scdensity: an R Package for Shape-Constrained Kernel Density Estimation

Mark A. Wolters
Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China

Summary
Adding shape constraints to a nonparametric estimator:• eliminates unrealistic waves and bumps in the estimate• maintains more shape flexibility than parametric families• improves statistical performance in small samples

The scdensity package implements two related methods for enforcing constraints on a kernel density estimator (KDE):1. The weighted KDE method (Hall and Huang, 2002)2. The adjusted KDE method (Wolters and Braun, 2018)

It unifies the methods under a common optimization scheme and makes estimation numerically stable.
The package makes it easy to get estimates with many different constraints, using familiar kernel methods.

How does it work?
A weighted KDE is \( f_\lambda(x|p) = \frac{1}{n} \sum_{i=1}^{n} p_i K \left( \frac{x - s_i}{h} \right) \), where \( s \) are the kernel centers and \( p \) are the weights.
The data are \( x \). We do not require kernel centers to be located at \( x \).
We can express the integrated squared error (ISE) between any two weighted KDEs as a quadratic form in the weights.

1. Set up a uniform grid of kernel centers, \( s \).
2. Find \( w \) that minimizes ISE(\( f_\lambda(x|w), f_\lambda(x|p_{0w}) \)) (this is a QP with no shape constraints).
3. Subdivide any intervals where \( f_\lambda \) is poorly approximated.
4. Repeat 2 & 3 until approximation is good.

Which Constraints Can it Handle?

Unimodal
scdensity(\( x, b, h, \text{constraint}="\text{unimodal}" \))

Two inflection points
scdensity(\( x, b, h, \text{constraint}="\text{twoInflections}" \))

Three inflection points
scdensity(\( x, b, h, \text{constraint}="\text{threeInflections}" \))

Symmetric
scdensity(\( x, b, h, \text{constraint}="\text{symmetric}" \))

Bimodal
scdensity(\( x, b, h, \text{constraint}="\text{bimodal}" \))

Examples
Constraints: twoInflections+, symmetric around zero.
Smooth tails
No restriction on tail weight
Fixed mode location

Axon diameters (Sepehrband et al., 2016).
Constraint: twoInflections+

Q & A
Is it fast?
- A fraction of a second for \( N(0,1) \) data with unimodal constraint.
- Several seconds for \( N \) data with twoInflections+
Is it robust?
- The QP problem is convex, but can be ill-conditioned.
The package checks for problems and remedies them.
- Constraint systems are occasionally infeasible. The package checks feasibility and handles problems gracefully.

What about asymptotics?
- Because we use the usual kernel density estimator, we can borrow its asymptotic behavior. If the constraints are valid, necessary shape adjustments should shrink to zero.

References

Turnbull, BC & Ghosh, SK (2014), "Unimodal density estimation using Bernstein polynomials," Computational Statistics and Data Analysis, 72, 13–29