



## Multiscale Inference for Blind Demixing with Applications in Cancer Genetics

Merle Behr\*

University of Göttingen, Göttingen, Germany - merle.behr@stud.uni-goettingen.de

Chris Holmes

University of Oxford, Oxford, UK - cholmes@stats.ox.ac.uk

Axel Munk

University of Göttingen, Göttingen, Germany - munk@math.uni-goettingen.de

We discuss a particular kind of blind source separation (BSS) embedded in a change-point regression setting. In BSS problems one observes a mixture of source functions, and aims to recover the original sources from the available observations. The blindness refers to the fact that neither the sources nor the mixing weights are known.

We provide a new statistical methodology SESAME (SEparateS-finite-Alphabet-MixturEs) for recovery of linear mixtures of piecewise constant functions taking values in a known finite alphabet with unknown mixing weights from noisy observations.

In the regression setting we obtain superpolynomial deviation bounds for estimating the mixing weights and the change-point locations of the mixture which for the latter yields the minimax sampling rate up to a log square term. Further, we show that the source functions are estimated correctly, up to the uncertainty in the change-point locations, at the same rate. Uniform confidence statements for all quantities are derived.

Our theory relies on new results on identifiability and exact recovery in BSS problems.

SESAME is motivated from cancer genetics where one aims to assign copy-number variations from genetic sequencing data to different tumor-clones and their corresponding proportions in the tumor.

We analyze such data using the proposed method in order to estimate the number of clones, their proportion in the tumor, and the corresponding copy number variations.

**Keywords:** Blind source separation; Exact recovery; Change point analysis; Genetic sequencing.