Kernel Density Estimator (KDE)

Mentee: Gefei Shen
Mentor: Yuhan Qian

DRP Spring 2024

Agenda

- Introduction/Motivation
- Definition and Properties of KDE
- Common Kernels
- Rate of convergence (Mean square Error)

Introduction

Motivation

- In general experiments, we would assume the data follows some distribution
- However, in real cases, we do not know the true distribution of our data
- That's why we will introduce kernel density estimator (KDE) to find the most suitable distribution for a given data.

Introduction

Kernel Density Estimator is a non-parametric method used to estimate the probability density function of a random variable. It works by placing a kernel function at each data point and then summing these functions to create a smooth estimate of the overall data distribution.

Definition

- General Form of KDEs

$$\hat{f}_h : x \mapsto \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x),$$

where $K_h : u \mapsto \frac{1}{h}K(u/h)$.

- K(x) is the kernel function, h is the bandwidth, x is fixed, X_i is only randomness the observed data points

Definition of Kernel

- 1. $\int K(x)$ dx= 1 (Definition of Kernel)

Some facts of Kernel

- 2. An S-th order kernel K satisfies

$$\int u^r K(u)du = 0, \text{ where } r = 1, \ldots, s-1$$

$$|\int u^r K(u)du| \text{ is finite}$$

- 3. If K(u) = K(-u), the 2^{nd} moment is finite (K is symmetric)
  - At least 2^{nd} order

Common Kernels

Kernel	$K(u)$
Uniform	$\frac{1}{2} I\{\lvert u\rvert \leq 1\}$
Epanechnikov	$\frac{3}{4}(1-u^2) I\{\lvert u\rvert \leq 1\}$
Biweight	$\frac{15}{16}(1-u^2)^2 I\{\lvert u\rvert \leq 1\}$
Triweight	$\frac{35}{32}(1-u^2)^3 I\{\lvert u\rvert \leq 1\}$
Gaussian	$\frac{1}{\sqrt{2\pi}} \exp\{-u^2/2\}.$

Mean Square Error (MSE)

$$E[\{\hat{f}_h(x_0) - f(x_0)\}^2] = \underbrace{\{E[\hat{f}_h(x_0)] - f(x_0)\}^2}_{\text{bias}^2} + \underbrace{\text{var}(\hat{f}(x_0))}_{\text{variance}}.$$

- Assume f' is L Lipshitz
- Assume K is nonnegative, 2nd order, with bounded support
- At a fixed point x_0

Bias

$$
\begin{aligned}
\mathbb{E}[\hat{f}_h(x_0)] - f(x_0) &= \frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x_0)}{h} - f(x_0) \\
&= \frac{1}{nh}\sum_{i=1}^{n} \mathbb{E}\left[K\left(\frac{X_i - x_0}{h}\right)\right] - f(x_0) \\
&= \frac{1}{h}\mathbb{E}\left[K\left(\frac{X_1 - x_0}{h}\right)\right] - f(x_0) \\
&= \frac{1}{h}\int K\left(\frac{X_i - x_0}{h}\right) f(x)\,dx - f(x_0) \\
&= \frac{1}{h}\int K(u) f(x_0 + uh)\,du + f(x_0) \quad (\text{using } u = \frac{x_i - x_0}{h}) \\
&= \int K(u) f(x_0 + uh)\,du + f(x_0) \\
&= \int K(u)[f(uh + x_0) - f(x_0)]\,du \\
&= \int K(u)(f'(\tilde{x}_{uh}) - f'(x_0))uh + K(uf'(x_0)uh\,du \\
&= \int K(u)(f'(\tilde{x}_{uh}) - f'(x_0))uh\,du \\
|Bias| &= |\int K(u)(f'(\tilde{x}_{uh}) - f'(x_0))uh\,du| \\
&\leq \int |K(u)(c)uh|\,du \\
&= \int K(u)\cdot h|u|\cdot|f'(\tilde{x}_{uh}) - f'(x_0)|\,du \\
&\leq \int K(u)\cdot h|u||\tilde{x}_{uh} - x_0|\cdot L\,du \\
&\leq h^2 \int K(u)^2 L\,du \\
&= Lh^2\sigma_k^2
\end{aligned}
$$

$$
\begin{aligned}
\text{Bias}^2 &\leq L^2 h^4 \sigma_k^4 \\
&= O(h^4)
\end{aligned}
$$

Variance

$$
\begin{aligned}
\operatorname{Var}(\hat{f}_h(x_0)) &= \operatorname{Var}\left(\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x_0}{h}\right)\right) \\
&= \frac{1}{(nh)^2}\sum_{i=1}^{n} \operatorname{Var}\left(K\left(\frac{X_i - x_0}{h}\right)\right) \\
&= \frac{1}{nh^2}\operatorname{Var}\left(K\left(\frac{X_1 - x_0}{h}\right)\right) \\
&\leq \frac{1}{nh^2}\mathbb{E}\left[K^2\left(\frac{X_1 - x_0}{h}\right)\right] \\
&= \frac{1}{nh^2}\int_{-\infty}^{\infty} K^2\left(\frac{x_1 - x_0}{h}\right) f(x_1)\, dx_1 \\
&= \frac{1}{nh^2}\int_{-\infty}^{\infty} K^2(u) f(x_0 + uh) h\, du \quad (\text{using } u = \frac{x_1 - x_0}{h}) \\
\text{Let } k_1 &= \inf\{x : k(x) > 0\} \\
\text{Let } k_2 &= \sup\{x : k(x) > 0\} \\
&= \frac{1}{nh}\int_{k_1}^{k_2} K(u)^2 f(uh + x_0)\, du \\
&\leq \frac{1}{nh} S \cdot C \\
&= O(\frac{1}{nh})
\end{aligned}
$$

Choice of h

$$\begin{aligned}\mathrm{MSE}(\hat{f}_h(x_0)) &= \mathrm{Bias}^2(\hat{f}_h(x_0)) + \mathrm{Var}(\hat{f}_h(x_0)) \\ &= O(h^4) + O\left(\frac{1}{nh}\right),\end{aligned}$$

I want them to converge in the same rate

$$\begin{aligned}h^4 &= \frac{1}{nh} \\ h &= n^{-\frac{1}{5}}\end{aligned}$$

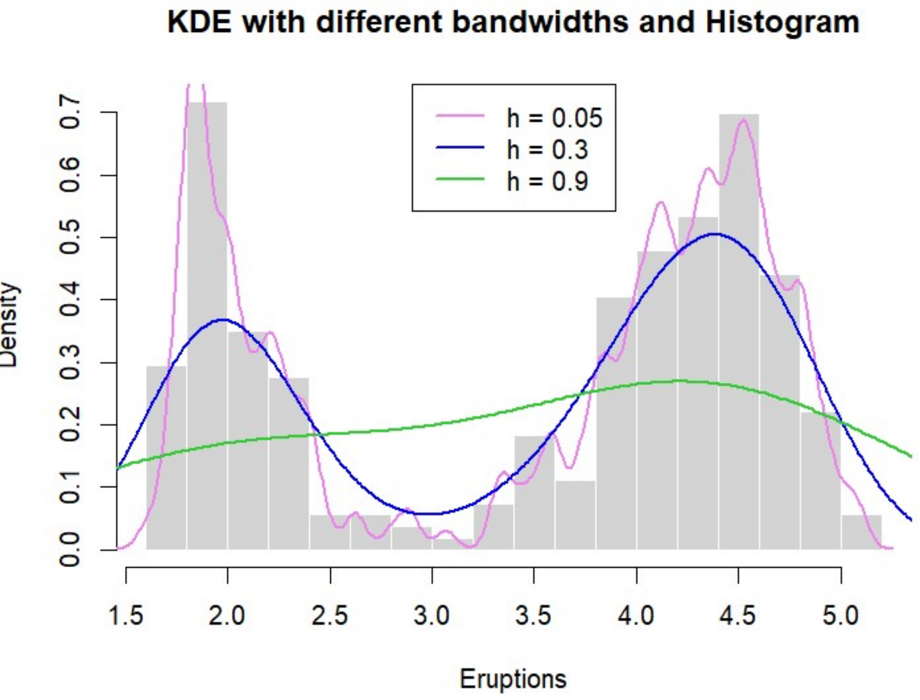
Mean Square Error (result)

$$h_{\text{opt}} = O(n^{-1/5}),$$

$$\text{Bias}^2(\hat{f}_h(x_0)) = O((n^{-1/5})^4) = O(n^{-4/5}),$$
$$\text{Var}(\hat{f}_h(x_0)) = O\left(\frac{1}{n \cdot n^{-1/5}}\right) = O(n^{-4/5}),$$

$$\text{MSE}(\hat{f}_h(x_0)) = O(n^{-4/5}).$$

Demo




```
dat = faithful$eruptions

kde1 = density(dat, bw=0.05)
kde2 = density(dat, bw=0.3)
kde3 = density(dat, bw=0.9)

hist(dat, probability = TRUE, col = "lightgray", border = "white",
     main = "KDE with different bandwidths and Histogram",
     xlab = "Eruptions", ylab = "Density", breaks = 20)

lines(kde1, col = 'violet', lwd = 2)
lines(kde2, col = 'blue', lwd = 2)
lines(kde3, col = 'limegreen', lwd = 2)

legend("top", legend = c("h = 0.05", "h = 0.3", "h = 0.9"),
       col = c("violet", "blue", "limegreen"), lwd = 2)
```

Overfitting and Underfitting

- When bandwidth is too small, it would overfitting
- When bandwidth is too large, it would underfitting