Il criterio di fattorizzazione

Luca La Rocca

13 ottobre 2018

Sia $T = g(Z)$ una statistica basata sul dato discreto Z. Siano $\mathcal{Z}$ lo spazio dei possibili valori di Z e $\mathcal{T}$ lo spazio dei possibili valori di T, di modo che $g : \mathcal{Z} \to \mathcal{T}$. Consideriamo il modello statistico definito da

$$\mathbb{P}_\theta\{Z = z\} = f_{Z|\Theta}(z|\theta), \quad z \in \mathcal{Z}, \tag{1}$$

al variare di θ in H. Possiamo senz'altro supporre che, comunque preso $z \in \mathcal{Z}$, esista $\theta \in H$ tale che $f_{Z|\Theta}(z|\theta) > 0$. Il **teorema di fattorizzazione** afferma che T è sufficiente per (1) se e solo se esistono $p : \mathcal{T} \times H \to \mathbb{R}_+$ e $q : \mathcal{Z} \to \mathbb{R}_+$ tali che

$$f_{Z|\Theta}(z|\theta) = p(g(z);\theta)q(z) \tag{2}$$

per ogni $z \in \mathcal{Z}$ e $\theta \in H$. Chiamiamo condizione di fattorizzazione la (2) e notiamo che in essa, per quanto supposto, necessariamente $q(z) > 0$ per ogni $z \in \mathcal{Z}$.

Mostriamo in primo luogo che la sufficienza di T implica la condizione di fattorizzazione. In effetti, se T è sufficiente, possiamo scrivere

$$\begin{aligned} f_{Z|\Theta}(z|\theta) &= \mathbb{P}_\theta\{Z = z\} \\ &= \mathbb{P}_\theta\{Z = z, T = g(z)\} \\ &= \mathbb{P}_\theta\{T = g(z)\}\mathbb{P}_\theta(Z = z|T = g(z)) \\ &= f_{T|\Theta}(g(z)|\theta)f_{Z|T}(z|g(z)), \end{aligned}$$

dove non importa come definiamo $f_{Z|T}(z|t)$ quando $f_{T|\Theta}(t|\theta) = 0$, perché nel caso $f_{T|\Theta}(g(z)|\theta) = 0$ abbiamo necessariamente $f_{Z|\Theta}(z|\theta) = 0$; per esempio, volendo fissare le idee, quando $f_{T|\Theta}(t|\theta) = 0$ sceglieremo arbitrariamente $z_0 \in \mathcal{Z}$

e porremo $f_{Z|T}(z|t) = \mathbb{I}_{\{z_0\}}(z)$. Abbiamo così verificato che vale la condizione di fattorizzazione con $p(t;\theta) = f_{T|\Theta}(t|\theta)$ e $q(z) = f_{Z|T}(z|g(z))$.

Mostriamo ora che la condizione di fattorizzazione implica la sufficienza di T. In effetti, se vale la condizione di fattorizzazione, possiamo ricavare

$$
\begin{aligned}
f_{T|\Theta}(t|\theta) &= \mathbb{P}_\theta\{T = t\} \\
&= \sum_{z|g(z)=t} \mathbb{P}_\theta\{Z = z\} \\
&= \sum_{z|g(z)=t} f_{Z|\Theta}(z|\theta) \\
&= \sum_{z|g(z)=t} p(g(z);\theta)q(z) \\
&= p(t;\theta) \sum_{z|g(z)=t} q(z) \\
&= p(t;\theta)k(t),
\end{aligned}
$$

dove $k(t) = 0$ implica $f_{T|\Theta}(t|\theta) = 0$. Allora, nel caso interessante in cui abbiamo $f_{T|\Theta}(t|\theta) > 0$ e $t = g(z)$, troviamo

$$
\begin{aligned}
\mathbb{P}_\theta(Z = z|T = t) &= \frac{\mathbb{P}_\theta\{Z = z, T = t\}}{\mathbb{P}_\theta\{T = t\}} \\
&= \frac{\mathbb{P}_\theta\{Z = z, T = g(z)\}}{\mathbb{P}_\theta\{T = g(z)\}} \\
&= \frac{\mathbb{P}_\theta\{Z = z\}}{\mathbb{P}_\theta\{T = g(z)\}} \\
&= \frac{f_{Z|\Theta}(z|\theta)}{f_{T|\Theta}(g(z)|\theta)} \\
&= \frac{p(g(z);\theta)q(z)}{p(g(z);\theta)k(g(z))} \\
&= \frac{q(z)}{k(g(z))}
\end{aligned}
$$

che non dipende da θ. Se invece $f_{T|\Theta}(t|\theta) > 0$, ma $t \neq g(z)$, troviamo subito $\mathbb{P}_\theta(Z = z|T = t) = 0$ che di nuovo non dipende da θ. Se infine $f_{T|\Theta}(t|\theta) = 0$, possiamo scegliere arbitrariamente $z_0 \in \mathcal{Z}$ e porre $\mathbb{P}_\theta(Z = z|T = t) = \mathbb{I}_{\{z_0\}}(z)$ che ancora non dipende da θ. Abbiamo così verificato che T è sufficiente.

Osserviamo che $k(t) = \sum_{z|g(z)=t} q(z)$ si annulla se e solo se t non appartiene all'immagine di $\mathcal{Z}$ tramite g. Possiamo dunque ottenere $k(t) > 0$, per ogni $t \in \mathcal{T}$, semplicemente supponendo $\mathcal{T}$ privo di elementi superflui (come già fatto all'inizio per $\mathcal{Z}$). In questo modo potremo sempre scrivere

$$p(g(z);\theta) = c(z) f_T(g(z)|\theta)$$

con $c(z) = 1/k(z) > 0$ e vedere la funzione p nella condizione di fattorizzazione come un nucleo di verosimiglianza sia per Z che per T.

Se Z è un dato continuo, per semplicità, prendiamo il criterio di fattorizzazione come definizione di sufficienza. Questa estensione ci permette di riscontrare la **sufficienza approssimata** della coppia formata dallo stimatore di massima verosimiglianza $\hat{\Theta} = \hat{\theta}(Z)$ e dalla corrispondente informazione di Fisher $\hat{I} = i(\hat{\Theta}; Z)$ nel caso $H = \mathbb{R}$, vale a dire la sufficienza di $T = (\hat{\Theta}, \hat{I})$ per l'approssimazione normale della verosimiglianza che corrisponde all'approssimazione quadratica della log-verosimiglianza. Infatti, scrivendo

$$f_{Z|\Theta}(z|\theta) \simeq \sqrt{\frac{i(\hat{\theta}(z);z)}{2\pi}} \exp\left\{-\frac{i(\hat{\theta}(z);z)}{2}(\theta - \hat{\theta}(z))^2\right\},$$

troviamo la condizione di fattorizzazione con

$$p(t_1, t_2; \theta) = \sqrt{t_2} \exp\left\{-\frac{t_2}{2}(\theta - t_1)^2\right\}, \quad q(z) = \frac{1}{\sqrt{2\pi}},$$

per esempio, dove $t_1 = \hat{\theta}(z)$ e $t_2 = i(\hat{\theta}(z); z)$. Pertanto $T = (\hat{\Theta}, \hat{I})$ è sufficiente, in senso approssimato, ogni qual volta si abbia $H = \mathbb{R}$.

Sia ora $\mathcal{I}(z)$ un risultato inferenziale basato sul dato z (per esempio una stima puntuale o per intervallo). I due principi seguenti vincolano la funzione $\mathcal{I}$ sullo spazio dei possibili dati $\mathcal{Z}$. Il primo principio trova giustificazione nel fatto che, se abbiamo osservato una statistica sufficiente, possiamo completare l'osservazione del dato senza conoscere il valore del parametro. Il secondo principio trova invece giustificazione nella pratica dell'inferenza di verosimiglianza. Vedremo che, per il criterio di fattorizzazione, si tratta in effetti di **due principi equivalenti**.

Principio di sufficienza Se z' e z'' sono due dati in $\mathcal{Z}$ tali che $g(z') = g(z'')$ con $g(Z)$ statistica sufficiente, allora $\mathcal{I}(z') = \mathcal{I}(z'')$.

Principio debole di verosimiglianza Se z' e z'' sono due dati in $\mathcal{Z}$ che soddisfano la condizione di proporzionalità $f_{Z|\Theta}(z'|\theta) = k(z', z'')f_{Z|\Theta}(z''|\theta)$, per una qualche costante $k(z', z'') > 0$ e ogni $\theta \in H$, allora $\mathcal{I}(z') = \mathcal{I}(z'')$.

Possiamo verificare la condizione di proporzionalità nel principio debole di verosimiglianza per mezzo di un qualsiasi nucleo di verosimiglianza $c(z)f_{Z|\theta}(z|\theta)$, $\theta \in H$, dove $c(z) > 0$. In particolare, se supponiamo per semplicità $f^{\max}_{Z|\Theta}(z) = \sup_{\theta\in H} f_{Z|\Theta}(z|\theta) < \infty$, per ogni $z \in \mathcal{Z}$, possiamo avvalerci della funzione di verosimiglianza relativa $\tilde{L}(\theta; z) = f_{Z|\Theta}(z|\theta)/f^{\max}_{Z|\Theta}(z)$, $\theta \in H$. In questo modo la condizione di proporzionalità nel principio debole di verosimiglianza diventerà una condizione di uguaglianza: $\tilde{L}(\theta; z') = \tilde{L}(\theta; z'')$, per ogni $\theta \in H$, vale a dire $\tilde{L}(\cdot; z') = \tilde{L}(\cdot; z'')$ come funzioni da H a $\mathbb{R}$ (elementi di $\mathbb{R}^H$).

Siamo pronti a verificare che il principio di sufficienza implica il principio debole di verosimiglianza. A tal fine, consideriamo come statistica $g(Z)$ il processo di verosimiglianza relativa $\tilde{L}(\theta; Z)$, $\theta \in H$, per il quale $g : \mathcal{Z} \to \mathbb{R}^H$ è definita da $g(z) = \tilde{L}(\theta; z)$, $\theta \in \mathbb{R}$. Posto $t = g(z)$, possiamo scrivere la condizione di fattorizzazione con $q(z) = f^{\max}_{Z|\Theta}(z)$ e $p(\theta; t) = t(\theta)$, $t \in \mathbb{R}^H$, $\theta \in H$. Resta così dimostrato che $g(Z)$ è sufficiente e ne segue immediatamente l'asserto.

Verifichiamo ora che il principio debole di verosimiglianza implica il principio di sufficienza. In effetti, se $T = g(Z)$ è una statistica sufficiente, la condizione di fattorizzazione comporta $f_{Z|\Theta}(z'|\theta) = k(z', z'')f_{Z|\Theta}(z''|\theta)$, per ogni $\theta \in H$, con la costante $k(z', z'') = q(z')/q(z'')$. Ne segue immediatamente l'asserto.

Il **principio di verosimiglianza** (detto forte in contrapposizione a quello debole) richiede l'equivalenza dei risultati inferenziali basati su $z' \in \mathcal{Z}'$ e $z'' \in \mathcal{Z}''$, anche quando $\mathcal{Z}' \neq \mathcal{Z}''$, se $f_{Z'|\Theta}(z'|\theta) = k(z', z'')f_{Z''|\Theta}(z''|\theta)$. Il teorema di Birnbaum afferma che questa richiesta è equivalente a combinare il principio di sufficienza con un opportuno *principio di condizionalità*; si veda per esempio l'Appendice B del libro "An Introduction to Bayesian Analysis: Theory and Methods" di Jayanta K. Ghosh, Mohan Delampady & Tapas Samanta, pubblicato da Springer nel 2006 (e disponibile a Modena presso la Biblioteca Scientifica Interdipartimentale di UniMoRe con collocazione SALA MATEM A.17/785).