MULTIVARIATE KERNEL DENSITY ESTIMATION

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

Let $\|x\|_2^2 = x_1^2 + \dots + x_d^2$ for $x \in \mathbb{R}^d$.

Consider a multivariate KDE of the form

$$\hat{f}_{n,h_1,\dots,h_d}(x) = \frac{1}{n(h_1 \cdot \dots \cdot h_d)} \sum_{i=1}^{n} \prod_{\ell=1}^{d} K\left(\frac{X_{i\ell} - x_\ell}{h_\ell}\right),$$

where $K: \mathbb{R} \to \mathbb{R}$ and $h_1, \dots, h_d$ are bandwidths.

<u>Exercise:</u> Check whether $\hat{f}_n(x)$ is a legitimate pdf. What assumptions are needed on K?

<u>Solution:</u> We see that $K \geq 0 \Rightarrow \hat{f}_n \geq 0$. Next,

$$\int_{\mathbb{R}^d} \hat{f}_n(x)\,dx = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \frac{1}{n \prod_{\ell=1}^{d} h_\ell} \sum_{i=1}^{n} \prod_{\ell=1}^{d} K\left(\frac{X_{i\ell} - x_\ell}{h_\ell}\right) dx_1 \cdot \dots \cdot dx_d$$

$$= \frac{1}{n} \sum_{i=1}^{n} \prod_{\ell=1}^{d} \int_{\mathbb{R}} \frac{1}{h_\ell} K\left(\frac{X_{i\ell} - x_\ell}{h_\ell}\right) dx_\ell$$

$u_\ell = \frac{X_{i\ell} - x_\ell}{h_\ell}$

$$= \frac{1}{n} \sum_{i=1}^{n} \prod_{\ell=1}^{d} \int_{\mathbb{R}} K(u_\ell)\,du_\ell \qquad dx_\ell = -h_\ell\,du_\ell$$

$$= 1$$

provided $\int_{\mathbb{R}} K(u)\,du = 1$.

E.g. $K(z) = (2\pi)^{-1/2} e^{-z^2/2}$ or $K(z) = \frac{3}{4}(1 - z^2)\,\mathbb{1}(|z| \leq 1)$.

More general version of multivariate KDE is

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

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

We will consider this estimator under $H = h \cdot I_d$, which becomes

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

<u>Bounding MSE $\hat{f}_n(x_0)$ at each $x_0 \in \mathbb{R}^d$</u>

<u>Variance</u>: To get an upper bound for $\sigma_n^2(x_0) = \mathrm{Var}\, \hat{f}_n(x_0)$, assume.

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

(F1) $\quad f(x) \leq f_{max} < \infty \quad$ for all $x \in \mathbb{R}^d$.

<u>Result</u>: Under (K2) and (F1), we have

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

<u>Proof</u>: We have

$$\mathrm{Var}\, \hat{f}_n(x_0) = \frac{1}{n^2 h^{2d}} \sum_{i=1}^{n} \mathrm{Var}\, K\left(h^{-1}(X_i - x)\right)$$

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

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

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

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

$u = h^{-1}(z - x) \quad z = x + hu$

$\frac{\partial z}{\partial u} = h \cdot I_d \,, \quad \left|\frac{\partial z}{\partial u}\right| = h^d$

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

$$\le \frac{1}{nh^d} K^2 \cdot f_{max}. \qquad \square$$

__Bias:__ Let $b_n(x_0) = \mathbb{E} \hat{f}_n(x_0) - f(x_0)$.

Again, the bias is much trickier, because we must consider the wiggliness of the true density.

In the multivariate setting, we need to expand our definition of the Hölder class to $d \ge 1$.

We will make use of Multi-index notation:

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

$$|\alpha| = \alpha_1 + \ldots + \alpha_d$$

$$\alpha! = \alpha_1! \cdot \ldots \cdot \alpha_d!$$

$$x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}, \qquad \text{for} \quad x \in \mathbb{R}^d$$

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

Now we can define the multivariate Hölder class.

<u>Defn:</u> Let $T \subset \mathbb{R}^d$, $\beta > 0$ be an integer, and $L > 0$.
The Hölder class $\mathcal{H}(\beta, L)$ on T is the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ of which all partial derivatives of order $\ell = \beta - 1$ exist and satisfy

$$|D^\alpha f(x) - D^\alpha f(x')| \leq L \| x - x' \|_2 \qquad \forall \; x, x' \in T.$$

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

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

$$\nabla f(x) = \left(\frac{\partial}{\partial t_1} f(t), \ldots, \frac{\partial}{\partial t_d} f(t) \right)^T \Big|_{t = x}.$$

Let the set of all pdfs belonging to $\mathcal{H}(\beta, L)$ on $\mathbb{R}^d$ be represented by

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

To analyze the bias, we need a multivariate Taylor expansion:

Multivariate Taylor Expansion:

Suppose $f : \mathbb{R}^d \to \mathbb{R}$ has partial derivatives of order $k+1$ defined on a convex set S. Then if $x_0 \in S$ and $x_0 + t \in S$,

$$f(x_0+t) = \sum_{|\alpha| \le k} \frac{D^\alpha f(x_0)}{\alpha!} t^\alpha + R_{x_0,k}(t),$$

where the Lagrange form of the remainder is

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

Exercise: Let $f(x) = x_1^2 e^{x_2} + x_1 x_3$. Find 1^{st} order Taylor expansion around $x_0 = (1,1,1)^T$ evaluated at $x = (2,2,2)^T$.

Defn: 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

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

In the following we will use the assumptions

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

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

where $|u|^\alpha = |u_1|^{\alpha_1} \cdot \ldots \cdot |u_d|^{\alpha_d}$,

with ℓ and β taken from the relevant Hölder class.

Result: Under (K1) and (K3) and if $f \in \mathcal{P}_H^d(\beta, L)$, then

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

for some constant $C > 0$ depending on β, L, and the kernel K.

Proof: We have

$$b_n(x_0) = \mathbb{E}\, \frac{1}{nh^d} \sum_{i=1}^n K\left(h^{-1}(X_i - x_0)\right) - f(x_0)$$

$$= \frac{1}{h^d}\, \mathbb{E}\, K\left(h^{-1}(X_1 - x_0)\right) - f(x_0)$$

$$= \frac{1}{h^d} \int_{\mathbb{R}^d} K\left(h^{-1}(z - x_0)\right) f(z)\, dz - f(x_0)$$

$u = h^{-1}(z - x_0) \iff z = x_0 + hu$, $\left|\frac{\partial z}{\partial u}\right| = |h \cdot I_d| = h^d$

$$= \int_{\mathbb{R}^d} K(u)\, f(x_0 + hu)\, du - f(x_0)$$

by (K1): $\int_{\mathbb{R}^d} K(u)\, du = 1$

$$= \int_{\mathbb{R}^d} K(u) \left[f(x_0 + hu) - f(x_0) \right] du$$

Taylor expansion

$$= \int_{\mathbb{R}^d} K(u) \left[\sum_{1 \leq |\alpha| \leq \ell - 1} \frac{D^\alpha f(x_0)}{\alpha!} (hu)^\alpha + \sum_{|\alpha| = \ell} \frac{D^\alpha f(x_0 + \tau hu)}{\alpha!} (hu)^\alpha \right] du$$

(for some $\tau \in (0,1)$)

By (K1): $\int_{\mathbb{R}^d} u^\alpha K(u)\, du = 0$ for $|\alpha| = 1, \dots, \ell$

$$= \int_{\mathbb{R}^d} K(u) \sum_{|\alpha| = \ell} \frac{D^\alpha f(x_0 + \tau hu)}{\alpha!} (hu)^\alpha\, du$$

$$= h^\ell \sum_{|\alpha| = \ell} \frac{1}{\alpha!} \int_{\mathbb{R}^d} u^\alpha K(u) \left[D^\alpha f(x_0 + \tau hu) - D^\alpha f(x_0) \right] du$$

$f \in \mathcal{H}(\beta, L)$

$$\leq h^{\ell} \sum_{|\alpha|=\ell} \frac{1}{\alpha!} \int_{\mathbb{R}^d} |u|^{\alpha} |K(u)| \, L \, \|\tau h u\|_2 \, du$$

$\beta = \ell + 1$

$$\leq h^{\beta} \cdot L \cdot \underbrace{\sum_{|\alpha|=\ell} \frac{1}{\alpha!} \int_{\mathbb{R}^d} \|u\|_2 \, |u|^{\alpha} |K(u)| \, du}_{=:C}$$

$$= h^{\beta} \cdot C. \qquad \square$$

The next result gives the MSE of $\hat{f}_n(x_0)$.

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

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

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

Exercise: Find the optimal choice of h.

Solution: $$\frac{\partial}{\partial h}\left(\text{Bound on MSE } \hat{f}_n(x_0)\right) = 2\beta \, h^{2\beta-1} \cdot C^2 - \frac{d}{nh^{d+1}} f_{max} \cdot K^2 = 0.$$

$$\Longleftrightarrow \quad h^{2\beta+d} = \frac{1}{n} \frac{d \cdot f_{max} \cdot K^2}{2\beta \cdot C^2}$$

$$\Longleftrightarrow \quad h_{opt} = n^{-\frac{1}{2\beta+d}} \left(\frac{d \cdot f_{max} \cdot K^2}{2\beta \cdot C^2} \right)^{\frac{1}{2\beta+d}}.$$

Exercise: Plug in the optimal h to obtain the minimum MSE bound.

Solution: We obtain $\text{MSE } \hat{f}_n(x_0) \leq n^{-\frac{2\beta}{2\beta+d}} \cdot C_1$ for some $C_1 > 0$.

What happens to $\text{MSE } \hat{f}_n(x_0)$ as d increases?

Understanding the effects of $d > 1$

Consider $\beta = 2$:

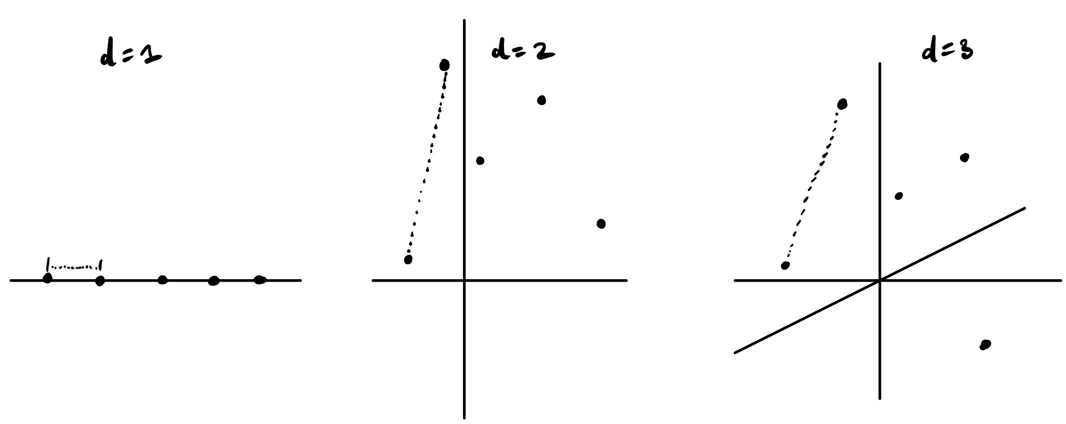
	$d=1$	$d=2$	$d=4$	$d=8$
Optimal Rate	$n^{-4/5}$	$n^{-4/6}$	$n^{-4/8}$	$n^{-4/12}$
MSE - Equivalent sample sizes to n_1 for $d \geq 1$	n_1	$n_1^{6/5}$	$n_1^{8/5}$	$n_1^{12/5}$
	100	251	1,585	63,096

For the second and third rows of the table, we ask what sample size under $d \geq 1$ would be equivalent in terms of the optimal MSE (ignoring constants) under $d = 1$. We solve for n_d below:

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

The above is meant to illustrate the "curse of dimensionality," the phenomenon that a fixed number of points positioned in space of higher and higher dimension, will be much further apart from each other.

The points do not "fill" the higher dimensional spaces.



The following exercise, from pg 30 of Györfi (2006) explores this idea:

__Exercise:__ Let $\underset{\sim}{X}, \underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n \overset{ind}{\sim} U([0,1]^d)$. Show that

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

__Hint:__ Use the fact that the volume of a ball in $\mathbb{R}^d$ with radius t is

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

to show that

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

__Solution:__ We have

$$\begin{aligned}
P\left(\min_{1 \le i \le n} \|\underset{\sim}{X} - \underset{\sim}{X}_i\|_2 \le t\right) &= P\left(\bigcup_{i=1}^{n} \{\|\underset{\sim}{X} - \underset{\sim}{X}_i\|_2 \le t\}\right) \\
&\le n\, P\left(\|\underset{\sim}{X} - \underset{\sim}{X}_1\|_2 \le t\right) \\
&= n \int_{\{\underset{\sim}{x} \in [0,1]^d\}} \int_{\{\underset{\sim}{x}_1 \in [0,1]^d :\, \|\underset{\sim}{x} - \underset{\sim}{x}_1\|_2 \le t\}} 1 \cdot d\underset{\sim}{x}_1\, d\underset{\sim}{x} \\
&\le n \int_{\{\underset{\sim}{x} \in [0,1]^d\}} \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} t^d \, d\underset{\sim}{x} \\
&= n \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} t^d.
\end{aligned}$$

Note that for large enough t, this inequality loses relevance, since the bound will exceed 1. So it is useful only when

$$n \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)} t^d \le 1 \iff t \le n^{-\frac{1}{d}} \frac{\Gamma(\frac{d}{2}+1)^{\frac{1}{d}}}{\sqrt{\pi}} .$$

Now for a continuous rv $T \in [0,\infty)$ with cdf F and which possesses a density, we have

$$\mathbb{E}\, T = \int_0^\infty [1 - F(t)]\, dt .$$

Thus

$$\mathbb{E}\left[\min_{1 \le i \le n} \| X - X_i \|_2\right] = \int_0^\infty \left[1 - P\left(\min_{1\le i\le n} \|X - X_i\|_2 \le t\right)\right] dt$$

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

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

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

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