Bandwidth selection for kernel density estimators of multivariate level sets and highest density regions

Charles R. Doss
(Joint work with Guangwei Weng)

Shape-Constrained Methods: Inference, Applications, and Practice

Jan 31, 2018

Talk outline

- Literature review: bandwidth selection for KDEs, and level sets
- Level set risk approximation
- Highest density regions (HDRs) and HDR risk approximation

KDEs

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}$ and for fixed $h > 0$ define the univariate KDE

$$\widehat{f}_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^{n} K(h^{-1}(x - X_i)).$$

KDEs

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}$ and for fixed $h > 0$ define the univariate KDE

$$\widehat{f}_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^{n} K(h^{-1}(x - X_i)).$$

or, based on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}^d$, for $\boldsymbol{H} > 0$, the multivariate KDE

$$\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i) = \frac{1}{n|\boldsymbol{H}|^{1/2}} \sum_{i=1}^{n} K(\boldsymbol{H}^{-1/2}(\boldsymbol{x} - \boldsymbol{X}_i)).$$

(In both cases, K is some (univariate, multivariate, respectively) kernel density function.)

Bandwidth choice

We have to select h: Least-squares cross-validation (Rudemo, 1982; Bowman, 1984), biased least-squares cross-validation (Scott and Terrell, 1987), smoothed cross validation (Müller, 1985; Staniswalis, 1989; Hall, Marron, and Park, 1992), bootstrap (Taylor, 1989; Faraway and Jhun, 1990; Hall, 1990), direct plug-in methods (Park and Marron, 1992), solve-the-equation direct plug-in methods (Scott, Tapia, and Thompson, 1977; Sheather 1986; Park and Marron, 1990; Sheather and Jones, 1991; Engel, Herrman, and Gasser, 1995).

Bandwidth choice

We have to select h: Least-squares cross-validation (Rudemo, 1982; Bowman, 1984), biased least-squares cross-validation (Scott and Terrell, 1987), smoothed cross validation (Müller, 1985; Staniswalis, 1989; Hall, Marron, and Park, 1992), bootstrap (Taylor, 1989; Faraway and Jhun, 1990; Hall, 1990), direct plug-in methods (Park and Marron, 1992), solve-the-equation direct plug-in methods (Scott, Tapia, and Thompson, 1977; Sheather 1986; Park and Marron, 1990; Sheather and Jones, 1991; Engel, Herrman, and Gasser, 1995).

Many of these methods can be extended to selecting $\boldsymbol{H}$, although there are sometimes added complications. The methods of direct plug-in (Wand and Jones, 1994; Duong and Hazelton, 2003; Chacon and Duong, 09), and smoothed, unbiased, and biased cross-validation (Duong and Hazelton, 2005ab) have been studied.

Density level sets and highest density regions

Density level sets are interesting for many reasons. They have been used for discriminant analysis (classification) (Mammen and Tsybakov, 1999; Duong, Koch, and Wand, 2009), clustering analysis (Hartigan, 1975; Cuevas, Febrero and Fraiman, 2001; Rinaldo and Wasserman, 2010; Jang, 2006), outlier/novelty detection (Lichman and Smyth, 2014; Park, Huang, and Ding, 2010) and in topological data analysis (Wasserman, 2016).

Level set estimation and inference

Level set estimation: Hartigan (1987), Müller and Sawitzki (1991), Polonik (1995), Tsybakov (1997), Walther (1997).

KDE plug-in estimators: Cuevas and Fraiman (1997), Baíllo, Cuesta-Albertos, and Cuevas (2001), Baíllo (2003), Cadre (2006), Mason and Polonik (2009), Jankowski and Stanberry (2012), Mammen and Polonik (2013), Chen, Genovese, and Wasserman (2016).

Level set estimation and inference

Level set estimation: Hartigan (1987), Müller and Sawitzki (1991), Polonik (1995), Tsybakov (1997), Walther (1997).

KDE plug-in estimators: Cuevas and Fraiman (1997), Baíllo, Cuesta-Albertos, and Cuevas (2001), Baíllo (2003), Cadre (2006), Mason and Polonik (2009), Jankowski and Stanberry (2012), Mammen and Polonik (2013), Chen, Genovese, and Wasserman (2016).

In the $d = 1$ case Samworth and Wand (2010) study the problem of highest density region (HDR) estimation. Very recently, Qiao (2017) studies a problem related to level set estimation when $d \geq 1$.

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}$ and for fixed $h > 0$ define the univariate KDE

$$\widehat{f}_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^{n} K(h^{-1}(x - X_i)).$$

Recall: if f_0 is twice continuously differentiable

$$\begin{aligned}
\text{Var}(\widehat{f}_{n,h}(x)) &= \frac{1}{nh} \int K^2(z) f_0(x - hz)dz - \frac{1}{n} E(\widehat{f}_{n,h}(x))^2 \\
&= \frac{1}{nh} f_0(x) R(K) + o(nh)^{-1} \qquad \text{and} \\
E(\widehat{f}_{n,h}(x)) &= \int K(z) f_0(x - hz)dz = f_0(x) + 2^{-1} h^2 f_0''(x) \mu_2(K) + o(h^2)
\end{aligned}$$

as $nh \to \infty$ and $h \searrow 0$. (Here, $R(K) := \int K^2(z)dz$ and $\mu_2(K) := \int zK(z)dz$ depend only on K.)

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}$ and for fixed $h > 0$ define the univariate KDE

$$\widehat{f}_{n,h}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^{n} K(h^{-1}(x - X_i)).$$

Recall: if f_0 is twice continuously differentiable

$$\begin{aligned} \operatorname{Var}(\widehat{f}_{n,h}(x)) &= \frac{1}{nh} \int K^2(z) f_0(x - hz)dz - \frac{1}{n} E(\widehat{f}_{n,h}(x))^2 \\ &= \frac{1}{nh} f_0(x) R(K) + o(nh)^{-1} \qquad \text{and} \\ E(\widehat{f}_{n,h}(x)) &= \int K(z) f_0(x - hz)dz = f_0(x) + 2^{-1} h^2 f_0''(x) \mu_2(K) + o(h^2) \end{aligned}$$

as $nh \to \infty$ and $h \searrow 0$. (Here, $R(K) := \int K^2(z)dz$ and $\mu_2(K) := \int zK(z)dz$ depend only on K.) Thus

$$\text{MSE}(\widehat{f}_{n,h}(x)) = (nh)^{-1} f(x) R(K) + \frac{1}{4} h^4 f''(x)^2 \mu_2(K) + o((nh)^{-1} + h^4) \text{ and so}$$

$$\text{MISE}(\widehat{f}_{n,h}) = (nh)^{-1} R(K) + \frac{1}{4} h^4 \mu_2(K)^2 R(f'') + o((nh)^{-1} + h^4).$$

Bandwidth selection for squared error loss

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} f_0$ on $\mathbb{R}^d$ and for $\boldsymbol{H} \equiv \boldsymbol{H}_{d\times d} > 0$, define the KDE

$$\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) = n^{-1} \sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i) = \frac{1}{n|\boldsymbol{H}|^{1/2}} \sum_{i=1}^{n} K(\boldsymbol{H}^{-1/2}(\boldsymbol{x} - \boldsymbol{X}_i)).$$

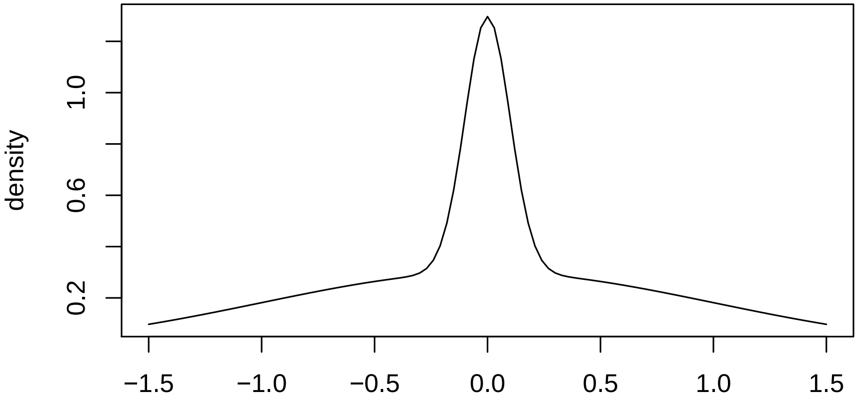
If f_0 is twice continuously differentiable then

$$\begin{aligned} \mathrm{Var}(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x})) &= (n|\boldsymbol{H}|^{1/2})^{-1} f_0(\boldsymbol{x}) R(K) + o(n|\boldsymbol{H}|^{1/2})^{-1} \\ E(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x})) &= f_0(\boldsymbol{x}) + 2^{-1}\,\mathrm{tr}(\boldsymbol{H}\nabla^2 f_0(\boldsymbol{x}))\mu_2(K) + o(\mathrm{tr}(\boldsymbol{H})). \end{aligned}$$

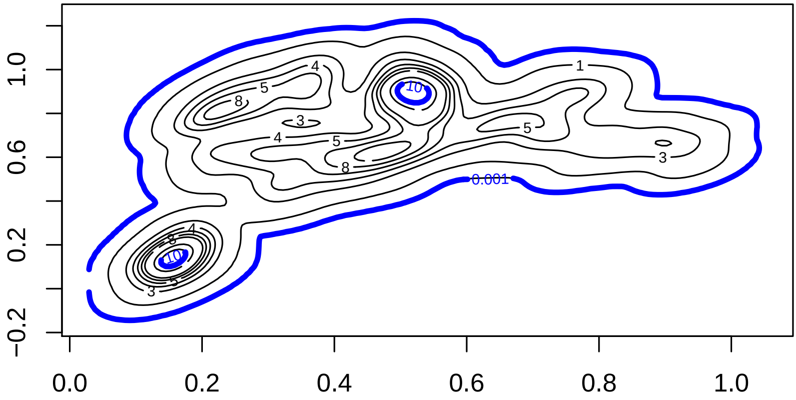
Thus

$$\begin{aligned} \text{MISE}(\widehat{f}_{n,\boldsymbol{H}}) = (n|\boldsymbol{H}|^{1/2})^{-1} R(K) + 4^{-1}\mu_2(K)^2 \int \mathrm{tr}(\boldsymbol{H}\nabla^2 f_0(\boldsymbol{x}))^2 d\boldsymbol{x} \\ + o((n|\boldsymbol{H}|^{1/2})^{-1} + \mathrm{tr}(\boldsymbol{H})^2). \end{aligned}$$

A density ($d = 1$)



A density ($d = 2$)



Bandwidths for KDE level sets

Fix the level $0 < c < \max f_0$. Let

$$\mathcal{L}_c := \{\boldsymbol{x} : f_0(\boldsymbol{x}) \geq c\} \quad \text{and} \quad \widehat{\mathcal{L}}_c := \left\{\boldsymbol{x} : \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq c\right\}.$$

We let

$$L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) = \int_{\mathcal{L}_c \Delta \widehat{\mathcal{L}}_c} f_0(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathbb{R}^d} |\mathbb{1}_{\widehat{\mathcal{L}}_c} - \mathbb{1}_{\mathcal{L}_c}|.$$

The risk is

$$\mathbb{E}L(\widehat{\mathcal{L}}_c, \mathcal{L}_c)$$

Bandwidths for KDE level sets

Fix the level $0 < c < \max f_0$. Let

$$\mathcal{L}_c := \{\boldsymbol{x} : f_0(\boldsymbol{x}) \geq c\} \quad \text{and} \quad \widehat{\mathcal{L}}_c := \left\{\boldsymbol{x} : \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq c\right\}.$$

We let

$$L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) = \int_{\mathcal{L}_c \Delta \widehat{\mathcal{L}}_c} f_0(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathbb{R}^d} |\mathbb{1}_{\widehat{\mathcal{L}}_c} - \mathbb{1}_{\mathcal{L}_c}|.$$

The risk is

$$\mathbb{E}L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) = \int_{\mathbb{R}^d} f_0(\boldsymbol{x}) |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) < c) - \mathbb{1}_{\{f_0(\boldsymbol{x}) < c\}}| d\boldsymbol{x}$$

Bandwidths for KDE level sets

Fix the level $0 < c < \max f_0$. Let

$$\mathcal{L}_c := \{\boldsymbol{x} : f_0(\boldsymbol{x}) \geq c\} \quad \text{and} \quad \widehat{\mathcal{L}}_c := \left\{\boldsymbol{x} : \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq c\right\}.$$

We let

$$L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) = \int_{\mathcal{L}_c \Delta \widehat{\mathcal{L}}_c} f_0(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathbb{R}^d} |\mathbb{1}_{\widehat{\mathcal{L}}_c} - \mathbb{1}_{\mathcal{L}_c}|.$$

The risk is

$$\begin{aligned}\mathbb{E}L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) &= \int_{\mathbb{R}^d} f_0(\boldsymbol{x}) |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) < c) - \mathbb{1}_{\{f_0(\boldsymbol{x}) < c\}}| d\boldsymbol{x} \\ &= f_\tau \int_{\beta^{\delta_n}} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) < c) - \mathbb{1}_{\{f_0(\boldsymbol{x}) < c\}}| d\boldsymbol{x} + \text{error}\end{aligned}$$

where $\beta_c = \{\boldsymbol{x} : f_0(\boldsymbol{x}) = c\}$ and $\beta_c^{\delta_n} := \{\boldsymbol{x} : |\boldsymbol{x} - \beta_c| \leq \delta_n\}$. Here $\delta_n \searrow 0$.

Bandwidths for KDE level sets

We then show that

$$\int_{\beta_c^{\delta_n}} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) < c) - \mathbb{1}_{\{f_0(\boldsymbol{x})<c\}}| d\boldsymbol{x}$$
$$= \int_{\beta_c} \int_{J_n} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y}(\boldsymbol{x},t)) < c) - \mathbb{1}_{\{f_0(\boldsymbol{y}(\boldsymbol{x},t))<c\}}| dt d\mathcal{H}(\boldsymbol{x}) + O(\delta_n^2)$$

where

- $\mathcal{H}$ is $(d-1)$-dimensional Hausdorff measure,
- $\boldsymbol{y}(\boldsymbol{x},t) := \boldsymbol{x} + t u_{\boldsymbol{x}}$ where $u_{\boldsymbol{x}}$ is the outer normal vector to β_c at $\boldsymbol{x}$,
- and $J_n \subset \mathbb{R}$.

Bandwidths for KDE level sets

We then show that

$$\int_{\beta_c^{\delta_n}} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) < c) - \mathbb{1}_{\{f_0(\boldsymbol{x})<c\}}| d\boldsymbol{x}$$
$$= \int_{\beta_c} \int_{J_n} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y}(\boldsymbol{x},t)) < c) - \mathbb{1}_{\{f_0(\boldsymbol{y}(\boldsymbol{x},t))<c\}}| dt d\mathcal{H}(\boldsymbol{x}) + O(\delta_n^2)$$

where

- $\mathcal{H}$ is $(d-1)$-dimensional Hausdorff measure,
- $\boldsymbol{y}(\boldsymbol{x},t) := \boldsymbol{x} + tu_{\boldsymbol{x}}$ where $u_{\boldsymbol{x}}$ is the outer normal vector to β_c at $\boldsymbol{x}$,
- and $J_n \subset \mathbb{R}$.

Then $P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y}(\boldsymbol{x},t) < c)$ can be approximated by a normal probability since $\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y}(\boldsymbol{x},t)) - c$ is approximately $N\left(\mu, \sigma^2\right)$ where

- $\mu = t\|\nabla f_0(\boldsymbol{x})\| + 2^{-1}\mu_2(K)\,\mathrm{tr}(\boldsymbol{H}\nabla^2 f_0(\boldsymbol{x}))$,
- $\sigma^2 = \frac{R(K)c}{n|\boldsymbol{H}|^{1/2}}$

Theorem 1

Let Assumptions D, K, and H hold. Then

$$\mathbb{E}L(\widehat{\mathcal{L}}_c, \mathcal{L}_c) = \mathrm{LS}(\boldsymbol{H}) + o\left\{(n|\boldsymbol{H}|^{1/2})^{-1/2} + \mathrm{tr}(\boldsymbol{H})\right\}$$

as $n \to \infty$*, where*

$$\mathrm{LS}(\boldsymbol{H}) := \frac{c}{\sqrt{n|\boldsymbol{H}|^{1/2}}} \int_{\beta_c} \frac{2\phi(B_{\boldsymbol{x}}(\boldsymbol{H})) + 2\Phi(B_{\boldsymbol{x}}(\boldsymbol{H}))B_{\boldsymbol{x}}(\boldsymbol{H}) - B_{\boldsymbol{x}}(\boldsymbol{H})}{A_{\boldsymbol{x}}} \, d\mathcal{H}(\boldsymbol{x}),$$

$$A_{\boldsymbol{x}} := \frac{\|\nabla f_0(\boldsymbol{x})\|}{\sqrt{R(K)c}}, \quad \textit{and} \quad B_{\boldsymbol{x}}(\boldsymbol{H}) := -\frac{\sqrt{n|\boldsymbol{H}|^{1/2}} D_1(\boldsymbol{x}, \boldsymbol{H})}{\sqrt{R(K)c}},$$

with $D_1(\boldsymbol{x}, \boldsymbol{H}) := \frac{1}{2}\mu(K)\,\mathrm{tr}(\boldsymbol{H}\nabla^2 f_0(\boldsymbol{x}))$*.*

Here, ϕ and Φ are the standard normal density and CDF, respectively.

Level set expansion assumptions

Assumption K: Assumptions on the kernel K.

Level set expansion assumptions

Assumption K: Assumptions on the kernel K.

Assumption D:

1. Assume f_0 has two bounded continuous partial derivatives for $\boldsymbol{x} \in \mathbb{R}^d$ and that f_0 has a continuous third derivative in a neighborhood of β_c.
2. Assume $\inf \|\nabla f_0\| > 0$, where the inf is over an open neighborhood of β_c.
3. The level set β_c is a disjoint union of sets that are each diffeomorphic to the sphere S^{d-1}.

Level set expansion assumptions

Assumption K: Assumptions on the kernel K.

Assumption D:

1. Assume f_0 has two bounded continuous partial derivatives for $\boldsymbol{x} \in \mathbb{R}^d$ and that f_0 has a continuous third derivative in a neighborhood of β_c.
2. Assume $\inf \|\nabla f_0\| > 0$, where the inf is over an open neighborhood of β_c.
3. The level set β_c is a disjoint union of sets that are each diffeomorphic to the sphere S^{d-1}.

Assumption H: Let $\lambda_- \equiv \lambda_{-,n}$, $\lambda_+ \equiv \lambda_{+,n}$ be two positive sequences converging to 0 such that $\lambda_- \leq \lambda_+$. Assume that $n\lambda_-^{d/2}/(\log n)^3 \to \infty$ and $\lambda_+^{(d+8)} n(\log n)^2 = O(1)$. We assume that $\lambda_- \leq \lambda_{\max}(\boldsymbol{H}_n) \leq \lambda_+$.

Highest density regions

What if the level is unknown, specified only through the super level set probability content? Fix $\tau \in (0,1)$. Let $f_{\tau,0}$ be such that

$$\int f_0(\boldsymbol{x}) \mathbb{1}_{\{f_0(\boldsymbol{x}) \geq f_{\tau,0}\}} = 1 - \tau.$$

We refer to estimation of

$$\mathcal{L}_\tau := \{\boldsymbol{x} : f_0(\boldsymbol{x}) \geq f_{\tau,0}\}$$

based on knowledge of τ (but not $f_{\tau,0}$) as *HDR estimation*.

HDR estimation

Like in the LS problem, we can approximate $\mathbb{E}L(\widehat{\mathcal{L}}_\tau, \mathcal{L}_\tau)$ to studying

$$\int_{\beta_c} \int_{J_n} |P(\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y}(\boldsymbol{x}, t)) < \widehat{f}_{\tau,n}) - \mathbb{1}_{\{f_0(\boldsymbol{y}(\boldsymbol{x},t)) < f_{\tau,0}\}}| dt d\mathcal{H}(\boldsymbol{x}).$$

HDR estimation

We estimate $f_{\tau,0}$ by $\widehat{f}_{\tau,n}$, satisfying

$$\int \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \mathbb{1}_{\{\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq \widehat{f}_{\tau,n}\}} = 1 - \tau.$$

HDR estimation

We estimate $f_{\tau,0}$ by $\widehat{f}_{\tau,n}$, satisfying

$$\int \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \mathbb{1}_{\{\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq \widehat{f}_{\tau,n}\}} = 1 - \tau.$$

We can derive

$$\mathbb{E}\widehat{f}_{\tau,n} - f_{\tau,0} = w_0(V_1(\boldsymbol{H}) + V_2(\boldsymbol{H})) + o(\mathrm{tr}(\boldsymbol{H}))$$

where

$$w_0 := \left\{ \int_{\beta_\tau} \frac{1}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}) \right\}^{-1},$$

$$V_1(\boldsymbol{H}) := \int_{\beta_\tau} \frac{D_1(\boldsymbol{x}, \boldsymbol{H})}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}), \quad \text{and} \quad V_2(\boldsymbol{H}) := \frac{1}{f_{\tau,0}} \int_{\mathcal{L}_\tau} D_1(\boldsymbol{x}, \boldsymbol{H}) \, d\boldsymbol{x}.$$

HDR estimation

We estimate $f_{\tau,0}$ by $\widehat{f}_{\tau,n}$, satisfying

$$\int \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \mathbb{1}_{\{\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq \widehat{f}_{\tau,n}\}} = 1 - \tau.$$

We can derive

$$\mathbb{E}\widehat{f}_{\tau,n} - f_{\tau,0} = w_0(V_1(\boldsymbol{H}) + V_2(\boldsymbol{H})) + o(\mathrm{tr}(\boldsymbol{H}))$$

where

$$w_0 := \left\{ \int_{\beta_\tau} \frac{1}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}) \right\}^{-1},$$

$$V_1(\boldsymbol{H}) := \int_{\beta_\tau} \frac{D_1(\boldsymbol{x}, \boldsymbol{H})}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}), \quad \text{and} \quad V_2(\boldsymbol{H}) := \frac{1}{f_{\tau,0}} \int_{\mathcal{L}_\tau} D_1(\boldsymbol{x}, \boldsymbol{H}) \, d\boldsymbol{x}.$$

We can also show that $\mathrm{Var}\widehat{f}_{\tau,n} = o(n^{-1}|\boldsymbol{H}|^{-1/2})$.

HDR estimation

We estimate $f_{\tau,0}$ by $\widehat{f}_{\tau,n}$, satisfying

$$\int \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \mathbb{1}_{\{\widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{x}) \geq \widehat{f}_{\tau,n}\}} = 1 - \tau.$$

We can derive

$$\mathbb{E}\widehat{f}_{\tau,n} - f_{\tau,0} = w_0(V_1(\boldsymbol{H}) + V_2(\boldsymbol{H})) + o(\mathrm{tr}(\boldsymbol{H}))$$

where

$$w_0 := \left\{ \int_{\beta_\tau} \frac{1}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}) \right\}^{-1},$$

$$V_1(\boldsymbol{H}) := \int_{\beta_\tau} \frac{D_1(\boldsymbol{x}, \boldsymbol{H})}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}), \quad \text{and} \quad V_2(\boldsymbol{H}) := \frac{1}{f_{\tau,0}} \int_{\mathcal{L}_\tau} D_1(\boldsymbol{x}, \boldsymbol{H}) \, d\boldsymbol{x}.$$

We can also show that $\mathrm{Var}\widehat{f}_{\tau,n} = o(n^{-1}|\boldsymbol{H}|^{-1/2})$.

This implies that $\mathrm{Cov}(\widehat{f}_{\tau,n}, \widehat{f}_{n,\boldsymbol{H}}(\boldsymbol{y})) = o(n^{-1}|\boldsymbol{H}|^{-1/2})$ for fixed $\boldsymbol{y}$.

Theorem 2

Let Assumptions D, K, and H hold. Fix $0 < \tau < 1$. Then as $n \to \infty$, $\mathbb{E}L(\widehat{\mathcal{L}}_\tau, \mathcal{L}_\tau)$ equals, up to an $o((n|\boldsymbol{H}|^{1/2})^{-1/2} + \mathrm{tr}(\boldsymbol{H}))$ error term,

$$\mathrm{HDR}(\boldsymbol{H}) = \frac{f_{\tau,0}}{\sqrt{n|\boldsymbol{H}|^{1/2}}} \int_{\beta_\tau} \frac{2\phi(C_{\boldsymbol{x}}(\boldsymbol{H})) + 2\Phi(C_{\boldsymbol{x}}(\boldsymbol{H}))C_{\boldsymbol{x}}(\boldsymbol{H}) - C_{\boldsymbol{x}}(\boldsymbol{H})}{A_{\boldsymbol{x}}} \, d\mathcal{H}(\boldsymbol{x})$$

Theorem 2

Let Assumptions D, K, and H hold. Fix $0 < \tau < 1$. *Then as* $n \to \infty$, $\mathbb{E}L(\widehat{\mathcal{L}}_\tau, \mathcal{L}_\tau)$ *equals, up to an* $o((n|\boldsymbol{H}|^{1/2})^{-1/2} + \mathrm{tr}(\boldsymbol{H}))$ *error term,*

$$\mathrm{HDR}(\boldsymbol{H}) = \frac{f_{\tau,0}}{\sqrt{n|\boldsymbol{H}|^{1/2}}} \int_{\beta_\tau} \frac{2\phi(C_{\boldsymbol{x}}(\boldsymbol{H})) + 2\Phi(C_{\boldsymbol{x}}(\boldsymbol{H}))C_{\boldsymbol{x}}(\boldsymbol{H}) - C_{\boldsymbol{x}}(\boldsymbol{H})}{A_{\boldsymbol{x}}} \, d\mathcal{H}(\boldsymbol{x})$$

where $A_{\boldsymbol{x}} := \dfrac{\|\nabla f_0(\boldsymbol{x})\|}{\sqrt{R(K) f_{\tau,0}}}$ *and* $C_{\boldsymbol{x}}(\boldsymbol{H}) := B_{\boldsymbol{x}}(\boldsymbol{H}) + \sqrt{\dfrac{n|\boldsymbol{H}|^{1/2}}{R(K) f_{\tau,0}}} D_2(\boldsymbol{H})$,

$D_2(\boldsymbol{H}) := w_0 \left\{ V_1(\boldsymbol{H}) + V_2(\boldsymbol{H}) \right\}$ *and* $B_{\boldsymbol{x}}(\boldsymbol{H}) := -\dfrac{\sqrt{n|\boldsymbol{H}|^{1/2}} D_1(\boldsymbol{x}, \boldsymbol{H})}{\sqrt{R(K) f_{\tau,0}}}$,

with $D_1(\boldsymbol{x}, H) := \frac{1}{2}\mu_2(K)\, \mathrm{tr}(\boldsymbol{H}\nabla^2 f_0(\boldsymbol{x}))$, *and*

$$w_0 := \left\{ \int_{\beta_\tau} \frac{1}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}) \right\}^{-1},$$

$$V_1(\boldsymbol{H}) := \int_{\beta_\tau} \frac{D_1(\boldsymbol{x}, \boldsymbol{H})}{\|\nabla f_0(\boldsymbol{x})\|} \, d\mathcal{H}(\boldsymbol{x}), \quad \textit{and} \quad V_2(\boldsymbol{H}) := \frac{1}{f_{\tau,0}} \int_{\mathcal{L}_\tau} D_1(\boldsymbol{x}, \boldsymbol{H}) \, d\boldsymbol{x}.$$

Here, we simplify to the case where $\boldsymbol{H} = h^2 \mathrm{Id}_{d\times d}$.

Corollary 3

Let Assumptions D, K, and H hold. Assume further that $f_0(x) = g(\|x\|)$ *and that the function* $g(r)$ *defined for* $r > 0$ *is strictly decreasing on* $[0, \infty)$*. Then there exists a constant* c_0 *depending on* f_0 *and* K *(but not on* n*) such that there is a unique positive number* $h_0 = \operatorname{argmin}_{h\in[0,\infty]} \mathrm{HDR}(h)$ *satisfying*

$$h_0 = c_0 n^{-1/(d+4)} \quad \textit{and} \quad h_{00} = h_0(1 + o(1)) \qquad \textit{as } n \to \infty,$$

where h_{00} *is any minimizer of* $\mathbb{E}L(\widehat{\mathcal{L}}_\tau, \mathcal{L}_\tau)$.

Both LS and HDR depend on unknown constants. To use them, we form pilot estimators. For HDR, we estimate

- $f_{\tau,0}$, β_τ (with bandwidth h_1)
- ∇f_0 (with bandwidth h_2),
- and $\nabla^2 f_0$ (with bandwidth h_3).

Both LS and HDR depend on unknown constants. To use them, we form pilot estimators. For HDR, we estimate

- $f_{\tau,0}$, β_τ (with bandwidth h_1)
- ∇f_0 (with bandwidth h_2),
- and $\nabla^2 f_0$ (with bandwidth h_3).

Corollary 4

Let Assumptions D, D2, K, K2, H, and H2 hold. Assume there exists a constant c_0 (depending on f_0 and K but not on n) such that $h_0 = \operatorname{argmin}_{h \in [0,\infty]} \mathrm{HDR}(h) = c_0 n^{-1/(d+4)}$. Then $\widehat{h}_{\mathrm{HDR}}/h_0 = 1 + O_p(\log n / n^{2/d+4})$.

Assumptions

Assumption D2: Assume f_0 is four times continuously differentiable.

Assumptions

Assumption D2: Assume f_0 is four times continuously differentiable.

Assumption H2: We assume that h_1 is of the order $n^{-1/d+4}$, h_2 is order $n^{-1/d+7}$, h_3 is order $n^{-1/d+9}$.

Computation

Hyndman (1996) proposes a “quantile method” for computing $f_{\tau,0}$ from a known density by Monte Carlo.

Computation

Hyndman (1996) proposes a "quantile method" for computing $f_{\tau,0}$ from a known density by Monte Carlo. Simulate $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M \overset{\text{iid}}{\sim} f_0$; let $\boldsymbol{Y}_i = f_0(\boldsymbol{Z}_i)$; by definition, $P(f_0(Z_i) \geq f_{\tau,0}) = 1 - \tau$, hence $f_{\tau,0}$ is the τ quantile of the distribution of $\boldsymbol{Y}_i$; estimate this by letting $M \to \infty$.

Computation

Hyndman (1996) proposes a "quantile method" for computing $f_{\tau,0}$ from a known density by Monte Carlo. Simulate $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M \overset{\text{iid}}{\sim} f_0$; let $\boldsymbol{Y}_i = f_0(\boldsymbol{Z}_i)$; by definition, $P(f_0(Z_i) \geq f_{\tau,0}) = 1 - \tau$, hence $f_{\tau,0}$ is the τ quantile of the distribution of $\boldsymbol{Y}_i$; estimate this by letting $M \to \infty$.

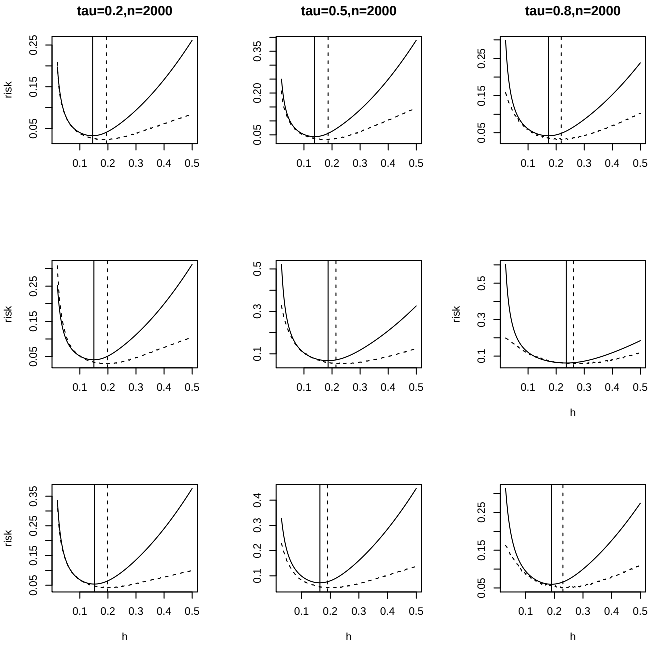
When $d = 2$, we use numerical integration and a binary search.

Computation

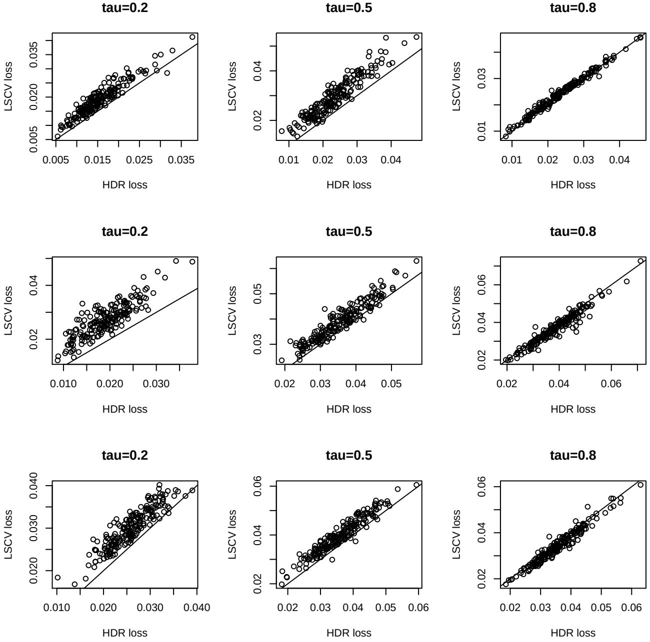
Hyndman (1996) proposes a "quantile method" for computing $f_{\tau,0}$ from a known density by Monte Carlo. Simulate $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M \overset{\text{iid}}{\sim} f_0$; let $\boldsymbol{Y}_i = f_0(\boldsymbol{Z}_i)$; by definition, $P(f_0(Z_i) \geq f_{\tau,0}) = 1 - \tau$, hence $f_{\tau,0}$ is the τ quantile of the distribution of $\boldsymbol{Y}_i$; estimate this by letting $M \to \infty$.

When $d = 2$, we use numerical integration and a binary search.

Optimizing over $\mathrm{LS}(\boldsymbol{H})$ or $\mathrm{HDR}(\boldsymbol{H})$?



tau=0.2,n=2000
tau=0.5,n=2000
tau=0.8,n=2000
risk
h
0.1 0.2 0.3 0.4 0.5


tau=0.2
tau=0.5
tau=0.8
LSCV loss
HDR loss
0.005
0.015
0.020
0.025
0.035
0.01
0.02
0.03
0.04
0.05
0.06
0.010
0.030
0.040

Notes

For some τ's the LSCV will generally be comparable to our method, but for some it will be worse, and don't know which in advance. Unfortunately there were some cases where LSCV beat our method.

Notes

For some τ's the LSCV will generally be comparable to our method, but for some it will be worse, and don't know which in advance. Unfortunately there were some cases where LSCV beat our method.

The V_2 term depends on $\mathcal{L}_\tau$.

Notes

For some τ's the LSCV will generally be comparable to our method, but for some it will be worse, and don't know which in advance. Unfortunately there were some cases where LSCV beat our method.

The V_2 term depends on $\mathcal{L}_\tau$.

The bias of $\widehat{f}_{\tau,n}$ seems to in some cases be poorly approximated (simulation evidence; dependent on τ?). Related to V_2?

Notes

For some τ's the LSCV will generally be comparable to our method, but for some it will be worse, and don't know which in advance. Unfortunately there were some cases where LSCV beat our method.

The V_2 term depends on $\mathcal{L}_\tau$.

The bias of $\widehat{f}_{\tau,n}$ seems to in some cases be poorly approximated (simulation evidence; dependent on τ?). Related to V_2?

Pilot estimators improvable?

Conclusions

- Improvements in our expansion (super level-set integral term)?

- Implementation when $d \geq 3$?
- Multi-stage bandwidth selector?

- Methodology using (bandwidth selection in) level set or HDR estimators?
- Regression?
- $\vdots$

“Fubini Theorem”

Lemma 5

Assume Assumption D holds for f_0*. Let* $\beta := f_0^{-1}(f_{\tau,0})$ *and let* $\delta > 0$ *be no larger than the* reach *of* β*. Let* h *be a bounded function on* β^δ*. Let* $H(\boldsymbol{x}) := \int_{-\delta}^{\delta} h(\boldsymbol{x} + tu_{\boldsymbol{x}})dt$*. Then*

$$\left| \int_{\beta^\delta} h(\boldsymbol{x})d\boldsymbol{x} - \int_\beta H(\boldsymbol{z})d\mathcal{H}(\boldsymbol{z}) \right| \leq C \sup_{\boldsymbol{x}\in\beta} \int_{-\delta}^{\delta} th(\boldsymbol{x} + tu_{\boldsymbol{x}})dt,$$

where C *is a constant depending on* f_0*.*