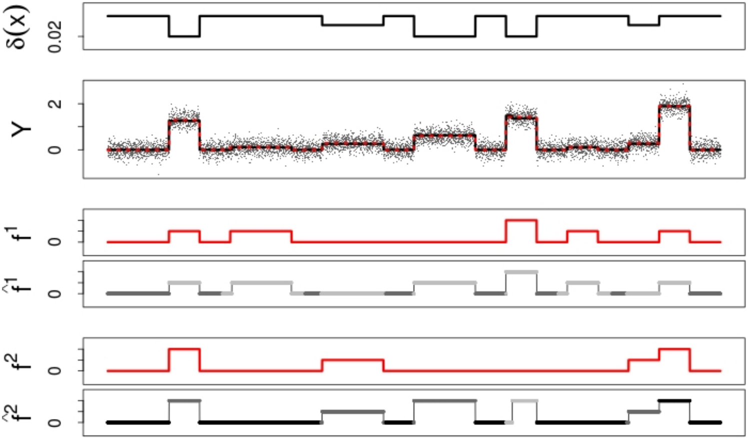
1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \dots, \hat{f}^m$, but uncertainty is captured in confidence bands.

Local finite alphabet separation boundary:

$$0 < \delta(x) := \min_{a \neq f(x) \in \mathfrak{A}^m} \left| \omega^\top a - \omega^\top f(x) \right|$$

Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.



Violation of identifiability

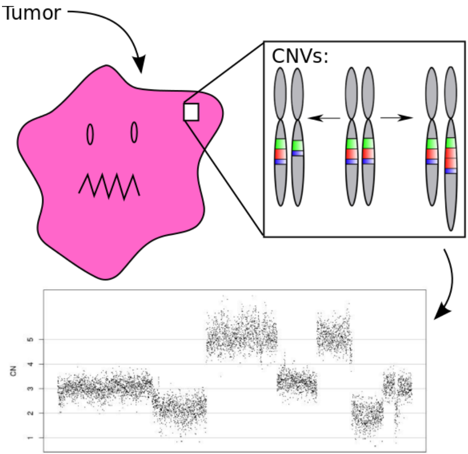
1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.

2. (VS) condition violated, i.e., too little variation of $f^1, ..., f^m$:
   2.1 Big influence on $\hat{\omega}$.
   2.2 Big influence on $\hat{f}^1, ..., \hat{f}^m$ as estimate is based on $\hat{\omega}$.

   $\longrightarrow$ Simulation study (Behr et al.'15)

When f comes from a Markov chain, probability that variation is rich enough converges exponentially fast to 1.

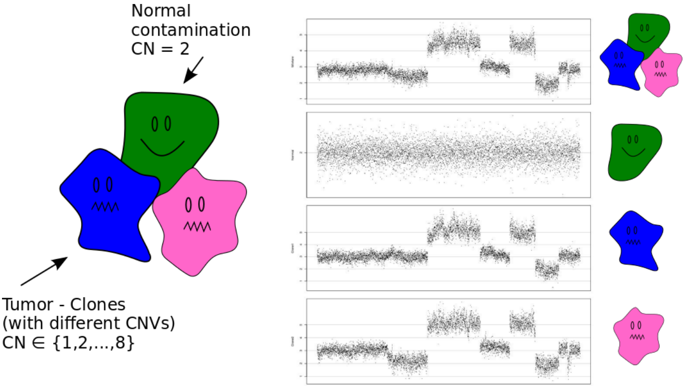
Inferring intra-tumor heterogeneity [6]



CNVs := Copy-number variations

[6] [Beroukhim et al., 2010, Greaves and Maley, 2012, Shah et al., 2012]

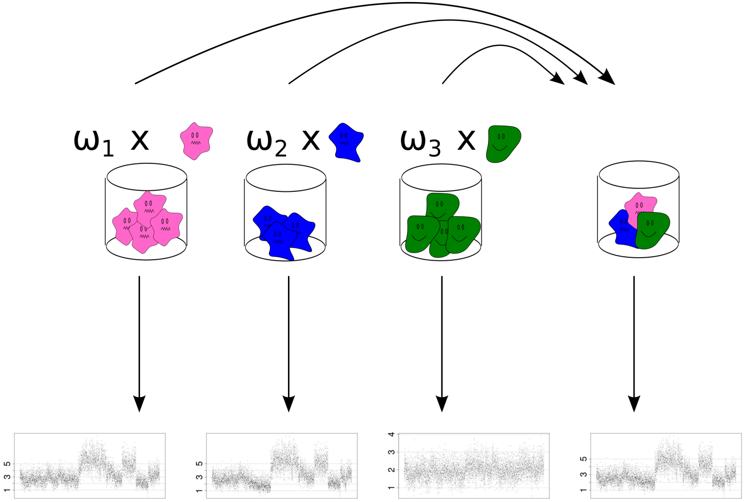
Inferring intra-tumor heterogeneity [6]



$$f^1, \ldots, f^m \sim \text{ CNVs of tumor-clones / normal contamination.}$$
$$\omega \sim \text{ proportion of the clone in the tumor.}$$

[6] [Beroukhim et al., 2010, Greaves and Maley, 2012, Shah et al., 2012]

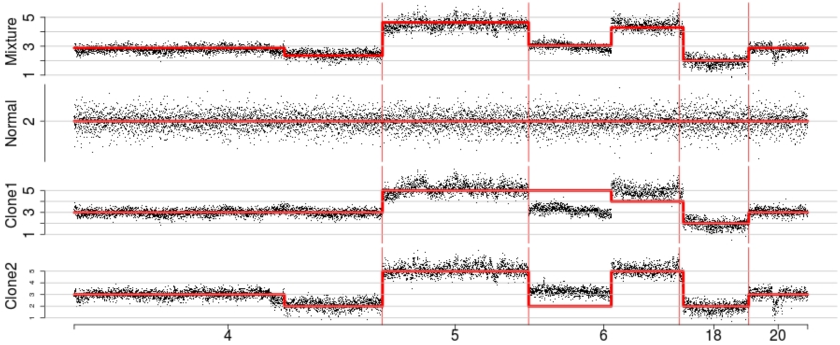
Generating test data for CNV characterization [7]



[7] Sequencing was done through a collaboration of Complete Genomics with the Welcome Trust Center for Human Genetics at the University of Oxford.

Characterizing CNVs in tumors

For $(\omega_{\text{Normal}}, \omega_{\text{Clone 1}}, \omega_{\text{Clone 2}}) = (0.2, 0.35, 0.45)$ SESAME estimated $(\hat{\omega}_{\text{Normal}}, \hat{\omega}_{\text{Clone 1}}, \hat{\omega}_{\text{Clone 2}}) = (0.12, 0.35, 0.53)$.



Summary

1. Statistical Blind Source Separation Regression (SBSSR) model [1].
2. Complete (not shown) characterization of identifiability[2].
3. SESAME:
    - Optimal estimators (up to log-factors) for the mixing weights and the source functions under very weak identifiability conditions.
    - Honest confidence statements[1] for all quantities.
    - Algorithms[1] for efficient computations (DP based, not shown).

---

[1] Behr, M., Holmes, C., and Munk, A., Multiscale blind source separation, prepint 2015

[2] Behr, M. and Munk, A. (2015). Identifiability for blind separation of multiple finite alphabet linear mixtures, arXiv:1505.05272.

Discussion

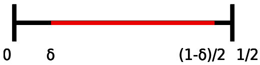
- 'Objective' parameter choice for α, β.
  Data driven choices targeting minimizing risk possible (not shown)
- Simulations studies show stability in choice of confidence parameters α and β and reasonable robustness against normality and heteroscedasticity[1].

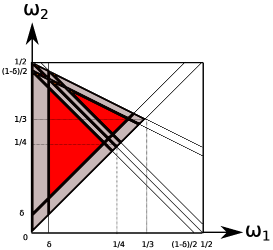
Discussion

- How big is the set $ASB(\omega) \geq \delta$?

Finite alphabet separation boundary:

$$0 < \delta := \min_{a \neq a' \in \mathfrak{A}^m} \left| \omega^\top a - \omega^\top a' \right| \qquad \textbf{(ASB)}$$





We can show:
$P(\omega \text{ identifiable}) = 1 - O(\delta)$

Discussion/Outlook

- linear model

$$Y = F\omega + \epsilon, \quad F = (f^i(x_j))_{1 \le i \le m, 1 \le j \le n}$$

  compressive sensing: F known, ω sparse
  matrix completion: here we sample from one linear functional (mixture), no low rank assumption, rather large rank is beneficial $\rightarrow$ identifiablity, finite alphabet is crucial
- nonnegative matrix factorization $F, \omega \geq 0$, (Donoho/Stodden'03) simpliciality condition $\leftrightarrow$ ASB-condition, $M > 1$, finite alphabet is crucial again.
- Open issue: unknown m (number of mixture components), unknown alphabet

Behr, M., Holmes, C., and Munk, A. (2015).
Blind multiscale demixing of step functions.
In preparation.

Behr, M. and Munk, A. (2015).
Identifiability for blind source separation of multiple finite alphabet linear mixtures.
arXiv preprint arXiv:1505.05272.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J. S., Dobson, J., Urashima, M., et al. (2010).
The landscape of somatic copy-number alteration across human cancers.
Nature, 463(7283):899–905.

Davies, P. and Kovac, A. (2001).
Local extremes, runs, strings and multiresolution.
Annals of Statistics, 29(1):1–65.

Diamantaras, K. I. (2006).
A clustering approach for the blind separation of multiple finite alphabet sequences from a single linear mixture.
Signal Processing, 86(4):877–891.

Dümbgen, L. and Spokoiny, V. (2001).
Multiscale testing of qualitative hypotheses.
The Annals of Statistics, 29(1):124–152.

Dümbgen, L. and Walther, G. (2008).
Multiscale inference about a density.
The Annals of Statistics, 36(4):1758–1785.

Frick, K., Munk, A., and Sieling, H. (2014).
Multiscale change point inference.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(3):495–580.
With discussion and rejoinder.

Greaves, M. and Maley, C. C. (2012).
Clonal evolution in cancer.
Nature, 481(7381):306–313.

Shah, S. P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012).
The clonal and mutational evolution spectrum of primary triple-negative breast cancers.
Nature, 486(7403):395–399.

Siegmund, D. and Yakir, B. (2000).
Tail probabilities for the null distribution of scanning statistics.
Bernoulli, 6(2):191.

Technical Material

Multiscale statistic

As the jump locations may occur at any place, a natural way for inferring the function values of g is to use local log-likelihood ratio test statistics in a multiscale fashion[8]. For the local test problem

$$H_0 : g|_{[x_i,x_j]} \equiv g_{ij} \quad \text{vs.} \quad H_1 : g|_{[x_i,x_j]} \not\equiv g_{ij}$$

we employ the test statistic

$$T_i^j(Y_i,\ldots,Y_j,g_{ij}) = \frac{(\sum_{l=i}^{j} Y_l - g_{ij})^2}{\sigma^2(j-i+1)},$$

in a multiscale fashion

$$T_n(Y,\tilde{g}) := \max_{\substack{1\le i\le j\le n \\ \tilde{g}|_{[i/n,j/n]}\equiv\tilde{g}_{ij}}} \frac{|\sum_{l=i}^{j} Y_l - \tilde{g}_{ij}|}{\sigma\sqrt{j-i+1}} - \sqrt{2\ln\left(\frac{en}{j-i+1}\right)}.$$

---

[8] [Siegmund and Yakir, 2000, Dümbgen and Spokoiny, 2001, Davies and Kovac, 2001, Dümbgen and Walther, 2008]

Geometric interpretation of the statistic T_n

$$T_n(Y,\tilde{g}) \leq q \;\Leftrightarrow \tilde{g}_{ij} \in B(i,j)\; \forall 1 \leq i \leq j \leq n \text{ with } \tilde{g}|_{[i/n,j/n]} \equiv \tilde{g}_{ij},$$

for $q \in \mathbb{R}$, with intervals

$$B(i,j) := \left[\overline{Y}_i^j - \frac{q + pen(j-i+1)}{\sqrt{j-i+1}/\sigma}, \overline{Y}_i^j + \frac{q + pen(j-i+1)}{\sqrt{j-i+1}/\sigma}\right].$$

From simulations one obtains $q_n(\alpha)$, $\alpha \in (0,1)$, the $1-\alpha$ quantile of $T_n = T_n(Y,0)$, i.e.,

$$\inf_g \mathbf{P}(T_n(Y,g) \leq q_n(\alpha)) \geq 1-\alpha.$$

Hence, for $B(i,j)$ with $q = q_n(\alpha)$,

$$\inf_g \mathbf{P}(g_{ij} \in B(i,j)\; \forall 1 \leq i \leq j \leq n \text{ with } g|_{[i/n,j/n]} \equiv g_{ij}) \geq 1-\alpha.$$

Confidence boxes

Let $\mathfrak{B} = \{B(i,j) : 1 \leq i \leq j \leq n\}$ with $q = q_n(\alpha)$ and assume $B^\star := B(i_1^\star, j_1^\star) \times \ldots \times B(i_m^\star, j_m^\star) \in \mathfrak{B}^m$ has been constructed, such that

$$f|_{[i_r^\star, j_r^\star]} \equiv [A]_r, \tag{2}$$

with A as in **(VS)**. Then

$$\{\omega \in A^{-1}B^\star\} \supset \bigcap_{1 \leq r \leq m} \{g|_{[i_r^\star j_r^\star]} \equiv \omega^\top [A]_r \in B(i_r^\star, j_r^\star)\}$$

and

$$\{T_n(Y,g) \leq q_n(\alpha)\} = \bigcap_{\substack{1 \leq i \leq j \leq n \\ g|_{[i/n, j/n]} \equiv g_{ij}}} \{g_{ij} \in B(i,j)\}$$

which implies

$$\{\omega \in A^{-1}B^\star\} \supset \{T_n(Y,g) \leq q_n(\alpha)\}$$

Confidence boxes

One cannot obtain $B^\star$ directly as $f^1, \ldots, f^m$ are unknown. $\lightning$

$\Rightarrow$ Construct $\mathfrak{B}^\star \subset \mathfrak{B}^m$, with $\mathbf{P}\left(B^\star \in \mathfrak{B}^\star \middle| T_n \leq q_n(\alpha)\right) = 1$ and define

$$\boxed{\mathcal{C}_{1-\alpha} := \bigcup_{B \in \mathfrak{B}^\star} A^{-1}B.}$$

$$\begin{aligned}
&\mathbf{P}\left(\omega \in \mathcal{C}_{1-\alpha}\right) \\
\geq &\mathbf{P}\left(\omega \in \mathcal{C}_{1-\alpha} \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
= &\mathbf{P}\left(\omega \in \bigcup_{B \in \mathfrak{B}^\star} A^{-1}B \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
\geq &\mathbf{P}\left(\omega \in A^{-1}B^\star \middle| T_n \leq q_n(\alpha)\right) \mathbf{P}\left(T_n \leq q_n(\alpha)\right) \\
\geq &1 - \alpha.
\end{aligned}$$

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 4. Delete $B \in \mathfrak{B}^m$ if there exists an $r \in \{1, \ldots, m\}$, s.t. $\text{proj}_r(B) \in$

$\mathfrak{B}_{\text{nc}} := \{B(i,j) \in \mathfrak{B} : \exists [s,t],[u,v] \subset [i,j] \text{ with } B(s,t) \cap B(u,v) = \emptyset\}.$

$\rightarrow$ Exploring the fact that $f = (f_1, \ldots, f^m)^\top$ is constant on $[i_r^\star, j_r^\star]$, with $B^\star := B(i_1^\star, j_1^\star) \times \ldots \times B(i_m^\star, j_m^\star)$, conditioned on $\{T_n \leq q_n(\alpha)\}$.

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 5. Delete $B \in \mathfrak{B}^m$, with $[\underline{b}_r, \overline{b}_r] := \mathrm{proj}_r(B)$,

1. for any $2 \leq r \leq m$

$$\frac{a_2 + (m-1)a_1 - \sum_{k=1}^{r-1} \underline{b}_k}{m-r+1} \leq \underline{b}_r \quad \text{or} \quad \underline{b}_{r-1} \geq \overline{b}_r, or$$

2. ...

$\rightarrow$ Exploring the structure of $\Omega(m)$, e.g., $\omega_{i-1} < \omega_i < (1 - \sum_{j=1}^{i-1} \omega_j)/(m-i+1)$, ..., together with the specific choice of the matrix A in **(VS)**.

Construction of $\mathfrak{B}^\star$

Apply reduction rules R1. - R3. on $\mathfrak{B}^m$ reducing it to a smaller set $\mathfrak{B}^\star \subset \mathfrak{B}^m$:

R 6. Delete $B \in \mathfrak{B}^m$, if there exists a $k \in \{1, \ldots, n\}$ such that for all $[i,j] \in \{[i,j] : k \in [i,j] \text{ and } B(i,j) \notin \mathfrak{B}_{\text{nc}}\}$

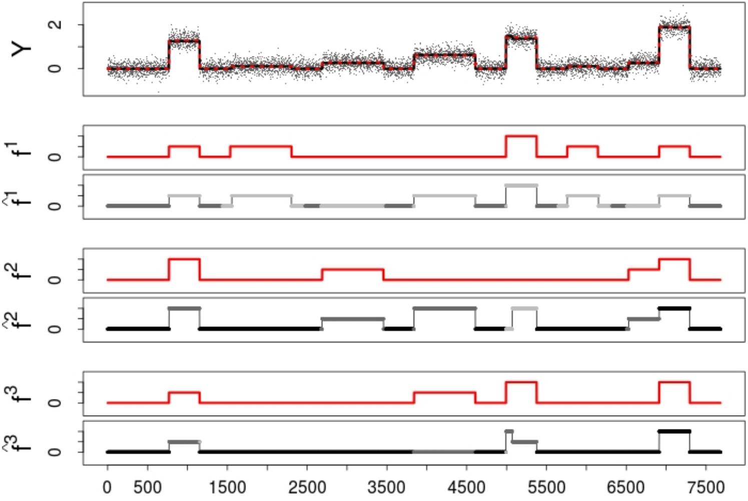
$$\left[\max_{i \leq u \leq v \leq j} \underline{b}_{uv}, \min_{i \leq u \leq v \leq j} \overline{b}_{uv} \right] \cap \left\{ \tilde{\omega}^\top a : a \in \mathfrak{A}^m \;\text{ and } \tilde{\omega} \in A^{-1} B \right\}$$

is empty, with $B(u,v) = [\underline{b}_{uv}, \overline{b}_{uv}] \in \mathfrak{B}$.

$\to$ Exploring the fact that $g = \omega^\top f$ maps to $\{\tilde{\omega}^\top a : a \in \mathfrak{A}^m \text{ and } \tilde{\omega} \in A^{-1} B^\star\}$ conditioned on $\{T_n \leq q_n(\alpha)\}$.

Example

$m = 3$, $\mathfrak{A} = \{0, 1, 2\}$, $\sigma = 0.22$, and $n = 7680$, with $\omega = (0.11, 0.29, 0.6)$. We estimated $\hat{\omega} = (0.11, 0.26, 0.63)$ (with $\mathcal{C}_{0.9} = [0.00, 0.33] \times [0.07, 0.41] \times [0.39, 0.71]$).



Color code: deviation with confidence $\{0, 1, 2\}$

Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \dots, \hat{f}^m$, but uncertainty is captured in confidence bands.

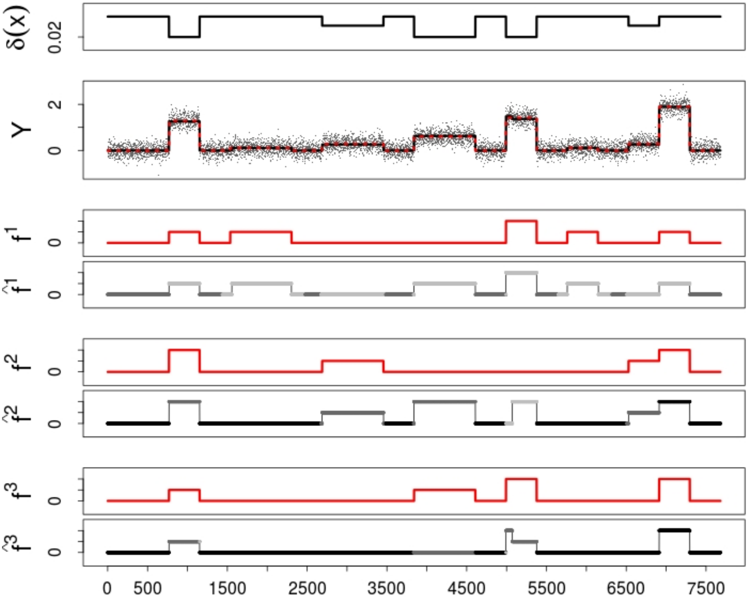
Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
    1.1 Little influence on $\hat{\omega}$.
    1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.

Local finite alphabet separation boundary:

$$0 < \delta(x) := \min_{a \neq f(x) \in \mathfrak{A}^m} \left| \omega^\top a - \omega^\top f(x) \right|$$

Violation of identifiability



Violation of identifiability

For f as in our example, but ω choose randomly, uniformly distributed on $\Omega(3)$, we compute 10.000 realizations of $\hat{\omega}$, $\mathcal{C}_{1-\alpha}$, $\hat{f}^1, \ldots, \hat{f}^3$, and $\tilde{\mathcal{H}}(\beta)$, for $\sigma = 0.05$, $n = 1280$, and $\alpha = \beta = 0.1$. Consequently, for each run we get a different ω and δ, respectively.

$\delta \in$	MAE($\hat{\omega}$) [10^{-3}]	dist($\omega, \mathcal{C}_{1-\alpha}$) [$10^{-3}$]
$[0, 0.0001]$	$(6, 4, 5)$	29
$[0.0001, 0.01]$	$(7, 4, 7)$	34
$[0.01, 0.02]$	$(4, 4, 4)$	30
$[0.02, 0.03]$	$(4, 4, 4)$	29
$[0.03, 0.04]$	$(4, 3, 4)$	31
$[0.04, 0.05]$	$(4, 3, 4)$	31
$[0.05, 0.06]$	$(4, 3, 5)$	31
$[0.06, 0.07]$	$(3, 3, 4)$	31

$\rightarrow$ SESAME's performance of $\hat{\omega}$ and $\mathcal{C}_{1-\alpha}$, respectively, is not much influenced by the ASB δ

Violation of identifiability

For f as in our example, but ω choose randomly, uniformly distributed on $\Omega(3)$, we compute 10.000 realizations of $\hat{\omega}$, $\mathcal{C}_{1-\alpha}$, $\hat{f}^1, \ldots, \hat{f}^3$, and $\tilde{\mathcal{H}}(\beta)$, for $\sigma = 0.05$, $n = 1280$, and $\alpha = \beta = 0.1$. Consequently, for each run we get a different ω and δ, respectively.

$\delta \in$	MIAE$(\hat{f}^i)$ $[10^{-4}]$	med$(\lvert\tilde{\mathcal{H}}_x(0.1)\rvert)$	$\delta(x) \in$
$[0, 0.0001]$	$(1916, 1067, 483)$	3	$[0, 0.001]$
$[0.0001, 0.01]$	$(1536, 923, 354)$	3	$[0.001, 0.01]$
$[0.01, 0.02]$	$(671, 474, 147)$	3	$[0.01, 0.02]$
$[0.02, 0.03]$	$(236, 164, 40)$	3	$[0.02, 0.03]$
$[0.03, 0.04]$	$(96, 37, 7)$	2	$[0.03, 0.04]$
$[0.04, 0.05]$	$(100, 7, 2)$	2	$[0.04, 0.05]$
$[0.05, 0.06]$	$(42, 1, 0)$	2	$[0.05, 0.1]$
$[0.06, 0.07]$	$(16, 4, 0)$	1	$[0.1, 0.33]$

$\rightarrow$ Uncertainty is captured in the confidence bands.

Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.

2. (VS) condition violated, i.e., too little variation of $f^1, ..., f^m$:
   2.1 Big influence on $\hat{\omega}$.
   2.2 Big influence on $\hat{f}^1, ..., \hat{f}^m$ as estimate is based on $\hat{\omega}$.

Violation of identifiability

1. (ASB) condition violated, i.e., $\delta = 0$:
   1.1 Little influence on $\hat{\omega}$.
   1.2 Big influence on $\hat{f}^1, \ldots, \hat{f}^m$, but uncertainty is captured in confidence bands.

2. (VS) condition violated, i.e., too little variation of $f^1, ..., f^m$:
   2.1 Big influence on $\hat{\omega}$.
   2.2 Big influence on $\hat{f}^1, ..., \hat{f}^m$ as estimate is based on $\hat{\omega}$.

When f comes from a Markov chain, probability that variation is rich enough converges exponentially fast to 1.[9]

[9] [Behr and Munk, 2015]