

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

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

Multivariate kernel density estimator

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

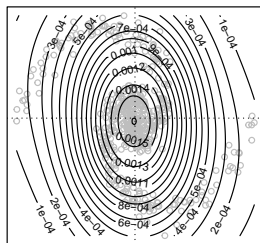
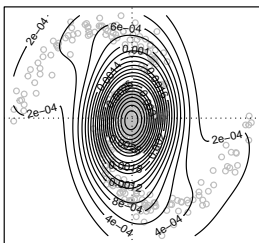
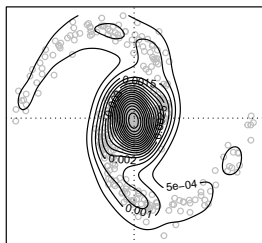
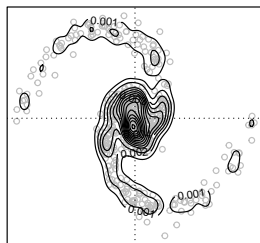
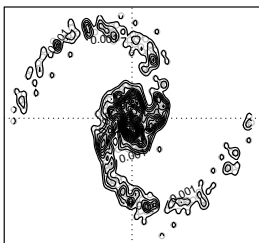
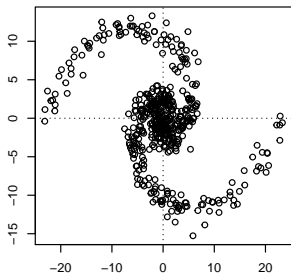
$$\hat{f}_{n,H}(x) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^n K(H^{-1/2}(X_i - x)),$$

for some function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ and positive definite bandwidth matrix H .

For simplicity we focus on the above estimator under $H = h^2 \mathbf{I}_d$, 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.



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

$$X = R \cdot \cos(U) + \varepsilon_1$$

$$Y = R \cdot \sin(U) + \varepsilon_2,$$

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.

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 $\alpha = (\alpha_1, \dots, \alpha_d)^T$, let

$$|\alpha| = \alpha_1 + \dots + \alpha_d,$$

$$\alpha! = \alpha_1! \dots \alpha_d!$$

$$x^\alpha = x_1^{\alpha_1} \dots x_d^{\alpha_d} \text{ for all } x \in \mathbb{R}^d,$$

$$D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}$$

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

Can use multi-index notation to:

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

Multivariate Taylor Expansion

Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $x_0 = (1, 1, 1)^T$.

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 $f : T \rightarrow \mathbb{R}$ for which all partial derivatives of order up to $\ell = \beta - 1$ exist and

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

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

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

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

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 $K : \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $K : \mathbb{R}^2 \rightarrow \mathbb{R}$ 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.

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

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

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

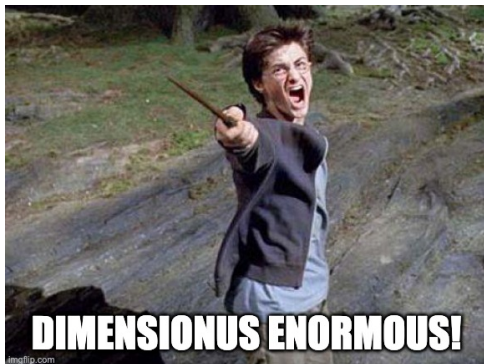
for each $x_0 \in \mathbb{R}$.

Follows from bounds on the bias and variance.

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

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

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



The curse of dimensionality:

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

Bound on $\text{MSE } \hat{f}_n(x_0)$ 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

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

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

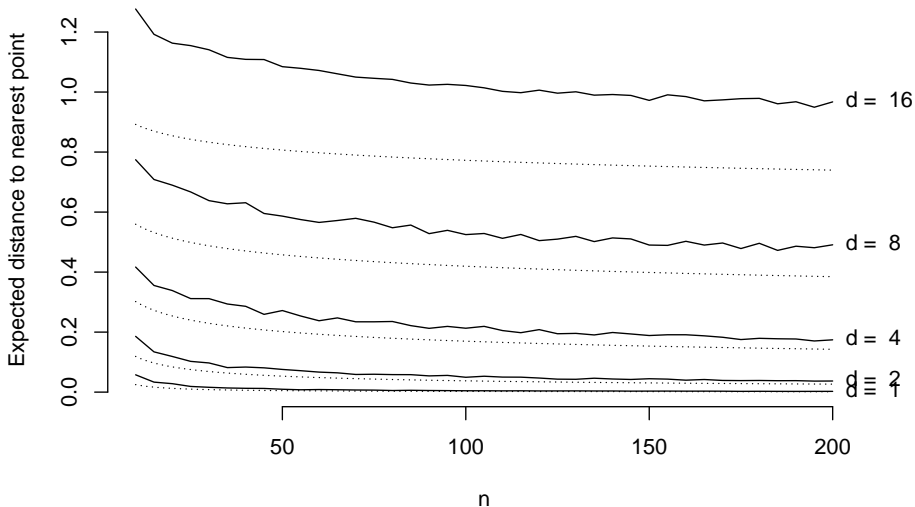
$$\mathbb{E} \left[\min_{1 \leq i \leq n} \|X - X_i\|_2 \right] \geq \frac{d}{d+1} \left[\frac{\Gamma(d/2 + 1)^{1/d}}{\sqrt{\pi}} \right] \frac{1}{n^{1/d}}.$$

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

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





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