Discussione su Eventi scambiabili e Paradossi di Simpson

Fabio Spizzichino

Università "La Sapienza", Roma

Univ. di Padova, Novembre 2025

Outline

Notazioni

$$E_1, ..., E_N$$

eventi aleatori

$$X_1, ..., X_N$$

variabili binarie, indicatrici degli eventi: per $i = 1, 2, ..., N$

$$X_i := \begin{cases} 1 & \text{se } E_i \text{ vero} \\ 0 & \text{se } E_i \text{ falso} \end{cases}$$

per $n = 1, ..., N$

$$\sum_{i=1}^{n} X_i = \text{numero di successi sulle prime } n \text{ prove}$$

Per $n = 1, ..., N$ e $k = 0, ..., n$

$$\omega_k^{(n)} := \mathbb{P}(\sum^{n} X_i = k)$$

Definizione

$E_1, ..., E_N$ *scambiabili* quando la distribuzione di probabilità congiunta di $\boldsymbol{X} \equiv (X_1, ..., X_N)$

è invariante per permutazioni degli indici i :

Se $\mathbf{x}' \equiv (x_1', ..., x_n')$, $\mathbf{x}'' \equiv \left(x_1'', ..., x_n''\right) \in \{0, 1\}^n$ sono tali che

$$x_1' + ... + x_n' = x_1'' + ... + x_n'' = k$$

(presentano lo stesso numero k di successi)
allora

$$\mathbb{P}(\mathbf{X} = \mathbf{x}') = \mathbb{P}(\mathbf{X} = \mathbf{x}'')$$

$$\mathbb{P}(\mathbf{X} = \mathbf{x}') = \mathbb{P}(\mathbf{X} = \mathbf{x}'') = \frac{\omega_k^{(n)}}{\binom{n}{k}}$$

$\mathbb{P}(\mathbf{X} = \mathbf{x})$ dipende soltanto da n e da k (con $n = 2, ..., N$; $k = 0, 1, ..., n$)

Casi Notevoli

Da un'urna contenente N biglie,
di cui B *bianche* e $(N - B)$ non bianche,
si effettuano n $(n \leq N)$ estrazioni casuali
e poniamo
$E_i =$
(biglia bianca alla i-esima estrazione)

$E_1, ..., E_n$ risultano scambiabili in ciascuno dei seguenti casi

- i) B assume un valore b **noto** $(0 \leq b \leq N)$ e le estrazioni sono effettuate *con reinserimento*

  $E_1, ..., E_n$ sono stocasticamente indipendenti ed equiprobabili con $\mathbb{P}(E_i) = \frac{b}{N}$

  Le $\omega_k^{(n)}$ sono probabilità *binomiali*

- ii) Le estrazioni sono ancora effettuate *con reinserimento*
  Ma il valore di B **non è noto**
  B viene vista come una variabile aleatoria a valori in $\{0, 1, ..., N\}$
  $E_1, ..., E_n$ sono condizionatamente indipendenti ed equiprobabili, dato B
  le $\omega_k^{(n)}$ sono misture di probabilità binomiali

- iii) Le estrazioni sono effettuate *senza reinserimento* e B è uguale ad un numero b **noto**

  Per $n < N$, le $\omega_k^{(n)}$ sono probabilità *ipergeometriche*

- iv) Le estrazioni sono effettuate *senza reinserimento* e B è una variabile aleatoria

  le $\omega_k^{(n)}$ sono misture di probabilità ipergeometriche

Vi sono molti altri casi di interesse probabilistico in cui si incontrano famiglie di eventi scambiabili,
sia in questioni teoriche che applicative

I quattro precedenti casi sono quelli che suggeriscono la definizione di scambiabilità

Ma c'è un ulteriore caso notevole che ne rivela il ruolo nella statistica inferenziale, dal punto di vista di Bruno de Finetti :

N successivi lanci di una stessa moneta

Facile capire che, in tale caso, è naturale **richiedere** la condizione di scambiabilità

Implicazioni della scambiabilità I

- Per $n < N$ e $j_1 \neq j_2 \neq \dots \neq j_n$,

$$X_{j_1}, \dots, X_{j_n}$$

  sono ancora scambiabili e la loro distribuzione congiunta è indipendente dalla scelta di $j_1 \neq j_2 \neq \dots \neq j_n$

- Per $n = 2, \dots, N$, $h = 0, 1, \dots, n-1$, e $x_1, \dots, x_n$ con

$$x_1 + \dots + x_n = h$$

  si ha

$$\mathbb{P}\left(\boldsymbol{X} = \boldsymbol{x} | \sum_{i=1}^{n} X_i = h\right) = \frac{1}{\binom{n}{h}}$$

- Per $n < N$,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i = h | \sum_{i=1}^{N} X_i = k\right) = \frac{\binom{k}{h} \cdot \binom{N-k}{n-h}}{\binom{N}{n}}$$

Implicazioni della scambiabilità II

- L'intera distribuzione di probabilità congiunta di $(X_1, ..., X_N)$ è determinata dalla distribuzione di probabilità marginale di $\sum_{i=1}^{N} X_i$
- Se, per qualche k,
$$\mathbb{P}\left(\sum_{i=1}^{N} X_i = k\right) = 1$$
allora si ha una situazione isomorfa a quella del caso iii) (estrazioni *senza reinserimento* da urna con composizione nota)
- *Rappresentazione delle famiglie finite di eventi scambiabili:* **Qualunque** situazione di eventi scambiabili è isomorfa a quella del precedente caso iv)
(estrazioni *senza reinserimento* da urna con composizione aleatoria)

Implicazioni della scambiabilità sull'inferenza

- Lo stato di informazione soggettivo si esprime dichiarando la distribuzione di probabilità marginale di $\sum_{i=1}^{N} X_i$
(Tale variabile prende il ruolo di un "parametro")

- In particolare (ragionamento predittivo) rimane determinata la previsione su $(X_{n+1}, ..., X_N)$
condizionatamente all'osservazione di $(X_1, ..., X_n)$

- La previsione su $(X_{n+1}, ..., X_N)$ è riassunta da quella su $\sum_{i=n+1}^{N} X_i$
Il dato necessario da recepire da $(X_1, ..., X_n)$ è tutto contenuto nell'osservazione di $\sum_{i=1}^{n} X_i$
(Tale statistica assume il ruolo di *statistica sufficiente*)

Conclusioni circa la Statistica

Anche nel caso dei lanci di una moneta (in cui non esiste nessun'urna) il problema dell'inferenza si può impostare come nel caso iv)

Ruolo centrale della distribuzione ipergeometrica

Fatto fondamentale: il campionamento **senza** reinserimento converge al campionamento **con** reinserimento
(quando la numerosità della popolazione tende a diventare molto grande)

Ciò permette di ottenere il famoso
"Teorema di de Finetti sulla rappresentazione delle famiglie infinite di eventi scambiabili":

Gli eventi in una famiglia scambiabile infinita sono indipendenti ed equiprobabili rispetto ad una variabile aleatoria a valori nell'intervallo $[0, 1]$
Tale variabile esiste (per una specifica Legge Grandi Numeri) ed è definita come

$$\lim_{N \to \infty} \frac{S_N}{N}$$

A parte difficoltà tecniche, la stessa linea di pensiero si estende in modo diretto ai problemi della Statistica Parametrica, più in generale

Outline

Definizione

Consideriamo due tabelle 2×2, dove gli elementi sono numeri naturali:

$$\mathcal{T}' \equiv \begin{pmatrix} n' & N' - n' \\ m' & M' - m' \end{pmatrix}, \mathcal{T}'' \equiv \begin{pmatrix} n'' & N'' - n'' \\ m'' & M'' - m'' \end{pmatrix}.$$

Consideriamo quindi la tabella 2×2 costruita, casella per casella, aggregando le caselle omologhe di $\mathcal{T}'$ e di $\mathcal{T}''$:

$$\begin{pmatrix} n' + n'' & (N' + N'') - (n' + n'') \\ m' + m'' & (M' + M'') - (m' + m'') \end{pmatrix}.$$

Diciamo che le due tabelle $\mathcal{T}'$ e $\mathcal{T}''$ *realizzano il paradosso di Simpson* quando si verificano, contemporaneamente, le seguenti diseguaglianze:

$$\frac{n'}{N'} > \frac{m'}{M'}, \frac{n''}{N''} > \frac{m''}{M''} \tag{1}$$

$$\frac{n' + n''}{N' + N''} < \frac{m' + m''}{M' + M''}.$$

Esempio numerico

Come riportato in molti testi divulgativi e siti on line, un semplicissimo esempio di tale situazione può essere il seguente

$$\mathcal{T}' \equiv \begin{pmatrix} 50 & 60 \\ 30 & 40 \end{pmatrix}, \mathcal{T}'' \equiv \begin{pmatrix} 60 & 30 \\ 90 & 50 \end{pmatrix}.$$

dove cioè

$$\frac{n'}{N'} = \frac{5}{11} > \frac{3}{7} = \frac{m'}{M'},$$

$$\frac{n''}{N''} = \frac{2}{3} > \frac{9}{14} = \frac{m''}{M''},$$

$$\frac{n' + n''}{N' + N''} = \frac{11}{20} < \frac{4}{7} = \frac{m' + m''}{M' + M''}.$$

Da un punto di vista puramente aritmetico non c'è molto altro da aggiungere!

Esempio statistico

Caso della sperimentazione relativa ad un nuovo farmaco

Il farmaco viene somministrato ad un gruppo di $N' + N'' = 200$ pazienti

Ad altri $M' + M'' = 210$ pazienti viene somministrato un placebo

I pazienti si suddividono inoltre in due diverse categorie:
$\mathcal{H}' \equiv${pazienti ospedalizzati, in condizioni di salute non ottimali}

e

$\mathcal{H}'' \equiv${pazienti sani, lin condizioni generali normali}.

Per i pazienti di qualunque dei due diversi tipi viene rilevato se, in un tempo debito, vi sia stata o meno una remissione del disturbo in esame.

Per il campione complessivo dei pazienti coinvolti, si ha una suddivisione in $2^3 = 8$ diverse classi

Le relative numerosità vengono riportate nelle Tabelle T' e T''
in T': dati relativi ai pazienti appartenenti a $\mathcal{H}'$,
in T'' dati relativi ai pazienti appartenenti a $\mathcal{H}''$

T'' mostra:
fra i pazienti in $\mathcal{H}''$, il farmaco è stato somministrato a 90 soggetti con 60 guariti
il placebo è stato somministrato a 140 soggetti, con 90 guariti spontaneamente .

Analogamente,

T' riporta i dati relativi ai pazienti in $\mathcal{H}'$.

Commento

Il dato

$$\frac{n'}{N'} > \frac{m'}{M'}, \frac{n''}{N''} > \frac{m''}{M''}$$

mostra che, in proporzione, il farmaco produce un effetto benefico, se confrontato con il placebo, per entrambe le categorie $\mathcal{H}'$ e $\mathcal{H}''$.
Ma, in base alla disuguaglianza

$$\frac{n' + n''}{N' + N''} < \frac{m' + m''}{M' + M''},$$

emergerebbe invece un dato più favorevole al placebo che non al farmaco, se i dati relativi ai due sottogruppi venissero aggregati.

Spiegazione aritmetica

L' apparente discrepanza non dovrebbe sorprendenci

Facendo comparire le coppie di rapporti

$$\left(\frac{n'}{N'}, \frac{n''}{N''}\right), \left(\frac{m'}{M'}, \frac{m''}{M''}\right),$$

i termini

$$\frac{n'+n''}{N'+N''}$$

e

$$\frac{m'+m''}{M'+M''}$$

possono essere decomposti come segue:

$$\frac{n'+n''}{N'+N''} = \frac{n'}{N'} \cdot \frac{N'}{N'+N''} + \frac{n''}{N''} \cdot \frac{N''}{N'+N''}, \tag{2}$$

$$\frac{m'+m''}{M'+M''} = \frac{m'}{M'} \cdot \frac{M'}{M'+M''} + \frac{m''}{M''} \cdot \frac{M''}{M'+M''}. \tag{3}$$

Si è assunto

$$\frac{n'}{N'} > \frac{m'}{M'}; \frac{n''}{N''} > \frac{m''}{M''}$$

Se risultasse

$$\frac{N'}{N' + N''} = \frac{M'}{M' + M''}$$

sarebbe banalmente valida anche la disuguaglianza

$$\frac{n' + n''}{N' + N''} > \frac{m' + m''}{M' + M''}$$

La disuguaglianza qui viene invertita in quanto

$$\frac{n' + n''}{N' + N''} = \frac{11}{20} < \frac{4}{7} = \frac{m' + m''}{M' + M''}$$

Formula delle probabilità totali

$\mathcal{H}$ *partizione* dell'*evento certo* negli eventi $H_1, ..., H_n$

$H_1, ..., H_n$ *esaustivi* e *a due a due incompatibili*:

E' certo che se ne verifica uno ed uno soltanto
Per E evento generico si ottiene la Formula delle Probabilità Totali:

$$\mathbb{P}(E) = \sum_{j=1}^{n} \mathbb{P}(E|H_j)\mathbb{P}(H_j). \tag{4}$$

Nel caso particolare in cui $n = 2$:

$$\mathbb{P}(E) = \mathbb{P}(E|H)\mathbb{P}(H) + \mathbb{P}(E|\widetilde{H})\mathbb{P}(\widetilde{H}) = \mathbb{P}(E|H)\mathbb{P}(H) + \mathbb{P}(E|\widetilde{H})\left(1 - \mathbb{P}(H)\right). \tag{5}$$

Formula di decomposizione per probabilità subordinate

Consideriamo ancora:

E evento generico e partizione $\{H_1, ..., H_n\}$ (con $\mathbb{P}(H_j) > 0$, per ogni)

e anche

D ulteriore evento generico con $\mathbb{P}(D) > 0$

Volendo decomporre $\mathbb{P}(E|D)$, otteniamo (vedere conteggio, successivamente)

$$\mathbb{P}(E|D) := \frac{\mathbb{P}(E \cap D)}{\mathbb{P}(D)} =$$

$$= \sum_{j=1}^{n} \mathbb{P}(E|D \cap H_j)\mathbb{P}(H_j|D). \tag{6}$$

Notare:
Se l'evento D è indipendente da ciascun H_j, la formula (8) si riduce a

$$\mathbb{P}(E|D) = \sum_{j=1}^{n} \mathbb{P}(E|D \cap H_j)\mathbb{P}(H_j)$$

Nel caso $n = 2$, (8) si riduce a

$$\mathbb{P}(E|D) = \mathbb{P}(E|D \cap H)\mathbb{P}(H|D) + \mathbb{P}(E|D \cap \widetilde{H})\,(1 - \mathbb{P}(H|D))\,. \tag{7}$$

In particolare, nel caso di indipendenza stocastica fra D ed H:

$$\mathbb{P}(E|D) = \mathbb{P}(E|D \cap H)\mathbb{P}(H) + \mathbb{P}(E|D \cap \widetilde{H})\mathbb{P}(\widetilde{H})$$

Per ottenere (8):

$$\mathbb{P}(E|D) = \frac{\mathbb{P}(E \cap D)}{\mathbb{P}(D)} = \frac{\sum_{j=1}^{n} \mathbb{P}(E \cap D \cap H_j)}{\mathbb{P}(D)} =$$

$$= \sum_{j=1}^{n} \frac{\mathbb{P}(E|D \cap H_j)\mathbb{P}(D \cap H_j)}{\mathbb{P}(D)} = \sum_{j=1}^{n} \mathbb{P}(E|D \cap H_j)\mathbb{P}(H_j|D). \tag{8}$$

Spiegazione probabilistica

Torniamo al caso $n = 2$

Fissiamo una volta per tutte un evento E

Dato un secondo evento D, confrontiamo

$\mathbb{P}(E|D)$ con $\mathbb{P}(E|\widetilde{D})$

supponendo di aver assegnato le probabilità condizionate

$$\mathbb{P}(E|D \cap H), \mathbb{P}(E|D \cap \widetilde{H}), \mathbb{P}(E|\widetilde{D} \cap H), \mathbb{P}(E|\widetilde{D} \cap \widetilde{H}), \mathbb{P}(H|D), \mathbb{P}(E|\widetilde{D}),$$

relativamente ad un terzo evento H

Scriviamo

$$\mathbb{P}(E|D) = \mathbb{P}(E|D \cap H)\mathbb{P}(H|D) + \mathbb{P}(E|D \cap \widetilde{H})\mathbb{P}(\widetilde{H}|D),$$

E confrontiamolo con

$$\mathbb{P}(E|\widetilde{D}) = \mathbb{P}(E|\widetilde{D} \cap H)\mathbb{P}(H|\widetilde{D}) + \mathbb{P}(E|\widetilde{D} \cap \widetilde{H})\mathbb{P}(\widetilde{H}|\widetilde{D}).$$

Supponiamo

$$\mathbb{P}(E|D \cap H) > \mathbb{P}(E|\widetilde{D} \cap H)$$

e

$$\mathbb{P}(E|D \cap \widetilde{H}) > \mathbb{P}(E|\widetilde{D} \cap \widetilde{H})$$

?? Che cosa possiamo concludere circa il confronto fra

$$\mathbb{P}(E|D) \quad \text{e} \quad \mathbb{P}(E|\widetilde{D}) \text{ ??}$$

Pensare ai diversi casi di *indipendenza*, *correlazione positiva*, *correlazione negativa*

fra H e D !

Riferimenti Bibliografici circa il Paradosso di Simpson

[1] A. F. Beardon. Unravelling Simpson's paradox. *Math. Gaz.* 102, no. 555, 534-535, 2018.

[2] A. Chambaz, I. Drouet, S. Memetea. Simpson's paradox, a tale of causality. *J. SFdS* 161, no. 1, 42-66, 2020.

[3] N. Falletta. *Il libro dei paradossi*. Ed. TEA, Milano, 1994.

[4] J. Gou, F. Zhang. Experience Simpson's paradox in the classroom. *Amer. Statist*. 71, no. 1, 61-66, 2017.

[5] D. G. Saari. Mathematics motivated by the social and behavioral sciences. CBMS-NSF Regional Conference Series in Applied Mathematics, 91. SIAM, Philadelphia, PA, 2018.

[6] M. L. Samuels. Simpson's paradox and related phenomena. *J. Amer. Statist. Assoc.*, 88, 81-88, 1993.

[7] M. Scarsini, F. Spizzichino. Simpson-type paradoxes, Dependence, and Ageing. *J. Appl. Prob.* 36, 119-131,1999.

[8] A. Selvitella. The Simpson's paradox in quantum mechanics. *J. Math. Phys.* 58, no. 3, 032101, 37, 2017.

[9] R. Steyer, W. Nagel. *Probability and conditional expectation. Fundamentals for the empirical sciences*. Wiley Series in Probability and Statistics. J. Wiley & Sons, Ltd., Chichester, 2017.