Привет :) 2021-11-26

Непараметрическое оценивание

① таблицы сопр-сти и всё что (почему-то) наз-т непараметрикой ④ – (1B) перестановочные тесты.

② оценка плотности

③ лок. линейная регрессия

Про таблички

	$X_i \in A$	$X_i \in B$	$X_i \in C$
	A	B	C
кол-во набл.	56	67	80
	N_1	N_2	N_3

$$H_0: \begin{cases} p_1 = \frac{1}{3} \\ p_2 = \frac{1}{3} \\ p_3 = \frac{1}{3} \end{cases} \qquad H_A: \left\{ p \neq \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix} \right.$$

① как протестировать H_0?

A. Сработает троица LR / LM / W.

[и критерий Пирсона]

UR: $\hat p_1 = 56/203$, $\hat p_2 = 67/203$, $\hat p_3 = 80/203$

R: $p_1 = p_2 = p_3 = \frac{1}{3}$

$$LR = 2\Big(\max_{\substack{p_1, p_2, p_3 \\ [p_1+p_2+p_3=1]}} \ell - \max_{p_1=p_2=p_3=\frac{1}{3}} \ell\Big) =$$

$$= 2\Big(\ln\big(\hat p_1^{N_1} \cdot \hat p_2^{N_2} \cdot \hat p_3^{N_3}\big) - \ln\big(p_1^{N_1} p_2^{N_2} p_3^{N_3}\big)\Big)$$

$$LR \xrightarrow[n\to\infty]{dist, H_0} \chi^2_2$$

в общем: χ^2_{g-1}, g – число групп

вар-т предп-ия о законе распр-ия X_i

Пример: $X_i \sim Exp(\lambda)$ незав.

$X_i \in [0;1)$	$X_i \in [1;2)$	$X_i \in [2;+\infty)$
56	67	80

Ⓐ $H_0: \lambda = 1$, $X_i \sim Exp(1)$; $H_A: X_i \sim Exp(\lambda)$

Ⓑ $H_0: X_i \sim Exp(\lambda)$; $H_A: X_i \sim law$

Ⓒ $H_0: X_i \sim Exp(1)$; $H_A: X_i \sim law$

LR

Способ В

$$LR = 2\left(\max_{UR} \ell - \max_{R} \ell\right) \xrightarrow[n\to\infty]{dist,\ H_0} ? \quad \chi^2_{k_{UR}-k_R} \quad \chi^2_1$$

!!

здесь 2 пар-ра оцениваем

$\hat{p}_1, \hat{p}_2$

$\hat{p}_3 = 1 - \hat{p}_1 - \hat{p}_2$

$k_{UR} = 2$

$\hat{\lambda}$ [один пар]

$p_1(\hat{\lambda}), p_2(\hat{\lambda})\ p_3(\hat{\lambda})$

$k_R = 1$

$$P(X_i \in [0;1)) = 1 - \exp(-\lambda) = p_1$$

$X_i \sim \exp(\lambda)$ $\quad f(x) = \lambda \cdot \exp(-\lambda x)$

$$P(X_i \geq x) = \exp(-\lambda x)$$

$$P(X_i \in [1;2)) = \exp(-\lambda) - \exp(-2\lambda) = p_2$$

$$P(X_i \in [2;+\infty)) = \exp(-2\lambda) - 0 = p_3$$

$$\max_{\lambda} \left(1 - \exp(-\lambda)\right)^{N_1} \cdot \left(\exp(-\lambda) - \exp(-2\lambda)\right)^{N_2} \cdot \left(\exp(-2\lambda)\right)^{N_3}$$

[в явном виде !!]

пол	нен свр	инд к свр	люб свр
М	N_{11}	N_{12}	N_{13}
Ж	N_{21}	N_{22}	N_{23}

H_0: пол и отношение к свету незав-мы

H_A: зависимы.

0,1	0,15	0,11
0,07	0,09	.

Q. Как тестировать?

A. Пицца-Тройка LR/LM/W

[критерий Пирсона]

$$LR = 2\left(\max_{UR} \ell - \max_{R} \ell\right) \longrightarrow \chi^2_{k_{UR}-k_R} = (g_1 \cdot g_2 - 1) - (g_1 - 1) - (g_2 - 1) = (g_1 - 1)\cdot(g_2 - 1)$$

$$C_{UR} = \ln\left(\hat{p}_{11}^{N_{11}} \cdot \hat{p}_{12}^{N_{12}} \cdot \ldots \cdot \hat{p}_{23}^{N_{23}}\right) \qquad \hat{p}_{ij} = \frac{N_{ij}}{n}$$

$$\max_{p} \quad \sum p_{ij} = 1$$

$K_{UR} = 5$

в общем случае

$K_{UR} = g_1 \cdot g_2 - 1$,

где g_1 – число групп в 1-м признаке, g_2 – во втором.

R-модель

	0,2	0,3	
	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$
0.1 $p_{1\cdot}$	p_{11}	p_{12}	p_{13}
$p_{2\cdot}$	p_{21}	p_{22}	p_{23}

Н. верно $P(A \cap B) = P(A) \cdot P(B)$

$$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$$

$p_{1\cdot} = 0,1$
$p_{\cdot 1} = 0,2$
$p_{\cdot 2} = 0,3$

$\Rightarrow p_{2\cdot} = 0,9 = 1 - 0,1$

$p_{\cdot 3} = 0,5 = 1 - 0,2 - 0,3$

$\Rightarrow$ все p_{ij}

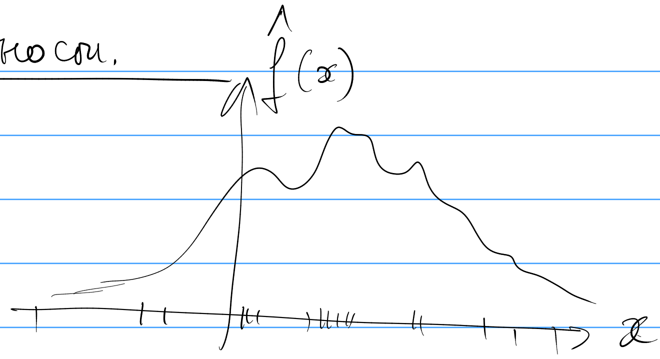
$$K_R = 3 = (g_1 - 1) + (g_2 - 1)$$

$$\hat{p}_{1\cdot} = \frac{N_{11} + N_{12} + N_{13}}{n} \qquad \hat{p}_{\cdot 1} = \frac{N_{11} + N_{21}}{n} \quad \ldots$$

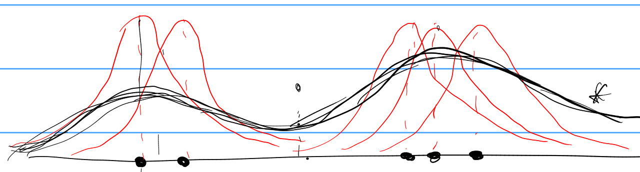
$X_i \sim$ нез, Law (непр)

Оценивание плотности.

x_1
x_2
$\vdots$
x_n



Идея



* есть точки на прямой
* бросим на место $\forall$ точки станд. плотность со станд. откл-ие h.
* сложим эти плотности и поделим на число набл-ий.

$\boxed{\text{формула}}$

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} pdf(x, \mu = x_i, \sigma = h)$$

$\boxed{\text{ядерная ф-ция}}$

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \leftarrow \text{плотность для } N(0;1)$$

$$\hat{f}(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - x_i}{h}\right)$$

$\boxed{\text{Вопрос: как выбрать } h\text{?}}$

более важен, чем выбор ядра.

$$K(u) = \begin{cases} \frac{1}{2} & u \in [-1;1] \\ 0 & u \notin [-1;1] \end{cases}$$

парам. задача

$X_i \sim U[0;\theta]$ незав

$\hat{\theta} = \max\{X_1, \dots, X_n\}$

$$\hat{f}(x) = \frac{1}{nh} \cdot \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

f — истинная ляw для x_i

$\boxed{\text{оценка смещена!}}$

$$E\left(\hat{f}(x)\right) = E\left(\quad \right) =$$

$$= E\left(\frac{1}{h} K\left(\frac{x - x_1}{h}\right)\right) = \boxed{\int_{\mathbb{R}} \frac{1}{h} \cdot K\left(\frac{x-u}{h}\right) \cdot f(u)\, du}$$

$$\neq f(x)$$

$\boxed{\text{смещение} \to 0}$ если $h \to 0$

$$\operatorname{Var}\left(\hat{f}(x)\right) = \frac{1}{(nh)^2} \cdot n \cdot \operatorname{Var}\left(K\left(\frac{x - x_1}{h}\right)\right)$$

дисперсия $\to 0$ если $nh \to \infty$

$$MSE(x) = E\left[\left[\hat{f}(x) - f(x)\right]^2\right] = \operatorname{Var}\left(\hat{f}(x)\right) + \left(E\left(\hat{f}(x)\right) - f(x)\right)^2$$

$\boxed{\begin{matrix} h \to 0 \\ nh \to \infty \end{matrix}}$

$$\operatorname{Var}\left(\hat{f}(x)\right) \to 0, \qquad \left(E\left(\hat{f}(x)\right) - f(x)\right)^2 \to 0$$

h выбирают с помощью CV.

$$MISE = \int_{\mathbb{R}} \left(\hat{f}(x) - f(x)\right)^2 dx =$$

mean integrated squared error

зависит от выбора h

$$= \boxed{\underbrace{\int \hat{f}^2(x)\, dx}_{\text{я могу посчитать}} - 2 \int \hat{f}(x) \cdot f(x)\, dx} + \underbrace{\int f^2(x)\, dx}_{\text{не нужно, но плевать!}}$$

проблема в оценке $\boxed{\int \hat{f}(x) f(x)\,dx}$

$\hat{f}^{-i}(x)$ $\leftarrow$ та же оценка, полученная исп-уя все набл-ия кроме i-го.

$\boxed{\text{Чв}}$ $\boxed{E\left(\frac{1}{n}\sum_{i=1}^{n}\hat{f}^{-i}(x)\right) = \ldots = E\left(\int \hat{f}(x) f(x)\,dx\right)}$

$$\hat{h}_{CV} = \arg\min_h \left[\int \hat{f}^2(x)\,dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}^{-i}(x)\right]$$

$$\hat{h}_A = \frac{1}{\sqrt{n}} \Rightarrow \begin{matrix} h \to 0 \\ nh \to \infty \end{matrix}$$