Lezione n.4 (a cura di Federica Campanella)

MASSIMA VEROSIMIGLIANZA

$X \sim f(x, \theta)$

X1,X2,….,Xn

L(θ)$\rightarrow l(\theta) \rightarrow max_{\theta}\, l(\theta)$

$X \sim N(\mu, \sigma^2)$

$\theta = (\mu, \sigma^2)$

$$f(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2}} \left(\frac{x-\mu}{\sigma 2}\right)^2$$

L$(\theta) = \Pi_{i=1}^{n} f(xi,\theta) = (1/\sqrt{(2\pi\sigma^2)})^{n}\ e^{((-1)/2)\sum_{i=1}^{n}(((x-\mu)/\sigma 2)^2)}$

$$l(\theta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2}\sum\left(\frac{x-\mu}{\sigma^2}\right)^2$$

$$\frac{\partial}{\partial\mu} l(\theta) = -\frac{1}{2}\,(-2)\sum(xi-\mu) = \frac{1}{\sigma 2}\left[\left(\sum (x_i) - n\mu\right)\right] = 0$$

$$\sum x_i = n\mu \ \rightarrow \mu = -\frac{\sum x_i}{n\sigma^2}$$

$$\frac{\partial^2}{\partial\mu} l(\mu) = -\frac{n}{\sigma^2} < 0$$

Se il modello è normale lo stimatore di massima verosimiglianza di μ è la media campionaria ($\overline{x}$).

$$l(\theta) = -\frac{n}{2}\log 2\pi\sigma^2 - \frac{1}{2}\sum\left(\frac{x-\mu}{\sigma^2}\right)^2$$

↓ ↓

Non dipende da μ Dipende da μ

$$max_{\mu}\, l(\theta) \Leftrightarrow min_{\mu} \sum_i (x_i - \mu)^2$$

Lo stimatore di massima verosimiglianza della varianza è:

$$\widehat{\sigma^2} = \frac{1}{n}\sum_i (x_i - \overline{x})^2$$

$S^2 - \frac{1}{n-1}\sum_i (x_i - \overline{x})^2$

PROPRIETA' DEGLI STIMATORI DI MASSIMA VEROSIMIGLIANZA:

1) **PROPRIETA' DI INAVARIANZA**

   Se conosco lo stimatore di massima verosimiglianza di θ e voglio stimare τ applico la trasformazione direttamente allo stimatore iniziale.
   Supponiamo di conoscere lo stimatore di un parametro se voglio stimare una sua trasformazione:
   $\tau = g(\theta)$
   $\hat{\tau} = g(\theta)$

   Stimo $\sigma 2$ con σ $\rightarrow$ $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_i (x_i - \overline{x})^2}$

2) **PROPRIETA' DI EFFICIENZA**

   Lo stimatore di massima verosimiglianza è generalmente quello più efficiente.
   Se n$\rightarrow \infty$ gode della proprietà asintotica della consistenza.

3) **PROPRIETA' DI NORMALITA' ASINTOTICA**

   Lo stimatore ha distribuzione asintotica normale per qualsiasi tipo di popolazione se n è grande.
   $\hat{\theta} \sim N(\theta, V(\theta))$
   $X \sim N(\mu, \sigma^2)$
   $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

   Intervallo di confidenza: $\hat{\theta} - z_{\frac{\alpha}{2}}\sqrt{V\left(\hat{\vartheta}\right)}$, : $\hat{\theta} + z_{\frac{\alpha}{2}}\sqrt{V\left(\hat{\theta}\right)}$

   $X \sim bern(p)$
   $\overline{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$

   Intervallo di confidenza: $\overline{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x}\overline{1-(x)}}{n}}$. $\overline{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{x}\overline{1-(x)}}{n}}$

VERIFICA DELLE IPOTESI:

$X \sim f(x, \theta)$

θ non lo conosciamo, dobbiamo stabilire se una certa ipotesi è accettabile:

H_0: ipotesi nulla con $H_0 : \theta \in H_0$

H_1 :ipotesi alternativa con $H_1 : \theta \in H_1$

$H_0 : \mu = 178$ $H_1 : \mu \neq 178$

$H_0 : \mu \geq 175$ $H_1 : \mu < 175$

$X_1, X_2, \ldots, X_n$

Accetto $H_0 \rightarrow$ sulla base del campione H_0 è più plausibile dell'ipotesi alternativa

Rifiuto $H_0 \rightarrow$ scarto l'ipotesi nulla

Facendo questa valutazione incorro in 2 errori:

1) Rifiuto H_0 se H_0 è vera $\rightarrow$ $\alpha =$ probabilità errore primo tipo
2) Accetto H_0 se H_0 è falsa $\rightarrow \beta =$ probabilità errore secondo tipo

POTENZA:

capacità di rifiutare H_0 quando H_0 è falsa.

Se aumenta la potenza aumenta la capacità di verifica dell'ipotesi quindi il metodo migliora.

$\prod$ =probabilità(R H_0 | H_0 falsa)

METODO DEL RAPPORTO DI VEROSIMIGLIANZA:

$L(\theta)$

$\lambda = max_{\theta \in H_0} (L(\theta) / \max_1 \theta L(\theta)) \leq 1$

λ è una misura di evidenza a favore dell'ipotesi nulla.

$\lambda \to 1$ ipotesi nulla plausibile

$\lambda \to 0$ ipotesi nulla poco plausibile

$H_0 : \theta \geq \theta_0$

$H_1 : : \theta < \theta_0$

$\lambda = \frac{90}{110}$=0,818 $\to$ c'è forte evidenza rispetto H_0

Rifiuto se $\lambda \leq k$

$p(\lambda \leq k | \ H_0$ vera)=α

$X \sim N(\mu, \sigma^{\uparrow}2$)

Con σ^2 nota $\to$ $\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \geq z_\alpha$ deriva dal metodo del rapporto di verosimiglianza

IL RAPPORTO DI VEROSIMIGLIANZA IN GENERALE E' IL Più POTENTE.

<u>CONFRONTO TRA MEDIE CON VARIANZE NOTE:</u>

supponiamo di avere 2 popolazioni

P_1 : X$\sim N$ (μ_1 , σ_1^2)

P_2 : Y$\sim N$ (μ_2 , σ_2^2)

H_0 : $\mu_{1=}\mu_2 \begin{cases} H_{1:} \ \mu_1 > \mu_2 \\ H_1 : \mu_1 < \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$

$\bar{X}$-$\bar{Y}$ con $\bar{X} = \frac{1}{n_1}\sum_i x_i$ $\bar{Y} = \frac{1}{n_2}\sum_i y_i$

$Z = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ standardizzazione → $Z \sim N(0,1)$ *se* H_0

Se

$H_1: \mu_1 > \mu_2 \rightarrow rifiuto\ se\ z \geq z_\alpha$

$H_1: \mu_1 < \mu_2 \rightarrow rifiuto\ se\ z \leq -z_\alpha$

$H_1: \mu_1 \neq \mu_2 \rightarrow rifiuto\ se\ |z| \geq z_{\frac{\alpha}{2}}$

<u>SE NON CONOSCO LE VARIANZE:</u>

σ_1^2, σ_2^2 non note

ho 2 casi:

1) $\sigma_1^2 = \sigma_2^2$

$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1+n_2-2}$

$T = \frac{X-Y}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t(n_1+n_2-2)$ *sotto* H_0

2) *GRANDI CAMPIONI*

$Z = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \sim N(0,1)$ *sotto* H_0