

CALCOLO DELLE PROBABILITÀ E STATISTICA 2021/2022

STATISTICA INFERENZIALE
STIMA PUNTUALE



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

1 Modello statistico di un esperimento aleatorio

Consideriamo un esperimento aleatorio descritto da uno spazio campionario Ω , che può essere un insieme finito o infinito (numerabile o più che numerabile). Accade spesso che la probabilità \mathbb{P} che descrive l'esperimento aleatorio non sia conosciuta, come nei seguenti esempi.

Esempio 1.1. Viene lanciato un dado, perciò $\Omega = \{1, 2, 3, 4, 5, 6\}$, ma si ignora se il dado sia equilibrato oppure no, cioè se $\mathbb{P}(\{\omega\}) = 1/6$ per ogni $\omega \in \Omega$ oppure no. Quindi possiamo affermare solamente che \mathbb{P} sia data da

$$\mathbb{P}(\{1\}) = p_1, \quad \mathbb{P}(\{2\}) = p_2, \quad \mathbb{P}(\{3\}) = p_3, \quad \mathbb{P}(\{4\}) = p_4, \quad \mathbb{P}(\{5\}) = p_5, \quad \mathbb{P}(\{6\}) = p_6,$$

con

$$0 \leq p_i \leq 1, \quad \sum_{i=1}^6 p_i = 1.$$

Esempio 1.2. Vengono effettuati tre lanci di una moneta, quindi

$$\Omega = \{TTT, TTC, TCT, CTT, TCC, CTC, CCT, CCC\}.$$

Sia p la probabilità che esca testa. Ipotizziamo che tale probabilità sia la stessa in ogni lancio e che i lanci siano indipendenti. Segue allora che la probabilità \mathbb{P} è data da

$$\begin{aligned} \mathbb{P}(\{TTT\}) &= p^3, & \mathbb{P}(\{TTC\}) &= p^2(1-p), & \mathbb{P}(\{TCT\}) &= p(1-p)^2, \\ \mathbb{P}(\{CTT\}) &= p^2(1-p), & \mathbb{P}(\{TCC\}) &= p(1-p)^2, & \mathbb{P}(\{CTC\}) &= p(1-p)^2, \\ & & \mathbb{P}(\{CCT\}) &= p(1-p)^2, & \mathbb{P}(\{CCC\}) &= (1-p)^3. \end{aligned}$$

In entrambi gli esempi notiamo che \mathbb{P} è descritta da un numero finito di parametri (ciò accade sempre nel caso in cui Ω sia un insieme *finito*).

- Nell'Esempio 1.1 la probabilità \mathbb{P} è parametrizzata dal vettore

$$\theta = (p_1, p_2, p_3, p_4, p_5, p_6),$$

appartenente all'insieme $\Theta = \{\theta \in \mathbb{R}^6: 0 \leq \theta_i \leq 1, \sum_{i=1}^6 \theta_i = 1\}$. Notiamo che ad ogni $\theta \in \Theta$ corrisponde una e una sola probabilità su Ω che indicheremo con \mathbb{P}_θ .

- Nell'Esempio 1.2 la probabilità \mathbb{P} è invece completamente descritta dal parametro reale $\theta = p$, appartenente all'intervallo $\Theta = [0, 1]$. Anche in questo caso, indicheremo con \mathbb{P}_θ la corrispondente probabilità su Ω .

Generalizzando quanto visto nei due esempi precedenti, possiamo dare la seguente definizione.

Definizione 1.1. Si chiama **modello statistico (dell'esperimento aleatorio)** una coppia $(\Omega, (\mathbb{P}_\theta)_{\theta \in \Theta})$, dove Θ è un insieme arbitrario, detto **insieme dei parametri** e ad ogni $\theta \in \Theta$ è assegnata una probabilità \mathbb{P}_θ su Ω .

2 Stimatori

Come abbiamo visto, un modello statistico è una famiglia di modelli probabilistici per un esperimento aleatorio, in cui si assume che *esista un valore di θ che fornisce il modello corretto*. Un problema fondamentale di Statistica inferenziale è dunque il **problema della stima**, che consiste nell'approssimare il valore di θ a partire dai dati. Per ottenere dei dati con cui stimare θ si considera una variabile aleatoria legata all'esperimento in esame.

- Nell'Esempio 1.1 si potrebbe considerare la variabile aleatoria

$X =$ “numero riportato sulla faccia superiore del dado”.

- Nell'Esempio 1.2 si potrebbe considerare la variabile aleatoria

$X =$ “numero di volte in cui testa è uscita nei tre lanci”.

La variabile aleatoria X non è solo utile per stimare θ , ma spesso è la variabile aleatoria X stessa a cui noi siamo interessati. La distribuzione di probabilità di tale variabile aleatoria generalmente dipenderà ancora dal parametro θ , è dunque possibile ottenere una stima di θ a partire dall'osservazione di X . A tale scopo, immaginiamo di poter ripetere diverse volte l'esperimento aleatorio. Gli ipotetici valori assunti dalla variabile aleatoria di interesse nei vari esperimenti sono dunque descritti da una campione di variabili¹ aleatorie i.i.d. (*indipendenti e identicamente distribuite*):

$$X_1 \quad X_2 \quad X_3 \quad \cdots \quad X_n.$$

Definizione 2.1. Si chiama **campione casuale (di dimensione o numerosità n)** un insieme di n variabili aleatorie X_1, \dots, X_n i.i.d. (*indipendenti e identicamente distribuite*).

Per ottenere una stima di θ a partire da un campione casuale X_1, \dots, X_n si eseguono opportune trasformazioni. L'oggetto matematico per descriverle è chiamato **statistica campionaria** o **stimatore del parametro θ** .

¹Ricordiamo che la lettera maiuscola sta ad indicare che gli esperimenti devono ancora essere svolti e le quantità sono quindi aleatorie. Solo dopo aver eseguito gli esperimenti conosceremo i valori da esse assunti, che saranno indicati con le lettere minuscole $x_1, x_2, x_3, \dots, x_n, \dots$

Definizione 2.2. Si chiama *statistica (campionaria) o stimatore (del parametro θ)* una qualunque funzione del campione casuale X_1, \dots, X_n :

$$T_n = t_n(X_1, \dots, X_n),$$

con $t_n: \mathbb{R}^n \rightarrow \Theta$.

NOTAZIONE. Nel seguito considereremo variabili aleatorie, discrete o continue, aventi distribuzione dipendente da un parametro θ incognito. Per tale ragione, è utile fissare le seguenti notazioni.

- Nel caso in cui X sia una v.a. discreta denoteremo con $p(x|\theta)$ (o anche $p(x;\theta)$) la densità discreta di X dipendente dal parametro θ .
- Nel caso in cui X sia una v.a. continua denoteremo con $f(x|\theta)$ (o anche $f(x;\theta)$) la densità di X dipendente dal parametro θ .

3 Come trovare uno stimatore

Ci sono casi in cui è semplice trovare lo stimatore di un parametro. Ad esempio, se il parametro è la media, è naturale considerare come stimatore la media campionaria. In casi più complicati, è necessario avere una metodologia per derivare uno stimatore. Vediamo le due metodologie più importanti.

3.1 Stimatori di massima verosimiglianza

Consideriamo un modello statistico $(\Omega, (\mathbb{P}_\theta)_{\theta \in \Theta})$ e sia X_1, \dots, X_n un campione casuale. Supponiamo che le X_i siano variabili aleatorie discrete, con densità discreta $p(x|\theta)$. Consideriamo n numeri reali x_1, \dots, x_n , che interpretiamo come i valori osservati delle variabili X_1, \dots, X_n . È ragionevole pensare che il valore θ che meglio si accorda ai valori osservati sia quel valore che massimizza la probabilità

$$\mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = p(x_1|\theta) \cdots p(x_n|\theta),$$

dove il fatto che tale probabilità sia il prodotto di $p(x_1|\theta) \cdots p(x_n|\theta)$ è una conseguenza dell'indipendenza di X_1, \dots, X_n . Tale quantità è dunque particolarmente rilevante e prende il nome di *verosimiglianza*.

Definizione 3.1. Siano $(\Omega, (\mathbb{P}_\theta)_{\theta \in \Theta})$ un modello statistico e X_1, \dots, X_n un campione casuale.

- Se X_i è una v. a. discreta, allora, per ogni $\theta \in \Theta$, si chiama **verosimiglianza**^a la funzione $L_\theta: \mathbb{R}^n \rightarrow [0, +\infty)$ data da

$$L_\theta(x_1, \dots, x_n) = p(x_1|\theta) \cdots p(x_n|\theta).$$

- Se X_i è una v. a. continua, allora, per ogni $\theta \in \Theta$, si chiama **verosimiglianza** la funzione $L_\theta: \mathbb{R}^n \rightarrow [0, +\infty)$ data da

$$L_\theta(x_1, \dots, x_n) = f(x_1|\theta) \cdots f(x_n|\theta).$$

^aIn inglese *Likelihood*.

Definizione 3.2. Uno **stimatore di massima verosimiglianza** o **MLE**^a del parametro θ è dato da $T_n = t_n(X_1, \dots, X_n)$ con $t_n: \mathbb{R}^n \rightarrow \Theta$ tale che

$$L_{t_n(x_1, \dots, x_n)}(x_1, \dots, x_n) \geq L_\theta(x_1, \dots, x_n), \quad \forall \theta \in \Theta.$$

^a*Maximum Likelihood Estimator*.

OSSERVAZIONE. Fissati dunque x_1, \dots, x_n ad arbitrio, un punto di massimo della funzione $\theta \mapsto L_\theta(x_1, \dots, x_n)$ è dato da $t_n(x_1, \dots, x_n)$.

3.2 Approccio bayesiano alla Statistica inferenziale

La formula di Bayes è alla base del cosiddetto *approccio bayesiano* alla Statistica inferenziale, che si differenzia dall'*approccio frequentista* visto fino a ora.

Nell'*approccio frequentista* si assume che ciò che si osserva sia un campione casuale X_1, \dots, X_n la cui distribuzione di probabilità dipende da un parametro $\theta \in \Theta$. Si assume, inoltre, che il parametro θ sia un valore incognito ma fissato, il cui valore “vero” va ricercato all’interno dell’insieme Θ facendo uso dei dati a disposizione.

Nell'*approccio bayesiano* invece, il parametro incognito, indicato con ϑ , come prima caratterizza la distribuzione di probabilità di X_1, \dots, X_n , ma in tal caso è considerato una *variabile aleatoria* con distribuzione di probabilità definita sull’insieme Θ . Per fissare le idee, supponiamo che Θ sia un insieme *finito*, quindi ϑ è una variabile aleatoria *discreta*. Indichiamo con p_ϑ la *funzione di distribuzione di probabilità* di ϑ . Questa viene detta **distribuzione a priori** e descrive le aspettative dello statistico circa i possibili valori che il parametro può assumere.

La distribuzione a priori dovrebbe rispecchiare tutte le informazioni sul parametro che sono a disposizione dello statistico prima di effettuare le osservazioni campionarie. Nell’approccio bayesiano si cerca quindi, tramite la formula di Bayes, di aggiornare le informazioni a priori sul parametro sulla base delle informazioni contenute nel campione estratto. In tal modo si giunge alla definizione della **distribuzione a posteriori**, il cui studio è alla base dei metodi inferenziali bayesiani.

La scelta della distribuzione a priori è ovviamente il punto più delicato di tutto il procedimento. Spesso² la scelta ricade su una distribuzione a priori di tipo uniforme, che rispecchia il fatto che non si hanno a disposizione informazioni sufficienti per privilegiare determinati valori del parametro rispetto ad altri. Tuttavia altre distribuzioni possono essere utilizzate e la scelta effettuata può influenzare in maniera importante il risultato dell'analisi inferenziale.

Esempio 3.1. Consideriamo un esperimento aleatorio con X v.a. discreta con densità discreta dipendente da un parametro reale θ :

$$p_X(\cdot|\theta) = \text{densità discreta di } X.$$

Sia $\Theta \subset \mathbb{R}$ un insieme finito dato da

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}.$$

Il parametro incognito, indicato con ϑ , è una variabile aleatoria discreta $\vartheta: \Omega \rightarrow \Theta$. Sia p_ϑ la densità discreta di ϑ a priori. Sulla base di un campione di dati osservati di dimensione n , x_1, \dots, x_n , si vuole aggiornare tale densità discreta. Per farlo, utilizziamo la formula di Bayes che ci permette di ricavare la densità discreta di ϑ a posteriori:

$$p_\vartheta(\theta_i|x_1, \dots, x_n) = \frac{p_X(x_1|\theta_i) p_X(x_2|\theta_i) \cdots p_X(x_n|\theta_i) p_\vartheta(\theta_i)}{\sum_{j=1}^m p_X(x_1|\theta_j) p_X(x_2|\theta_j) \cdots p_X(x_n|\theta_j) p_\vartheta(\theta_j)},$$

per ogni $i = 1, \dots, m$.

La distribuzione a posteriori è una distribuzione condizionata rispetto al campione. L'espressione della distribuzione a posteriori in termini della distribuzione a priori e della distribuzione del campione è dunque data dalla formula di Bayes, come abbiamo visto nell'esempio precedente. Tale espressione cambia a seconda che X oppure ϑ siano variabili aleatorie discrete oppure continue.

a) X discreta e ϑ discreto (come nell'Esempio 3.1):

$$p_\vartheta(\theta_i|x_1, \dots, x_n) = \frac{p_X(x_1|\theta_i) \cdots p_X(x_n|\theta_i) p_\vartheta(\theta_i)}{\sum_j p_X(x_1|\theta_j) \cdots p_X(x_n|\theta_j) p_\vartheta(\theta_j)},$$

per ogni i .

b) X continua e ϑ discreto:

$$p_\vartheta(\theta_i|x_1, \dots, x_n) = \frac{f_X(x_1|\theta_i) \cdots f_X(x_n|\theta_i) p_\vartheta(\theta_i)}{\sum_j f_X(x_1|\theta_j) \cdots f_X(x_n|\theta_j) p_\vartheta(\theta_j)},$$

per ogni i .

²L'approccio bayesiano con distribuzione a priori uniforme è stato predominante fino all'affermazione dell'approccio frequentista (intorno agli anni Trenta del Novecento). Successivamente agli anni Sessanta del Novecento ha ripreso vigore l'approccio bayesiano, grazie anche agli strumenti informatici che hanno facilitato i calcoli della distribuzione del parametro tramite la formula di Bayes.

c) X discreta e ϑ continuo:

$$f_{\vartheta}(\theta|x_1, \dots, x_n) = \frac{p_X(x_1|\theta) \cdots p_X(x_n|\theta) f_{\vartheta}(\theta)}{\int_{\Theta} p_X(x_1|t) \cdots p_X(x_n|t) f_{\vartheta}(t) dt}, \quad (3.1)$$

per ogni $\theta \in \Theta$.

d) X continua e ϑ continuo:

$$f_{\vartheta}(\theta|x_1, \dots, x_n) = \frac{f_X(x_1|\theta) \cdots f_X(x_n|\theta) f_{\vartheta}(\theta)}{\int_{\Theta} f_X(x_1|t) \cdots f_X(x_n|t) f_{\vartheta}(t) dt},$$

per ogni $\theta \in \Theta$.

3.2.1 Stimatori bayesiani

Nell'approccio bayesiano si utilizza la distribuzione a posteriori per ottenere degli stimatori puntuali del parametro, ad esempio considerando il valore atteso della distribuzione a posteriori.

Esempio 3.2. Consideriamo un campione casuale X_1, \dots, X_n , con $X_i \sim B(\theta)$, per $0 \leq \theta \leq 1$. Supponiamo che la distribuzione a priori del parametro ϑ sia $Beta(a, b)$, con $a, b > 0$, ossia ϑ è una variabile aleatoria continua con densità

$$f_{\vartheta}(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} 1_{(0,1)}(\theta), \quad \forall \theta \in \mathbb{R},$$

dove $\Gamma: (0, +\infty) \rightarrow \mathbb{R}$ è la funzione Gamma di Eulero definita da

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx, \quad \forall a > 0.$$

Determinare lo stimatore bayesiano dato dal valore atteso della distribuzione a posteriori.

OSSERVAZIONE. Se $a = b = 1$, la distribuzione $Beta(1,1)$ coincide con la distribuzione $Unif(0,1)$.

Soluzione. La densità discreta di X_i può essere scritta come segue:

$$p(x_i|\theta) = \theta^{x_i} (1-\theta)^{1-x_i}.$$

Quindi la distribuzione del campione è data da

$$p(x_1|\theta) \cdots p(x_n|\theta) = \theta^{x_1} (1-\theta)^{1-x_1} \cdots \theta^{x_n} (1-\theta)^{1-x_n} = \theta^{n\bar{x}_n} (1-\theta)^{n(1-\bar{x}_n)}.$$

Utilizzando (3.1), abbiamo che la distribuzione a posteriori di ϑ è data da

$$f_{\vartheta}(\theta|x_1, \dots, x_n) = \frac{p(x_1|\theta) \cdots p(x_n|\theta) f_{\vartheta}(\theta)}{\int_{\Theta} p(x_1|t) \cdots p(x_n|t) f_{\vartheta}(t) dt}$$

$$\begin{aligned}
&= \frac{\theta^{n\bar{x}_n+a-1}(1-\theta)^{n(1-\bar{x}_n)+b-1}}{\int_0^1 t^{n\bar{x}_n+a-1}(1-t)^{n(1-\bar{x}_n)+b-1} dt} \mathbf{1}_{(0,1)}(\theta) \\
&= \frac{\Gamma(n+a+b)}{\Gamma(n\bar{x}_n+a)\Gamma(n(1-\bar{x}_n)+b)} \theta^{n\bar{x}_n+a-1}(1-\theta)^{n(1-\bar{x}_n)+b-1} \mathbf{1}_{(0,1)}(\theta).
\end{aligned}$$

Perciò la distribuzione a posteriori di ϑ è $Beta(n\bar{x}_n + a, n(1 - \bar{x}_n) + b)$. Sapendo che il valore atteso della distribuzione $Beta(a, b)$ è pari a

$$\int_0^1 \theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} d\theta = \frac{a}{a+b},$$

si ottiene il seguente stimatore bayesiano:

$$T_n = \frac{n\bar{X}_n + a}{n + a + b}.$$

□

OSSERVAZIONE. Consideriamo lo stimatore ottenuto nell'Esempio 3.2. La distribuzione a priori ha media

$$\frac{a}{a+b}, \tag{3.2}$$

che sarebbe lo stimatore del parametro se non avessimo il campione e usassimo solo l'informazione a priori. D'altra parte, ignorando l'informazione a priori, avremmo probabilmente utilizzato come stimatore la media campionaria

$$\bar{X}_n. \tag{3.3}$$

In effetti è possibile dimostrare che \bar{X}_n è lo stimatore di massima verosimiglianza. Lo stimatore bayesiano mette insieme queste due informazioni nel seguente modo:

$$T_n = \frac{n}{n+a+b} \bar{X}_n + \frac{a+b}{n+a+b} \frac{a}{a+b}$$

Quindi T_n è una combinazione lineare convessa dei due stimatori (3.2) e (3.3).