A New Class of Shape Constraints for Probability Density Estimates

Mark Wolters

Shanghai Center for
Mathematical Sciences

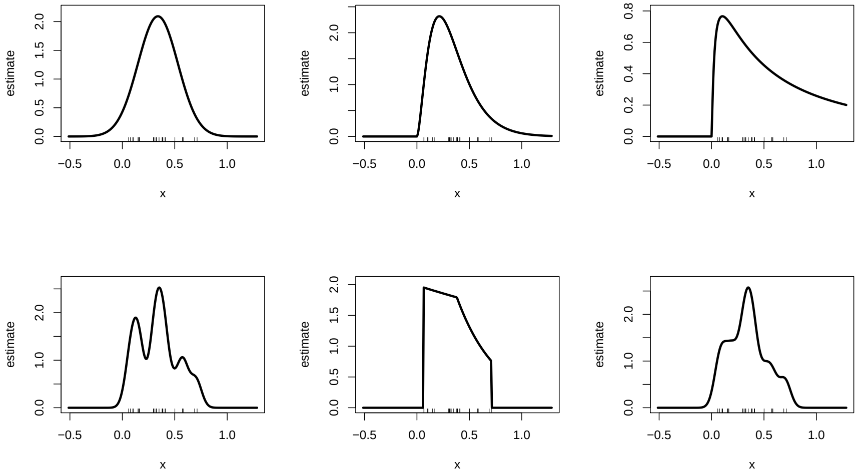
Fudan University

ICSA 2016, Shanghai
December 20, 2016

Motivation (1/3)

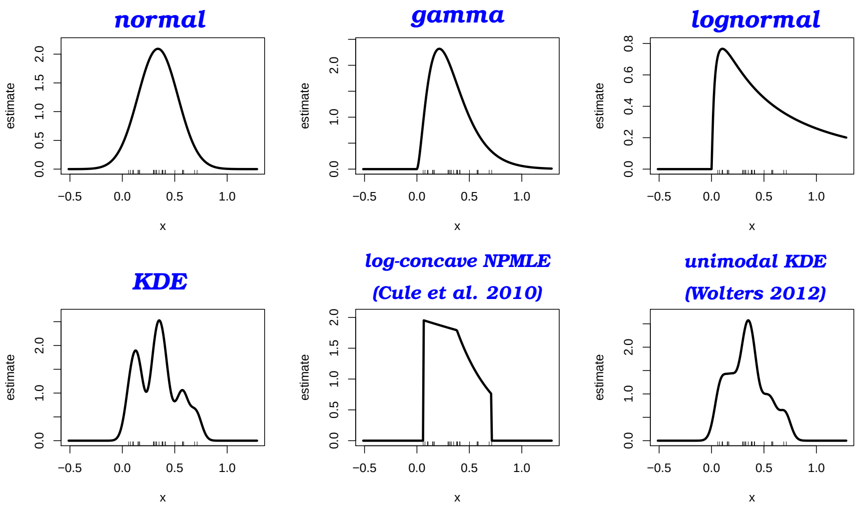
The parametric sniff test

Here are 6 density estimates. Which are nonparametric?



The parametric sniff test

Here are 6 density estimates. Which are nonparametric?



Motivation (2/4)

A typical example: axon diameter distribution

frontiers in Neuroanatomy

ORIGINAL RESEARCH
published: 26 May 2016
doi: 10.3389/fnana.2016.00059

CrossMark

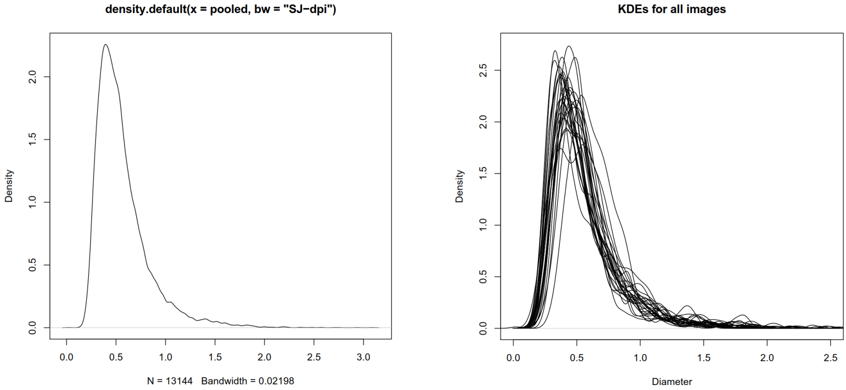
Parametric Probability Distribution Functions for Axon Diameters of Corpus Callosum

Farshid Sepehrband[1,2], Daniel C. Alexander[3], Kristi A. Clark[2], Nyoman D. Kurniawan[1], Zhengyi Yang[1,4,5] and David C. Reutens[1]*

[1] *Centre for Advanced Imaging, The University of Queensland, Brisbane, QLD, Australia,* [2] *Laboratory of Neuro Imaging, USC Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA,* [3] *Department of Computer Science, Centre for Medical Image Computing, University College London, London, UK,* [4] *Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China,* [5] *Faculty of Information Engineering, Southwest University of Science and Technology, Mianyang, China*

Axon diameter is an important neuroanatomical characteristic of the nervous system that alters in the course of neurological disorders such as multiple sclerosis. Axon diameters vary, even within a fiber bundle, and are not normally distributed. An accurate distribution function is therefore beneficial, either to describe axon diameters that are obtained from a direct measurement technique (e.g., microscopy), or to infer them

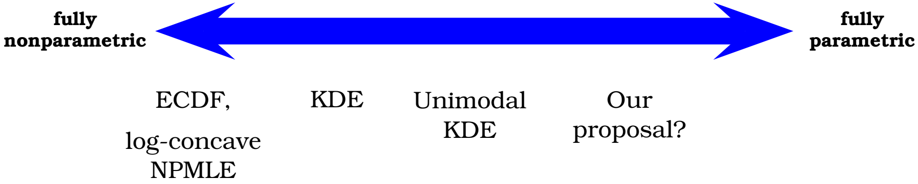
- Small to moderate n
- Smooth and regular histograms
- They tried 16 parametric forms and compared GOF.


density.default(x = pooled, bw = "SJ−dpi")
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
N = 13144 Bandwidth = 0.02198
KDEs for all images
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
Diameter

Motivation (4/4)

Claim: There is a market among data analysts for density estimators that have data-driven shape, but “look like” parametric densities.

- This is one motivation for shape restricted estimation
- Methods so far use constraints that are ***simple*** (e.g. unimodality) or ***convenient*** (e.g. log-concavity).

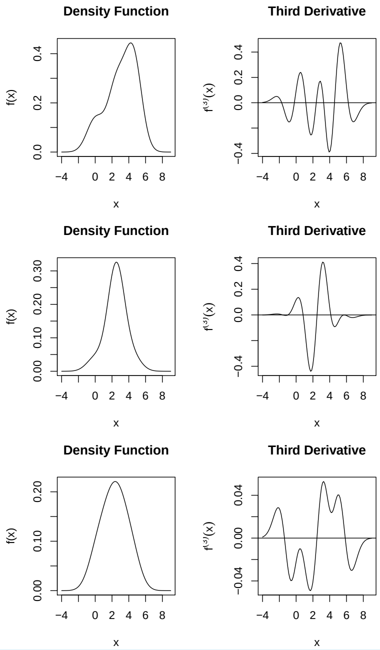


Proposal

Simple idea: restrict the number of inflection points

- "Waves" and "kinks" in the PDF: ***concavity changes***
- For increased smoothness, restrict inflections of ***derivatives*** of f, too
  - equivalently, control number of zeros of f', f'', f''', ...
  - After f''', restrictions become harder to notice.
- For a two-tailed density, maximal smoothness achieved when $f^{(r)}$ has r zero crossings.

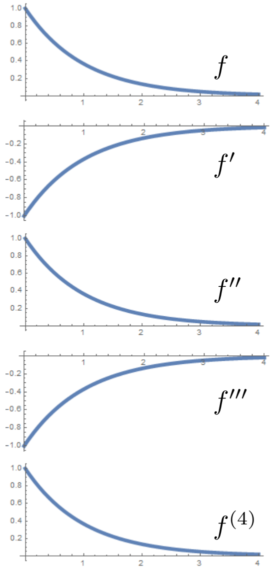
  (equivalently, when $f^{(r)}$ has $r+2$ inflection points)



k-monotonicity

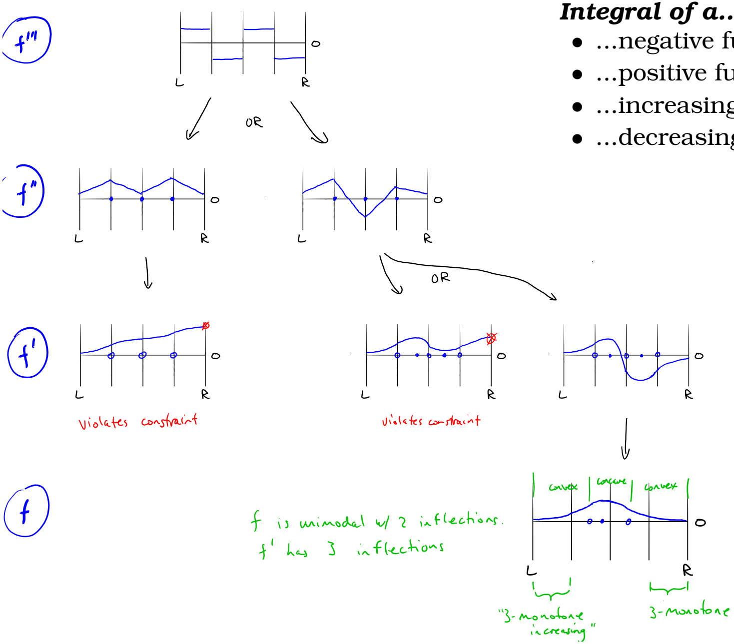
For decreasing densities, the ***k-monotonicity*** concept has already covered this idea*.

- Definition:
  - 1-monotone: f nonnegative, nonincreasing
  - 2-monotone: f nonnegative, nonincreasing, convex
  - k-monotone: $(-1)^j f^{(j)}$ nonnegative, nonincreasing, convex, $j = 0, 1, \ldots, k-2$.
- But, this only applies to convex, decreasing PDFs
- What about two-tailed densities?



*e.g., Balabdaoui & Wellner, An. Stat 2007; Chee & Wang CSDA 2014

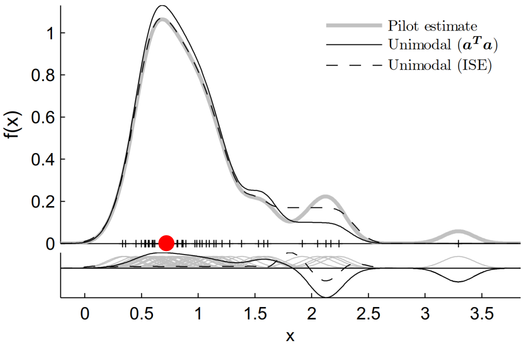
Extending k-monotone idea to two tails



Integral of a...

- ...negative function is decreasing
- ...positive function is increasing
- ...increasing function is convex
- ...decreasing function is concave

Method of Hall & Huang (Stat Sinica, 2002)



Change KDE weights.

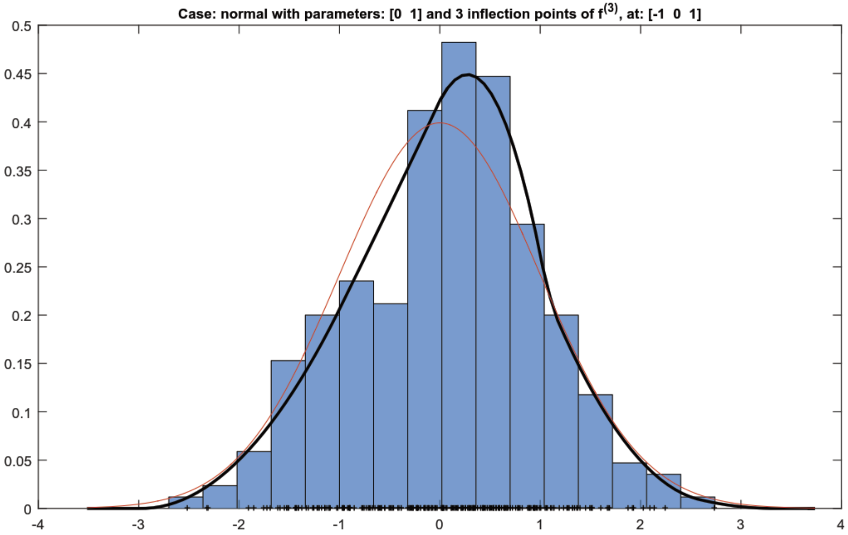
Developed for unimodality.

Can be used for other derivative constraints.

Can be generalized to work with other estimator constructions.

- Given locations of zeros, can solve by QP
- Need to search for zero locations
- Inherits KDE boundary issues.

Toward an Estimator (2/4)

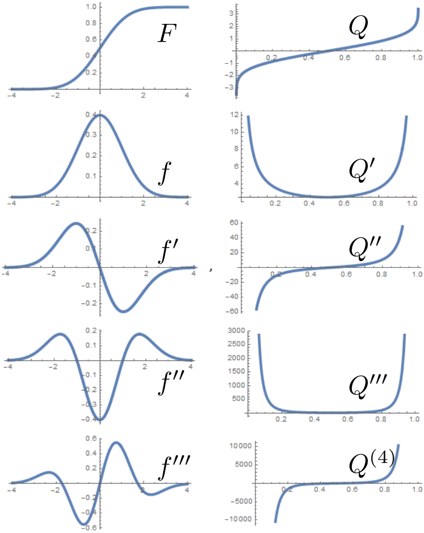


Quantile function approach

Let any point on the ***CDF*** be (x, p), where $p = F(x)$ and $x = Q(p)$.

Claim: constraint $Q''' > 0$ sufficient to ensure a nice looking density.

- $Q''' > 0$ guarantees unimodality (at most one zero of f').
- Observation: parametric PDFs with bounded f and convex tails have this property.
- Constructed examples ***not*** having this property have waves/kinks/knees
- Still working on formal results on zeros of f'', f'''.



Toward an Estimator (4/4)

A shape constraint based on Q'''

- Estimation of f via Q introduces complexities.
- But derivatives of F and Q have relationships (note $p = F(Q(p))$, then take derivatives)
- We find

$$Q''' > 0 \iff \boxed{3\left(f'\right)^2 - f f'' > 0}$$

  - Constraint applies uniformly (no searching for zero locations)
  - If f is a spline, the constraint is quadratic in parameters
    - Max likelihood is convex programming problem
    - Min discrepancy to ECDF is a QCQP problem
- Same constraint applies to unimodal, increasing, or decreasing densities
- Implementation of this approach is ongoing.

Conclusions

Summary

- Goal is nonparametric estimators that pass the parametric sniff test.
- Our ideal for qualitative smoothness: $f^{(r)}$ has $r+2$ inflection points, $r = 0, 1, 2, 3, \ldots$ (two-tailed case)
- Practical options to get close to this ideal:
  - A weighted-KDE method
  - Q'''-derived constraint looks promising
- Still many challenges:
  - Hard to avoid a smoothing parameter
  - Optimal construction, estimation
  - Theory about estimator quality
  - Confidence bounds
  - Testing validity of the constraint