

Partizioni sufficienti, sufficienti minimali e di verosimiglianza

Consideriamo un campione casuale (iid) $\mathbf{X}_n = (X_1, \dots, X_n)$ di dimensione n da una v.a. bernoulliana di parametro incognito θ . Il modello statistico è quindi

$$\left\{ \mathcal{X}^n = \{0, 1\}^n, \quad f_n(\mathbf{x}_n; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}, \quad \theta \in [0, 1] \right\}.$$

La funzione di verosimiglianza associata a un campione osservato $\mathbf{x}_n$ è

$$\begin{aligned} L(\theta) &= \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} \\ &= \theta^{n\bar{x}_n} (1 - \theta)^{n - n\bar{x}_n}, \quad \theta \in [0, 1]. \end{aligned}$$

Osserviamo preliminarmente quanto segue.

- Per il criterio di fattorizzazione, sia $T_1(\mathbf{x}_n) = \sum_{i=1}^n x_i$ che $T_2(\mathbf{x}_n) = \bar{x}_n$ sono **statistiche sufficienti** per il modello.
- Le statistiche T_1 e T_2 sono a valori scalari, ovvero: $T_j : \mathcal{X}^n \rightarrow \mathbb{R}^1$, $j = 1, 2$. L'informazione degli n numeri $x_1, \dots, x_n$ è esattamente la stessa di quella fornita dal numero $\sum x_i$ (o, equivalentemente, dal numero $\bar{x}_n$).
- Passando dai campioni alle statistiche sufficienti si ha una sostanziale **riduzione dimensionale** delle funzioni dei dati con cui dobbiamo avere a che fare (nell'esempio, da n a 1).
- T_1 e T_2 sono in **relazione biunivoca** (one-to-one).
- Le due statistiche T_1 e T_2 , oltre ad essere sufficienti, sono anche **minimali**. Ciò si evince dal **criterio di Lehmann-Scheffe**. Lo scopo di questo esempio è capire meglio questo aspetto.
- Per confronto consideriamo la statistica $T_3(\mathbf{x}_n) = (\sum x_i, x_1)$, che assume, per ogni campione, una coppia di valori, ovvero: $T_3 : \mathcal{X}^n \rightarrow \mathbb{R}^2$.
- Le statistiche T_1, T_2, T_3 sono tutte sufficienti; la statistica T_3 fornisce però un valore inutile: conoscere il valore di x_1 non serve per ottenere il nucleo della fdv.
- Le statistiche T_1 e T_2 si possono ottenere da T_3 ; T_3 non si può ottenere da T_1 o T_2 (se conosco solo quanto vale la somma o la media di tre numeri, non posso conoscere il primo dei tre valori). In altri termini, T_1 e T_2 sono funzioni di T_3 , ma non è vero il viceversa.
- Da quanto appena detto possiamo concludere che la statistica T_3 è sufficiente ma **non minimale**.

Per comprendere meglio la differenza profonda che sussiste tra le statistiche T_1 e T_2 da una parte (entrambi sufficienti minimali) e la statistica T_3 dall'altra (sufficiente ma non minimale), consideriamo il caso in cui $n = 3$ ed esaminiamo la seguente tabella, suddivisa in tre macrocolonne. In particolare, esaminiamo le **partizioni dello spazio dei campioni** indotte dalle tre statistiche sufficienti.

- La prima macrocolonna riporta la partizione dello spazio dei campioni indotta da T_1 e T_2 (la partizione è la stessa, in quanto le due statistiche sono in relazione biunivoca). Come vedremo, si tratta della **partizione sufficiente minimale**.
- La seconda macrocolonna riporta la partizione indotta dalla statistica sufficiente (non minimale) $T_3(\mathbf{x}_n) = (\sum x_i, x_1)$. Come vedremo, si tratta di una **partizione sufficiente non minimale**.
- La terza macrocolonna riporta la **partizione di verosimiglianza**, che coincide con quella indotta da T_1 e T_2 : proprio per questa coincidenza possiamo dire che T_1 e T_2 , benchè distinte tra loro (in quanto danno valori diversi), sono entrambe sufficienti minimali ed equivalenti in termini di informazione fornita su θ .

Partizione minimale				Partizione non minimale			Partizione di verosimiglianza	
		T_1	T_2			T_3		$L(\theta) = g(\theta; \mathbf{x}_n)$
$\mathcal{X}^n$		$\sum x_i$	$\bar{x}_n$	$\mathcal{X}^n$		$(\sum x_i, x_1)$		$\theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}$
(0, 0, 0)	A_1	0	0	(0, 0, 0)	A'_1	(0, 0)	A_1	$(1 - \theta)^3$
(1, 0, 0)	A_2	1	$\frac{1}{3}$	(1, 0, 0)	A'_2	(1, 1)	A_2	$\theta(1 - \theta)^2$
(0, 1, 0)				(0, 1, 0)	A'_3	(1, 0)		
(0, 0, 1)				(0, 0, 1)		(1, 0)		
(0, 1, 1)				(0, 1, 1)	A'_4	(2, 0)		
(1, 1, 0)	A_3	2	$\frac{2}{3}$	(1, 1, 0)	A'_5	(2, 1)	A_3	$\theta^2(1 - \theta)$
(1, 0, 1)				(1, 0, 1)		(2, 1)		
(1, 1, 1)	A_4	3	1	(1, 1, 1)	A'_6	(3, 1)	A_4	θ^3
n. gruppi	4	4	4	6	6	6	4	4

Vediamo nel dettaglio.

- Lo spazio dei campioni $\mathcal{X}^n$ è costituito da 8 campioni distinti (colonne 1 e 5).
- Le statistiche $T_1 = \sum_{i=1}^n x_i$ e $T_2 = \bar{x}_n$ inducono la partizione $\{A_1, \dots, A_4\}$ dello spazio dei campioni $\mathcal{X}^n$, dove
 $A_1 = \{(0, 0, 0)\}$,
 $A_2 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$,
 $A_3 = \{(0, 1, 1), (1, 1, 0), (1, 0, 1)\}$,
 $A_4 = \{(1, 1, 1)\}$.

Si tratta di una **partizione dello spazio dei campioni** $\mathcal{X}^n$, in quanto $\cup_{j=1}^4 A_j = \mathcal{X}^n$ e $\cap_{j=1}^4 A_j = \emptyset$.

- La partizione A_t indotta da T_1 e T_2 coincide con quella indotta dalla funzione di verosimiglianza:
 - tutti i campioni nello stesso insieme A_j , che hanno stesso valore della statistica, danno la stessa funzione di verosimiglianza;
 - tutti i campioni con stessa verosimiglianza sono nello stesso insieme A_j ;
 - i campioni che danno verosimiglianza diversa sono in insiemi A_j distinti.

In altri termini: campioni distinti portano allo stesso valore della statistica $\iff$ portano alla stessa verosimiglianza (criterio di Lehmann-Scheffe).

- La statistica $T_3 = (\sum_{i=1}^n x_i, x_1)$ induce la partizione $\{A'_1, \dots, A'_6\}$ dello spazio dei campioni $\mathcal{X}^n$, dove
 $A'_1 = \{(0, 0, 0)\} = A_1$,
 $A'_2 = \{(1, 0, 0)\}$,
 $A'_3 = \{(0, 1, 0), (0, 0, 1)\}$,
 $A'_4 = \{(0, 1, 1)\}$,
 $A'_5 = \{(1, 1, 0), (1, 0, 1)\}$,
 $A'_6 = \{(1, 1, 1)\} = A_4$.
- Il numero di elementi della partizione $\{A'_t\}$ è pari a 6; questo numero è superiore a quello della partizione $\{A_t\}$ di verosimiglianza e di T_1 e T_2 , pari a 4. Infatti, la presenza dell'elemento inutile (ai fini dell'individuazione della fdv) x_1 in T_3 fa sì che, nella corrispondente partizione, il campione (1, 0, 0) risulti in un insieme distinto da quello in cui si trovano (0, 1, 0) e (0, 0, 1), nonostante i tre campioni diano luogo alla stessa verosimiglianza. In altre parole, l'insieme A_2 viene diviso in due insiemi A'_2 e A'_3 nella partizione di $\mathcal{X}^n$ indotta da T_3 . Lo stesso succede per l'insieme A_3 che viene suddiviso nella partizione indotta da T_3 in A'_4 e A'_5 , sebbene i tre campioni diano luogo alla stessa verosimiglianza.
- La partizione indotta dalle statistiche sufficienti e minimali (qui T_1 e T_3) è costituita da 4 insiemi ed è **meno fine** di quella indotta dalla statistica sufficiente ma non minimale T_3 , costituita da 6 insiemi di campioni.¹

¹Dati due collezioni di insiemi, il primo è **meno fine** del secondo se contiene meno insiemi. Qui la partizione $\{A_t\}$ contiene 4 insiemi di campioni ed è quindi meno fine della partizione $\{A'_t\}$ che ne contiene 6.

- La partizione sufficiente minimale coincide con quella di verosimiglianza.
- In generale, la partizione indotta da una statistica sufficiente minimale è quella **meno fine** tra quelle indotte da tutte le possibili statistiche sufficienti non minimali di un modello.
- Le statistiche sufficienti minimali inducono una e una sola partizione, che coincide con quella di verosimiglianza. Questo dipende dal fatto che le statistiche sufficienti minimali sono in corrispondenza biunivoca tra loro.
- Per quanto detto, possiamo affermare che le statistiche sufficienti minimali sono infinite (come le funzioni biunivoche di una data quantità) ma **la partizione sufficiente minimale è unica**.