Local Orthogonal Polynomial Expansion (LOrPE) for Density Estimation

Alex Trindade

Dept. of Mathematics & Statistics, *Texas Tech University*

Igor Volobouev, *Texas Tech University* (Physics Dept.)

D.P. Amali Dassanayake, *University of Calgary* (former student)

July 2015

Outline

Kernel Density Estimation (KDE): Simple & Effective

Observe: $x_1, \ldots, x_n \sim IID$ from PDF $f(\cdot)$ on compact support $[a, b]$.

- Start with empirical density function (EDF):

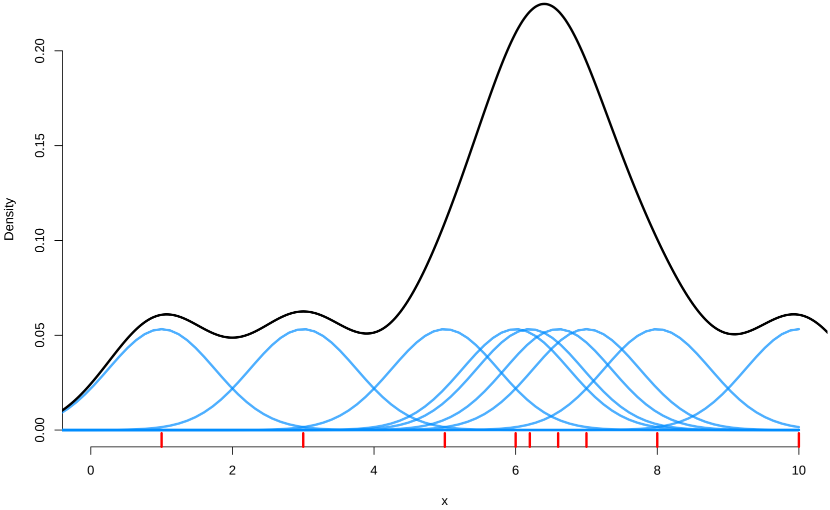
$$\hat{f}_{\text{EMP}}(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - x_i)$$

  where $\delta(\cdot)$ is the Dirac delta function.

- **Better:** If can assume PDF has first few derivatives or at most a few modes, a convolution of EDF with a kernel function $K(\cdot)$ gives a weighted average of points close to x:

$$\hat{f}_{\text{KDE}}(x) \equiv \int \frac{1}{h} K\left(\frac{x-y}{h}\right) \hat{f}_{\text{EMP}}(y) dy = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x-x_i}{h}\right)$$

Example: Explanation of KDE



KDE is Optimal, But has Problems...

Theorem

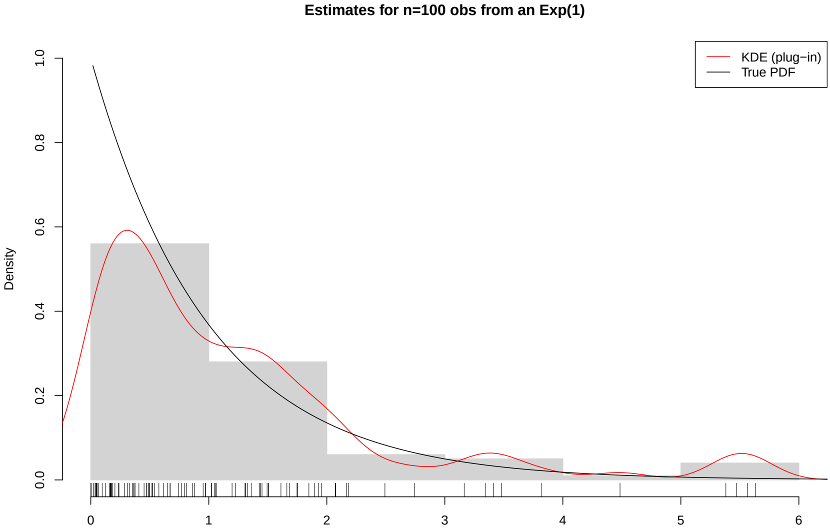
For kernel order r, and optimal bandwidth h_n^*, KDE's (asymptotic) MISE $\to 0$ at the fastest possible rate (for any estimator):

$$O(n^{-2r/(2r+1)}) = O(n^{-4/5}), \text{ for usual Gaussian kernel}$$

But:

- Although consistent for a sequence of bandwidths $h_n \to 0$ under mild conditions, KDE suffers from boundary bias, esp. for sharply truncated supports, e.g., Exp(1).
- In bounded supports KDE fails to attain this optimal convergence rate (Jones, 1993)...

Example: KDE Boundary Bias



Solution 1: Local and/or Adaptive Methods (1980's on)

- Adaptive kernels. Have local estimates in regions with ample data, but expand the neighborhood in regions where data is more scarce. (Juggle bias-variance trade-off to reduce MISE.) Achieved by:
  - changing bandwidths at different points, or
  - changing kernel shapes at different points (or both).
- Boundary corrections. Truncation and reflection of boundary kernels (data mirroring); optimal weighting schemes; data transforms; etc.
- Local polynomial/likelihood. Local Likelihood Density Estimation (LLDE) models PDF locally via a polynomial; estimation via ML (Hjort & Jones, 1996; Loader, 1996, 1999).
- Orthogonal polynomial expansions. Orthogonal Series Density Estimation (OSDE) (Čencov, 1962; Tarter & Lock, 1993; Efromovich, 1999).

Solution 2: LOrPE (under review. . .)

- Construct **truncated orthogonal polynomial series expansion for EDF** near each grid point x_{fit}.
- Like KDE, we need to choose some things:
  - $K(\cdot)$, the **kernel**;
  - h, the **bandwidth**;
  - M, the **polynomial order**. (An extra tuning parameter. . .)
- And like KDE, $K(\cdot)$ is less important than h & M.
- LOrPE estimate:

$$\tilde{f}_{\text{LOrPE}}(x) = \sum_{k=0}^{M} c_k(x_{\text{fit}}, h) P_k\left(\frac{x - x_{\text{fit}}}{h}\right)$$

LOrPE Construction

$$\tilde{f}_{\text{LOrPE}}(x) = \sum_{k=0}^{M} c_k(x_{\text{fit}}, h) P_k\left(\frac{x - x_{\text{fit}}}{h}\right) \quad \xrightarrow{\text{normalize}} \quad \hat{f}_{\text{LOrPE}}(x)$$

- **Polynomials** $P_k(\cdot)$ satisfy normalization condition:

$$\frac{1}{h}\int_a^b P_j\left(\frac{x - x_{\text{fit}}}{h}\right) P_k\left(\frac{x - x_{\text{fit}}}{h}\right) K\left(\frac{x - x_{\text{fit}}}{h}\right) dx = \delta_{jk}$$

- **Coefficients** $c_k(\cdot)$ satisfy integral equation:

$$c_k(x_{\text{fit}}, h) = \frac{1}{h}\int f(x) P_k((x - x_{\text{fit}})/h) K((x - x_{\text{fit}})/h) dx$$

- Subs $f(x) = \hat{f}_{\text{EMP}}(x)$ in above determines coefficients:

$$c_k(x_{\text{fit}}, h) = \frac{1}{nh}\sum_{i=1}^{n} P_k((x_i - x_{\text{fit}})/h) K((x_i - x_{\text{fit}})/h)$$

Connections With KDE

Theorem (Theorem 1)

LOrPE is linear combo of KDEs with varying kernels $K_k(z) \equiv P_k(z)K(z)$:

$$\tilde{f}_{LOrPE}(x) = \sum_{k=0}^{M} \hat{f}_{KDE}(x|h, K_k) P_k\left(\frac{x - x_{fit}}{h}\right)$$

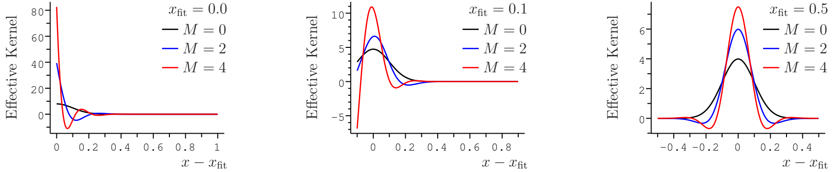
Theorem (Theorem 2)

When evaluated "away" from support boundaries, LOrPE is equivalent to KDE with a high-order ($r \approx M + 1$) *effective kernel:*

$$K_{eff}(x) = \sum_{k=0}^{M} P_k(0) P_k(-x) K(-x)$$

Example: LOrPE $K_{\text{eff}}(x - x_{\text{fit}})$ Plots

(a) (b) (c)



PDF support is $[0, 1]$ with sharp truncation at 0, and grid points are:

(a) exactly at boundary ($x_{\text{fit}} = 0$)

(b) close to boundary ($x_{\text{fit}} = 0.1$)

(c) away from boundary ($x_{\text{fit}} = 0.5$)

Connections With OSDE

- For PDF supported on $[a, b]$ with $\int f(x)^2 dx < \infty$, then for orthonormal basis $\{\phi_k\}$ classical OSDE is:

$$\hat{f}_{\text{OSDE}}(x) = \sum_{j=0}^{J} \hat{\theta}_j \phi_j(x), \qquad \hat{\theta}_j = \frac{1}{n} \sum_{i=1}^{n} \phi_j(x_i)$$

- Thus LOrPE is localized version of OSDE (basis functions $\{P_k\}$ are not global, but adjust locally depending on x_{fit}).
- Also:

Theorem (Theorem 3)

As $h \to \infty$ (fixed M), LOrPE reduces to OSDE with (scaled & shifted) orthonormal Legendre polynomials on $[-1, 1]$ as basis functions.

Connections With LLDE (heuristic. . .)

- LLDE: overcomes boundary bias by matching localized sample moments to population moments using log-polynomial PDF approx.
- LOrPE: matches localized moments of orthogonal polynomials to their sample values using polynomial PDF approx.
- **Comparison.**
  - LLDE: may be theoretically superior, but involves solving non-linear equations at every grid point.
  - LOrPE: enjoys pragmatic advantage of computational speed & numerical stability (is linear "smoother" of EDF).

Selection of Tuning Parameters

- Plug-in approach: similar to *Silverman's Rule* for KDE.
- Cross-validation methods. Choose h & M to minimize:
  - least squares cross-validation (LSCV)

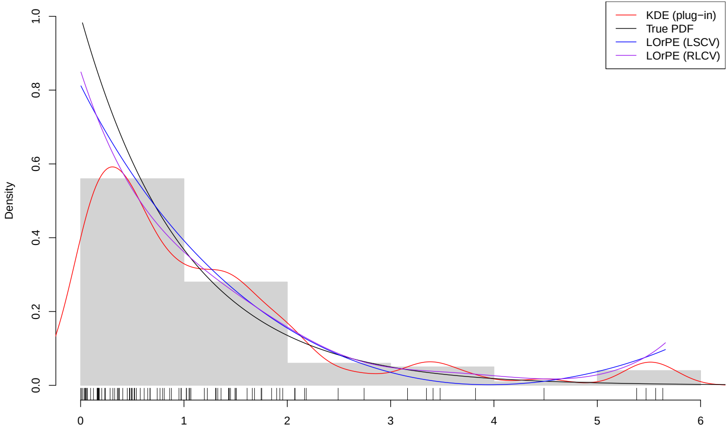
$$LSCV(h, M) = \int \tilde{f}_{\text{LOrPE}}(x|h, M)^2 dx - \frac{2}{n} \sum_{i=1}^{n} \tilde{f}^{(-i)}_{\text{LOrPE}}(x_i|h, M),$$

  - regularized $(\alpha = 0.5)$ likelihood cross-validation (RLCV)

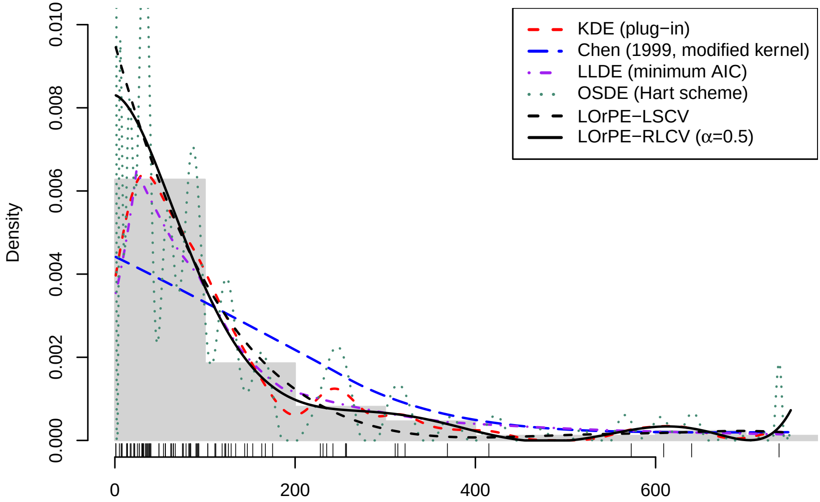
$$RLCV(h, M) = \prod_{i=1}^{n} \max \left\{ \tilde{f}^{(-i)}_{\text{LOrPE}}(x_i|h, M), \frac{\tilde{f}^{(+i)}_{\text{LOrPE}}(x_i|h, M)}{n^{\alpha}} \right\}$$

Revisit Exp(1) Example

Improved performance at boundary confirmed by extensive simulations; o/w competitive with (KDE, OSDE, LLDE).



Real Data Example: $n = 86$ Patients (Silverman, 1986)



Summary

- Away from boundary LOrPE acts like KDE with a high-order kernel.

$\Rightarrow$ **Faster asymptotic convergence rates!**

- Close to boundary LOrPE is adaptive:
  - effective kernels naturally change shape to accommodate endpoint;
  - reduces boundary bias.
- For fixed M, LOrPE inherits consistency from:
  - KDE, for $h \to 0$ (away from boundary);
  - OSDE, for $h \to \infty$.
- LOrPE extends to multivariate settings, need only switch to multivariate orthogonal polynomial systems.

Thank You!

[http://arxiv.org/abs/1505.00275]