Въпрос 19: Хипотези и Нейман-Пирсън

Martin Varbanov

Юни 2019

1 Дефиниции

1.1 Хипотези

ще въведем няколко означения

X_n - извадково пространство

$x = (x_1, .., x_n)$ - стойности на наблюденията

θ - параметър
$H_0 : \theta = \theta_0, \theta \in \mathbb{R}$ - Основна/нулева хипотеза

$H_1 : \theta = \theta_1, \theta \in \mathbb{R}$ - Алтернативна хипотеза

1.1.1 Функция на правдоподобност

$$L(x_1, .., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

1.1.2 проста хипотеза

Когато на H_0 съответства точно едно разпределение $L_0(x)$ се говори за проста основна хипотеза

1.1.3 сложна хипотеза

когато H_0 не се удовлетворява от едно единствено разпределение говорим за сложна основна хипотеза

1.2 Грешки от 1ви и 2ри род

	Нямаме основание да отхъвлим H_o	Отхвърляме H_0
H_0 е истина	OK	Грешка от 1ви род
H_1 е истина	Грешка от 2ри род	OK

1.3 Критична област

Тук ще дадем дефиниция за критична област:
$\overline{w} \subset X_n$ - мн-во от $(x_1, .., x_n)$, т.че H_0 да е вярно

$w = X_n \ \overline{w}$ - критична област

ВЪПРОС: ОСТАНА ДА НАПИШЕМ ЗА P-VALUE!!!

1.3.1 вероятност за грешка от 1ви род/Ниво на доверие

$$\alpha = P(x \in w|H_0) = \int_w L_0(x)dx$$

1.3.2 бетата

$$\beta = P(x \notin w|H_1)$$

1.3.3 Мощност

$$\Pi = P(x \in w|H_1)$$

1.3.4 Оптимална критична облас

w^* е оптимална критична област ако са изпълнени:

$$P(x \in w|H_0) \leq P(x \in w^*|H_0)) = \alpha, \forall w$$

$$P(x \in w^*|H_1) \geq P(x \in w|H_1)$$

2 Лема на Неймън-Пирсън

2.1 Дефиницийки

Нека w^* е оптимална критична област и:

$$(\int ... \int)_{w^* \subset X_n} L_1(x_1, ..., x_n)dx_1, .., dx_n = P(x \in w^*|H_1)$$

and

$$(\int ... \int)_{w^* \subset X_n} L_0(x_1, .., x_n)dx_1, ..., dx_n \leq \alpha$$

3 Лема на Неймън-Пирсън

M_α - множество от подмножества $w \in X_n$, тъй че:

$$P(x \in w|H_0) = \int_w L_0(x)dx = \alpha$$

$$P(x \in w^*|H_1) = \int_{w^*} L_1(x)dx \geq \int_{w^*} L_1(x)dx, \forall w^* \in M_\alpha$$

3.1 Лема

При проверка на хипотеза:

$$H_0 : L(x) = L_0(x)$$

$$H_1 : L(x) = L_1(x)$$

с ниво на съгласие α и $w^* \in M_\alpha$, така че да съществува константа $K = K_\alpha > 0$, т. че да е изпълнено:

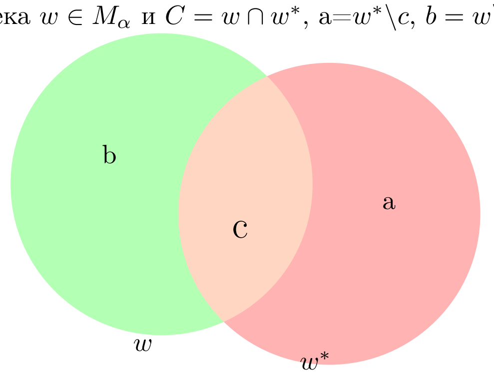
$$L_1 \geq KL_0, \forall x \in w^*$$

$$L_1 \leq KL_0, \forall x \notin w^*$$

$=>$ w^* е оптимална крит. обл.

3.2 Доказателство:

Нека $w \in M_\alpha$ и $C = w \cap w^*$, a$=w^* \backslash c$, $b = w \backslash c$



От $w, w^* \in M_\alpha$:

$$\alpha = \int_w L_0(x)dx = \int_b L_0(x)dx + \int_c L_0(x)dx$$

$$\alpha = \int_{w^*} L_0(x)dx = \int_a L_0(x)dx + \int_c L_0(x)dx$$

от това следва:

$$\int_b L_0(x)dx = \int_a L_0(x)dx$$

Нека сега разгледаме:

$$\Delta = \int_{w^*} L_1(x)dx - \int_w L_1(x)dx = \int_a L_1(x)dx - \int_b L_1(x)dx$$

Идеята по- долу дали следва от:

$$L_1(x) \geq K L_0 x, x \in w^* => \int_{w^*} L_1(x)dx \geq K \int_a L_0(x)dx$$

$$L_1(x) \leq K L_0 x, x \notin w^* => \int_w L_1(x)dx \leq K \int_b L_0(x)dx$$

и ако умножим второ по -1 и съберем с 1вото се получава:

$$\Delta = \int_{w^*} L_1(x)dx - \int_w L_1(x)dx \geq K(\int_a L_1(x)dx - \int_b L_1(x)dx)$$

$$=> \Delta \geq 0$$

от тука се вижда, че условието за оптимална критична област:

$$P(x \in w^* | H_1) = \int_{w^*} L_1(x)dx \geq \int_w L_0(x)dx$$

Следва, че w^* е оптимална критична област.

3.3 Следствия

$$\frac{L_1(x)}{L_0(x)} \geq K$$

$$P(\frac{L_1(x)}{L_0(x)} \geq K | H_0) = \alpha$$

$$\pi = P(x \in w^* | H_1) = \int_{w^*} L_1(x)dx$$