Lecture 11

Kernel density estimation

Q1
L V

Lecture Notes V – Non-parametric density estimation

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

January, 2025

Kernels

Kernel density estimators

Choosing h by Cross-Validation and the Bias-Variance trade-off

Reading: Ch. 7

So far:

Pdf: • discrete S ← ML (+ smoothing)

• continuous $S \subseteq (-\infty, \infty)$ — parametric ← ML

non-parametric models ← not ML ← $\mathcal{D} = \{x^1, \dots x^n\}$

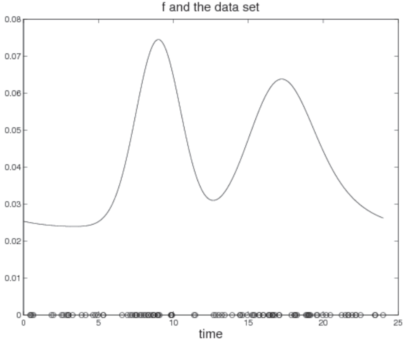
↳ no parameters estimated from

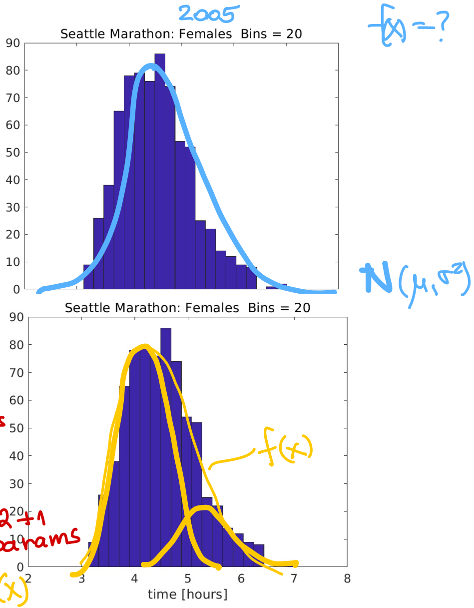
Why?

• can estimate $f(x)$ of arbitrary shape

if $n \nearrow$ grows

Why non-parametric?





1. $N(\mu, \sigma^2)$ by ML ← 2 params

1'. logistic? ... ← 2 params

2. Mixture of 2 Normals ← 2+2+1 params

$\pi_1 N(\mu_1, \sigma_1^2) + \pi_2 N(\mu_2, \sigma_2^2) = f(x)$

$\pi_1 + \pi_2 = 1$, $\pi_{1,2} > 0$ → Ex: Prove $f(x)$ is density

Kernel density estimators

3. KDE

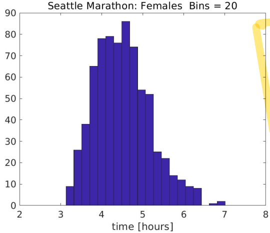
mixture of n elementary distributions = kernels

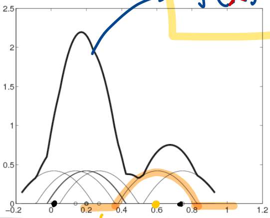
KDE Def

$= k(x)$

$$\boxed{f(x) = \frac{1}{n}\sum_{i=1}^{n} k(x - x^i)}$$

data $\{x^{1:n}\}$





$n=6$

x^6

$\frac{1}{n} k(x - x^6)$

x

Math footnote:

KDE is like convolution

$\hat{f} = \frac{1}{n}\sum \delta_{x^i}$

"empirical density"

$x^1 \quad x^2 \; x^3 \; \cdots \; x^n$

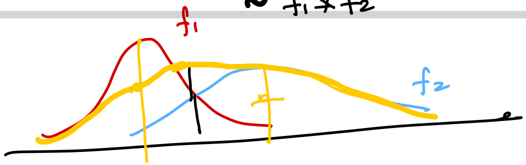
$f = k * \hat{f}$

$x_{1,2} \sim N(\mu_{1,2}, \sigma_{1,2}^2)$ $\qquad x_1 + x_2 \sim N(\mu_1 + \mu_2, \dots)$

$\sim f_1 * f_2$

$x \sim \pi_1 f_1 + \pi_2 f_2$

$\pi_1 = \pi_2 = \frac{1}{2}$



Sampling from Mixture:

1. sample $k \in \{1,2\} \sim (\pi_1, \pi_2)$ ← which Gaussian?
2. sample $x \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{if } k=1 \\ N(\mu_2, \sigma_2^2) & \text{if } k=2 \end{cases}$

Sampling from KDE

$$f_{(x)} = \frac{1}{n} \sum_{i=1}^{n} k(x - x^i)$$

"μ^i"

$\pi_1 = \pi_2 = \dots = \pi_n = \frac{1}{n}$

1. sample $i \sim$ uniform $\{1, 2, \dots n\}$
2. sample $x \sim k(x - x^i)$

fixed by data

sample near some x^i

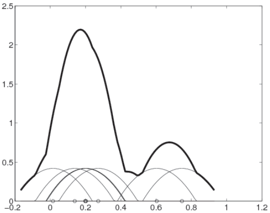
Kernel density estimators

another def

seen as fction calls

f Gaussian given x ← normal_pdf(x, mu, sigma)

f KDE given x ← kde(x, kernel_type, h, data)



in python

width

f(x) = kde(x, kernel_type, h, data)

array ← $x^1, \ldots x^n$

EXPENSIVE

1. data must be kept ⇒ memory O(n)
2. compute time O(n)

"memory based learning"

Kernels

∞ - derivable

2.

cont
k' almost everywh.

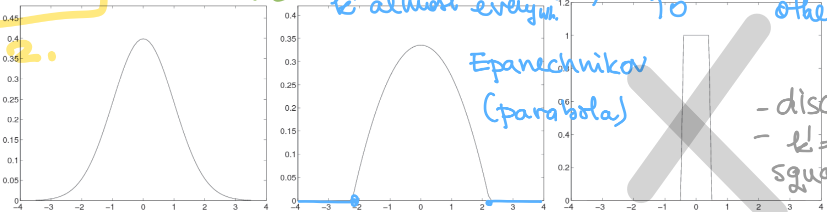
$k(z) \propto \begin{cases} a^2 - z^2 & \text{for } |z|<a \\ 0 & \text{otherwise} \end{cases}$

Epanechnikov (parabola)

- discontinuous
- $k' = \begin{cases} 0 \\ \text{undef} \end{cases}$

square

$k = \text{unif}[-0.5, 0.5]$



any density $k : (-\infty, \infty) \to [0, \infty)$

$k(z) \geq 0$ for all x

$$\int_{-\infty}^{\infty} k(z)\,dz = 1$$

"convenience" / typical wanted

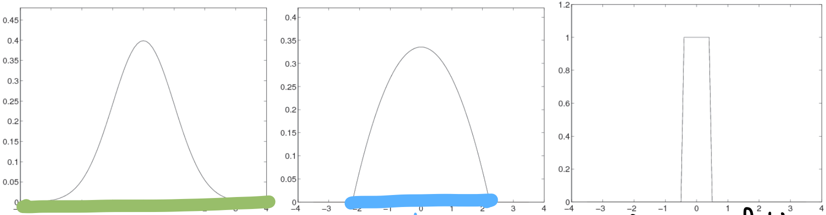
1. symmetric $k(z) = k(-z)$
2. monotonic ↘ with $|z|$

[3.] standardized $\int_{-\infty}^{\infty} k^2(z) = 1$

Options

4. smooth ? $\exists\, k', k''$, continuous
5. support compact ?

Kernels



Normal
supp $k^{\text{Normal}} = (-\infty, \infty)$

supp k^{Epa}

for any fction f
$\text{supp } f = \{z \mid f(z) \neq 0\}$

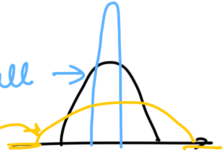
$\boxed{\text{compact support} \iff \text{supp } f \text{ Bounded}}$

Kernel width h

$$k_h(z) = k\left(\frac{z}{h}\right) \cdot \frac{1}{h}$$

h small

h large

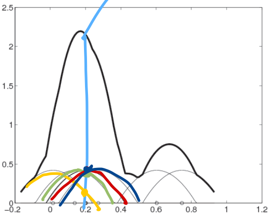


T.B. continued 2/12 - feb 13

Kernel density estimators

when $k()$ has compact support

$$f(x) = \frac{1}{6}\Big[k(x-x^1)+k(x-x^2) + k(x-x^3)+k(x-x^4) + \underbrace{k(x-x^5)}_{0}+\underbrace{k(x-x^6)}_{0}\Big]$$



in general

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} k(x-x^i)$$

sum over all n data points

compact supp → $\frac{1}{n}\sum_{x^i \text{ near } x} k(x-x^i)$

sum only over neighbors

Kernel density estimators
for marathon data (t.b. cont l12 - feb13)