Smoothing of Variable Bandwidth Kernel Estimate of Heavy-Tailed Density Function

Natalia M. Markovich

Dr.Sci., Senior Scientist
Institute of Control Sciences
Russian Academy of Sciences, Moscow, Russia

4th Conference on Extreme Value Analysis
Probabilistic and Statistical Models and their Applications
August 15-19, 2005, Gothenburg, Sweden

Heavy-tailed density kernel estimation.

Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of i.i.d. r.v. distributed with the heavy-tailed CDF $F(x)$ and the PDF $f(x)$.

Variable bandwidth kernel estimate, Abramson (1982)

$$\boxed{\hat{f}^A(x|h) = (nh)^{-1} \Sigma_{i=1}^n f(X_i)^{1/2} K\left((x - X_i) f(X_i)^{1/2}/h\right)}$$

Practical version

$$\boxed{\widetilde{f}^A(x|h_1, h) = (nh)^{-1} \Sigma_{i=1}^n \hat{f}_{h_1}(X_i)^{1/2} K\left((x - X_i) \hat{f}_{h_1}(X_i)^{1/2}/h\right)}$$

Non-variable bandwidth kernel estimate

$$\boxed{\hat{f}_h(x) = (nh)^{-1} \Sigma_{i=1}^n K((x - X_i)/h)}$$

Mean squared errors:

- $MSE(\hat{f}_h) \sim n^{-4/5}$ ($\boxed{\text{bias} \sim h^2}$; variance $\sim (nh)^{-1}$)

  if a non-variable bandwidth kernel estimator with a second-order kernel as $\hat{f}_h(x)$ is used,

  $$h \sim n^{-1/5},\ f \text{ has two continuous derivatives,}$$

- $MSE(\hat{f}^A(x|h)) \sim n^{-8/9}$ ($\boxed{\text{bias} \sim h^4}$; variance $\sim (nh)^{-1}$)

  if a variable bandwidth kernel estimator with a fourth-order kernel (non-positive) as $\hat{f}_h(x)$ is used,

  $$h \sim n^{-1/9},\ f \text{ has four continuous derivatives.}$$

Disadvantage of a variable bandwidth kernel estimators:

- **they are not intended for the estimation of the density at infinity,**

  at least with compactly supported kernels.

What's new?

A combination of

preliminary data transformation

+ a variable bandwidth kernel estimator

+ data-driven smoothing tool

to provide

- **the consistency of the estimation,**
- **the** MSE **of the fastest achievable order** $n^{-8/9}$**,**
- **good density estimation at infinity,**

is considered.

Smoothing methods and their reliability.

- **Over-smoothing bandwidth selection:**

$$\hat{h}_{OS} = \left(\frac{243R(K)}{35\mu_2(K)^2 n}\right)^{1/5} \cdot s,$$

s **is the sample standard derivation,**

$$\mu_2(K) = \int z^2 K(z)dz, \qquad R(K) = \int K^2(x)dx.$$

- **Cross-validation:**

$$\prod_{i=1}^{n} \hat{f}_{-i}(X_i; h) \longrightarrow \max_h,$$

$$\hat{f}_{-i}(x; h) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i}^{n} K(\frac{x - X_j}{h})$$

- **Least squares cross-validation:**

$$LSCV(h) = n^{-1} \sum_{i=1}^{n} \int \hat{f}_{-i}(x; h)^2 dx - 2n^{-1} \sum_{i=1}^{n} \hat{f}_{-i}(X_i; h) \rightarrow \min_h;$$

Non-consistency of cross-validation on heavy-tailed densities:

$$h \rightarrow \infty \text{ as } n \rightarrow \infty.$$

Consistency of cross-validation for compactly supported densities.

Cross-validation for a variable bandwidth kernel estimator (P.Hall (1992)).

- **Weighted integrated squared error**

$$WISE = \int \breve{f}_{-i}(x;h)^2\omega(x)dx - 2\int \breve{f}_{-i}(x;h)^2 f(x)\omega(x)dx,$$

where

$$\begin{aligned}\breve{f}_{-i}(x;h) &= \frac{1}{nh^p}\sum_{j=1,j\neq i}^{n}\hat{f}_{-i}(X_j,h_1)^{p/2}K\left(\frac{(x-X_j)\hat{f}_{-i}(X_j,h_1)^{1/2}}{h}\right)\\ &\cdot\mathbf{1}\left(|x-X_j|\leq Ah\right), \qquad \forall A>0,\end{aligned}$$

$\omega(x)$ **is a bounded, nonnegative function, e.g.,**

$$\omega(x) = \begin{cases} 1, & \textbf{for} \qquad \|\widehat{\Sigma}^{-1/2}(x-\widehat{\mu})\|^2 \leq z_\eta, \\ 0, & \textbf{otherwise}, \end{cases}$$

where $\widehat{\mu}$ and $\widehat{\Sigma}$ denote the sample mean and variance,

$\|\cdot\|$ is Euclidean distance,

z_η is the upper $(1-\eta)$-level critical point of the chi-squared distribution.

- **Practical version**

$$\widehat{WISE} = \int \breve{f}_{-i}(x;h)^2\omega(x)dx - \frac{2}{n}\sum_{i=1}^{n}\breve{f}_{-i}(X_i;h)^2\omega(X_i)$$

What is h?

General discrepancy method.

The bandwidth h is defined as the solution of the discrepancy equation

$$\boxed{\rho(\widehat{F}_h, F_n) = \delta,}$$

where

$$\widehat{F}_h(x) = \int_{-\infty}^{x} \widehat{f}_h(t)dt,$$

δ **is a known uncertainty of the estimation of the CDF** $F(x)$ **by the empirical CDF** $F_n(t)$**, i.e.** $\delta = \rho(F, F_n)$**,**

$\rho(\cdot, \cdot)$ **is a metric in the space of CDFs,**

*** Markovich (1989); Vapnik, Markovich and Stefanyuk (1992).**

$\boxed{\delta \text{ is a quantile}}$ **of the limit distribution of the Mises-Smirnov statistic**

$$\omega_n^2 = n \int (F_n(x) - F(x))^2 f(x)dx,$$

or Kolmogorov-Smirnov statistic

$$D_n = \sqrt{n} \sup_{-\infty<x<\infty} |F(x) - F_n(x)|$$

Consistency and convergence rate in L_2 of discrepancy method based on Mises-Smirnov statistic

is proved for projection estimators when

- **density is compactly supported,**
- **its kth derivative has a bounded variation.**

Practical version:

h :

$$n\omega_n^2(h) = n \int \left(F_n(x) - \widehat{F}_h(x)\right)^2 f(x)dx = 0.05$$

ω^2**-method**

h :

$$\sqrt{n}D_n(h) = \sqrt{n} \sup_{-\infty<x<\infty} |\widehat{F}_h(x) - F_n(x)| = 0.5$$

D**-method**

0.05 **and** 0.5 **are the maximum likelihood values of** ω_n^2 **and** D_n **statistics, respectively.**

Discrepancy method on finite and heavy-tailed densities. Examples.

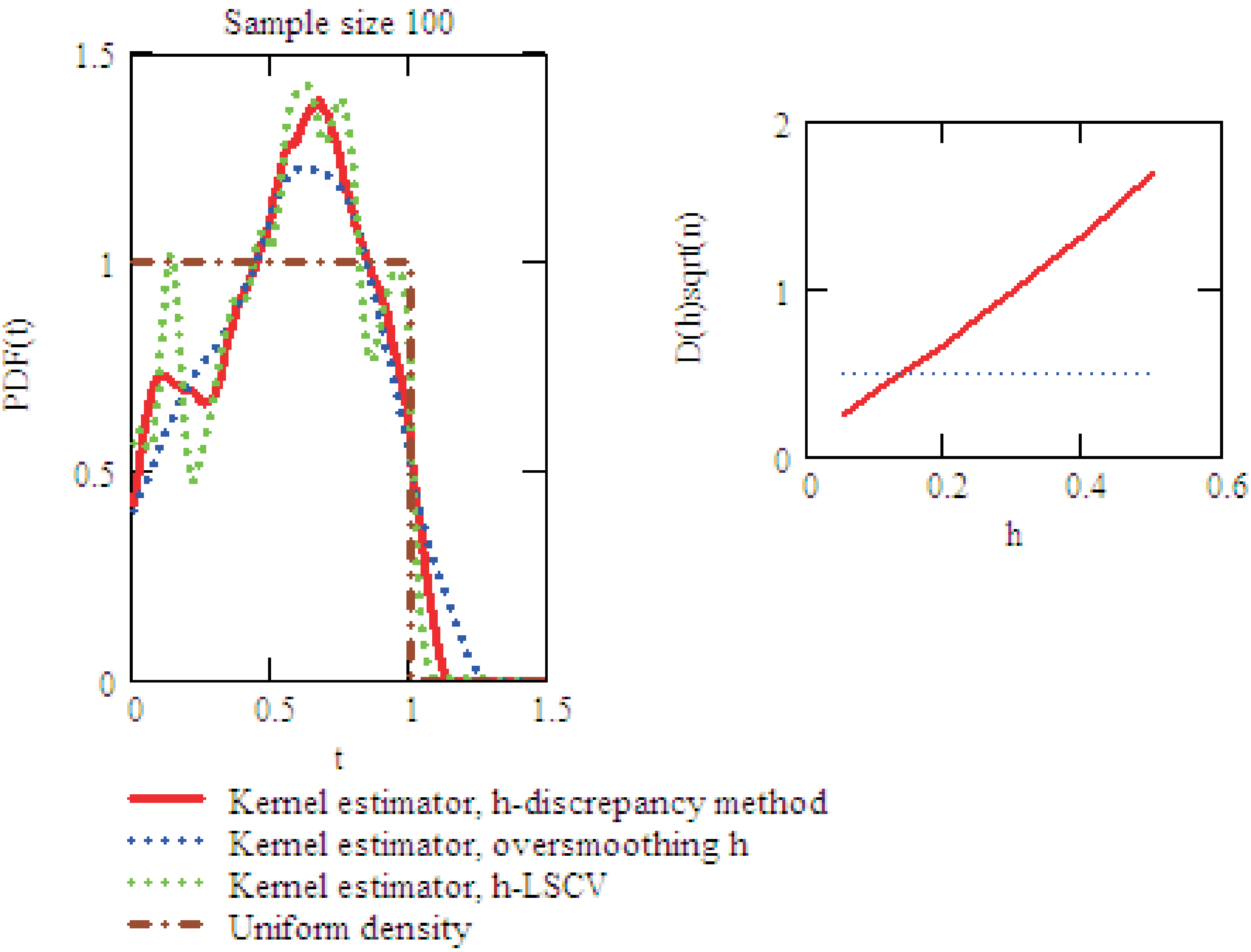

Figure 1: Standard kernel estimates with different smoothing for uniform distribution (left) and the dependence of the statistic $\sqrt{n}D_n(h)$ against h (right) distributions. Discrepancy method: $h = 0.14$. Normal kernel is used in the case of the Least squares cross-validation (LSCV), otherwise Epanechnikov's kernel.

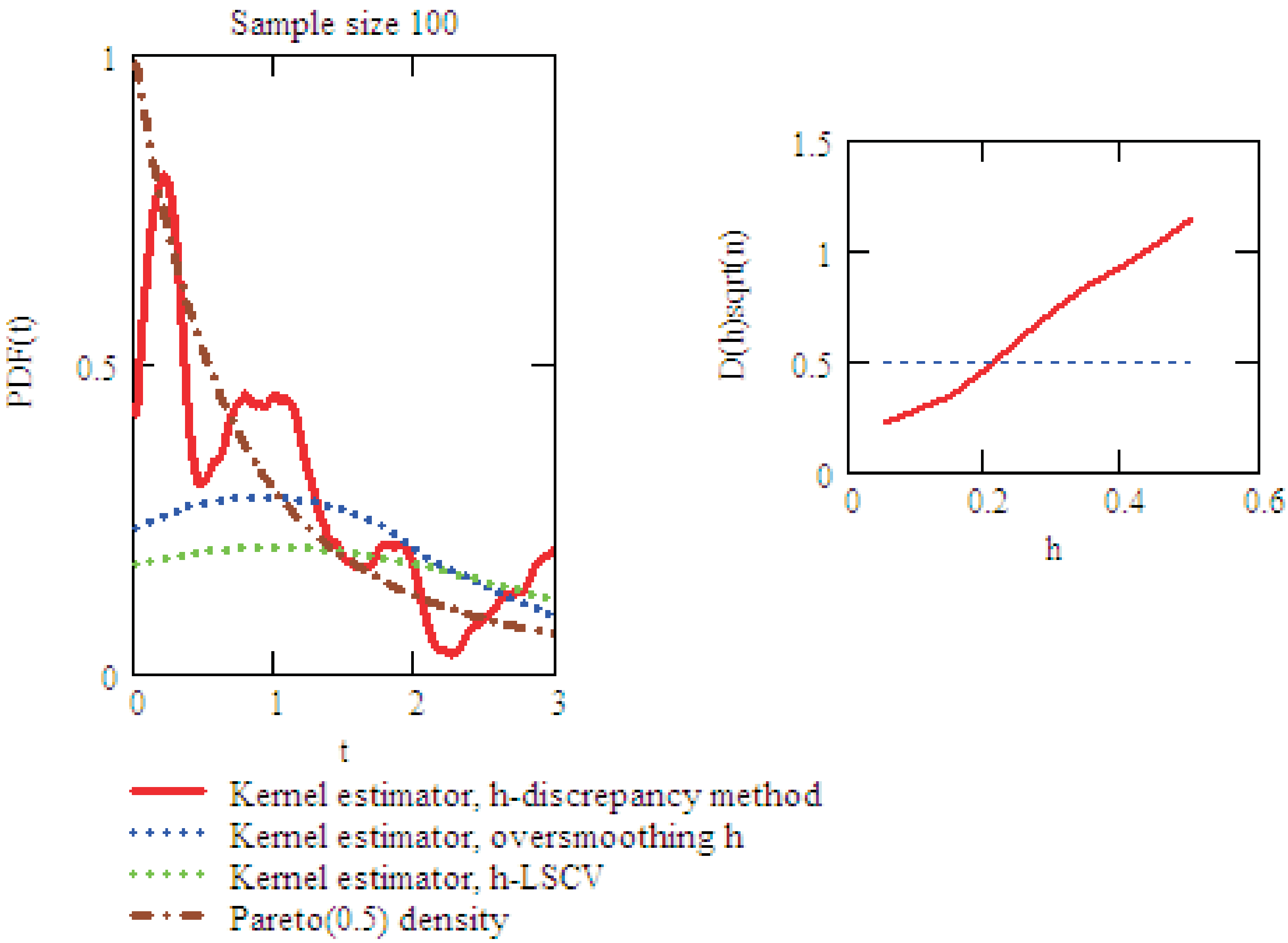

Figure 2: Standard kernel estimates with different smoothing for Pareto distribution (left) and the dependence of the statistic $\sqrt{n}D_n(h)$ against h (right) distributions. Discrepancy method: $h = 0.23$. Normal kernel is used in the case of the Least squares cross-validation (LSCV), otherwise Epanechnikov's kernel.

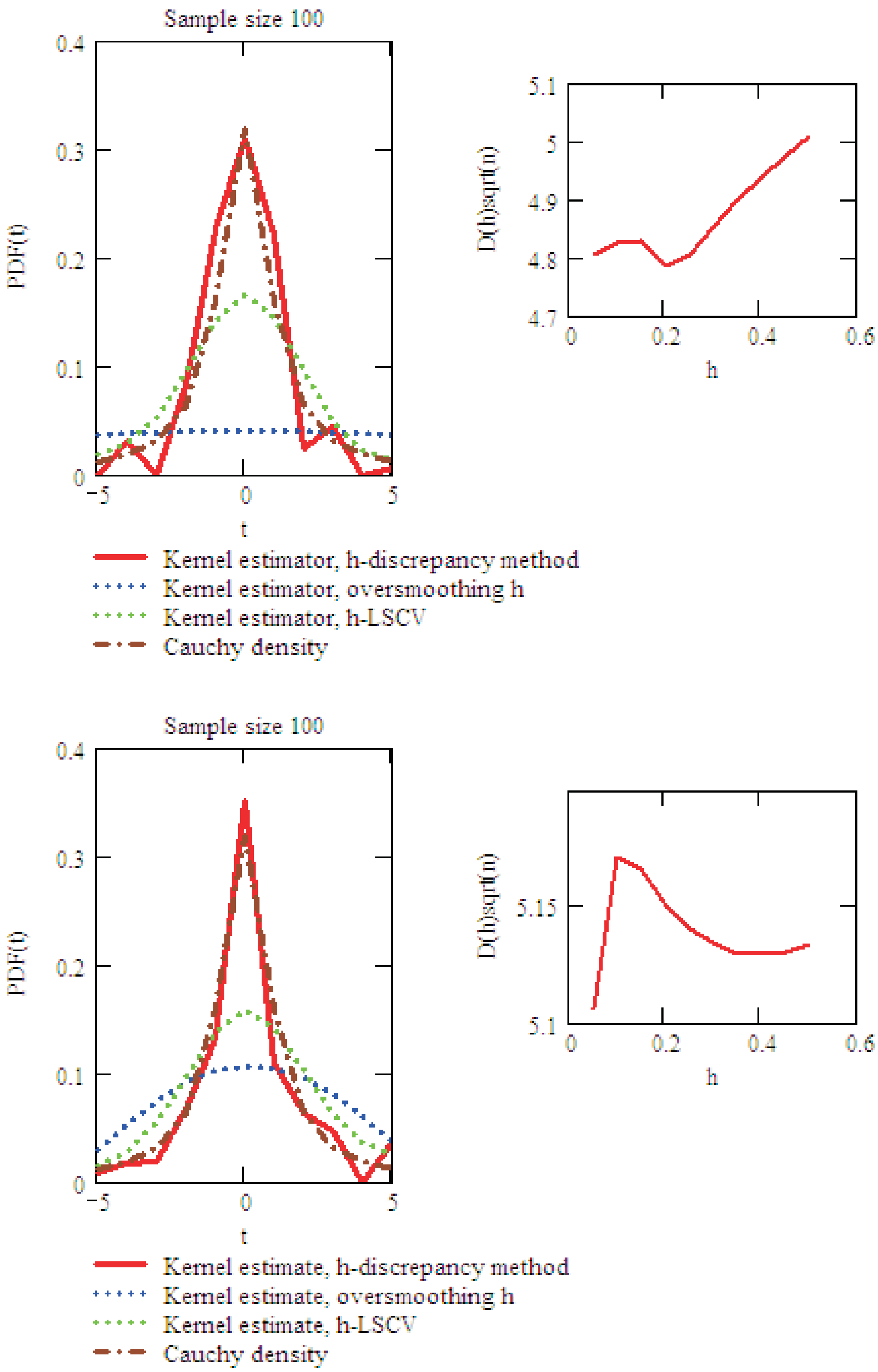

Figure 3: Standard kernel estimates for two Cauchy distributions (left), and the dependence of the statistic $\sqrt{n}D_n(h)$ against h. Discrepancy method, h corresponded to the largest minimum of $\sqrt{n}D_n(h)$ is selected: $h = 0.21$ (top), $h = 0.4$ (bottom). Normal kernel is used in the case of the Least squares cross-validation (LSCV), otherwise Epanechnikov's kernel.

Transformation to $[0,1]$ interval.

$$X_1, ..., X_n \to^T Y_1, ..., Y_n, \qquad Y_j = T(X_j), \qquad j = 1, \ldots, n$$

Let $T(x)$ be a monotone increasing "one-to-one" transformation function (T' is continuous).

The PDF of X_i is estimated by

$\hat{f}(x) = \hat{g}(T(x))T'(x), \qquad g(x)$ **is the PDF of the r.v.** Y_i.

The CDF of the r.v. Y_i is

$$G(x) = \mathbb{P}\{Y_i \le x\} = \mathbb{P}\{T(X_i) \le x\} = F(T^{-1}(x))$$

Fixed transformations: $\ln x$, $2/\pi \arctan x$.

Adapted transformation (Maiboroda & Markovich (2004))

from the Pareto CDF

$$\Psi_{\hat{\gamma}}(x) = \begin{cases} 1 - (1 + \hat{\gamma}x)^{-1/\hat{\gamma}}, & \textbf{if} \quad x \ge 0, \\ 0, & \textbf{if} \quad x < 0. \end{cases}$$

to the triangular distribution

$\Phi^{+tri}(x) = (2x - x^2)\mathbf{1}\{x \in [0,1]\} + \mathbf{1}\{x > 1\}$ **is**

$$\boxed{T_{\hat{\gamma}}(x) = 1 - (1 + \hat{\gamma}x)^{-1/(2\hat{\gamma})},}$$

where $\hat{\gamma}$ is some estimate of the extreme value index γ.

Comparison of re-transformed kernel estimate and variable bandwidth kernel estimate.

Pure variable bandwidth kernel estimator

does not fit the density at infinity

at least with compact supported kernels

in contrast to variable bandwidth kernel estimator

that uses transformation of the data.

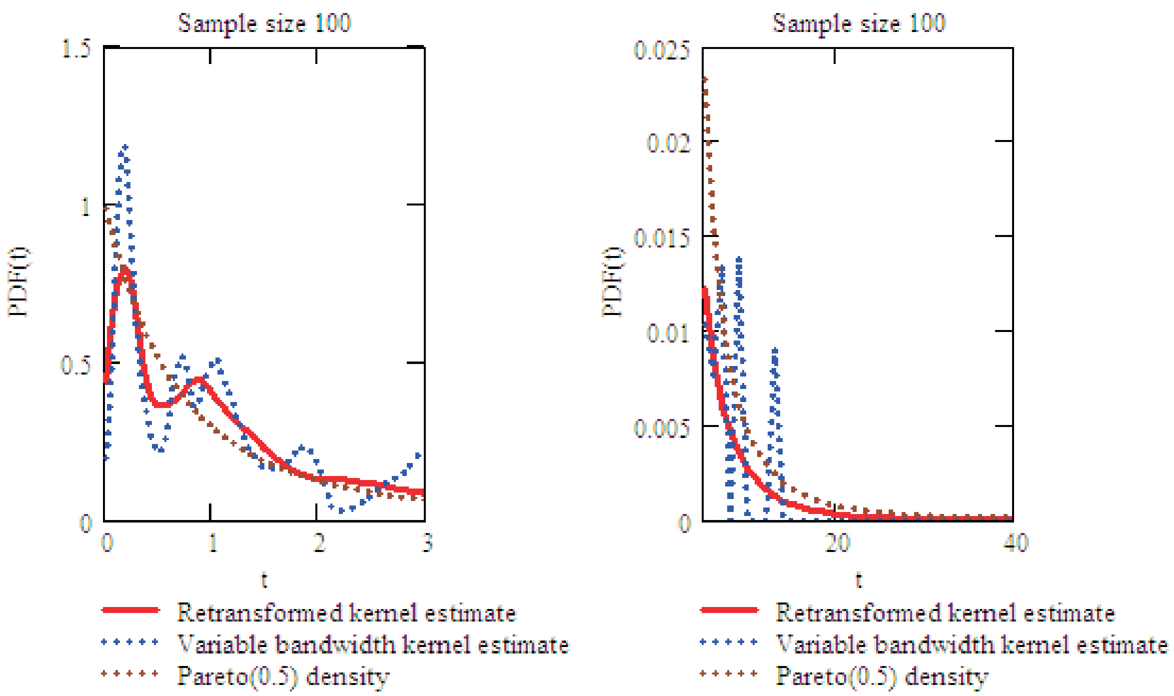

Figure 4: Retransformed standard kernel estimate and variable bandwidth kernel estimate with Epanechnikov's kernel for Pareto distribution: body (left) and tail (right). h is selected by D-method.

Discrepancy method for variable bandwidth kernel estimator.

Let h_* be a solution of the equation

$$\sup_{x \in \Omega^*} |F_n(x) - F^A_{h,h_1}(x)| = \delta n^{-1/2}, \qquad (1)$$

where $\Omega^* \subseteq (-\infty, \infty)$ is some finite interval,

$$F^A_{h,h_1}(x) = \int_{(-\infty,x] \cap \Omega^*} \widetilde{f}^A(t \mid h_1, h) dt,$$

$\widetilde{f}^A(t \mid h_1, h)$ is a variable bandwidth kernel estimator.

The application of (1) requires the **preliminary transform** of the data to some finite interval.

Accuracy of the discrepancy method for variable bandwidth kernel estimator.

Theorem 1 .

Let $X^n = \{X_1, \ldots, X_n\}$ ***be i.i.d. r.v.s with a density*** $f(x)$ ***that is supported at*** $\Omega^* = [0,1]$***. Suppose that*** $f(x)$ ***and*** $1/f(x)$ ***have four continuous derivatives of all types and*** $f(x)$ ***is bounded away from zero, on*** $\Re^\varepsilon$ ***for some*** $\varepsilon > 0$***. We assume that*** K ***is symmetric, continuous and satisfies***

$$K_3 = \int x^4 K(x)dx < \infty, \qquad \sup_x \mid K(x) \mid < \infty, \qquad \int K(x)dx = 1. \tag{2}$$

Let the non-random bandwidth h_1 ***in a pilot standard kernel estimator*** $\hat{f}_{h_1}(x)$ ***be*** $cn^{-1/5}$***. Then at least one of the solutions*** h_* ***of equation (1) obeys the condition***

$$\eta \le h_* n^{1/9} \le \lambda, \qquad \lambda > \eta > 0, \tag{3}$$

with probability 1***.***

Denotations:

$\Re$ **is a compact set of** R**,**

$$\Re^\varepsilon \equiv \{x \in R : \textbf{for some} \qquad y \in \Re, \|x - y\| \le \varepsilon\}, \\ \varepsilon > 0,$$

where $\| \cdot \|$ **is the usual Euclidean norm**

Theorem 2 .

Let the density $f(x)$ ***be estimated by the variable bandwidth kernel estimate*** $\widetilde{f}^A(x|h_1, h)$***.***

Assume the conditions on $f(x)$ ***and*** $K(x)$ ***given in Theorem 1 hold.***

In addition, we assume that $K(x)$ ***vanishes outside a compact set and has two bounded derivatives.***

Let us assume that $\mathbb{E}(Z \cdot \hat{f}^A(x|h)) = 0$***, where*** Z ***is a standard normal r.v., a non-random bandwidth*** h_1 ***in non-variable kernel estimator*** $\widehat{f}_{h_1}(x)$ ***obeys*** $h_1 = cn^{-1/5}$***.***

Then at least one solution h_* ***of the discrepancy equation (1) exists such that***

$$MSE(\widetilde{f}^A(x|h_1, h_*)) = O\left(n^{-8/9}\right)$$

as $n \to \infty$***.***