Nikita Kazeev

Kernel density estimation

Aka advanced KNN

2021

MLHEP 2021

NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS

LAMBDA • HSE

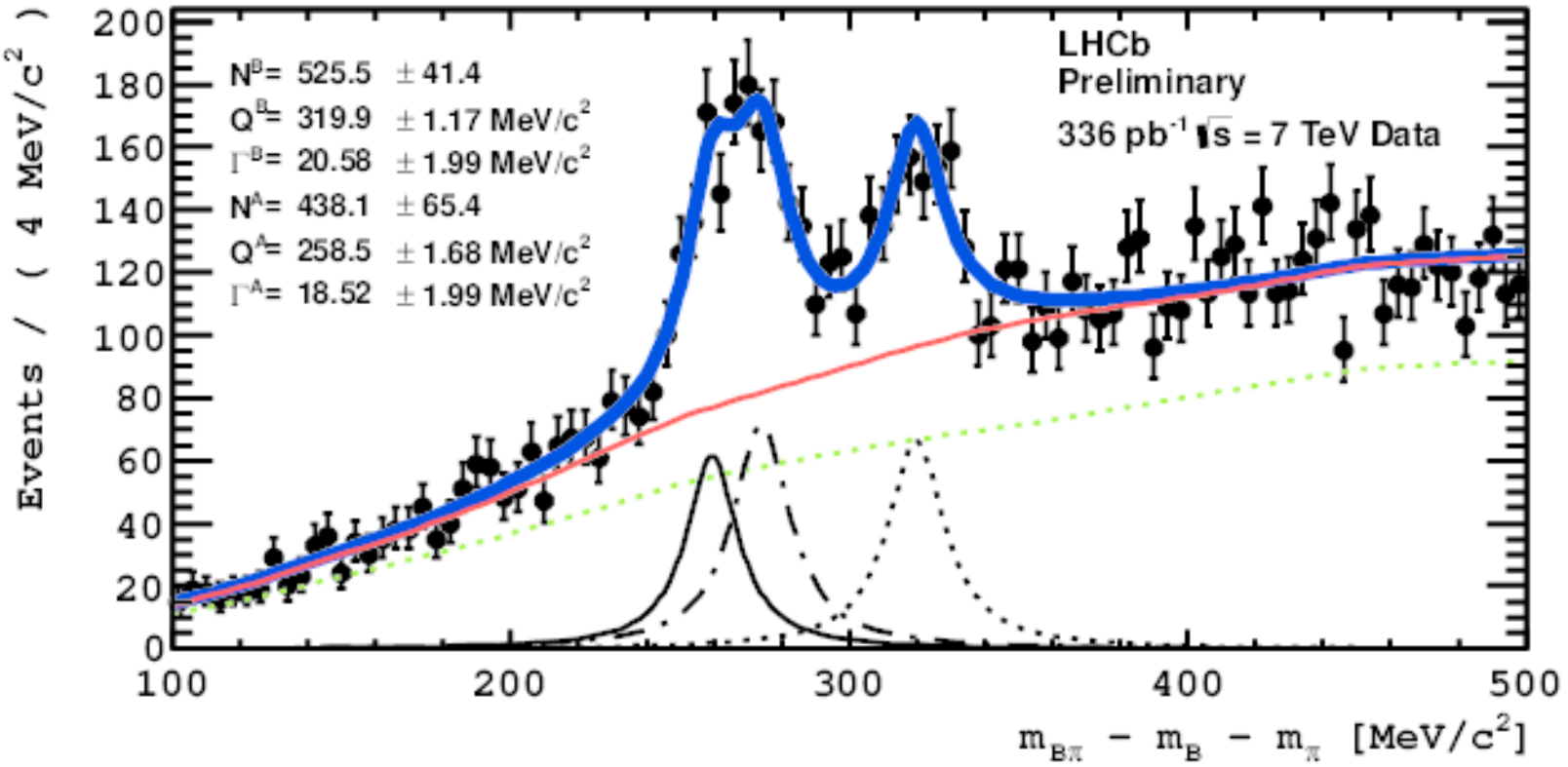
Yandex

SCHOOL OF DATA ANALYSIS

EPFL

SIT Schaffhausen Institute of Technology

Good ol' distribution fitting



Good ol' distribution fitting

Given:

- data points $x_1, \ldots, x_n \in \mathbb{R}^m$
- a parametrized distribution $P(x|\theta)$

Find a set of parameters to maximize the empirical likelihood:

$$\max L(\theta|x) = \max \prod_i P(x|\theta)$$

Kernel density

- Place many Gaussians on the data points and call their sum a PDF
- “Histogram interpretation”: fuzzy bins
- “KNN interpretation”: take into account the distances

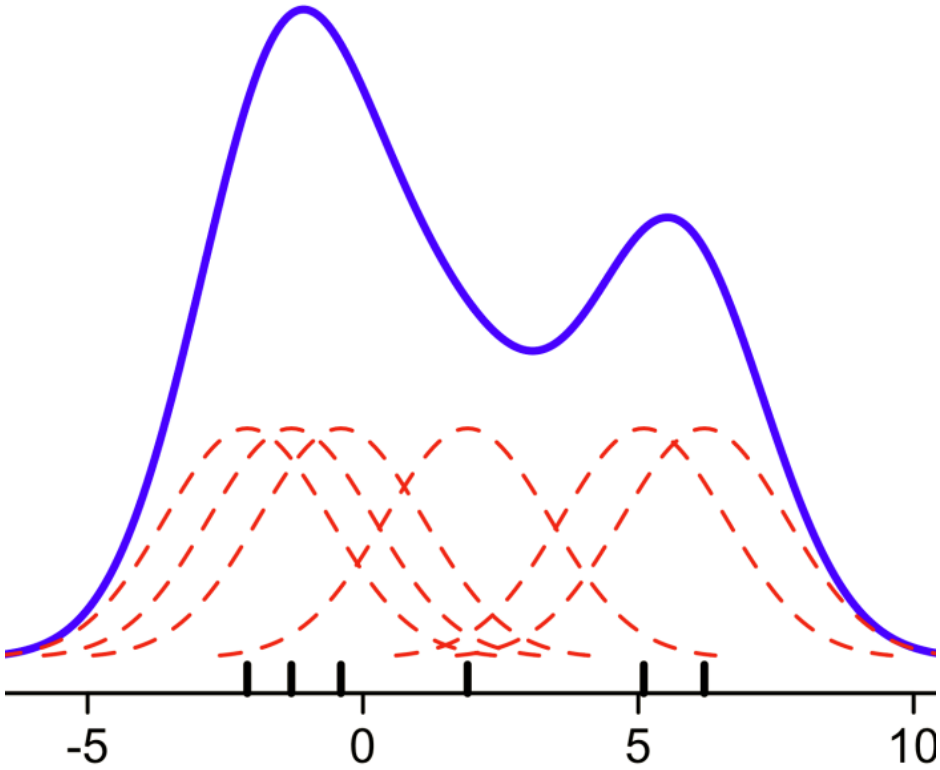
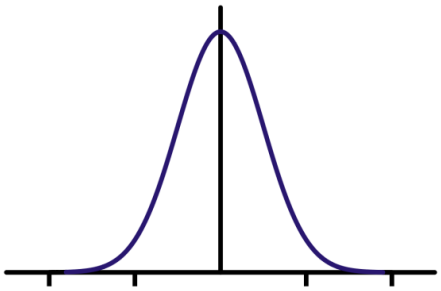


Image: http://www.spiderfinancial.com/support/documentation/numxl/users-guide/descriptive-statistics/kernel-density-estimation-kde-tutorial

Kernel density

Kernel function, usually looks like this:



$$P_{\mathrm{KDE}}(x) = \sum_i K\,(x - x_i)/N$$

Number of points in the training dataset

Sum over all points in the training dataset

Image: Wikipedia

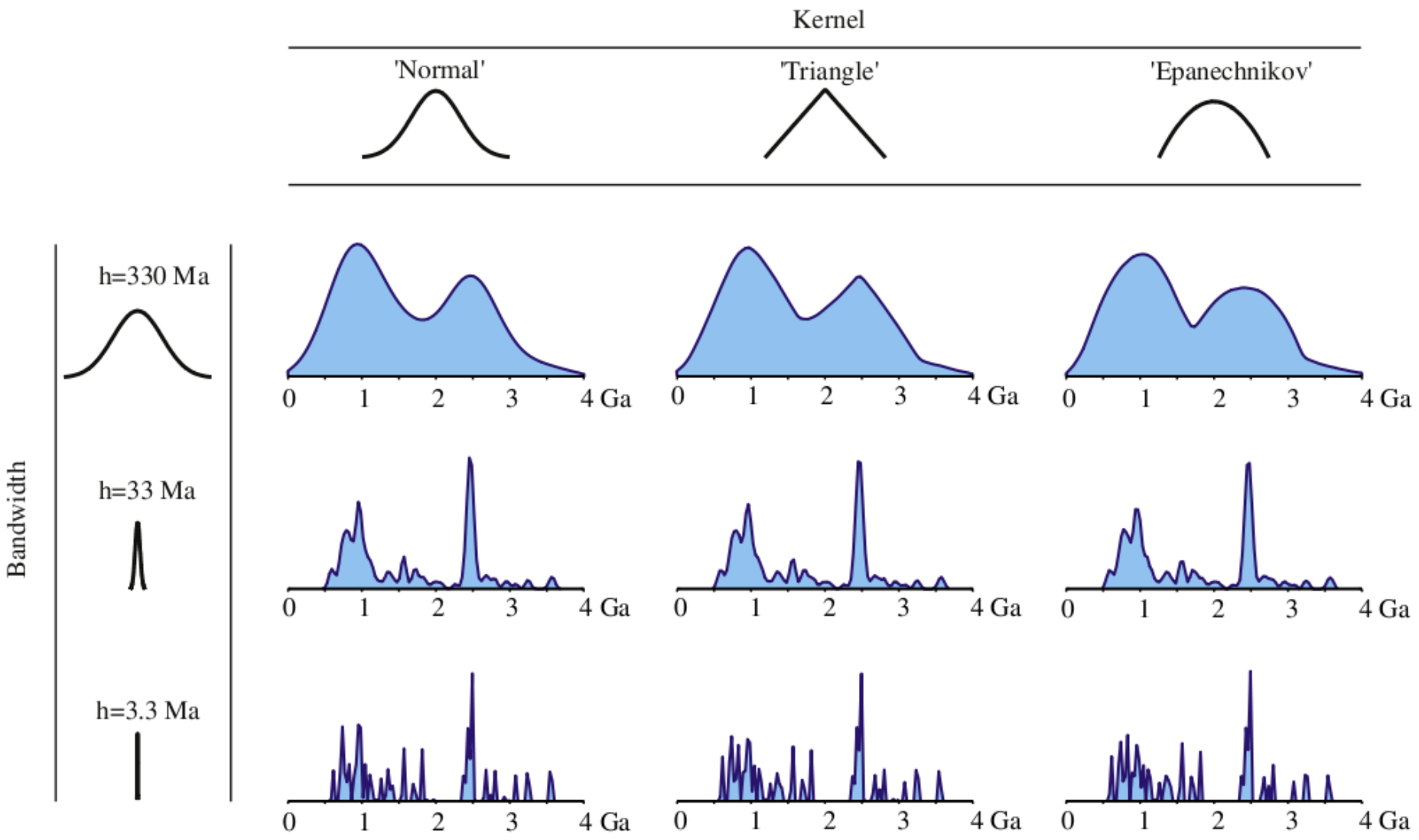


Image: http://www.ucl.ac.uk/~ucfbpve/papers/VermeeschChemGeol2012/

Kernel Density vs Histogram

Kernel Density	Histogram
Smooth PDF	Discrete binning
With number of data points approaching infinity, the value in a point approaches the convolution of the PDF with the kernel function	With number of data points approaching infinity, the value in a bin approaches the unbiased mean PDF in that bin
No easy way to estimate the uncertainty	Straightforward uncertainty estimation of bin values
User-defined parameter: kernel shape and width	User-defined parameter: bins
Requires storing the full training dataset Finite support kernels allow for fast lookup	Lookup time and memory are proportional to the number of bins

Kernel Density: summary

- Go-to way for an easy 1-2D PDF approximation
- Histograms competitor
- Has heuristic parameters: kernel shape
- Memory expensive
- Doesn’t scale for high dimensions
- Nice demo: https://mathisonian.github.io/kde/

Thank you!


nkazeev@hse.ru


kazeevn


hse_lambda

Nikita Kazeev