A novel density estimator and its use for LHC signal detection

Alessandro Morandini (SISSA,INFN)

DM@LHC 2020, 4th June 2020

Based on *2006.xxxxx*

with A. De Simone

Motivation

What are people doing now?

Big data → Machine Learning analysis

Bigger data → More complicated architecture

PROS	CONS
it works very well	time consuming
popular and hot	black box
many implementations	hard to interpret

Markov Chain Density Estimator (MCDE)

We are looking for understandable + easy to optimize

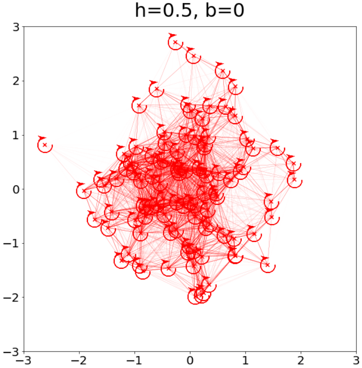
Idea: process can be described by $p(x)$

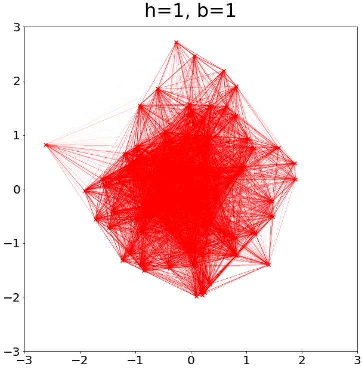
$\{x_i\} \to \hat{p}(x) \sim p(x)$: make estimate with Markov Chain

1. Calculate the distance matrix d_{ij}
2. Build the weight matrix $W_{ij} = g_h(d_{ij})(1 - b\delta_{ij})$
3. Consider Markov Chain on $\{x_i\}$ with $P_{ij} \sim W_{ij}$
4. Find eigenvector of the matrix P_{ij}: $\pi_i \propto p(x_i)$

Role of h and b

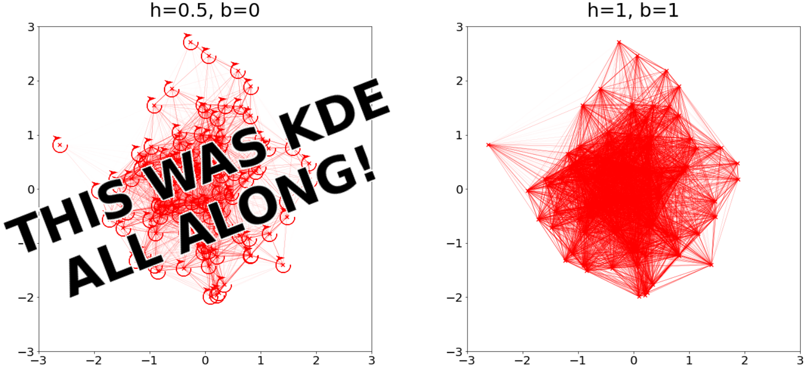
$$W_{ij} = g_h(d_{ij})(1 - b\delta_{ij}), \text{ with } g_h(d_{ij}) = \exp\left(-\frac{d_{ij}^2}{2h^2}\right)$$



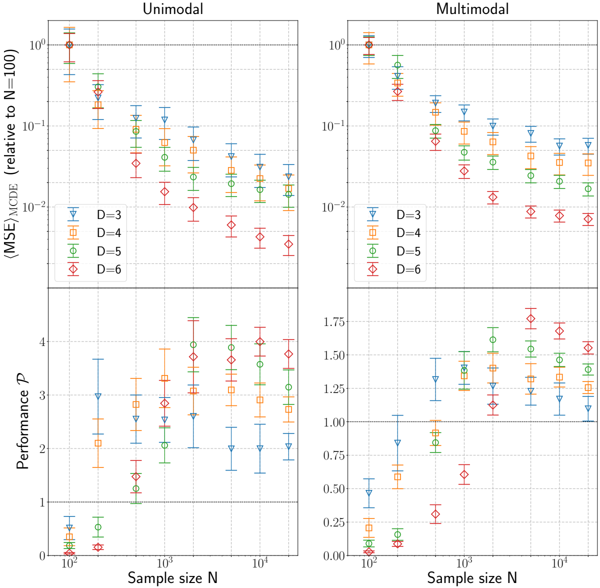


Role of h and b

$$W_{ij} = g_h(d_{ij})(1 - b\delta_{ij}), \text{ with } g_h(d_{ij}) = \exp\left(-\frac{d_{ij}^2}{2h^2}\right)$$



Consistency and performance



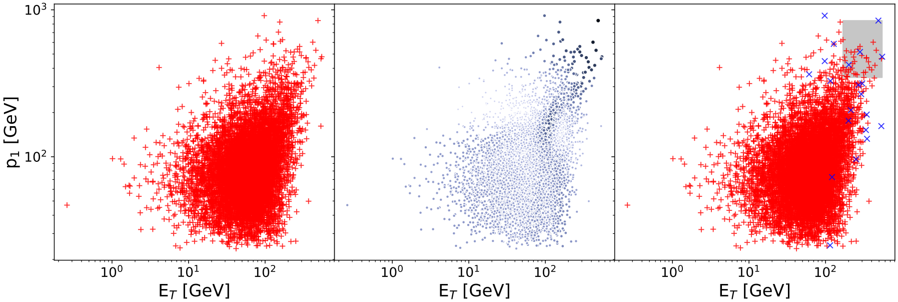
Anomaly detection → DarkMachines → Z' (1TeV)

$$X_b \longrightarrow \hat{p}_b(x)$$
$$X_s \longrightarrow \hat{p}_s(x)$$

$$S(x) = \frac{\langle \hat{p}_s \rangle_k}{\langle \hat{p}_b \rangle_k}$$

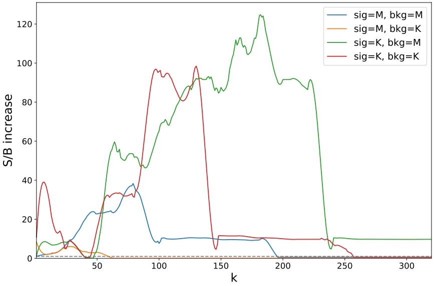
4 ways to do this!

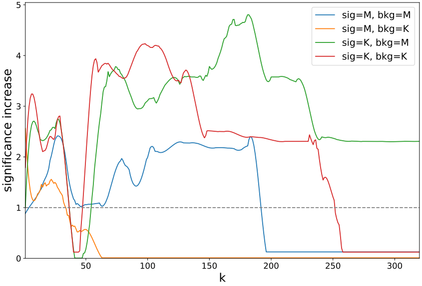
Cut on $S(x)$



Z' benchmark

- particle cut: bottom, lepton required
- $\mathscr{L} = 0.1\text{fb}^{-1} \to N_{\text{SM}} \sim 8k,\ N_{\text{BSM}} = 21$
- cut on 0.2 percentile of $S(x)$





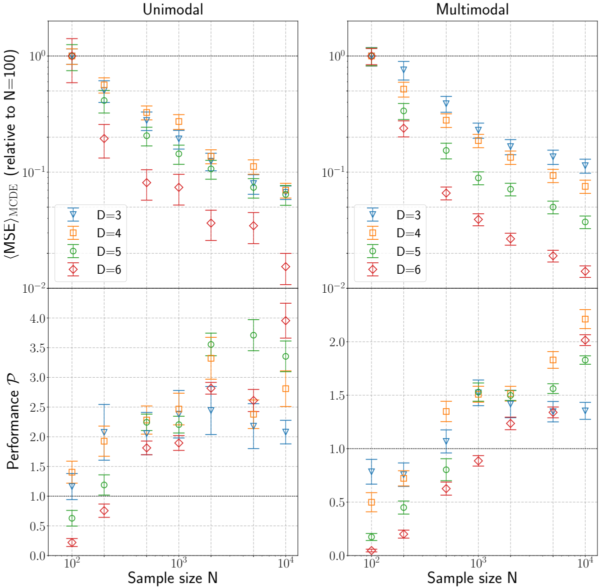
Conclusions and outlook

- MCDE: new idea and works well
- Different weight to tails → interesting for outlier
- Outlier detection with MCDE can be used at LHC (but more work is needed to make it robust!)

Thank you!

BACKUP

Different optimization on h for KDE and MCDE



Why small k bad?
$S(x)$ very unbalanced among points!

This is $k = 1$ with the same cut as the other plots: