Lec 9-11 More on densities.

Lec 9. other DE

- 近邻k-nearest neighbor $\hat{f}(x) = \frac{1}{n d_k(x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{d_k(x)}\right)$

- Penalized MLE: $\underset{\substack{\uparrow \\ \text{log-likelihood}}}{\ell_\lambda(f) = \ell(f)} - \lambda \underset{\substack{\uparrow \\ \text{penalty} = \int (f'')^2}}{R(f)}$

- Ortho-series: $f = \sum_{j=0}^{\infty} \theta_j \underset{\substack{\uparrow \\ \text{basis}}}{\varphi_j}$, estimate $\hat{\theta}_j = \frac{1}{n} \sum \varphi_j(X_i)$

  since $\theta_j = \int \varphi_j f = E[\varphi_j]$.

- Mixture model $f = \sum_{i=1}^{c} \pi_i f_i$ 有 C个组成部分 (components)

  $\triangle$ EM algorithm: $(Y_i, Z_i)_{i=1}^{n}$, Z_i are missing data.

  E: taking expectations $\hat{Z}_i = E_\theta[Z_i \mid Y]$

  M: maximizing ... likelihood, $\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \ell(y_i, \hat{z}_i)$

  In GMM: (y_i, z_i): y_i : observations (Sometimes denote as x_i)

  z_i : which component y_i belongs to.

  parameters $\theta = \{(\pi_i, \mu_i, \Sigma_i)\}_{i=1}^{c}$

- FFT*: computes fast, but not a novel estimation

Lec 10. Multivariate KDE (d-dim)

$$\hat{f}_h(x) \overset{(简单)}{=} \frac{1}{n} \sum_{i=1}^{n} K_{\vec{h}}(\vec{x} - \vec{X_i}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_1 \cdots h_d} K\left(\frac{x_1 - X_{i1}}{h_1}\right) \cdots K\left(\frac{x_d - X_{id}}{h_d}\right)$$

$$\doteq \frac{1}{n} \sum_{i=1}^{n} K_H(\vec{x} - \vec{X_i}), \text{ 其中 } \mu_2(K) I_d \triangleq \int_{\mathbb{R}^d} u u^T K(u)\, du$$

$$\|K\|^2 \triangleq \int K^2(u)\, du$$

$E \hat{f}_H(x) - f(x) \approx \frac{1}{2} \overset{K_{21}}{\mu_2(K)} \overset{h^2 f''(x)}{tr(H' D^2 f(x) H)}$, $Var(\hat{f}_H(x)) \approx \frac{1}{n|H|} \overset{K_{02}}{\|K\|^2} f(x)$ Taylor.

$h_{opt} \sim n^{-\frac{1}{4+d}}$, $AMISE_{opt} \sim n^{-\frac{4}{4+d}}$

- extensions: ①* estimate level-set $\{\vec{x}: f(\vec{x}) \geq c\}$ for $c \geq 0$.

  ②* clustering

  ③ estimate conditional density $f(y|x)$ by $\frac{\hat{f}(y,x)}{\hat{f}(x)}$

  bias, var, MSE, h_{opt}, CV

  … CDF $F(y|x)$

Lec 11 Hypothesis Testing

- general form:

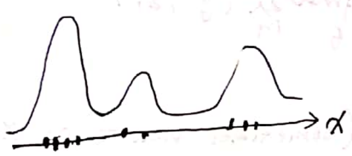
  $H_0: f=g \longleftrightarrow H_1: f \neq g$

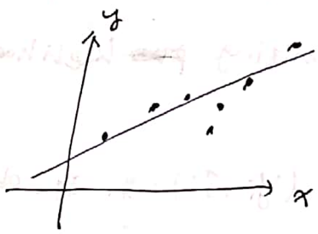
I: an information statistic that compares f with g, smaller means more similar.

Asymptotic: $T = \frac{n\sqrt{|H|}\,(\hat{I} - C(n))}{\hat{\sigma}} \rightsquigarrow N(0,1)$.

plug-in　bias-correction

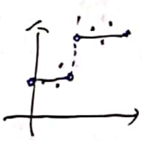
Lec 12-16 Smoothers.

- 



	density	v.s. smoother
data	只有 x, $x \sim f$	有 x 和 y, $y = S(x) + \varepsilon$.
aim	找出 $x \sim f$?	找出 $y = S(x)$? 特别, 找 smooth $m \triangleq (S(x_1), \cdots, S(x_n))^T$.

Lec 12 Introduction.

	density	v.s.	smoother
①	Para: e.g. $N(\mu, \sigma^2)$		para. e.g. Linear regression.
②	histogram		regressogram / bin smoother 
③	naive → KDE: $f(x) = \sum_i \frac{1}{n} W_h(x - X_i)$ weights with X_i at x. weights 求积为1		linear smoother $S(x) = \frac{1}{n}\sum_{i=1}^{n} W_{h_i}(x) Y_i$ 则 $\vec{m} = W_h \vec{Y}$ weights 求和为1或n

$E[Y|x] = \int y f(y|x) dy$

对 KDE plug-in

Kernel regression.

Lec 13 Kernel Regression.

• N-W: $\hat{m}(x) = \sum_i \left(\frac{K_h(x-X_i)}{\sum_j K_h(x-X_j)} \right) Y_i$ 有时记 $W_{n_i}(x)$ 为 $\frac{K_h(x-X_i)}{\sum_j K_h(x-X_j)}$ 或多乘 n 在分母

$$= \frac{1}{n} \sum_i \left(\frac{K_h(x-X_i)}{\frac{1}{n}\sum_j K_h(x-X_j)} \right) Y_i$$

$$\Leftrightarrow \min_{m(x)} \sum (Y_i - m(x))^2 K_h(x-X_i)$$

Thm. Bias, Var similar to KDE of conditional estimate Taylor

$\uparrow$ K_{21} (Bias) $\uparrow$ K_{02} (Var)

$h_{opt} \sim n^{-\frac{1}{5}}$, $AMISE_{opt} \sim n^{-\frac{4}{5}}$

• Similarly extend to ① multivariate $\vec{x} \in \mathbb{R}^d$.

② Bandwidth selection

③* mixture model

Lec 14 Local methods ($\overset{\text{local}}{\text{Taylor}}$ expansion)

• Loess: a computational method fitting local polynomials.
(Local Regression)

• Local linear kernel regression: loess with kernel method.

$$\min_{m(x) = m + (X_i - x)^T \beta} \sum (Y_i - m(x))^2 K_H(x-X_i)$$

Ex. Write in matrix form

Thm. similar to N-W.

• Extension: $\min_{m(x) = \beta_0 + \beta_1(X_i - x) + \cdots + \beta_p(X_i - x)} \sum (Y_i - m(x))^2 K_H(x-X_i)$

Ex. Write in matrix form

consider x is 1-dim, since it's complicate enough with polynomials…
multi-dim are analogous.
with $X = \begin{pmatrix} 1 & X_{11}-x_1 & X_{12}-x_2 & \cdots & (X_{11}-x_1)^2 \cdots & (X_{11}-x_1)(X \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & & & & & \end{pmatrix}$