Lezione n. 3 (a cura di Ubaldi Marica)

NORMALE MULTIVARIATA

E'l'unica distribuzione relativa alle variabili casuali multiple.
Una VARIABILE CASUALE MULTIPLA può essere indicata con il vettore $\underline{X} = (X_1, X_2, ...X_n)$ contenente n elementi di variabili casuali semplici.
I singoli valori sono dati dal vettore $\underline{x} = (x_1, x_2, ...x_n)$

Il 1° aspetto da considerare è la FUNZIONE DI DENSITA', che può essere scritta come:
$f(x_1, x_2, ...x_n) = f(\underline{x})$ calcolata in un certo valore x.
La formula rappresenta un'estensione della funzione di densità della Normale:

$$f(\underline{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \cdot e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^1 \Sigma^{-1}(\underline{x}-\underline{\mu})}$$ dove:

- $\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_n. \end{pmatrix}$ è un vettore n*1 variabili e gli elementi sono le medie delle singole variabili.
- $\sum = \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & . & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & . & Cov(X_2, X_n) \\ . & . & . & . \\ Cov(X_n, X_1) & . & . & Var(X_n) \end{pmatrix}$

è la matrice VARIANZA –COVARIANZA di dimensione n*n.

Proprietà della matrice:
- *Quadrata*: ha lo stesso numero di righe e di colonne
- *Simmetrica*: poiché non dipende dall'ordine degli elementi. La $Cov(X_i, X_j) = Cov(X_j, X_i)$
- *È sempre definita positiva*
  Def: Una matrice si dice definita positiva quando:
  $$\underbrace{a^1 \sum a}_{forma\_quadratica} > 0 \ , \ \forall a \neq 0$$

Es:

$$\underline{X} = \begin{pmatrix} X_1 \\ X_n \end{pmatrix}$$ con n=2 , X_1=altezza , X_2=peso

Scrivendo $\underline{X} \sim N_2\ (\mu, \Sigma)$ si può imporre che $\mu = \begin{pmatrix} 175 \\ 70 \end{pmatrix}$ e $\Sigma = \begin{pmatrix} 50 & 20 \\ 20 & 30 \end{pmatrix}$ allora:

$$\rho = \frac{20}{\sqrt{50 \cdot 30}} > 0$$

La condizione di " sempre definita positiva" implica che le varianze sulla diagonale di Σ siano positive.

Confrontiamo la formula sopra di $f(\underline{x})$ con la formula della densità della normale:

$$f(\underline{x}) = \underbrace{\frac{1}{\sqrt{|2\pi\Sigma|}}} \cdot e^{\overbrace{-\frac{1}{2}(\underline{x}-\underline{\mu})^1\Sigma^{-1}(\underline{x}-\underline{\mu})}}$$

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

Considerando i vettori: $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ $\underline{x} - \mu = \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}$

Per la forma quadratica:

$$\begin{pmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} =$$

$$= \begin{pmatrix} (x_1 - \mu_1)a_{11} + (x_2 - \mu_2)a_{21} & (x_1 - \mu_1)a_{12} + (x_2 - \mu_2)a_{22} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} =$$

$$= (x_1 - \mu_1)^2 a_{11} + (x_2 - \mu_2)(x_1 - \mu_1)a_{21} + (x_1 - \mu_1)(x_2 - \mu_2)a_{12} + (x_2 - \mu_2)^2 a_{22}$$

PROPRIETA':

1) Σ è DIAGONALE: ha elementi =0 fuori dalla diagonale

$$\Sigma = \begin{pmatrix} 50 & 0 \\ 0 & 30 \end{pmatrix}$$

Allora $Cov(X_i, X_j) = 0$ per $\forall i, j$ e le variabili di Σ sono *incorrelate* e *indipendenti.*

Pur essendo concetti diversi, in caso di normale multivariata non c'è differenza tra indipendenza ed incorrelazione.

2) $X \sim N(\mu,\Sigma) \Rightarrow: X_i \sim N(\mu_i,\sigma_{ii})$, $\forall_j$.

Ciò vuol dire che se la normale multivariata "è vera", ogni variabile ha distribuzione Normale e posso ricavare direttamente la distribuzione marginale.

3) Considerando le *distribuzioni condizionate* $X_2|X_1 = 180 \sim N(\mu_2^*, (\sigma^2)^*)$

4) Date n variabili, la *combinazione lineare* è:

$Y = a_1X_1 + a_2X_2 + ...a_nX_n = \underline{a}'\underline{X}$ (generale)

utilizzando la nozione matriciale, si ha:

$$Y = \underline{a}'X \quad \text{dove} \quad a = \begin{pmatrix} a_1 \\ a_2 \\ . \\ a_n \end{pmatrix}$$

Ritornando a livello generale:

$X \sim N(\underline{a}'\underline{\mu}, \underline{a}'\Sigma\underline{a})$

Ogni combinazione lineare di variabili casuali normali ha ancora distribuzione normale.

CONCETTI DI INFERENZA STATISTICA

Si è interessati ad una variabile X (es. reddito) con una certa distribuzione nella popolazione di riferimento: $X \sim f(x,\theta)$ dove θ è il parametro.
Il problema dell'inferenza è che si conosce la distribuzione ma non il parametro.
Si procede estraendo un campione dalla popolazione e si fa inferenza attribuendo un valore a θ (metodo della stima puntuale).

Vi sono *3 metodi* per effettuare l'inferenza:
-*Stima puntuale*
-*Stima per intervallo*
-*Verifica per ipotesi*

METODO DELLA STIMA PUNTUALE

Si assegna un unico valore a θ applicando una certa formula ai dati del campione (es. media):
$T = t(X_1, X_2, \ldots, X_n)$ dove T = stimatore

<u>*Criterio di massima verosimiglianza*</u>:
è un metodo generale per costruire uno stimatore di σ.
Questo si basa sulla *funzione di verosimiglianza*:

$$L(\theta) = \prod_i f(x_i;\theta)$$

è la probabilità o densità del campione osservato vista come funzione di θ.
Non è altro che una misura di evidenza fornita dal campione a favore di ogni valore di θ.

Esempio:
Se $L(\theta_1) > L(\theta_2) \Rightarrow \theta_1$ è più plausibile/attendibile di θ_2

X ~ Bernulli(p) con $\theta = p$

$$X = \begin{cases} 0, 1-p \\ 1, p \end{cases} \qquad f(x,p) = p^x (1-p)^{1-x}$$

La verosimiglianza sarà:

$$L(p) = \prod_i p^{x_i} (1-p)^{(1-x_i)} = p^{\sum_i x_i} (1-p)^{\sum_i (1-x_i)} = p^s (1-p)^{n-s} \qquad \text{con} \quad S = \sum_i x_i$$

Es: Si considera un campione n=100 (0,0,1,0,1…)
S=25 (tra i 100 studenti almeno 25 hanno sostenuto più di 5 esami il 1° anno di università)
Se, osservando 2 valori, L(0,5)>L(0.7) allora il valore più plausibile è 0,5.

Il valore più plausibile in assoluto:
$\max_\theta L(\theta) \rightarrow$ il valore corrispondente è $\hat{\theta}$ = STIMATORE DI MASSIMA VEROSIMIGLIANZA

Per trovare $\hat{\theta}$ si deve risolvere il problema di <u>*ottimizzazione*</u> attraverso le seg. fasi:
1) Passo dalla verosimiglianza alla log-verosimiglianza :
$$l(\theta) = \log L(\theta) = \sum_i \log f(x_i;\theta)$$
2) Calcolo la derivata prima e la pongo =0: $\frac{d}{d\theta} l(\theta) = 0$

3) Verifico che è un punto di max; calcolando la derivata seconda e osservando che sia negativa:

$$\frac{d^2}{d\theta^2} l(\theta) < 0$$

Esempio: $L(p) = p^s (1-p)^{n-s}$

1) $l(p) = S \log(p) + (n-S)\log(1-p)$

2) $\frac{d}{dp} l(p) = \frac{s}{p} - \frac{n-S}{1-p} = \frac{S(1-p) - p(n-S)}{p(1-p)} = \frac{s-pn}{p(1-p)} = 0$

Il denominatore lo posso trascurare poiché sempre positivo e avrò:

$S - pn = 0 \;\Rightarrow\; p = \frac{S}{n}$ è un punto di ottimo, ora per capire se è min o max facciamo la derivata seconda.

3) $\frac{d^2}{dp^2} l(p) = -\frac{S}{p^2} - \frac{n-S}{(1-p)^2} < 0$ è sempre negativo quindi è un punto di max.