Data Reduction

Vers. 1.0.3

Gianluca Mastrantonio

gianluca.mastrantonio@polito.it

Outline

Qualche cenno in più sull'inferenza statistica I

La statistica nasce come tentativo di dare obiettività e universalità alla ricerca scientifica e come i dati, o i risultati degli esperimenti, dovessero essere analizzati e trattati.

La statistica si differenzia dalla altre scienze matematica per il modo in cui l'inferenza viene fatta. Nella matematica (logica, algebra, geometria), dove le premesse sono certe e generali, come le conclusioni, a cui si arriva grazie a un ragionamento deduttivo (dal generale al particolare). Nella statistica le basi sono certe (i dati, il campione) e si arriva a considerazioni generali sul processo generativo dei dati (e.g. parametri della distribuzione) o la popolazione, tramite un processo induttivo (dal particolare al generale) che sono incerte, affette da errore, che può essere quantificato grazie alla probabilità.

Qualche cenno in più sull'inferenza statistica II

Modello statistico

Dato un campione $\mathbf{X} = (X_1, \ldots, X_n)^T$, un modello statistico $\mathcal{F}$ è una famiglia di leggi di probabilià F per $\mathbf{X}$, indicizzata da un parametro $\theta \in \Theta$ (non per forza scalare), dove Θ è lo spazio parametrico:

$$\mathcal{F} = \{F(.;\theta); \theta \in \Theta\}$$

oppure utilizzando la pdf o pmf

$$\mathcal{F} = \{f(.;\theta); \theta \in \Theta\}$$

Avendo osservato dei dati, si decide un modello statistico, e lo scopo è quello di dire qualcosa su θ

Qualche cenno in più sull'inferenza statistica III

Esempio - Esperimento di Michelson

Nell 1987 Michelson fece 5 esperimenti in cui in ognonu fece 20 misurazione della velocità della luce. Assumiamo che i 5 esperimenti siano stati condotto nelle stesse identiche condizioni, e assumiamo che i dati ($20 \cdot 5$ misurazioni) provengono da una normale $N(\mu, \sigma^2)$, con μ e σ^2 non noti.

Esempio - Esperimento di Michelson - 2

Stesso esperimento di prima, ma in questo caso assumiamo che i dati provengano da una normale $N(\mu, \sigma^2)$, con σ^2 noto. Questo può succedere nella realtà visto che spesso la varianza nelle misurazioni è data dal produttore dello strumento.

Esempio - Esperimento di Michelson - 3

Stesso esperimento di prima, ma in questo caso assumiamo che i dati provengano da una gamma $G(a, b)$, con a e b non noti.

In tutti gli esempi, stiamo facendo delle assunzioni, sia sulla distribuzione, sia sul fatto che i 5 esperimenti siano identici, e una volta che le assunzioni sono state fatte, il nostro scopo è dire qualcosa sui parametri. Le assunzioni fatta andrebbero verificate, ma questo è un argomento di cui non ci occuperemo.

Attenzione! Per semplicità noi assumeremo sempre che i dati siano iid ma, almeno che non specificato diversamente (per esempio nella regressione), tutti i risultati valgono in situazioni più generiche, con dati dipendenti e/o che provengono da distribuzioni diverse, per esempio un campione $\mathbf{X} = (X_1, X_2)$, con $X_1 \sim \text{Pois}(\lambda)$ e $X_2 \sim N(X_1, 1)$. Questo perchè potete sempre vedere il campione $\mathbf{X}$ come una **singola realizzazione** di un modello multivariato con componenti dipendenti.

Sufficienza I

Riprendiamo il concetto di *statistica* $T(\mathbf{X})$ che è una funzione del campione. Ogni statistica può essere vista come una partizione dello spazio dei campioni $\mathcal{X}$ (spazio di tutti i possibili risultati dell'esperimento).
Più precisamente definiamo $\mathcal{T} = \{t : t = T(\mathbf{x}) \text{ per } \mathbf{x} \in \mathcal{X}\}$ e $A_t = \{\mathbf{x} : T(\mathbf{x}) = t\}$, allora $\mathcal{T}$ partizione $\mathcal{X}$ nei set A_t dove campioni differenti che hanno lo stesso valore della statistica, cadono nello stesso subset. Pet esempio se due campioni $\mathbf{x}$ e $\mathbf{y}$ hanno lo stesso valore della statistica, $T(\mathbf{x}) = T(\mathbf{y}) = t$, allora $\mathbf{x} \in A_t$ e $\mathbf{y} \in A_t$.

Sufficienza II

Esempio - Partizione

Supponiamo di avere un campioni $\mathbf{X} = (X_1, X_2, X_3)^T$ di variabili iid da una bernulli di parametri p. Se utilizziamo la statistica $T(\mathbf{X}) = \sum_{i=1}^{3} X_i$, allora i possibili subset A_t sono

- A_0, corrispondente a $T(\mathbf{x}) = 0$, composto dal set $(0, 0, 0)^T$
- A_1, corrispondente a $T(\mathbf{x}) = 1$, composto dai sets $(1, 0, 0)^T$, $(0, 1, 0)^T$ e $(0, 0, 1)^T$
- A_2, corrispondente a $T(\mathbf{x}) = 2$, composto dai sets $(1, 1, 0)^T$, $(1, 0, 1)^T$ e $(0, 1, 1)^T$
- A_3, corrispondente a $T(\mathbf{x}) = 3$, composto dal set $(1, 1, 1)^T$

La partizione prodotta da $T(\mathbf{X})$ è uguale a quella di $\mathbf{X}$ o più **grossolana**.

Il campione $\mathbf{x}$ contiene informazioni sui parametri incogniti del modello statistico, e quando produciamo una partizione grossolana potremmo star perdendo informazione circa il parametro. Ci chiediamo quanto e in che modo possiamo scendere nel rendere grossolana una partizione, senza perdere informazione o, in altre parole, trovare la

Sufficienza III

maniera più sintetica per sintetizzare i dati (data reduction), senza perdere l'informazione su un parametro contenuta nel campione.

Esempio - Brexit

Si vuole conoscere l'esito del voto sulla Brexit, quindi la proporzione p di leave rispetto al totale. Vengono intervistati telefonicamente n soggetti a cui viene chiesto cosa voteranno. Ognuno di questo viene scelto a caso e abbiamo una probabilità p di intervistare un leave e $(1-p)$ un remain (assumiamo che siano le uniche due opzioni possibili, quindi niente indecisi o persone che non rispondono). Possiamo allora assumere che i dati X_i siano iid da una Bern(p).

Se vogliamo conoscere p, abbiamo veramente bisogno di conoscere il valore di ogni x_i oppure ci interessa meno, magari la statistica $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$, oppure $T(\mathbf{x}) = \sum_{i=1}^{n}(1-x_i)$?
Nell'esempio di Morley invece, dove assumiamo $N(\mu, \sigma^2)$, con σ^2 noto, per imparare μ potrebbe bastare

Sufficienza IV

- la media campionaria $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$
- la mediana $T(\mathbf{x}) = \hat{Q}(0.5)$;
- la differenza tra massimo e minimo $T(\mathbf{x}) = X_{(n)} - X_{(1)}$
- la coppia minimo e massimo $T(\mathbf{x}) = (X_{(1)}, X_{(n)})$.

Diamo una definizione più formale di questi concetti con il **principio di sufficienza** e le **statistiche sufficienti**.

Definizione - Principio di Sufficienza

La statistica $T(\mathbf{X})$ è sufficiente per un parametro θ, se ogni inferenza (quindi deduzione/induzione) su θ dipende dal campione $\mathbf{X}$ solo tramite il valore $T(\mathbf{X})$. In altre parole, se $\mathbf{x}$ e $\mathbf{y}$ sono due possibili campioni con $T(\mathbf{x}) = T(\mathbf{y})$, allora l'inferenza su θ deve essere la stessa sia se osserviamo $\mathbf{X} = \mathbf{x}$ o $\mathbf{X} = \mathbf{y}$.

Dal punto di vista matematico possiamo definire la statistica sufficiente nel seguente modo

Sufficienza V

Definizione - Statistica sufficiente

La statistica $T(\mathbf{X})$ è sufficiente per un parametro θ se la distribuzione condizionata di $\mathbf{X}$ dato $T(\mathbf{X})$, non dipende più da θ $\forall \mathbf{x} \in \mathcal{X}$ e e $T(\mathbf{X}) \in \mathcal{T}$.

Quello che dice la definizione è che la statistica $T(\mathbf{X})$ contiene tutta l'informazione sul parametro che possiamo avere dai dati. Fate attenzione che la statistica è sufficiente per un parametro in un determinato modello statistico, cioè la media campionaria potrebbe essere sufficiente per la media di una normale, ma non per il parametro di una $G(a, 1)$.

ATTENZIONE! la definizione di statistica sufficiente non deve valore per un particolare valore di $\mathbf{X}$ e $T(\mathbf{X})$, ma per tutti i possibili valori $\mathbf{X} \in \mathcal{X}$ e $T(\mathbf{X}) \in \mathcal{T}$

La verifica dalla sufficienza utilizzando la definizione può essere molto complicato, ma fortunatamente ci sono diversi teoremi che ci permettono di determinare se una specifica $T(\mathbf{X})$ è sufficiente.

Sufficienza VI

Teorema - Sufficienza

Se $p(\mathbf{x}|\theta)$ è la pmf o pdf di $\mathbf{X}$ e $q(t|\theta)$ è la pmf o pdf di $T(\mathbf{X})$, allora $T(\mathbf{X})$ è sufficiente per θ se e solo se per ogni $\mathbf{x} \in \mathcal{X}$ il rapporto

$$\frac{p(\mathbf{x}|\theta)}{q(t|\theta)}$$

è costante come funzione di θ.

Dimostrazione:

Per semplicità ipotizziamo che $\mathbf{X}$ e $T(\mathbf{X})$ siano discrete. ma la stessa dimostrazione si può fare nel caso continuo o misto. Per la definizione di statistica sufficiente abbiamo che

$$P_\theta(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t)$$

Sufficienza VII

non dipende da θ. $P_\theta(\mathbf{X} = x|T(\mathbf{X}) = t)$ è diversa da zero solo se $t = T(\mathbf{x})$ e quindi possiamo scrivere

$$P_\theta(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = T(\mathbf{x})) = \frac{P_\theta(\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = T(\mathbf{x}))}{P_\theta(T(\mathbf{X}) = T(\mathbf{x}))} = \frac{P_\theta(\mathbf{X} = \mathbf{x})}{P_\theta(T(\mathbf{X}) = T(\mathbf{x}))} = \frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

Quindi il rapporto $\frac{p(\mathbf{x}|\theta)}{q(t|\theta)}$ non dipende da θ.

$\square$

Vediamo qualche esempio

Esercizio - 1

Dimostrare che nell'esampio Brexit, $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ è sufficiente per p se le X_i sono iid da Bern(p)

Soluzione:

Sappiamo che la congiunta di $\mathbf{X}$ si può scrivere come

$$p(\mathbf{x}|p) = \prod_{i=1}^{n} p(x_i|p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Per trovare la distribuzione di $T(\mathbf{X})$ possiamo ricordarci che somma di n Bernulliane è distribuita come una binomiale di parametri (n, p), quindi

$$q(t|p) = \binom{n}{t} p^t (1-p)^{n-t}$$

Sufficienza IX

visto che $t = \sum_{i=1}^{n} x_i$. Possiamo scrivere il rapporto come

$$\frac{p(\mathbf{x}|p)}{q(T(\mathbf{x})|p)} = \frac{p^t(1-p)^{n-t}}{\left(\begin{array}{c} n \\ t \end{array}\right) p^t(1-p)^{n-t}} = \left(\begin{array}{c} n \\ t \end{array}\right)^{-1}$$

che non dipende da p, quindi $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ è sufficiente per p. $\square$

Questo esempio ci dice che per poter dire qualcosa su p non dobbiamo conoscere le risposte date da ognuno degli intervistati, ma ci basta sapere quante persone hanno detto di votare leave. Il che è abbastanza intuitivo.

Sufficienza X

Esercizio - 2

Riprendiamo l'esperimento di Michelson e dimostriamo che la media campionaria è sufficiente per la media della normale se σ^2 è noto.

Soluzione:

Prima di procedere con la dimostrazione vediamo un modo diverso per poter scrivere $\sum_{i=1}^{n}(x_i - \mu)^2$:

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - \mu)^2 + 2(\bar{x} - \mu)\sum_{i=1}^{n}(x_i - \bar{x}$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - \mu)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

dove abbiamo usato il fatto che $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$. La congiunta di $\mathbf{x}$ è

$$f(\mathbf{x}|\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

Abbiamo visto precedentemente che $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ e quindi

$$f(\bar{x}|\mu) = (2\pi\sigma^2/n)^{-\frac{1}{2}} \exp\left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right)$$

e il loro rapporto è dato da

$$\frac{f(\mathbf{x}|\mu)}{f(\bar{x}|\mu)} = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right)}{(2\pi\sigma^2/n)^{-\frac{1}{2}} \exp\left(-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right)} = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2}\right)}{(2\pi\sigma^2/n)^{-\frac{1}{2}}}$$

che non dipende più da μ dimostrando che la media campionaria è sufficiente per la media di una normale nel caso in cui σ^2 sia noto. $\square$

Nei due esempi precedenti era abbastanza facile trovare una statistica sufficiente, ma in casi generali può essere complicato. Per esempi, che statistica sufficiente potremmo testare per il parametro a di una $G(a, b)$? Quindi, sebbene il teorema sia utile, è limitante e abbiamo bisogno di qualcosa che ci dia indicazioni su quali sono le statistiche sufficienti per un parametro. Ci viene in aiuto il seguente teorema

Teorema di Fattorizzazione

Indichiamo con $f(\mathbf{x}|\theta)$ la congiunta, sia pmf che pdf, di un campione $\mathbf{X}$. Allora una statistica $T(\mathbf{X})$ è sufficiente per θ **se e solo se** esistono due funzioni $g(t|\theta)$ e $h(\mathbf{x})$ tale che per ogni campione $\mathbf{x} \in \mathcal{X}$ e parametro $\theta \in \Theta$, abbiamo

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

Dimostrazione:

Anche in questo caso diamo la dimostrazione nel caso discreto.

(i) dimostriamo prima che se $T(\mathbf{X})$ è sufficiente $\Longrightarrow$ esiste la fattorizzazione. In questo caso basta definire

$$h(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = T(\mathbf{x}))$$

che non dipende da θ perchè $T(\mathbf{X})$ è sufficiente, e

$$g(T(\mathbf{x})|\theta) = P_\theta(T(\mathbf{X}) = T(\mathbf{x}))$$

Teorema di fattorizzazione II

Per il “Teorema - Sufficienza” abbiamo che

$$h(\mathbf{x}) = \frac{f(\mathbf{x}|\theta)}{g(T(\mathbf{x})|\theta)} \Rightarrow f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

(ii) Assumiamo che la fattorizzazione esiste e quindi dobbiamo dimostrare che $\rightarrow T(\mathbf{X})$ è sufficiente.
Se definiamo $q(T(\mathbf{x}))|\theta)$ come la pmf di $T(\mathbf{X})$, dobbiamo dimostrare che il rapporto $f(\mathbf{x}|\theta)/q(T(\mathbf{x}))|\theta)$ non dipende da θ. Possiamo notare che si può scrivere

$$q(T(\mathbf{x}))|\theta) = \sum_{\mathbf{y}\in A_{T(\mathbf{x})}} f(\mathbf{y}|\theta)$$

cioè la probabilità di osservare $T(\mathbf{x})$ è uguale alla probabilità di osservare un campione $\mathbf{y}$ per cui $T(\mathbf{x}) = T(\mathbf{y})$. Visto ceh la fattorizzazione esiste, possiamo scrivere

$$\sum_{\mathbf{y}\in A_{T(\mathbf{x})}} f(\mathbf{y}|\theta) = \sum_{\mathbf{y}\in A_{T(\mathbf{x})}} g(T(\mathbf{y})|\theta)h(\mathbf{y}) = g(T(\mathbf{x})|\theta) \sum_{\mathbf{y}\in A_{T(\mathbf{x})}} h(\mathbf{y})$$

Teorema di fattorizzazione III

Siamo adesso pronti a dimostrare la sufficienza di $T(\mathbf{X})$ facendo vedere che il rapporto seguente non dipende da θ:

$$\frac{f(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)} = \frac{g(T(\mathbf{x})|\theta)h(\mathbf{x})}{g(T(\mathbf{x})|\theta)\sum_{\mathbf{y}\in A_{T(\mathbf{x})}} h(\mathbf{y})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{y}\in A_{T(\mathbf{x})}} h(\mathbf{y})}$$

che non dipende da θ. □

Quindi, data la congiunta del campione, questo teorema ci permette di trovare una statistica sufficiente.

Esercizio - Statistiche sufficienti Gamma

Trovare una statistica sufficiente per i parametri di una gamma $G(a, b)$, assumendo di aver osservato un campione $\mathbf{X}$ di dimensione n

Soluzione:

Possiamo usare il teorema di fattorizzazione, e facciamo vedere che esistono soluzioni differenti. Scriviamo intanto la congiunta che è

$$f(\mathbf{x}|a,b) = \prod_{i=1}^{n} \frac{b^a}{\Gamma(a)} x_i^{a-1} \exp\left(-bx_i\right) = \frac{b^{na}}{\Gamma(a)^n} \left(\prod_{i=1}^{n} x_i^{a-1}\right) \exp\left(-b\sum_{i=1}^{n} x_i\right) =$$
$$\frac{b^{na}}{\Gamma(a)^n} \exp\left((a-1)\sum_{i=1}^{n} \log(x_i)\right) \exp\left(-b\sum_{i=1}^{n} x_i\right)$$

Possiamo trovare diverse soluzione:

(1) definiamo $h(\mathbf{x}) = 1$ e

$$g(T(\mathbf{x})|\theta) = f(\mathbf{x}|a,b)$$

e questo ci permette di dimostrare che $T(\mathbf{X}) = (X_1, \dots X_n)$, cioè l'intero campione, è sufficiente (soluzione banale).

(2) Un'altra soluzione è il campione ordinato $T(\mathbf{X}) = (X_{(1)}, \ldots X_{(n)})$.
(3) la terza soluzione è quella più interessante: $T(\mathbf{X}) = (\sum_{i=1}^n \log(X_i), \sum_{i=1}^n X_i)$. □

Esercizio - Statistiche sufficienti Gamma II

Stesso problema precedente, ma in questo caso abbiamo che a è noto

Soluzione:

Come prima abbiamo che la congiunta si può scrivere

$$f(\mathbf{x}|a,b) = \frac{b^{na}}{\Gamma(a)^n} \exp\left((a-1)\sum_{i=1}^n \log(x_i)\right) \exp\left(-b\sum_{i=1}^n x_i\right)$$

ma siccome siamo interessati solo a b possiamo definire

$$h(\mathbf{x}) = \exp\left((a-1)\sum_{i=1}^{n}\log(x_i)\right)$$

e

$$g(T(\mathbf{x})|\theta) = \frac{b^{na}}{\Gamma(a)^n}\exp\left(-b\sum_{i=1}^{n}x_i\right)$$

E vedere che in questo caso la statistica sufficiente è $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$.

$\square$

Il caso in cui a sia noto e bisogna trovare la statistica sufficiente per b lo lascio come esercizio.

Due cose si possono notare nell'esempio precedente:

Teorema di fattorizzazione VII

- anche $T(\mathbf{X}) = (\sum_{i=1}^n \log(X_i), \sum_{i=1}^n X_i)$ è sufficiente per a sebbene portarsi dietro $\sum_{i=1}^n \log(X_i)$ non serve a niente. Come è sufficiente qualsiasi altra statistica $T(\mathbf{X}) = (\sum_{i=1}^n X_i, T'(\mathbf{x}))$.
- ogni funzione biunivoca di $\sum_{i=1}^n X_i$ è ancora sufficiente, per esempio $n\sum_{i=1}^n X_i$, $\left(\sum_{i=1}^n X_i\right)^2$ etc.

Funzioni biunivoche I

Concentriamoci per un momento su trasformazioni di statistiche sufficienti

Proposizione - Trasformazioni biunivoche di statistiche sufficienti

Se $T(\mathbf{X})$ è una statistica sufficiente per un parametro θ e $T'(\mathbf{X}) = r(T(\mathbf{X}))$ per ogni $\mathbf{x} \in \mathcal{X}$, e r è una funzione biunivoca con inversa $r^{-1}(\cdot)$, allora $T'(\mathbf{X})$ è una statistica sufficiente per θ.

Dimostrazione:

Se $T(\mathbf{X})$ è sufficiente per il teorema di fattorizzazione abbiamo

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x}) = f(\mathbf{x}|\theta) = g(r^{-1}(T(\mathbf{x}))|\theta)h(\mathbf{x}) = g^*(T'(\mathbf{x})|\theta)h(\mathbf{x})$$

e quindi $T'(\mathbf{x})$ è sufficiente per θ.

$\square$

Funzioni biunivoche II

Tra la moltitudine di statistiche sufficienti che possiamo trovare, dobbiamo definire un modo per valutare la "bonta" di uan statistica sufficiente. Ricordiamo l'idea alla base delle statistiche sufficienti è di avere una riduzione dei dati ma senza perdere informazioni riguardo il parametro θ che abbiamo nei dati. Per capire meglio il concetto riprendiamo un esempio fatto precedentemente e espandiamolo

Esempio - Partizione

Supponiamo di avere un campioni $\mathbf{X} = (X_1, X_2, X_3)^T$ di variabili iid da una bernulli di parametri p e vediamo le partizioni indotte da $T(\mathbf{X}) = \sum_{i=1}^{x} X_i$ e $T'(\mathbf{X}) = (\sum_{i=1}^{x} X_i, X_1)$.

Soluzione:

Ricordiamo che abbiamo visto che $T(\mathbf{X})$ è sufficiente per p e quindi lo è anche $T'(\mathbf{X})$. La partizione indotta da $T(\mathbf{X})$ è la seguente

- A_0, corrispondente a $T(\mathbf{x}) = 0$, composto dal set $(0, 0, 0)^T$

Funzioni biunivoche III

- A_1, corrispondente a $T(\mathbf{x}) = 1$, composto dai sets $(1,0,0)^T$, $(0,1,0)^T$ e $(0,0,1)^T$
- A_2, corrispondente a $T(\mathbf{x}) = 2$, composto dai sets $(1,1,0)^T$, $(1,0,1)^T$ e $(0,1,1)^T$
- A_3, corrispondente a $T(\mathbf{x}) = 3$, composto dal set $(1,1,1)^T$

mentre quella prodotta da $T'(\mathbf{X})$ è la seguente

- $A'_{0,0}$, corrispondente a $T'(\mathbf{x}) = (0,0)$, composto dal set $(0,0,0)^T$
- $A'_{1,0}$, corrispondente a $T'(\mathbf{x}) = (1,0)$, composto dai sets $(0,1,0)^T$ e $(0,0,1)^T$
- $A'_{1,1}$, corrispondente a $T'(\mathbf{x}) = (1,1)$, composto dal set $(1,0,0)^T$
- $A'_{2,0}$, corrispondente a $T'(\mathbf{x}) = (2,0)$, composto dal set $(0,1,1)^T$
- $A'_{2,1}$, corrispondente a $T'(\mathbf{x}) = (2,1)$, composto dai sets $(1,1,0)^T$ e $(1,0,1)^T$
- $A'_{3,1}$, corrispondente a $T'(\mathbf{x}) = (3,1)$, composto dal set $(1,1,1)^T$

Funzioni biunivoche IV

Entrambe le statistiche hanno la stessa informazione su θ, visto che sono entrambe sufficienti, ma $T'(\mathbf{X})$ produce una partizione meno grossolana $\square$

Prendiamo per esempio i valori $(1,0)$ and $(1,1)$ di $T'(\mathbf{x})$, anche se questi hanno valori differenti, i campioni associati devono produrre la stessa inferenza su θ perchè hanno lo stesso valore $T(\mathbf{x}) = 1$ (per il principio di sufficienza). Quindi $T'(\mathbf{x})$ produce una partizione meno grossolana.

Minimalità I

Quello che vogliamo è trovare la statistica sufficiente che produce la partizione più grossolana possibili o, in altre parole, trovare la massima riduzione dei dati. Introduciamo quandi il concetto di **minimalità**.

Definizione - Minimalità

Una statistica sufficiente $T(\mathbf{x})$ si dice statistica sufficiente minimale se per ogni altra statistica sufficiente $T'(\mathbf{x})$, $T(\mathbf{x})$ è una funzione di $T'(\mathbf{x})$.
In altre parole abbiamo che ogni volta che $T'(\mathbf{x}) = T'(\mathbf{y})$, dobbiamo avere anche che $T(\mathbf{x}) = T(\mathbf{y})$, ma non deve valere il viceversa.

Nell'esempio precedente avevamo che $\mathbf{x} = (0, 1, 0)^T$ e $\mathbf{y} = (0, 0, 1)^T$ hanno lo stesso valore $T'(\mathbf{x}) = c(1, 0)$ e $T'(\mathbf{y}) = c(1, 0)$ e stesso valore di $T(\mathbf{x}) = 1$ e $T(\mathbf{y}) = 1$, e questo è vero per ogni coppia di campioni $\mathbf{x}$ per cui vale $T'(\mathbf{x}) = T'(\mathbf{y})$, ma possiamo trovare due campione per cui $T(\mathbf{x}) = T(\mathbf{y})$ e per cui non vale $T'(\mathbf{x}) = T'(\mathbf{y})$, per esempio $\mathbf{x} = (1, 0, 0)^T$ e $\mathbf{y} = (0, 1, 0)^T$.

Minimalità II

Come per la sufficienza, abbiamo bisogno di qualcosa che ci aiuti a trovare una statistica sufficiente e minimale.

Teorema - Minimalità

Sia $f(\mathbf{x}|\theta)$ la pmf o pdf di un campione $\mathbf{X}$ e supponiamo esista una funzione $T(\mathbf{x})$ tale per cui per coppia di campioni $\mathbf{x} \in \mathcal{X}$ e $\mathbf{y} \in \mathcal{X}$ il rapporto

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = C_{x,y}$$

è costante come funzione di θ **se e solo se** $T(\mathbf{x}) = T(\mathbf{y})$, allora $T(\mathbf{X})$ è sufficiente e minimale per θ.

In altre parole

$$\left(\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = C_{x,y} \Leftrightarrow T(\mathbf{x}) = T(\mathbf{y})\right) \Rightarrow T() \text{ sufficiente e minimale}$$

Dimostrazione:

(i) Dimostriamo prima la sufficienza di $T(\mathbf{X})$, che richiede solo che

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = C_{x,y}$$

se $T(\mathbf{x}) = T(\mathbf{y})$.

Indichiamo con $A_t = \{\mathbf{x} : T(\mathbf{x}) = t\}$ la partizione indotta da $T(\mathbf{X})$. Per ogni possibile set A_t scegliamo un valore rappresentativo e indichiamolo con $\mathbf{x}_{T(\mathbf{x})}$.

Visto che $T(\mathbf{x}) = T(\mathbf{x}_{T(\mathbf{x})})$ sappiamo per ipotesi che

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{x}_{T(\mathbf{x})}|\theta)} = C_{x,x_T} = h(\mathbf{x})$$

non dipende da θ.

Possiamo scrivere

$$f(\mathbf{x}|\theta) = \frac{f(\mathbf{x}|\theta)}{f(\mathbf{x}_{T(\mathbf{s})}|\theta)} f(\mathbf{x}_{T(\mathbf{x})}|\theta) = h(\mathbf{x}) f(\mathbf{x}_{T(\mathbf{x})}|\theta)$$

Minimalità IV

Per costruzione abbiamo che $\mathbf{x}_{T(\mathbf{x})}$ è funzione solo di $T(\mathbf{x})$, i.e., per ogni valore di $T(\mathbf{x})$ abbiamo un associato valore di $\mathbf{x}_{T(\mathbf{x})}$, quindi possiamo definire

$$f(\mathbf{x}_{T(\mathbf{x})}|\theta) = g(T(\mathbf{x})|\theta)$$

de cui possiamo scrivere

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g(T(\mathbf{x})|\theta)$$

e per il teorema di fattorizzazione è sufficiente.

(ii) Dimostriamo che è anche minimale.
Prendiamo una qualsiasi altra statistica sufficiente $T'(\mathbf{x})$, e consideriamo due campioni $\mathbf{x}$ e $\mathbf{x}$ per cui vale $T'(\mathbf{x}) = T'(\mathbf{y})$. Per il teorema di fattorizzazione abbiamo che

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{g'(T'(\mathbf{x})|\theta)h'(\mathbf{x})}{g'(T'(\mathbf{y})|\theta)h'(\mathbf{y})}$$

Minimalità V

dove sia al numeratore che al denominatore abbiamo usato il teorema di fattorizzazione. Visto che $T'(\mathbf{x}) = T'(\mathbf{y})$ abbiamo che $g'(T'(\mathbf{x})|\theta) = g'(T'(\mathbf{y})|\theta)$ e quindi

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{h'(\mathbf{x})}{h'(\mathbf{y})}$$

non dipende da θ. Per ipotesi il rapporto $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$ non dipende da θ se e solo se $T(\mathbf{x}) = T(\mathbf{y})$, quindi abbiamo che ogni volta che $T'(\mathbf{x}) = T'(\mathbf{y})$ per ogni altra statistica sufficiente, dobbiamo anche avere $T(\mathbf{x}) = T(\mathbf{y})$, che è la definizione di statistica sufficiente minimale. $\square$

Questo teorema ci da un modo trovare e testare se una statistica è minimale.

Esercizio - Statistica sufficiente e minimale per una bernulliana

Trovare una statistica sufficiente e minimale per il parametro p di una bernulliana per un campione $\mathbf{X}$ di dimensione n.

Soluzione:

Ricordiamo che la congiunta per una binomiale è

$$f(\mathbf{x}|p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

Calcoliamo il rapporto

$$\frac{f(\mathbf{x}|p)}{f(\mathbf{y}|p)} = \frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i}} = p^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i}(1-p)^{n-n+\sum_{i=1}^n y_i - \sum_{i=1}^n x_i}$$

e questo rapporto non dipende da p se e solo se $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$, quindi $T(\mathbf{X}) = \sum_{i=1}^n X_i$ è sufficiente e minimale. $\square$

Notare che anche in questo caso una trasformazione biunivoca di una statistica sufficiente minimale è ancora una statistica sufficiente minimale. Per esempio nell'esempio precedente il rapporto non dipende da θ anche se $\bar{x} = \bar{y}$.

Esercizio - Statistica sufficiente e minimale per una Poisson

Trovare la statistica sufficiente e minimale per i parametri di una $N(\mu, \sigma^2)$ basata su un campione $\mathbf{X}$ di dimensione n.

Soluzione:

Avevamo visto precedentemente che

$$f(\mathbf{x}|\mu, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2}\right).$$

adesso calcoliamo il rapporto

$$\frac{f(\mathbf{x}|\mu,\sigma^2)}{f(\mathbf{y}|\mu,\sigma^2)} = \frac{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2}{2\sigma^2}\right)}{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i-\bar{y})^2 + n(\bar{y}-\mu)^2}{2\sigma^2}\right)}$$

che si può scrivere come

$$\exp\left(-\frac{\sum_{i=1}^n (x_i-\bar{x})^2 - \sum_{i=1}^n (y_i-\bar{y})^2 + n\left((\bar{x}-\mu)^2 - (\bar{y}-\mu)^2\right)}{2\sigma^2}\right)$$

o

$$\exp\left(-\frac{\sum_{i=1}^n (x_i-\bar{x})^2 - \sum_{i=1}^n (y_i-\bar{y})^2}{2\sigma^2}\right) \exp\left(-\frac{n\left((\bar{x}-\mu)^2 - (\bar{y}-\mu)^2\right)}{2\sigma^2}\right)$$

che non dipende da μ se e solo se $\exp\left(-\frac{n\left((\bar{x}-\mu)^2-(\bar{y}-\mu)^2\right)}{2\sigma^2}\right)$ non dipende da μ, che coincide con $\bar{x}=\bar{y}$, e non dipende da σ^2 se $\sum_{i=1}^n (x_i-\bar{x})^2 = \sum_{i=1}^n (y_i-\bar{y})^2$, quindi la statistica sufficiente e minimale è $T(\mathbf{X}) = (\bar{X}, \sum_{i=1}^{n}(X_i - \bar{X})^2)$ oppure, $T(\mathbf{X}) = (\bar{X}, S^2)$, cioè la media e la varianza campionaria sono minimali. □

Esercizio - Statistica sufficiente e minimale per una normal

Usare il teorema di fattorizzazione per trovare una statistica sufficiente per λ con un campione $\mathbf{X}$ di dimensione n iid da una Pois(λ). Verificare che la distribuzione condizionata di $\mathbf{X}$ dato la statistica non dipenda da λ e verificare se è minimale

Soluzione:

Iniziamo con scrivere la congiunta di $\mathbf{X}$

$$f(\mathbf{x}|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}\exp(-\lambda)}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i}\exp(-n\lambda)}{\prod_{i=1}^{n} x_i!}$$

Minimalità X

possiamo definire $h(\mathbf{x}) = \frac{1}{\prod_{i=1}^n x_i!}$ e $g(T(\mathbf{x})|\lambda) = \exp(-n\lambda)\lambda^{\sum_{i=1}^n x_i}$ che ci dice che $T(\mathbf{x}) = \sum_{i=1}^n x_i$ è sufficiente per λ.
Siccome $T(\mathbf{x})$ è somma di Poisson è ancora una Poisson con parametri $n\lambda$ e quindi, definiamo $t = T(\mathbf{x}) = \sum_{i=1}^n x_i$, abbiamo

$$f(t|\lambda) = \frac{n^t \lambda^t \exp(-n\lambda)}{t!}$$

Il rapporto tra la pmf di $\mathbf{X}$ e $T(\mathbf{X})$ è

$$\frac{f(\mathbf{x}|\lambda)}{f(t|\lambda)} = \frac{\lambda^t \exp(-n\lambda)}{\prod_{i=1}^n x_i!} \frac{t!}{n^t \lambda^t \exp(-n\lambda)} = \frac{t!}{n^t \prod_{i=1}^n x_i!}$$

che è costante rispetto a λ come ci aspettavamo.
Verifichiamo se è minimale calcolando il rapporto di verosimiglianze

$$\frac{f(\mathbf{x}|\lambda)}{f(\mathbf{y}|\lambda)} = \frac{\lambda^{\sum_{i=1}^n x_i} \exp(-n\lambda)}{\prod_{i=1}^n x_i!} \frac{\prod_{i=1}^n y_i!}{\lambda^{\sum_{i=1}^n y_i} \exp(-n\lambda)} = \frac{\lambda^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} \prod_{i=1}^n y_i!}{\prod_{i=1}^n x_i!}$$

Minimalità XI

che non dipende da λ se e solo se $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i$, quindi $T(\mathbf{X})$ è sufficiente e minimale.

$\square$