STATISTICA INFERENZIALE

Introduzione

Finora abbiamo visto che esistono modelli probabilistici che possiamo utilizzare per prevedere gli esiti di esperimenti aleatori. Naturalmente la previsione è di tipo probabilistico: diciamo infatti con quanta probabilità si verificano certi esiti.
Con la statistica inferenziale si utilizzano i dati di più esperimenti aleatori per dire qualcosa sul modello probabilistico che si suppone incognito. In tutti i problemi di stima dei parametri si seguono questi passi:

1. Scegliere un modello con un parametro incogniti
2. Fare n esperimenti
3. Raccogliere i dati e calcolare una funzione di questi dati
4. Usare questa funzione per stimare il parametro incognito

Modello statistico parametrico

Diamo ora una serie di definizioni utili allo studio della statistica inferenziale. Definiamo un modello statistico in due diversi modi:

- Utilizzando la densità = un modello statistico parametrico è una famiglia di leggi di variabili aleatorie dipendenti da uno o più parametri ϑ: lo indichimo tramite la funzione di densità $\{f(x;\vartheta):\vartheta \in \Theta\}$ dove Θ è lo spazio dei parametri. In altre parole, il modello statistico indica la scelta di un tipo di variabile aleatoria.
- In generale = un modello statistico parametrico è una famiglia di leggi di variabili aleatorie dipendenti da uno o più parametri ϑ: lo indichimo tramite la legge $\{\mathcal{L}_\vartheta:\vartheta \in \Theta\}$ dove Θ è lo spazio dei parametri. In altre parole, la legge $\mathcal{L}_\vartheta$ simboleggia la scelta di un tipo di legge (la bernoulliana, la binomiale, la poissoniana, normale etc)

Stimatore

Definiamo inoltre il campione casuale di dimensione n una n-upla di variabili aleatorie $X_1, \dots, X_n$. Dato un campione casuale $X_1, \dots, X_n$ si dice statistica di variabile aleatorie T funzione del campione casuale che non sia funzione di alcun parametro incognito. Assegnata la statistica T una volta estratto un particolare campione il numero $\tau = t(x_1, \dots x_n)$ si dice stima di $g(\vartheta)$. Uno stimatore si dice non distorto o corretto se $\mathbb{E}_\vartheta(T) = g(\vartheta)$ ovvero quando il valore atteso della statistica è uguale alla sua stima. In generale se ciò non accade si definisce distorsione la differenza tra il valore atteso della statistica e la sua stima: $\mathbb{E}_\vartheta(T) - g(\vartheta)$.
Spesso lo stimatore scelto dipende esplicitamente dal numero n ovvero dalla numerosità del campione. Una successione di stimatori si dice consistente se il limite per n che tende a infinito del valore atteso del quadrato della distorsione è nullo ovvero:

$$\lim_{n\to\infty} \mathbb{E}_\vartheta\left(\left(T_n - g(\vartheta)\right)^2\right) = 0$$

Il valore atteso del quadrato della distorsione $\mathbb{E}_\vartheta\left(\left(T_n - g(\vartheta)\right)^2\right)$ è detto rischio quadratico medio. Per uno stimatore consistente vale che:

$$\mathbb{E}_\vartheta\left(\left(T_n - g(\vartheta)\right)^2\right) = Var(T_n) + \left(\mathbb{E}_\vartheta\left(T_n - g(\vartheta)\right)\right)^2$$

Se lo stimatore oltre a essere consistente è anche non distorto allora vale che:

$$\mathbb{E}_\vartheta\left(\left(T_n - g(\vartheta)\right)^2\right) = Var(T_n)$$