Statistica Bayesiana

Christian Mancini

7 marzo 2023

Indice

Capitolo 1

Concetti preliminari

> Sotto il teorema di Bayes, nessuna teoria è perfetta. Piuttosto, è un lavoro in corso, sempre soggetto a ulteriori perfezionamenti e prove.
>
> *Nate Silver*

La statistica bayesiana è un sottocampo della statistica in cui la verità di un enunciato è espressa in termini di credibilità (*belief*) o più specificatamente di **probabilità bayesiana**.

Definizione 1.1 (Probabilità bayesiana)**.** La probabilità bayesiana è un'interpretazione del concetto di probabilità, in cui, anziché la frequenza di un qualche evento, la probabilità viene vista come aspettativa razionale che rappresenta un stato di conoscenza o come quantificazione di una convinzione personale.

Nella visione bayesiana assegniamo una probabilità ad un'ipotesi, che è la probabilità che l'ipotesi sia vera. Nell' approccio frequentista invece, l'ipotesi viene solo verificata senza che le venga assegnata una probabilità. Per valutare la probabilità di un'ipotesi si deve fornire una **probabilità iniziale**(*probabilità a priori*). Questa a sua volta viene aggiornata con l'arrivo di nuovi dati, che vengono utilizzati per calcolare la **probabilità aggiornata**(*probabilità a posteriori*), utilizzando la **regola di Bayes**.

Definizione 1.2 (Inferenza bayesiana)**.** L'inferenza bayesiana è un approccio alla statistica che utilizza la regola di Bayes per aggiornare la probabilità iniziale con l'arrivo di nuovi dati.

Più in generale, possiamo derivare dei metodi di analisi dei dati detti **metodi bayesiana**, che forniscono:

- stima dei parametri con buone proprietà statistiche;
- descrizione parsimoniosa dei dati osservati;
- previsione di dati mancanti e di dati futuri;
- una struttura computazionale (*computational framework*) per la stima, selezione e validazione per i modelli.

1.1 Apprendimento bayesiano

L'induzione statistica è un processo che consente di apprendere le caratteristiche generali di una popolazione a partire da un sottoinsieme di essa, ossia da un campione. I valori numerici delle caratteristiche di una popolazione sono espressi in termini di parametri θ, e le descrizioni numeriche del sottoinsieme forma il dataset y. Prima che venga ottenuto il dataset, i valori numerici sia delle caratteristiche della popolazione che del dataset sono incerti. Dopo che un dataset y è stato ottenuto, L'informazione contenuta nel dataset può essere utilizzata per ridurre la nostra incertezza sulle caratteristiche della popolazione, cioè su θ. Quantificare questo cambiamento di incertezza è il compito dell'inferenza bayesiana. I nostri spazi sono:

- $\mathcal{Y}$: spazio del campione;
- Θ: spazio dei parametri;

Lo spazio del campione è l'insieme di tutti i possibili dataset a cui appartiene un singolo dataset y. Lo spazio dei parametri è l'insieme di tutti i possibili valori dei parametri θ, da cui speriamo di identificare il valore che meglio descrive la popolazione.
L'apprendimento bayesiano inizia con una formulazione numerica della probabilità congiunta di y e θ, espressi in termini di distribuzione di probabilità su $\mathcal{Y}$ e Θ. In soldoni, la probabilità congiunta è la probabilità che y e θ si verifichino contemporaneamente. A questo punto inizia il processo ricorsivo:

1. Per ogni valore numerico $\theta \in \Theta$, la nostra **distribuzione iniziale** $p(\theta)$ descrive la nostra convinzione iniziale che θ rappresenti le vere caratteristiche della popolazione.
2. Per ogni $\theta \in \Theta$ e $y \in \mathcal{Y}$, il nostro **modello campionario** $p(y|\theta)$ descrive la nostra convinzione che y sia il risultato del nostro studio sapendo che θ è vero.
3. Per ogni valore numerico $\theta \in \Theta$, la nostra **distribuzione a posteriori** $p(\theta|y)$ descrive la nostra convinzione che θ rappresenti le vere caratteristiche della popolazione sapendo avendo osservato y.

La distribuzione a posteriore è ottenuta dalla distribuzione iniziale e dal modello campionario, tramite la **regola di Bayes** (che vedremo successivamente).

Nota. È importante notare che la regola di Bayes non indica come dovrebbero essere le nostre convinzioni, ma come dovrebbero cambiare con l'arrivo di nuove informazioni.

Capitolo 2

Credibilità e eventi

2.1 Funzioni belief e probabilità

La probabilità è un modo per esprimere numericamente un'aspettativa razionale. Mostriamo adesso che molte proprietà che hanno le nostre aspettative numeriche sono anche valide per le probabilità.

2.1.1 Funzione belief (belief function)

Definizione 2.1 (Funzione belief)**.** La funzione belief è una funzione che associa ad ogni possibile evento o enunciato un valore reale compreso tra 0 e 1.

Siano F, G e H tre possibili enunciati che possono sovrapporsi. Sia $Be()$ una funzione belief che assegna un numero agli enunciati tale che maggiore è il numero assegnato, maggiore è la credibilità dell'enunciato (*degree of belief*). Ad esempio, se $Be(F) > Be(G)$ allora F è più credibile di G.
Vogliamo che la funzione descriva le nostre credenze sotto certe condizioni:

- $Be(F|H) > Be(G|H)$ significa che dato H vero, allora è più credibile F che G.
- $Be(F|G) > Be(F|H)$ significa che se siamo costretti a scommettere su F, lo facciamo più volentieri sotto la condizione che G sia vero rispetto alla condizione che H sia vero.

2.1.2 Assiomi di credibilità

Ogni funzione che rappresenta le nostre credenze deve avere le seguenti proprietà:

$$Be(\neg H|H) \leq Be(F|H) \leq Be(H|H) \tag{2.1}$$

$$Be(F \cup G|H) \geq \max\{Be(F|H), Be(G|H)\} \tag{2.2}$$

$$Be(F \cap G|H) \textit{ può essere derivata da } Be(G|H) \textit{ e } Be(F|G \cap H) \tag{2.3}$$

Adesso vediamo cosa significano le proprietà sopra elencate.

- **Proprietà 2.1**: Il numero che assegno a $Be(F|H)$ è compreso tra i numeri che assegniamo all'assenza di credenza e alla piena credenza.
- **Proprietà 2.2**: La credibilità di F o G dato H è almeno uguale alla credibilità di F o G.
- **Proprietà 2.3**: Possiamo decidere se F e G sono entrambi credibili dato H, prima decidendo se G è credibile dato H e successivamente decidendo se F è credibile dato G e H.

Possiamo confrontare gli assiomi di credibilità con gli assiomi di probabilità:

$$0 = P(\neg H|H) \leq P(F|H) \leq P(H|H) = 1 \tag{2.4}$$

$$P(F \cup G|H) = P(F|H) + P(G|H) \textit{ se } F \cap G = \emptyset \tag{2.5}$$

$$P(F \cap G|H) = P(G|H)P(F|G \cap H) \tag{2.6}$$

Notiamo che una funzione che soddisfa gli assiomi di credibilità soddisfa anche gli assiomi di probabilità. Quindi se usiamo una funzione di probabilità per descrivere le nostre credenze, allora la funzione soddisfa gli assiomi di credibilità.

2.2 Partizioni e regola di Bayes

Definizione 2.2 (Partizione di un insieme)**.** Una partizione dell'insieme H è una famiglia di sottoinsiemi $\{H_1, H_2, \cdots H_k\}$ che soddisfa le seguenti proprietà:

1. $$H_i \cap H_j = \emptyset \; \forall i \neq j \tag{2.7}$$
2. $$\bigcup_{i=1}^{k} H_i = H \tag{2.8}$$

La famiglia $\{H_1, H_2, \cdots H_k\}$ è detta **disgiunta** se vale la proprietà 2.7, mentre è detta **completa** se vale la proprietà 2.8. Possiamo dire che $\{H_1, H_2, \cdots H_k\}$ è una **partizione di H** se è disgiunta e completa.
Prima di introdurre la regola di Bayes, vediamo due proprietà che derivano direttamente dagli assiomi di credibilità visti nella sezione 2.1.2.

Lemma 2.1. Sia $\{H_1, H_2, \cdots H_k\}$ un partizione di H, $P(H) = 1$ e sia E un evento specifico. Allora gli assiomi di probabilità ci dicono che:

$$\sum_{i=1}^{k} P(H_i) = 1. \tag{2.9}$$

$$P(E) = \sum_{i=1}^{k} P(E \cap H_i) = \sum_{i=1}^{k} P(E|H_i)P(H_i). \tag{2.10}$$

Dove 2.9 è detta **Regola di probabilità totale** e 2.10 è detta **Regola di probabilità marginale**.

Teorema 2.2 (Teorema di Bayes)**.** Sia $\{H_1, H_2, \cdots H_k\}$ un partizione di H e sia $E \subset H \mid P(E) \neq 0$. Il teorema afferma che:

$$P(H_j|E) = \frac{P(E|H_j)P(H_j)}{P(E)} \overset{2.10}{=} \frac{P(E|H_j)P(H_j)}{\sum_{i=1}^{k} P(E|H_i)P(H_i)}. \tag{2.11}$$

L'evento E implica che si è avverata almeno un'alternativa e quindi $P(E) \neq 0$. Cosa importante, altrimenti il teorema non avrebbe senso, perché non possiamo dividere per zero.

Dimostrazione. Il teorema deriva dalla definizione di probabilità condizionata e dalla regola di probabilità totale. La probabilità di un evento A dato un evento B è definita come:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.12}$$

Analogamente, la probabilità di un evento B dato un evento A è definita come:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \tag{2.13}$$

Usando la regola di probabilità marginale 2.10 possiamo scrivere la probabilità congiunta come:

$$P(A \cap B) = P(B|A)P(A). \tag{2.14}$$

Sostituendo la 2.14 in 2.12 troviamo il teorema di bayes:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}. \tag{2.15}$$

□