Teoria dell' Inferenza / 13-05

Principio di massima verosimiglianza

L'idea è come derivare l'informazione trasmessa dai dati

es lancio moneta con risultati 11011110100111010111

$n = 20$ $\quad$ $3/10$ freq 0 $\quad$ $7/10$ freq 1

posso inferire (dedurre) che la p di ottenere 1 è $7/10$
= dato da un'intuizione frequentista

$p_0 = p(\text{teste}) = 1$ $\quad$ 1 $\quad$ con un solo caso
$1 - p_0 = p(\text{croce}) = 0$ $\quad$ $p = 1$, difficile determinare altrimenti

Posso determinare se data una moneta "onesta", questa lo sia veramente → inferisco delle conclusioni

p_0 $\quad$ voglio inferire, dedurre un valore θ_0 con $\theta \in \Theta$ dagli n dati

$p \in [0,1]$

X utilizza valori di Bernoulli

es. $\quad$ 0 0 1 0

$p = 0{,}1 \quad 0{,}2 \quad 0{,}5 \quad 0{,}9$

calcolo $P_R(X \mid p) = P_R(x_1 x_2 x_3 x_4 \mid p)$

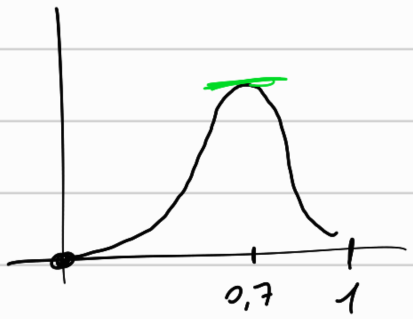
X: ciò che ho $\quad$ $x_1 x_2 x_3 x_4$: cioè di X non so che p ho utilizzato, assumiamo $p(1)=0{,}1$ $p(0)=0{,}9$

$P_?(\underbrace{00 1 0}_{\text{indip.}} \mid p=0{,}1) = 0{,}9 \cdot 0{,}9 \cdot 0{,}1 \cdot 0{,}9 = 0{,}0729$ probabilità della sequenza

$P_?(0010 \mid p=0{,}9) = 0{,}1 \cdot 0{,}1 \cdot 0{,}9 \cdot 0{,}1 = 0{,}0009$

diremo che p è il parametro che massimizza $P_?$
e questo è detto massimo della verosomiglianza

Posso garantire la sicurezza di quanto affermo, ma non che sia giusto o sbagliato



ovvero posso affermare quale valore di p determina un risultato più probabile e che quindi si avvicina di più a quello reale

$f_\theta(X) \longrightarrow$ verosomiglianza

θ: valore del parametro che la genera

X: la stringa

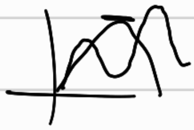
$$f_p(x) = p^{x}(1-p)^{(1-x)}$$

var di Bernoulli

$x_i \in X$ sono eventi INDIPENDENTI

$x = 1 \qquad f_p(x=1) = p$

$x = 0 \qquad f_p(x=0) = 1-p$



$$L(x_1, \dots, x_n \mid p) = \prod_{i=1}^{n} f_p(x_i)$$

preserva i massimi (non rischio di cambiarli)

Applicando il logaritmo rendo la funzione monotona:

↳ e trasformo il prodotto in somme

$$\ln L = \sum_{i=1}^{n} \ln\left(p^{x_i}(1-p)^{1-x_i}\right)$$

$$\frac{d \ln L}{dp} = 0 \longrightarrow p_0 = \frac{1}{n} \sum x_i$$

annullo la derivata rispetto a p

applico il principio di massima verosomiglianza su una base frequentista

ovvero determino p sulla base della frequenza di x_i (ovvero osservo il valore medio)

$$= \sum_{i=1}^{n} \left(x_i \ln p + (1-x_i)\ln(1-p)\right)$$

$$= \sum_{i=1}^{n} \left(\frac{x_i}{p} - (1-x_i)\frac{1}{1-p}\right) = 0$$

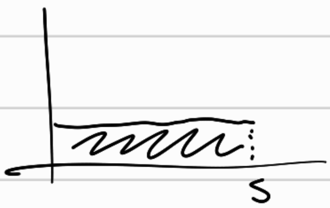
$$\rightarrow \frac{1}{p}\sum_i x_i - \underbrace{\sum_i \frac{1}{1-p}}_{\frac{n}{1-p}} + \frac{1}{1-p}\sum_i x_i = 0$$

$$\rightarrow (1-\cancel{p})\sum x_i - np + \cancel{p\sum x_i} = 0 \longrightarrow \hat{p} = \frac{\sum x_i}{n}$$

dove $\hat{p}$ è una <u>stima</u> di p_0 (perchè è probabile ma non sicuro)

- 57, 112, 5 n° di taxi passati, p di vedere un taxi è maggiore in 112

si prende una distribuzione uniforme



$f_S(x)$ è ora una distr. uniforme e non bernoulliana

$$f_S(x) = \begin{cases} 1/S & \text{se } 0 \leq x \leq S \quad (x \text{ è un taxi}) \\ 0 & \text{altrimenti} \end{cases}$$

dove S è il parametro θ, che determina L

$L(x_1, x_2, x_3 \mid S)$

$2 = 0 \cdot 0 \cdot 0$ non è verosimile che vedo 2 taxi se ne vedo di più

$L(57, 112, 5 \mid 40) = 0 \cdot 0 \cdot \frac{1}{40}$

$L(57, 112, 5 \mid 1000) = \frac{1}{1000} \cdot \frac{1}{1000} \cdot \frac{1}{1000} = 10^{-9}$ piccola, infatti è poco prob. di vedere 1000

che valore di S per ottenere il max? 112

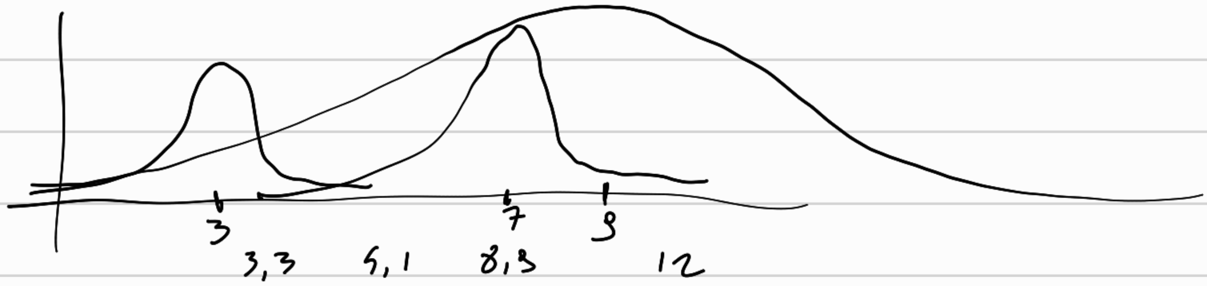
quindi $\hat{S} = x_{max}$

Gaussiana

$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

ma in generale non sono noti μ e σ^2

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Cerco μ e σ^2 che determinano il max di verosimiglianza

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$ infatti:

$$\ln L = \ln\left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right]$$

$$= -\frac{1}{2}\ln(2\pi\sigma^2]^n - \sum^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

prima dell'esp.

$$\frac{dL}{d\mu} = \underset{\text{non dipende da } \mu}{0} - \sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$\Downarrow$

derivata parziale, ciò che non è μ diventa una costante

$\longrightarrow \sum x_i = n\mu \longrightarrow \hat{\mu} = \frac{\sum x_i}{n}$

ma ora voglio considerare anche σ^2

$$-\frac{1}{2}\ln(2\pi\sigma^2)^n - \sum^n \frac{(x_i-\hat{\mu})^2}{2\sigma^2}$$ so che in derivata $\Rightarrow \mu$ vale $\hat{\mu}$

DELTA

$$\frac{\partial \ln L}{\partial \sigma^2} \rightarrow -\frac{n}{2\sigma^2}+\left[\frac{1}{2}\sum_i (x_i-\hat{\mu})^2\right]\frac{1}{\sigma^4}=0$$

$$= -1 + \frac{\sum(x_i-\hat{\mu})^2}{\sigma^2} = 0 \rightarrow \hat{\sigma}^2 = \sum(x_i-\hat{\mu})^2$$

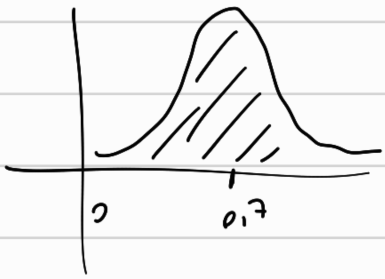
Ho trovato i migliori $\hat{\mu}$ e $\hat{\sigma}^2$ per cui la gaussiana mi produce i valori verosimiglianti a quelli dati

La media da gaussiana e' uno stimatore corretto, per più "run" sicuramente oscilla ma comunque attorno a p_0
Altrimenti lo stimatore e' distorto (bias), e' improprio dire "sbagliato"

$$E[\hat{p}] = p_0$$ il principio ci garantisce che lo stimatore oscilla fra il valore vero

/ 13-05

Non stiamo parlando di probabilità



l'area totale sottesa alla gaussiana **non** è $= 1$

si osservano, su un numero alto di frequenze, i parametri che meglio rappresentano le probabilità con cui ottengo un certo evento A

Teorema di Bayes

2 persone su 1000 → 0,2%   prob. afflitto dalla patologia

Carlo fa di un test

SENSIBILITÀ del 95% → probabilità che un affetto dalla patologia risulti positivo al test

SPECIFICITÀ del 99% → probabilità che un individuo sano risulti negativo

Senza test, p (Carlo sia malato) = 0,2%

$P(\text{malato} \mid \text{positivo}) = ?$

$P(\text{positivo} \mid \text{malato}) = 0{,}95$

$P(\text{negativo} \mid \text{malato}) = 0{,}05$

$P(\text{positivo} \mid \text{sano}) = 0{,}01$

$P(\text{negativo} \mid \text{sano}) \approx 0,99$

Carlo risulta positivo: con che p è malato?

VEROSOMIGLIANZA (valore più vicino alla realtà)

$$P(\text{mal} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{mal}) \cdot P(\text{mal})}{P(\text{pos})}$$

$$P(\text{pos}) = P(\text{pos} \mid \text{sano})\, P(\text{sano}) + P(\text{pos} \mid \text{mal}) \cdot P(\text{mal})$$

$$P(\text{mal} \mid \text{pos}) = \frac{0,95 \cdot 0,002}{0,95 \cdot 0,002 + 0,01 \cdot (0,998)} = 1,6\%$$

(8 volte tanto rispetto p essere malato senza test, ovvero 0,2%.)

(inferisco delle informazioni)

Allora il 98,4% delle persone dichiarate pos non sono malate

• 

A A B C

Sacchetto con 4 monete, di tipo A, B, C

$P(A) = \frac{1}{2}$

$P(B) = \frac{1}{4}$

$P(C) = \frac{1}{4}$

$$\left.\begin{array}{l} P(T \mid A) = \frac{1}{2} \\ P(T \mid B) = \frac{3}{5} \\ P(T \mid C) = \frac{9}{10} \end{array}\right\} \text{verosomiglianze}$$

Presa una moneta a caso: inferisco quale tipo ho preso } più probabilmente A senza alcun tiro

1) Sapendo che tirando la moneta è uscito testa? → scelgo la più probabile

2) se ottengo nuovamente testa? ↗

$P(A|T) + P(B|T) + P(C|T) = 1$

p. a posteriori (osservo il verificarsi di un evento, poi scelgo)

$$P(A|T) = \frac{P(T|A)\,P(A)}{P(T|A)\cdot P(A) + P(T|B)\cdot P(B) + P(T|C)\cdot P(C)}$$

per $P(B|T)$ e $P(C|T)$ il denominatore è lo stesso, quindi cerco il NUMERATORE più grande

H	p priori	verosimiglianza	p. posteriori non normalizzato (PRODOTTO)	normalizzo e "promuovo" a una p priori
A	0,5	$\frac{1}{2}$	0,25	0,4
B	0,25	$\frac{3}{5}$	0,15	0,24
C	0,25	$\frac{9}{10}$	0,225	0,36
	1		0,625	1

→ applico verosim. e ottengo la nuova scelta x il 2° lancio (vedi sotto) *

scelgo A (p priori)

scelgo A (p. posteriori non normalizzato)

Il frequentista lancia tante volte e conta chi è uscito maggiormente; il bayesiano osserva con la verosomiglianza e ripete

Lancio nuovamente la moneta e ottengo testa:

H	p priori *	verosimiglianze	posteriori non normalizzate	normalizza e "promuovo" a una priori
A	0,4	$\frac{1}{2}$	modello 0,2	
B	0,24	$\frac{3}{5}$	0,144	
C	0,36	$\frac{9}{10}$	0,324	
	1		$\neq 1$	1

dopo due teste scommetto su tipo C

se uscisse croce ripeto con la nuova p. a priori da quelle posteriori ma le verosimiglianza non è più su TESTA ma su CROCE

indipendentemente dalla moneta

$$P(T_1) = P(T_1|A)\cdot P(A) + P(T_1|B)\cdot P(B) + P(T_1|C)\cdot P(C)$$

$$= \frac{1}{2}\cdot\frac{1}{2} + \frac{3}{5}\cdot\frac{1}{4} + \frac{9}{10}\cdot\frac{1}{4} = \frac{10+6+9}{40} = \frac{25}{40} = \frac{5}{8} = 0{,}625$$

(effettivamente le monete sono più propense ad avere T, $P(T|\cdot) \geq \frac{1}{2}$)

TESTA (NON IMPORTA se prima/seconda ...)

$$P(T_2|T_1) = P(T_2|A)\cdot P(A|T_1) + P(T_2|B)\cdot P(B|T_1) + P(T_2|C)\cdot P(C|T_1)$$

sapendo che ho avuto testa prima, osservo con quale p ho avuto A, B, C

$= 0{,}5\cdot 0{,}4 + 0{,}6\cdot 0{,}24 + 0{,}9\cdot 0{,}36 = 0{,}668$ p una seconda testa

Relativamente alla Gaussiana

$$\hat{\mu} = \frac{\sum x_i}{n} \qquad \hat{\sigma}^2 = \sum \frac{(\hat{\mu} - x_i)^2}{n}$$

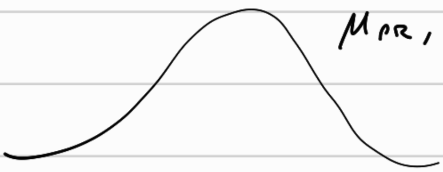
dal punto di vista frequentista

Dal punto di vista bayesiano determino le stime avendo osservato qualcosa di precedente, dandomi informazioni a priori

↓

ovvero che anche loro provengono da una gaussiana

$\hat{\mu}_{POST}$ $\hat{\delta}_{post}$



μ_{PR}, δ_{PR}

tale per cui i parametri:

$\mu_{PR} = 4$

$\delta^2_{PR} = 1$

avendo ora i dati 3,1 2,4 2,7 3,4 come δ

Che quindi è traslata a sx rispetto a quella reale (pregiudizio leggermente errato)

Se applico il prodotto dei dati rispetto alle verosomiglianze ottengo una nuova gaussiana.

μ_0, δ_0^2 sono veri e non noti

$\hat{\mu} \xrightarrow{\text{TENDE}} \mu_0$ in probabilità per $n \to \infty$ || frequentista

$\hat{\delta}^2 \to \delta_0^2/n$ " " "

$$\mu_{POST} = \frac{\delta_0^2 \mu_{PR} + \delta^2_{PR}\, n \hat{\mu}}{\delta_0^2 + n\, \delta^2_{PR}}$$

la somma è 1

$\left| \begin{array}{c} \mu_{PR} \\ \delta^2_{PR} \end{array} \right|$ || bayesiano

pregiudizio, fornito a priori

|

↓

quindi:
la μ_{post} ottenuta la si ha tenendo conto del pregiudizio $\delta_0^2 \mu_{PR}$
sommando i dati effettivamente ottenuti $\delta^2_{PR}\, n \hat{\mu}$

$$\delta^2_{post} = \frac{\delta_0^2\ \delta^2_{PR}}{\delta_0^2 + n\,\delta^2_{PR}}$$

in accordo al frequentista
per n grandi vince il giudizio a priori
$\left(\delta^2_{post} \rightarrow \delta_0^2/2\right)$

/ 20-05

Principio di max verosomiglianza
Principio di max probabilità a posteriori

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B)$$