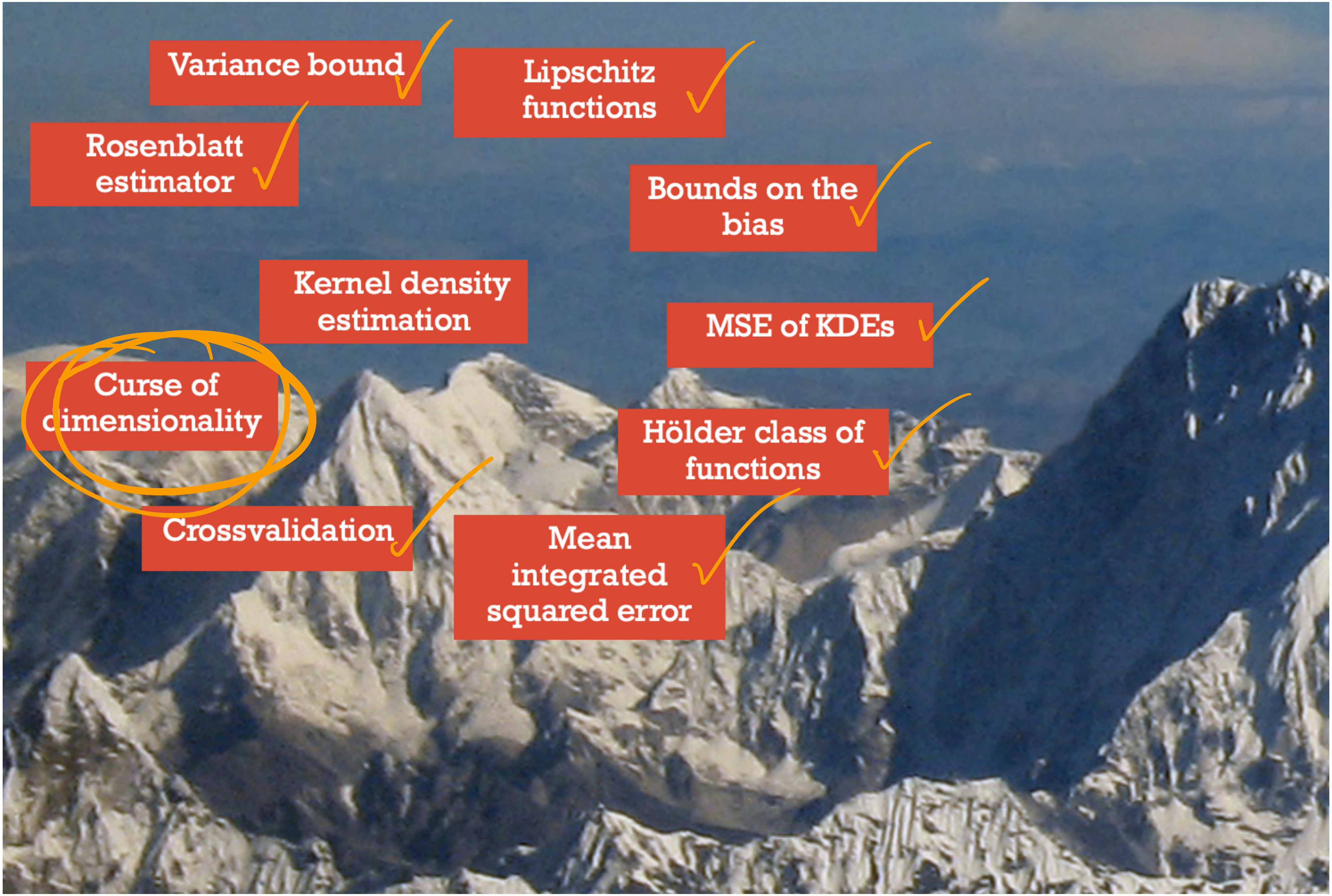
STAT 824 sp 2025 Lec 03 slides

Multivariate kernel density estimation

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.



Variance bound
Lipschitz functions
Rosenblatt estimator
Bounds on the bias
Kernel density estimation
MSE of KDEs
Curse of dimensionality
Hölder class of functions
Crossvalidation
Mean integrated squared error

$X_1, \dots, X_n \in \mathbb{R}^d, \quad d \geq 1$

$X_1, \dots, X_n \overset{ind}{\sim} f, \quad f$ a d-dimensional joint density.

KDE is like:

$$\hat{f}_n(x) = \frac{1}{n|H|^{\frac{1}{2}}} \sum_{i=1}^{n} K\left(H^{-1/2}(X_i - x)\right)$$

$$\int_{\mathbb{R}^d} \hat{f}_n(x)\,dx = \int_{\mathbb{R}^d} \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} K\left(H^{-1/2}(X_i - x)\right) dx$$

$$= \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} \int_{\mathbb{R}^d} K\left(\underbrace{H^{-1/2}(X_i - x)}\right) dx$$

$u = -H^{-1/2}(X_i - x)$

$\underset{\sim}{x} = X_i + H^{1/2} u$

$\left|\frac{d\underset{=}{x}}{d\underset{=}{u}}\right| = |H^{1/2}|$

$$= \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} \int_{\mathbb{R}^d} K(-u)\, |H^{1/2}|\, du$$

$|H|^{1/2} = |H^{1/2}|$

$$= \frac{1}{n} \sum_{i=1}^{n} \underbrace{\int_{\mathbb{R}^d} K(-u)\, du}_{=1}$$

$$= 1.$$

Let $X_1, \dots, X_n \in \mathbb{R}^d$, $d \geq 1$, with pdf $f : \mathbb{R}^d \to \mathbb{R}$.

Multivariate kernel density estimator

A multivariate kernel density estimator for a a pdf $f : \mathbb{R}^d \to \mathbb{R}$ takes the form

$$\hat{f}_{n,H}(x) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} K\left(H^{-1/2}(X_i - x)\right)$$

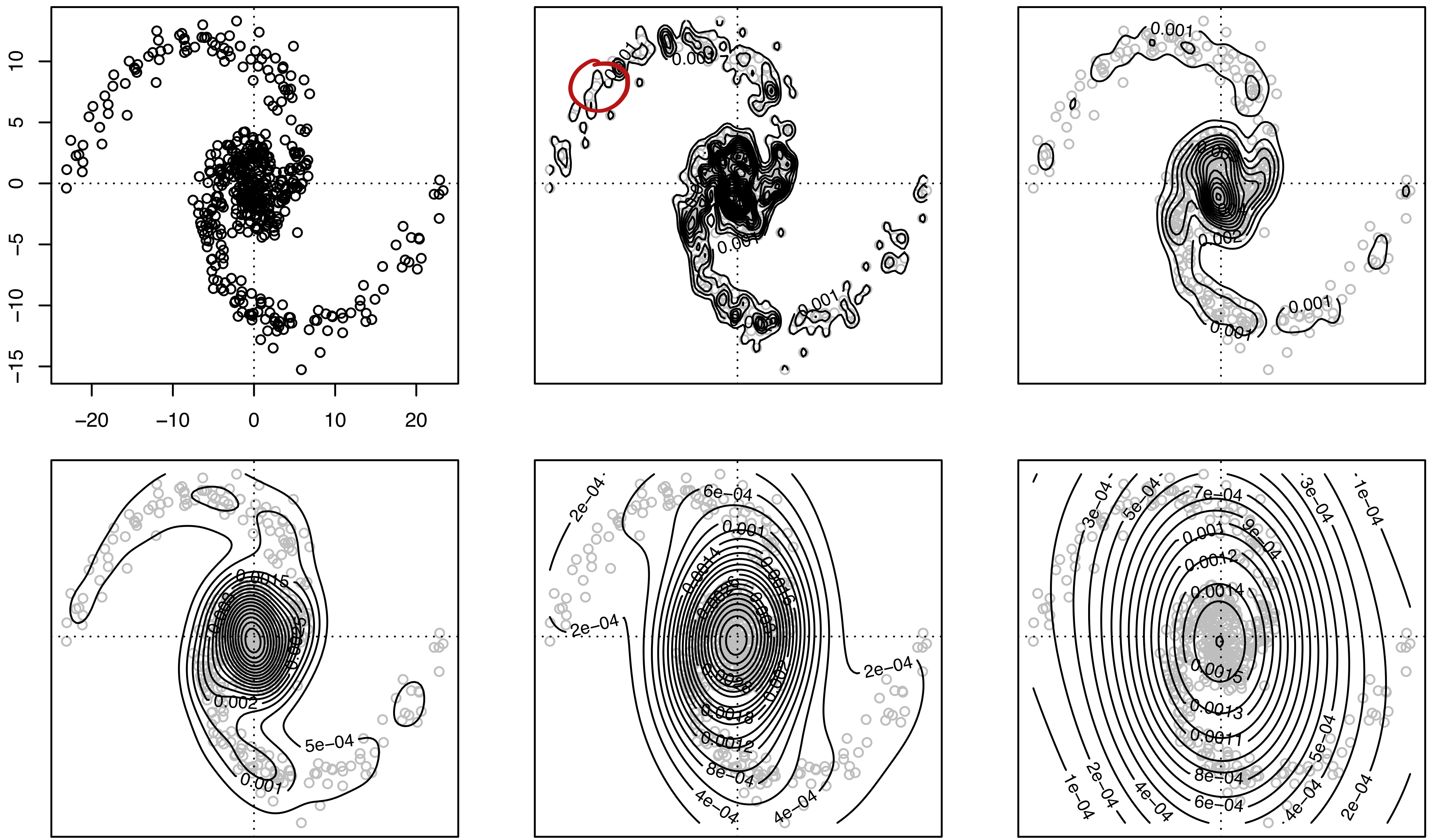
$$\hat{f}_{n,H}(x) = \frac{1}{n} \sum_{i=1}^{n} K_H(X_i - x),$$

where $K_H(u) = |H|^{-1/2} K(H^{-1/2} u)$ for a $K : \mathbb{R}^d \to \mathbb{R}$ and a bandwidth matrix H.

For simplicity we focus on the above estimator under $H = h^2 \cdot \mathbf{I}_d$, which is given by

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K(h^{-1}(X_i - x)).$$

Exercise: Check conditions on K needed to make $\hat{f}_n$ a valid pdf.


10
5
0
−5
−10
−15
−20
−10
0
10
20
0.001
0.002
0.0015
5e−04
2e−04
6e−04
4e−04
8e−04
0.0012
0.0018
3e−04
7e−04
1e−04
0.0014
0.0013
0.0011
0.0015

Example:

1. Generate $n = 500$ realizations of (X, Y) as follows:

   Let $U \sim \text{Uniform}(0, 2\pi)$ and $S \in \{-1, 1\}$ with $P(S = 1) = 1/2$ with $U \perp\!\!\!\perp S$. Then set $R = S \cdot \exp(U/2)$ and let

   $$\begin{aligned} X &= R \cdot \cos(U) + \varepsilon_1 \\ Y &= R \cdot \sin(U) + \varepsilon_2, \end{aligned}$$

   with $\varepsilon_1, \varepsilon_2 \stackrel{\text{ind}}{\sim} \text{Normal}(0, 1)$.

2. Plot data with bivariate KDE

   $$\hat{f}_n(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} \phi\left(\frac{X_i - x}{h}\right) \phi\left(\frac{Y_i - y}{h}\right)$$

   overlaid with bandwidths $h \in \{1/2, 1, 2, 4, 8\}$, $\phi(z) = 1/\sqrt{2\pi} e^{-z^2/2}$.

Now work on bounding $\text{MSE}\,\hat{f}_n(x_0) = b_n^2(x_0) + \sigma_n^2(x_0)$ for $x_0 \in \mathbb{R}^d$.

Assumptions that allow us to bound the variance

(F1) There exists $f_{\max} > 0$ such that $f(x) \leq f_{\max} < \infty$ for all $x \in \mathbb{R}^d$.

(K2) $\int_{\mathbb{R}^d} K^2(u)du \leq \kappa^2 < \infty$.

Bound for the variance of $\hat{f}_n(x_0)$

Under (F1) and (K2) we have

$$\sigma_n^2(x_0) \leq \frac{1}{nh^d} \cdot f_{\max} \cdot \kappa^2$$

for each $x_0 \in \mathbb{R}^d$.

Exercise: Prove the above.

$$\operatorname{Var} \hat{f}_h(x) = \operatorname{Var}\left(\frac{1}{nh^d} \sum_{i=1}^{n} K\left(h^{-1}(X_i - x)\right)\right)$$

$$= \left(\frac{1}{nh^d}\right)^2 \sum_{i=1}^{n} \operatorname{Var} K\left(h^{-1}(X_i - x)\right)$$

$$= \frac{1}{nh^{2d}} \operatorname{Var} K\left(h^{-1}(X_1 - x)\right)$$

$$= \frac{1}{nh^{2d}} \left(\mathbb{E}\, K^2\left(h^{-1}(X_1 - x)\right) - \left(\mathbb{E}\, K\left(h^{-1}(X_1 - x)\right)\right)^2 \right)$$

$$\leq \frac{1}{nh^{2d}} \mathbb{E}\, K^2\left(h^{-1}(X_1 - x)\right)$$

$$= \frac{1}{nh^{2d}} \int_{\mathbb{R}^d} K^2\left(h^{-1}(t - x)\right) f(t)\, dt$$

$$u = h^{-1}(t - x)$$

$$t = x + hu$$

$$\frac{dt}{du} = h I_d$$

$$\left|\frac{dt}{du}\right| = h^d$$

$$= \frac{1}{nh^{2d}} \int_{\mathbb{R}^d} K^2(u) \underbrace{f(x + hu)}_{\leq f_{max}} h^d \, du$$

$$\leq \frac{1}{nh^d} f_{max} \cdot \mathcal{K}^2$$

$$\int_{\mathbb{R}^d} K^2(u)\, du$$

To bound the bias $b_n(x_0)$ we must consider smoothness of f. But first...

Multi-index notation

For a vector of positive integers $\alpha = (\alpha_1, \ldots, \alpha_d)^T$, let

$$|\alpha| = \alpha_1 + \cdots + \alpha_d, \qquad \alpha! = \alpha_1! \cdots \alpha_d!$$

$$x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d} \text{ for all } x \in \mathbb{R}^d, \qquad D^\alpha = \boxed{\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}}$$

$$|x|^\alpha = |x_1|^{\alpha_1} \cdots |x_d|^{\alpha_d}$$

$D^\alpha f(x)$

Can use multi-index notation to:

- write down multivariate Taylor expansions
- define higher-dimensional Hölder classes

Multivariate Taylor Expansion

Let $f : \mathbb{R}^d \to \mathbb{R}$ have partial derivs of order $k+1$ on a convex set T, $x_0, x \in T$. Then

$$f(x) = f_{x_0,k}(x) + R_{x_0,k}(x),$$

where $f_{x_0,k}(x)$ is the *kth order Taylor expansion of f at x_0*, given by

$$f_{x_0,k}(x) = \sum_{|\alpha| \leq k} \frac{D^\alpha f(x_0)}{\alpha!}(x - x_0)^\alpha,$$

and where the remainder (in Lagrangian form) is

$$R_{x_0,k}(x) = \sum_{|\alpha|=k+1} \frac{D^\alpha f((1-\tau)x_0 + \tau x)}{\alpha!}(x - x_0)^\alpha \quad \text{for some } \tau \in (0,1).$$

Exercise: Give 1st-ord Taylor expansion of $f(x) = x_1^2 e^{x_2} + x_1 x_3$ at $x_0 = (1,1,1)^T$.

1^{st}-order Taylor expansion of $f(x) = x_1^2 e^{x_2} + x_1 x_3$ at $\underset{\sim}{x}_0 = (1,1,1)^T$.

$$|\alpha| = 1 \begin{cases} D^{(1,0,0)} f(x) = \frac{\partial}{\partial x_1} f(x) = 2x_1 e^{x_2} + x_3 \\ D^{(0,1,0)} f(x) = \frac{\partial}{\partial x_2} f(x) = x_1^2 e^{x_2} \\ D^{(0,0,1)} f(x) = \frac{\partial}{\partial x_3} f(x) = x_1 \end{cases}$$

$$f(x) = \sum_{|\alpha| \leq 1} \frac{D^{\alpha} f(x_0)}{\alpha!} (x - x_0)^{\alpha} + R$$

$$= \underbrace{f(x_0)}_{|\alpha|=0} + \sum_{|\alpha|=1} \frac{D^{\alpha} f(x_0)}{\alpha!} (x - x_0)$$

$(x_1 - 1)^1 (x_2 - 1)^0 (x_3 - 1)^0$

$$= f(x_0) + \frac{D^{(1,0,0)} f(x_0)}{1!\,0!\,0!} (x - x_0)^{(1,0,0)}$$

$$+ \frac{D^{(0,1,0)} f(x_0)}{0!\,1!\,0!} (x - x_0)^{(0,1,0)}$$

$$+ \frac{D^{(0,0,1)} f(x_0)}{0!\,0!\,1!} (x - x_0)^{(0,0,1)}$$

$$+ R$$

$$= \underbrace{(e+1)}_{f((1,1,1))} + (2e+1)(x_1 - 1) + e(x_2 - 1) + (x_3 - 1) + R$$

Hölder class of functions in d-dimensions

For $T \subset \mathbb{R}^d$, $\beta > 0$ an integer, and $L > 0$, the *Hölder class* of functions $\mathcal{H}(\beta, L)$ on T is the set functions $f : T \to \mathbb{R}$ for which all partial derivatives of up to order $\ell = \beta - 1$ exist, and that

$$|D^\alpha f(x) - D^\alpha f(x')| \leq L\|x - x'\|_2 \text{ for all } x, x' \in T.$$

for all vectors α of positive integers such that $|\alpha| = \ell$.

For $\beta = 2$, the condition is $|\nabla f(x) - \nabla f(x')| < L\|x - x'\|_2$.

We use the notation $\|x\|_2^2 = x_1^2 + \cdots + x_2^2$ for $x \in \mathbb{R}^d$.

Let $\mathcal{P}^d_{\mathcal{H}}(\beta, L)$ denote the set of densities in $\mathcal{H}(\beta, L)$ on $\mathbb{R}^d$, that is, let

$$\mathcal{P}^d_{\mathcal{H}}(\beta, L) = \left\{ f : f \geq 0, \int_{\mathbb{R}^d} f(x)dx = 1, \text{ and } f \in \mathcal{H}(\beta, L) \text{ on } \mathbb{R}^d \right\}.$$

To accommodate the Hölder class of functions in $\mathbb{R}^d$, we need:

d-dimensional Kernel of order ℓ

Let $\ell \geq 1$ be an integer. We call $K : \mathbb{R}^d \to \mathbb{R}$ a *d-dimensional kernel of order ℓ* if the functions $u \mapsto u^\alpha K(u)$, $|\alpha| = 0, 1, \dots, \ell$ are integrable and satisfy

$$\int_{\mathbb{R}^d} K(u)du = 1 \quad \text{and} \quad \int_{\mathbb{R}^d} u^\alpha K(u)du = 0, \quad |\alpha| = 1, \dots, \ell.$$

Exercise: Check the order of the kernel $K : \mathbb{R}^2 \to \mathbb{R}$ given by $K(u) = \phi(u_1)\phi(u_2)$.

We now analyze the bias when $f \in \mathcal{P}_{\mathcal{H}}^{d}(\beta, L)$. We need two assumptions:

Assumptions for bounding the bias when $f \in \mathcal{P}_{\mathcal{H}}^{d}(\beta, L)$

(K1) K is a d-dimensional kernel of order ℓ

(K3) $\int_{\mathbb{R}^d} \|u\|_2 |u|^\alpha |K(u)| du < \infty$ for all $|\alpha| = \ell$

Bound for the bias of $\hat{f}_n(x_0)$

Under (K1) and (K3) and if $f \in \mathcal{P}_{\mathcal{H}}^{d}(\beta, L)$, we have

$$|b_n(x_0)| \leq h^\beta \cdot C$$

for each $x_0 \in \mathbb{R}$, where $C > 0$ is a constant.

Exercise: Prove the above.

$$\text{Bias } \hat{f}_n(x) = \mathbb{E}\,\hat{f}_n(x) - f(x)$$

$$= \mathbb{E}\,\frac{1}{nh^d}\sum_{i=1}^{n} K\left(h^{-1}(X_i - x)\right) - f(x)$$

$$= \frac{1}{h^d}\,\mathbb{E}\,K\left(h^{-1}(X_1 - x)\right) - f(x)$$

$$= \frac{1}{h^d}\int_{\mathbb{R}^d} K\left(h^{-1}(t - x)\right) f(t)\,dt - f(x)$$

$u = h^{-1}(t-x)$
$t = x + hu$
$\left|\frac{dt}{du}\right| = h^d$

$$= \int_{\mathbb{R}^d} K(u)\, f(x+hu)\,du - f(x)$$

$$= \int_{\mathbb{R}^d} K(u)\,[f(x+hu) - f(x)]\,du$$

Taylor Expansion

$$R_{x_0,k}(x) = \sum_{|\alpha|=k+1} \frac{D^\alpha f((1-\tau)x_0 + \tau x)}{\alpha!}(x - x_0)^\alpha$$

$$f(x+hu) = f(x) + \sum_{|\alpha|\le \ell-1} \frac{D^\alpha f(x)}{\alpha!}(hu)^\alpha + \sum_{|\alpha|=\ell} \frac{D^\alpha f(x+\tau hu)}{\alpha!}(hu)^\alpha$$

$\tau \in [0,1]$

$$= \int_{\mathbb{R}^d} K(u)\left[\sum_{|\alpha|\le \ell-1} \frac{D^\alpha f(x)}{\alpha!}(hu)^\alpha + \sum_{|\alpha|=\ell} \frac{D^\alpha f(x+\tau hu)}{\alpha!}(hu)^\alpha\right] du$$

(K1) $\quad \int u^\alpha K(u)\,du = 0 \quad \forall\, |\alpha| \le \ell$

$$= \int_{\mathbb{R}^d} K(u) \sum_{|\alpha|=l} \frac{D^\alpha f(x+\tau h u)}{\alpha!} (hu)^\alpha \, du$$

$$= \sum_{|\alpha|=l} \int_{\mathbb{R}^d} K(u) \frac{(hu)^\alpha}{\alpha!} \left[D^\alpha f(x+\tau h u) - D^\alpha f(x) \right] du$$

$$|D^\alpha f(x) - D^\alpha f(x')| \leq L\|x - x'\|_2$$

$$\left| \text{Bias } \hat{f}_n(x) \right| \leq \sum_{|\alpha|=l} \int_{\mathbb{R}^d} K(u) \frac{h^{|\alpha|} |u|^\alpha}{\alpha!} \left| D^\alpha f(x+\tau h u) - D^\alpha f(x) \right| du$$

Hölder class

$$\leq h^l \sum_{|\alpha|=l} \int_{\mathbb{R}^d} \frac{K(u)|u|^\alpha}{\alpha!} L \|\tau h u\|_2 \, du$$

$$\leq \underbrace{h^{l+1} \sum_{|\alpha|=l} \frac{L}{\alpha!} \underbrace{\int_{\mathbb{R}^d} K(u) |u|^\alpha \|u\|_2 \, du}_{< \infty}}_{C}$$

$\beta = l+1$

$$= h^\beta \cdot C$$

Bound for the MSE of $\hat{f}_n(x_0)$

Under (K1), (K2), (K3), and (F1), and if $f \in \mathcal{P}^d_{\mathcal{H}}(\beta, L)$, we have

$$\text{MSE}\, \hat{f}_n(x_0) \leq h^{2\beta} \cdot C^2 + \frac{1}{nh^d} \cdot f_{\max} \cdot \kappa^2$$

for each $x_0 \in \mathbb{R}$.

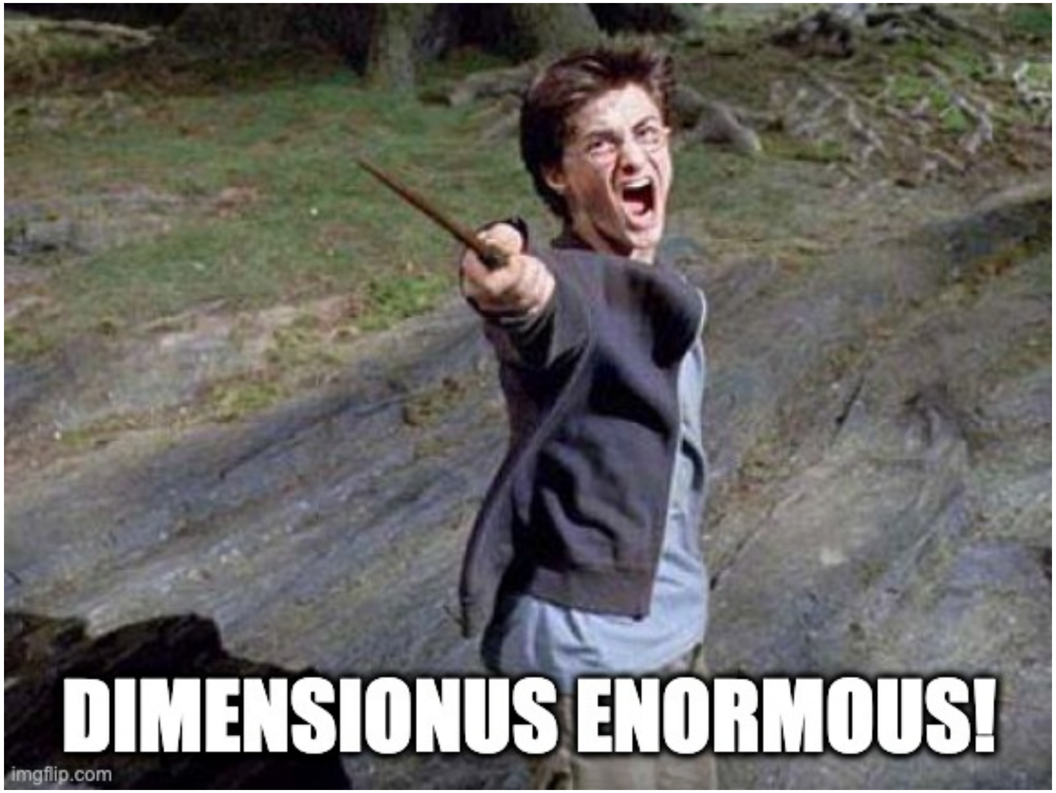
$h_{opt} = c^* n^{-\frac{1}{2\beta+d}}$

Follows from bounds on the bias and variance.

Exercise: Show that the optimal bound in the above result is of the form

$$\boxed{\text{MSE}\, \hat{f}_n(x_0) \leq C \cdot n^{-\frac{2\beta}{2\beta+d}}} \quad \text{for all } x_0 \in \mathbb{R}^d.$$

Discuss: What is the effect of increasing the dimension d on MSE $\hat{f}_n(x_0)$?



The curse of dimensionality:

- Points get very spread out in high-dimensional space.
- Variance explodes (or required sample size to keep variance fixed explodes).

$$\text{MSE } \hat{f}_n(x) \le C\, n^{-\frac{2\beta}{2\beta+d}}$$

$$(\beta=2) \quad n^{-\frac{4}{4+d}}$$

Bound on MSE $\hat{f}_n(x_0)$ when $\beta = 2$:

d	1	2	4	8
Optimal rate	$n^{-4/5}$	$n^{-4/6}$	$n^{-4/8}$	$n^{-4/12}$
To get same MSE bound as when $d = 1$ under n_1	n_1	$n_1^{6/5}$	$n_1^{8/5}$	$n_1^{12/5}$
$n_1 = 100$	100	251	1,585	63,096

sample size under $d=1$

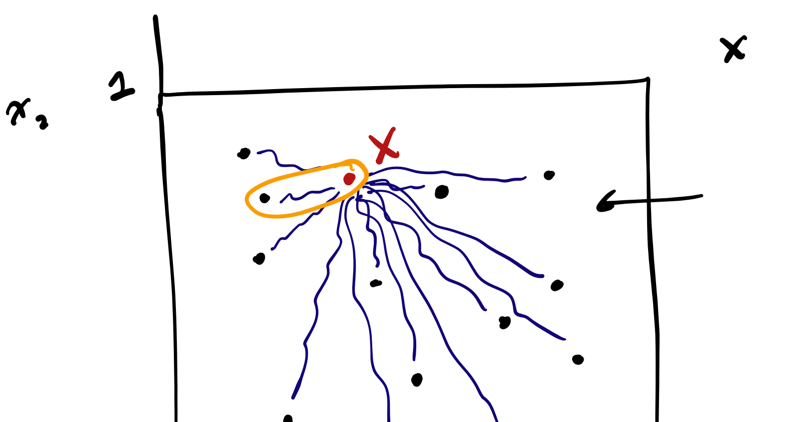
$d=1$

$$\text{MSE} = C \cdot n_1^{-\frac{2\beta}{2\beta+1}}$$

d

$$n_1^{-\frac{2\beta}{2\beta+1}} = n_d^{-\frac{2\beta}{2\beta+d}} \iff n_1^{\frac{2\beta+d}{2\beta+1}} = n_d$$

sample size needed in $d \ge 1$ dimension to match MSE when $d=1$.



$x \quad \underset{2\times 1}{X_1}, \ldots, \underset{2\times 1}{X_n} \in [0,1]^2 ,$

Uniform on $[0,1]^2$

$Y > 0, \quad \mathbb{E} Y = \int_0^\infty [1 - P(Y \le y)] dy$

The following, from [1], explores the far-between-ness of points in high-D space:

Exercise: Let $X, X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Uniform}([0,1]^d)$. Show that

$$\mathbb{E}\left[\min_{1 \le i \le n} \|X - X_i\|_2\right] \ge \frac{d}{d+1}\left[\frac{\Gamma(d/2+1)^{1/d}}{\sqrt{\pi}}\right]\frac{1}{n^{1/d}}.$$

Hint: Use fact that a ball in $\mathbb{R}^d$ with radius t has volume

$$\frac{\pi^{d/2}}{\Gamma(d/2+1)} t^d$$

to show that

$$P\left(\min_{1 \le i \le n} \|X - X_i\|_2 \le t\right) \le n \cdot \frac{\pi^{d/2}}{\Gamma(d/2+1)} t^d.$$

$$P\left(\min_{1 \le i \le n} \| x - x_i \|_2 \le t \right) = P\left(\text{" } \|x - x_i\| \le t \text{ for at least one } i \text{ "} \right)$$

$$= P\left(\bigcup_{i=1}^{n} \{ \|x - x_i\| \le t \} \right)$$

$$\le \sum_{i=1}^{n} P\left(\|x - x_i\| \le t \right)$$

$$= n \, P\left(\|x - x_1\| \le t \right)$$

$$= n \int_{\underset{\sim}{x} \in [0,1]^d} \int_{\underset{\sim}{x}_1 \in [0,1]^d \, \cap \, \|\underset{\sim}{x}_1 - \underset{\sim}{x}\| \le t} 1 \; d\underset{\sim}{x}_1 \, d\underset{\sim}{x}$$

$$\le n \int_{\underset{\sim}{x} \in [0,1]^d} \int_{\|\underset{\sim}{x}_1 - \underset{\sim}{x}\|_2 \le t} 1 \; d\underset{\sim}{x}_1 \, d\underset{\sim}{x}$$

$$= n \int_{x \in [0,1]^d} \left(\text{Volume of ball in } \mathbb{R}^d \text{ with radius } t \right) dx$$

$$= n \left(\text{Volume of ball in } \mathbb{R}^d \text{ with radius } t \right)$$

$$= n \, \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2} + 1\right)} \, t^d$$

Now

$$\mathbb{E}\left[\min_{1\le i\le n}\|X-X_i\|\right] = \int_0^\infty \left[1 - \underbrace{P\left(\min_{1\le i\le n}\|X-X_i\| \le t\right)}\right] dt$$

$$n\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}t^d \le 1$$

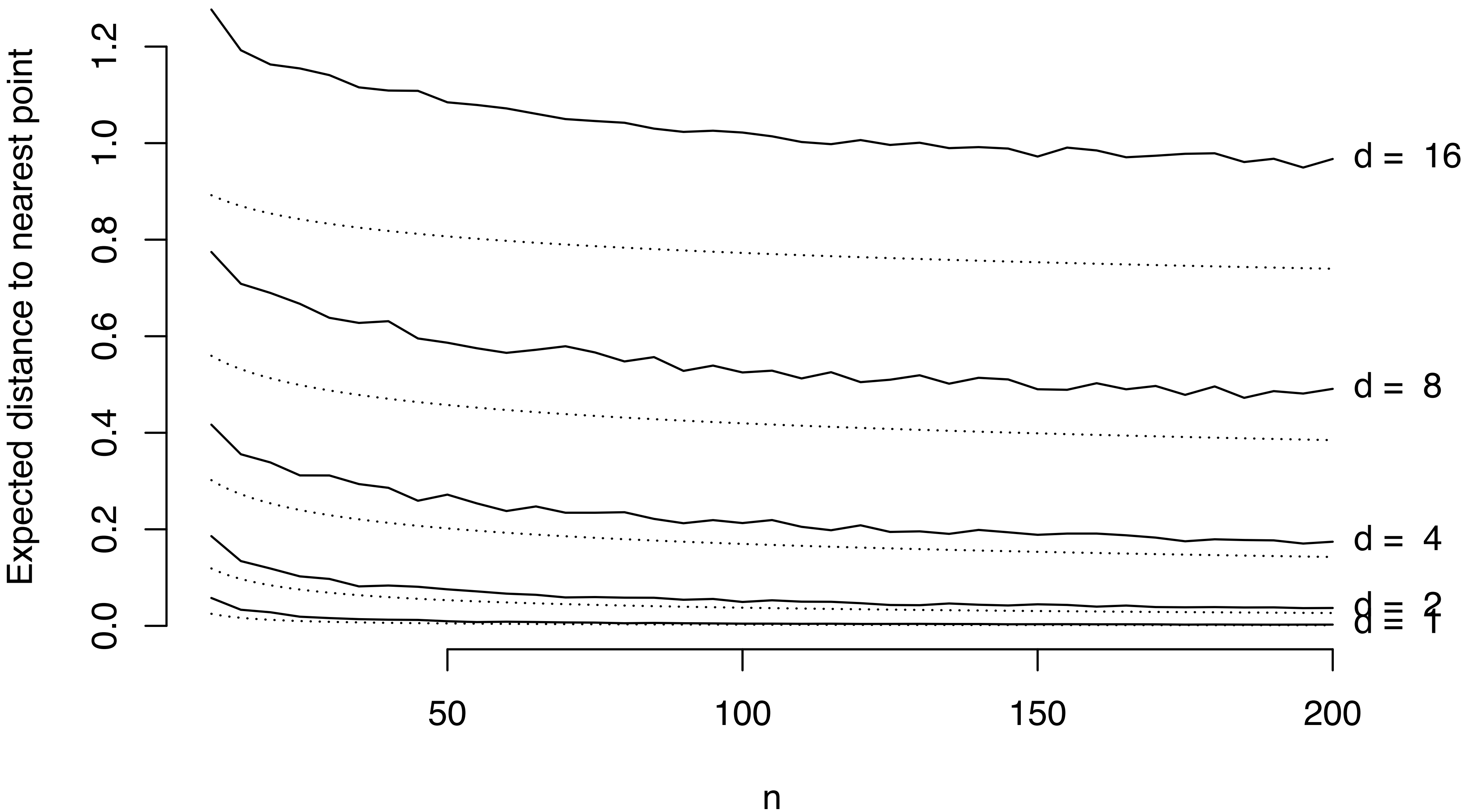
$$\le \int_0^{n^{-\frac{1}{d}}\frac{\left(\Gamma\left(\frac{d}{2}+1\right)\right)^{1/d}}{\pi^{1/2}}} \left[1 - n\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}t^d\right] dt$$

$$\Leftrightarrow t^d = \frac{1}{n}\frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{d/2}}$$

$$= n^{-\frac{1}{d}}\frac{\left(\Gamma\left(\frac{d}{2}+1\right)\right)^{1/d}}{\pi^{1/2}} - \frac{n\frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)}\left(n^{-\frac{1}{d}}\frac{\left(\Gamma\left(\frac{d}{2}+1\right)\right)^{1/d}}{\pi^{1/2}}\right)^{d+1}}{d+1}$$

$$\Leftrightarrow t = \frac{1}{n^{\frac{1}{d}}}\frac{\left[\Gamma\left(\frac{d}{2}+1\right)\right]^{1/d}}{\pi^{1/2}}$$

$\vdots$


Expected distance to nearest point
0.0
0.2
0.4
0.6
0.8
1.0
1.2
50
100
150
200
n
d = 16
d = 8
d = 4
d = 2
d = 1

László Györfi, Michael Kohler, Adam Krzyzak, and Harro Walk.
A distribution-free theory of nonparametric regression.
Springer Science & Business Media, 2006.