Statistica e Laboratorio

4. Calcolo delle probabilità: probabilità elementare

Paolo Vidoni

Dipartimento di Scienze Economiche e Statistiche
Università di Udine
via Tomadini 30/a - Udine
paolo.vidoni@uniud.it

https://elearning.uniud.it/

Sommario

Sommario

- **Introduzione**
- **Esperimenti aleatori ed eventi**
- **La definizione di probabilità: gli assiomi**
- **Probabilità condizionata**
- **Eventi indipendenti**
- **Teorema di Bayes**

Introduzione al calcolo delle probabilità

Fornisce strumenti matematici per lo studio degli esperimenti (fenomeni) casuali (aleatori), tra cui l'esperimento di campionamento.

Costituisce il fondamento teorico e il presupposto formale della statistica inferenziale.

La statistica inferenziale ha l'obiettivo di studiare le caratteristiche di interesse di una popolazione (fenomeno) di riferimento, utilizzando le informazioni contenute in un campione casuale di osservazioni.

La **definizione classica** specifica la probabilità di un evento è come il rapporto tra il numero di casi ad esso favorevoli e il numero di casi possibili, supposti tutti *egualmente probabili*: definizione applicabile solo in contesti molto specifici.

Non si presentano le varie definizioni di probabilità, si segue l'**approccio assiomatico** proposto dal matematico sovietico A.N. Kolmogorov nel 1933: approccio generale che definisce la probabilità di un evento come la *misura* della possibilità che esso si realizzi.

Sommario

Esperimenti aleatori

Un **esperimento** o **fenomeno casuale** (**aleatorio**) è un fenomeno (esperimento) in riferimento al quale le conoscenze inducono a ritenere possibile una pluralità di esiti.

Prima di eseguire l'esperimento, o di osservare il fenomeno, non è possibile individuare con certezza quale dei risultati ammissibili si realizzerà.

Sono esperimenti (fenomeni) aleatori:

a) il lancio di un dado;

b) il numero di giocate al lotto prima di vincere per la prima volta;

c) la misurazione della lunghezza di una barra d'acciaio di lunghezza nominale 20 cm con uno strumento affetto da errore;

d) il rendimento di un titolo azionario;

e) l'estrazione di un campione casuale da una popolazione finita.

Spazio fondamentale ed eventi

L'insieme di tutti i possibili risultati di un esperimento aleatorio, indicato con Ω, è chiamato **spazio fondamentale** (**spazio degli eventi elementari** o **spazio campionario**) è l'insieme di tutti i possibili risultati di un esperimento aleatorio.

I singoli risultati vengono chiamati **eventi elementari** e sono supposti disgiunti in senso insiemistico.

Nonostante lo spazio fondamentale sia noto, non si può individuare con certezza quale evento elementare si realizzerà.

Una volta osservato il fenomeno, o effettuato l'esperimento, uno e un solo evento elementare si sarà realizzato.

Ω è **discreto** se costituito da un **numero finito o da un'infinità numerabile** di punti. È invece detto **continuo** se è costituito da un **insieme continuo** di punti.

Esempio. Sono discreti gli spazi fondamentali generati dagli esperimenti a) e b), che corrispondono rispettivamente a $\Omega = \{1, 2, 3, 4, 5, 6\}$ e $\Omega = \{1, 2, 3, \ldots\} = \mathbf{N}^+$.

Sono continui gli spazi fondamentali generati dagli esperimenti c) ed d), che corrispondono rispettivamente a $\Omega = \mathbf{R}^+$ e $\Omega = \mathbf{R}$. $\Diamond$

Un **evento** è un sottoinsieme dello spazio fondamentale Ω, cioè ogni elemento dell'insieme delle parti (insieme di tutti i sottoinsiemi) di Ω, ovvero di $\mathcal{P}(\Omega)$.

Un **evento si realizza** se e solo se si realizza uno degli eventi elementari che lo definiscono.

Esempio. *Dado.* Nel caso del lancio del dado $\Omega = \{1, 2, 3, 4, 5, 6\}$ e sono eventi, ad esempio, $A =$ "Esce un numero dispari" $= \{1, 3, 5\}$, $B = \{1, 3, 6\}$, $C =$ "Esce il numero 5" $= \{5\}$, se interpretato come sottoinsieme di Ω e non come elemento di Ω. $\Diamond$

C'è una evidente analogia tra eventi di uno spazio fondamentale e sottoinsiemi di un dato insieme. È quindi possibile, come per questi ultimi, definire alcune **operazioni logiche sugli eventi**.

Dati due eventi $A, B \subseteq \Omega$

- A^c indica l'**evento complementare** ad A e contiene tutti gli eventi elementari che *non appartengono ad* A;
- $A \cup B$ indica l'**evento unione** tra A e B e contiene tutti gli eventi elementari che *appartengono o ad* A *o a* B;
- $A \cap B$ indica l'**evento intersezione** tra A e B e contiene tutti gli eventi elementari che *appartengono sia ad* A *che a* B;
- $A \setminus B$ indica l'**evento differenza** tra A e B e contiene tutti gli eventi elementari che *appartengono ad* A *ma non a* B.

Ω è detto anche **evento certo** (contiene tutti gli eventi elementari), mentre con il simbolo $\emptyset$ si indica l'**evento impossibile** (non contiene eventi elementari).

Se $A \cap B = \emptyset$, gli eventi A e B si dicono **incompatibili** (**disgiunti**), poiché *non hanno eventi elementari in comune* e quindi non si realizzano contemporaneamente.

Se $A \subseteq B$, allora A **implica** B, poiché *tutti gli eventi elementari di* A *cadono anche in* B (il viceversa non è necessariamente vero), quindi la realizzazione di A implica la realizzazione di B.

Se $A \subseteq B$ e $B \subseteq A$, gli **eventi** A e B vengono detti **equivalenti**.

Esempio. *Dado* (continua). Nel caso del lancio del dado, se $A = \{1, 3, 5\}$, $B = \{1, 3, 6\}$ e $C = \{5\}$, allora

$$A^c = \{2, 4, 6\}, \quad A \cup B = \{1, 3, 5, 6\}, \quad C \subseteq A,$$

$$A \cap B = \{1, 3\}, \quad A \setminus C = \{1, 3\}, \quad B \cap C = \emptyset.$$

$\diamondsuit$

Sommario

Gli assioni di Kolmogorov

Dato uno spazio fondamentale Ω, si considerano tutti gli eventi di interesse: Ω, $\emptyset$, A, B, C, ecc. In alcuni casi si considerano tutti i possibili eventi (sottoinsiemi di Ω).

La **probabilità è una misura** che associa ad ogni evento $A \subseteq \Omega$ un numero reale, che indica la sua possibilità di realizzazione.

Seguendo l'impostazione assiomatica di Kolmogorov, una **misura di probabilità** P deve essere tale che:

A1. per ogni evento A, $P(A) \geq 0$ (**assioma di non negatività**);

A2. $P(\Omega) = 1$ (**assioma di normalizzazione**);

A3. per ogni collezione finita o al più numerabile di eventi A_i, $i \in I \subseteq \mathbf{N}$, tali che $A_i \cap A_j = \emptyset$, $i \neq j$, si ha che $P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)$ (**assioma di σ-additività**).

Dall'assioma A3. discende che, se $A \cap B = \emptyset$, allora $P(A \cup B) = P(A) + P(B)$ (additività semplice).

Un **evento** A tale che $P(A) = 0$ è detto **trascurabile**. Un **evento** A tale che $P(A) = 1$ è detto **quasi certo**.

Esempio. *Dado* (continua). Nel caso del lancio di un dado regolare, lo spazio fondamentale è $\Omega = \{i : i = 1, \ldots, 6\}$ e ogni faccia ha la stessa probabilità di uscire.

In accordo con tale congettura, si associa ad ogni evento elementare i un peso $p_i = 1/6$, $i = 1, \ldots, 6$, e, dato un generico evento A, $P(A) = \sum_{i \in A} p_i$.

Se $A = \{1, 3, 5\}$, allora $P(A) = 1/6 + 1/6 + 1/6 = 1/2$, che corrisponde alla somma dei pesi degli eventi elementari che compongono A. $\diamondsuit$

Esempio. *Dado* (continua). Si consideri l'esperimento che consiste nel lanciare un dato regolare. Si è interessati al numero di lanci necessari per ottenere l'esito 6 per la prima volta. In questo caso $\Omega = \mathbf{N}^+$.

Si può pensare di associare ad ogni evento elementare $i \in \mathbf{N}^+$, "l'esito 6 si verifica per la prima volta al lancio i-esimo", il peso $p_i = (5/6)^{i-1}(1/6)$, che traduce il fatto che ci sono $i-1$ insuccessi prima di osservare l'esito 6 per la prima volta.

Anche in questo caso, dato un evento A, $P(A) = \sum_{i \in A} p_i$.

Se si ha l'evento $A = \{2, 4, 6, \ldots\}$, "l'esito 6 si verifica per la prima volta in un numero pari di lanci", allora $P(A) = \sum_{i=1}^{+\infty} p_{2i}$. $\Diamond$

Questi due esempi suggeriscono il seguente **criterio costruttivo per definire misure di probabilità** che soddisfano ai tre assiomi di Kolmogorov, nel caso di esperimenti con Ω **finito o numerabile**.

Ad ogni evento elementare $\omega_i \in \Omega$ si associa un peso p_i tale che $p_i > 0$ e $\sum_i p_i = 1$ e si definisce la misura di probabilità P tale che, per ogni evento A, $P(A) = \sum_{i \in A} p_i$.

In entrambi gli esempi vengono soddisfatte le condizioni sui pesi p_i.

Se Ω è **finito** e gli **eventi elementari** sono **equiprobabili**, come ad esempio nel caso del singolo lancio di un dado regolare, il criterio evidenziato in precedenza corrisponde alla **definizione classica di probabilità**.

Infatti, se Ω è costituito da n eventi elementari equiprobabili e $A = \{\omega_i, \, i \in I\}$, con $I \subseteq \{1, \ldots, n\}$, allora $p_i = 1/n$, $i = 1, \ldots, n$, e

$$P(A) = \sum_{i \in I} \frac{1}{n} = \frac{\text{no. casi favorevoli ad } A}{\text{no. casi possibili}}.$$

Quando si parla di "scelta a caso di un elemento da un insieme" Ω finito, si intende implicitamente che tutti gli eventi elementari sono ugualmente probabili.

In molti casi bisogna fare attenzione a definire in modo corretto gli eventi elementari.

Conseguenze degli assiomi

Si presentano alcuni risultati che sono *conseguenze* immediate *degli assiomi di Kolmogorov*.

1) $P(\emptyset) = 0$.

Infatti, per il secondo e il terzo assioma,

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset),$$

da cui $P(\emptyset) = 0$.

2) Per ogni evento A, $P(A^c) = 1 - P(A)$.

Infatti, per il secondo e il terzo assioma,

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c),$$

da cui $P(A^c) = 1 - P(A)$.

3) Se $A \subseteq B$, allora $P(A) \leq P(B)$ e $P(B \setminus A) = P(B) - P(A)$.

Infatti, per il terzo assioma

$$P(B) = P((B \setminus A) \cup A) = P(B \setminus A) + P(A),$$

da cui si ottengono entrambi i risultati.

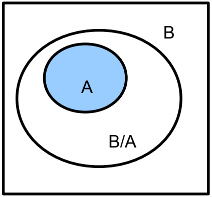
4) Dati gli eventi A e B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

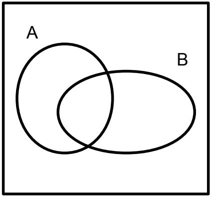
Poiché

$$A \cup B = (A \cap B) \cup [B \setminus (A \cap B)] \cup [A \setminus (A \cap B)],$$

il risultato si ottiene dalla seguente relazione

$$P(A \cup B) = P(A \cap B) + P(B) - P(A \cap B) + P(A) - P(A \cap B).$$

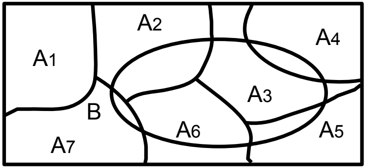




5) Dato un evento B e una partizione A_i, $i \in I \subseteq \mathbf{N}$, di Ω, allora $P(B) = \sum_{i \in I} P(B \cap A_i)$ (**formula di addizione**).

Poiché gli eventi A_i, $i \in I$, sono incompatibili e la loro unione dà Ω, anche gli eventi $B \cap A_i$, $i \in I$, sono incompatibili e, per il terzo assioma, si ha che

$$\begin{aligned} P(B) &= P\left(B \cap \Omega\right) = P\left(B \cap \bigcup_{i \in I} A_i\right) \\ &= P\left(\bigcup_{i \in I} B \cap A_i\right) = \sum_{i \in I} P\left(B \cap A_i\right). \end{aligned}$$



Esempio. *Lotteria*. Una lotteria è costituita 1000 biglietti, di cui 5 vincenti. Si scelgono a caso 10 biglietti. Si vuole determinare la probabilità di $A =$ "un biglietto è vincente".

Tutti i gruppi di dieci biglietti hanno la stessa probabilità di venire estratti. Quindi, utilizzando la definizione classica di probabilità,

$$P(A) = \frac{\binom{5}{1}\binom{995}{9}}{\binom{1000}{10}} \doteq 0.048.$$

Per $B =$ "almeno un biglietto è vincente", conviene determinare la probabilità dell'evento complementare "nessun biglietto vincente"

$$P(B^c) = \frac{\binom{5}{0}\binom{995}{10}}{\binom{1000}{10}} \doteq 0.951,$$

da cui si ricava che $P(B) = 1 - 0.951 = 0.049$. $\diamondsuit$

Esempio. *Indirizzi.* Una rete aziendale è costituita da un server e da dieci PC. Quando un PC accede alla rete, riceve un indirizzo IP scelto in modo casuale tra 200 disponibili.

Nell'ipotesi che tutti e dieci i PC accedano insieme alla rete, quale è la probabilità che il server abbia assegnato almeno due indirizzi IP identici?

I casi elementari equiprobabili sono le sequenze ordinate di 10 indirizzi scelti tra i 200 disponibili e corrispondono alle disposizioni con ripetizione di 200 elementi in gruppi di 10, cioè 200^{10}.

Poiché le configurazioni favorevoli all'evento complementare "tutti i PC hanno indirizzo diverso" sono date dalle disposizioni semplici di 200 elementi in gruppi di 10, si conclude che

$$1 - (200!/190!)/(200^{10}) \doteq 0.204$$

è la probabilità dell'evento cercato. $\Diamond$

Esempio. *Dadi.* Si lanciano tre dadi regolari, uno rosso, uno verde e uno nero, e si vuole calcolare la probabilità dell'evento $A =$"almeno due lati presentano la cifra sei".

Le possibili terne di risultati sono $6^3 = 216$. Dal momento che i dadi sono regolari, tutte le terne hanno la stessa probabilità di presentarsi e tale probabilità corrisponde a $1/216$.

Le terne favorevoli all'evento A sono date da:

- 1 terna con tutti sei;
- 5 terne con sei nei dadi rosso e verde e uno dei restanti cinque numeri nel dado nero;
- 5 terne con sei nei dadi rosso e nero e uno dei restanti cinque numeri nel dado verde;
- 5 terne con sei nei dadi nero e verde e uno dei restanti cinque numeri nel dado rosso.

Quindi la probabilità cercata è $P(A) = 16/216 = 0.06$. $\Diamond$

Sommario

Probabilità condizionata

Dati due eventi A e B, con $P(A) > 0$, può essere interessante specificare la probabilità di B nel caso sia noto il realizzarsi di A, ossia la probabilità dell'**evento condizionato** $B \mid A$.

La probabilità di $B|A$, chiamata **probabilità condizionata** di B dato A, è definita ponendo

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

Intuitivamente, se A si realizzato, l'unica parte di B che può ancora verificarsi è quella comune anche ad A.

La quantità $P(A)$ al denominatore permette di ristabilire le proporzioni, assicurando la normalizzazione.

Esempio. *Roulette*. Si giocano alla roulette i numeri 7, 23 e 32. Poiché la roulette è suddivisa in 37 settori, numerati da 0 a 36, la probabilità di vincere è $P(B) = 3/37 \doteq 0.081$, con $B = \{7, 23, 32\}$.

Se la roulette fosse truccata di modo che possano uscire soltanto i numeri compresi tra 0 e 15, posto $A = \{0, \ldots, 15\}$, la probabilità di vincita corrisponderebbe a

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{1/37}{16/37} = \frac{1}{16} \doteq 0.062.$$

$\Diamond$

Dalla definizione di probabilità condizionata si ottiene la **formula della probabilità composta** (**formula di moltiplicazione**)

$$P(B \cap A) = P(A)P(B|A),$$

con A, B eventi tali che $P(A) > 0$.

La formula di moltiplicazione si può estendere anche al caso di tre o più eventi. Ad esempio, dati A_1, A_2, A_3, tali che $P(A_1 \cap A_2) > 0$,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).$$

Esempio. *Due palline*. Si consideri l'estrazione, senza reinserimento, di due palline da un'urna contenente dieci palline nere e cinque bianche. Si vuole calcolare la probabilità che esca pallina nera in entrambe le estrazioni.

Indicati con A_1 e A_2 gli eventi "esce una pallina nera", rispettivamente, alla prima e alla seconda estrazione, si ha che $P(A_1) = 10/15$ e $P(A_2|A_1) = 9/14$.

Utilizzando la formula di moltiplicazione, la probabilità cercata è $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1) = 6/14$.

Se le singole estrazioni avvengono con reinserimento, si ha che $P(A_2|A_1) = 10/15 = P(A_1)$ e quindi $P(A_1 \cap A_2) = 4/9$. $\Diamond$

Utilizzando alcune relazioni considerante in precedenza, si ottiene il seguente risultato, che risulta molto utile nelle applicazioni.

Dato un evento B e una partizione A_i, $i \in I \subseteq \mathbf{N}$, di Ω, con $P(A_i) > 0$, vale la **formula della probabilità totale**

$$P(B) = \sum_{i \in I} P(A_i) P(B \mid A_i).$$

Infatti, considerando la formula di addizione, dal momento che, per la formula di moltiplicazione, $P(B \cap A_i) = P(A_i)P(B|A_i)$, $i \in I$, si ottiene immediatamente il risultato.

Esempio. *Spam.* Si suppone di possedere tre caselle di posta elettronica. È noto che il 70% della posta proviene dalla prima casella, il 20% dalla seconda e il 10% dalla terza

Dalla prima casella si riceve abitualmente l'1% di messaggi *spam*, mentre dalle altre due si riceve il 2% e il 5% di messaggi *spam*, rispettivamente.

Si vuole calcolare la probabilità di riceve un messaggio *spam*.

Si considerino gli eventi $B =$ "ricevere un messaggio *spam*" e $A_i =$ "ricevere posta dalla casella i", con $i = 1, 2, 3$.

Evidentemente $\{A_1, A_2, A_3\}$ è una partizione di Ω costituita da eventi di probabilità 0.7, 0.2 e 0.1, rispettivamente.

Poiché

$$P(B|A_1) = 0.01, \quad P(B|A_2) = 0.02, \quad P(B|A_3) = 0.05,$$

utilizzando la formula della probabilità totale, si ottiene che

$$\begin{aligned} P(B) &= \sum_{i=1}^{3} P(A_i)P(B|A_i) \\ &= 0.7 \cdot 0.01 + 0.2 \cdot 0.02 + 0.1 \cdot 0.05 \doteq 0.016. \end{aligned}$$

$\Diamond$

Esempio. *Dado e moneta*. Si suppone di lanciare un dado equilibrato e quindi di lanciare una moneta equilibrata tante volte quanto è il punteggio realizzato dal dado. Si vuole calcolare la probabilità dell'evento $B =$ "si ottengono cinque testa nei lanci della moneta".

Si considerino gli eventi $A_i =$ "il punteggio del dado è i", con $i = 1, \ldots, 6$. Evidentemente, $\{A_i, i = 1, \ldots, 6\}$ è una partizione di Ω costituita da eventi di probabilità $1/6$.

Poiché $P(B|A_i) = 0$, $i = 1, 2, 3, 4$, e

$$P(B|A_5) = \left(\frac{1}{2}\right)^5, \quad P(B|A_5) = \left(\begin{array}{c} 6 \\ 5 \end{array}\right) \left(\frac{1}{2}\right)^6,$$

utilizzando la formula della probabilità totale si ottiene che

$$P(B) = \sum_{i=5}^{6} P(A_i)P(B|A_i) = \frac{1}{6}\left(\frac{1}{2}\right)^5 + \frac{1}{6}\left(\begin{array}{c} 6 \\ 5 \end{array}\right)\left(\frac{1}{2}\right)^6 \doteq 0.021.$$

$\diamondsuit$

Sommario

Indipendenza tra eventi

Intuitivamente, due eventi si dicono indipendenti se il realizzarsi o meno di uno dei due non modifica la probabilità di realizzazione dell'altro.

Formalmente, due eventi A e B si dicono (**stocasticamente**) **indipendenti**, se

$$P(A \cap B) = P(A)P(B).$$

Se, invece, $P(A \cap B) \neq P(A)P(B)$, A e B sono detti **dipendenti**.

Si verifica che:

- se A e B sono non trascurabili, la definizione di indipendenza è equivalente a $P(B|A) = P(B)$ oppure $P(A|B) = P(A)$;
- se A e B sono indipendenti, allora lo sono anche A e B^c, A^c e B, A^c e B^c;
- Ω, $\emptyset$, ed anche ogni evento trascurabile, sono indipendenti da qualsiasi evento.

Esempio. *Dado* (continua). Si suppone di lanciare un dado equilibrato e si vuole verificare l'indipendenza tra $A = \{1, 2, 6\}$ e $B = \{3, 6\}$.

Nonostante A e B sembrino, a prima vista, dipendenti, si ha che $P(A \cap B) = 1/6$, $P(A) = 3/6$ e $P(B) = 2/6$, da cui segue invece l'indipendenza stocastica. $\Diamond$

L'indipendenza è un concetto diverso dall'incompatibilità. Ad esempio, se due eventi non trascurabili A e B sono incompatibili, allora $P(A \cap B) = 0$ e quindi necessariamente sono dipendenti, poiché $P(A \cap B) \neq P(A)P(B) > 0$.

La definizione di indipendenza può venire estesa al *caso di più di due eventi*. In particolare, A_1, A_2, A_3 sono indipendenti se

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3), \quad P(A_1 \cap A_2) = P(A_1)P(A_2),$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3), \quad P(A_1 \cap A_3) = P(A_1)P(A_3).$$

Esempio. *Circuito.* Si consideri un circuito con sei componenti dal funzionamento indipendente. La probabilità di rottura, in un certo intervallo di tempo, è 0.5 per il primo componente, 0.2 per il secondo e 0.1 per i rimanenti quattro.

Si determini la probabilità che il circuito si blocchi nell'intervallo di tempo prefissato, nel caso in cui i componenti siano in serie e nel caso siano in parallelo

Sia $A_i =$ "il componente i-esimo si rompe", $i = 1, \ldots, 6$, e $B =$ "il circuito si interrompe". Poiché gli eventi A_i sono indipendenti, lo sono anche i corrispondenti complementari.

Se i componenti sono in serie, il circuito si interrompe se almeno un componente si rompe, quindi

$$P(B) = 1 - P(B^c) = 1 - P(A_1^c \cap \ldots \cap A_6^c) = 1 - \prod_{i=1}^{6} P(A_i^c) \doteq 0.738.$$

Se i componenti sono in parallelo, il circuito si interrompe se tutti i componenti si rompono, quindi

$$P(B) = P(A_1 \cap \ldots \cap A_6) = \prod_{i=1}^{6} P(A_i) = 0.00001.$$

$\diamondsuit$

Esempio. *Uomini e donne*. In una stanza ci sono 5 uomini e 5 donne. Si scelgono a caso due persone (senza reinserimento). Quale è la probabilità che siano entrambe donne?

Sia $A_i =$ "l'i-esima persona scelta è donna", $i = 1, 2$. Visto che $P(A_1) = 5/10$ e $P(A_2 \mid A_1) = 4/9$, la probabilità cercata è

$$P(A_1 \cap A_2) = P(A_1)P(A_2 \mid A_1) \doteq 0.22.$$

Gli eventi non trascurabili A_1 e A_2 sono dipendenti poiché

$$P(A_2) = 1/2 \quad \text{e} \quad P(A_2 \mid A_1) = 4/9.$$

$\diamondsuit$

Esempio. *Banca*. Una filiale di un istituto bancario ha 1210 clienti titolari di conto corrente.

L'ufficio crediti distingue tra *buoni* e *cattivi* clienti, tenendo conto delle eventuali insolvenze. Inoltre, sono noti i dati sull'eventuale possesso della carta di credito.

Le informazioni disponibili vengono sintetizzate nella seguente tabella

	cattivo cliente	*buon* cliente	
con carta di credito	60	520	580
senza carta di credito	21	609	630
	81	1129	1210

Si sceglie casualmente un cliente e si vuole valutare l'eventuale indipendenza tra gli eventi $A =$ "si sceglie un *buon* cliente" e $B =$ "si sceglie un possessore di carta di credito".

Dalla tabella si ricava che

$$P(A) = \frac{1129}{1210} \doteq 0.933, \quad P(B) = \frac{580}{1210} \doteq 0.479,$$

$$P(A \cap B) = \frac{520}{1210} \doteq 0.430.$$

Poiché

$$P(A \cap B) = 0.43 \neq P(A)P(B) = 0.45,$$

si conclude che i due eventi sono dipendenti. Inoltre, la probabilità condizionata

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{520/1210}{580/1210} = \frac{520}{580} \doteq 0.897$$

risulta diversa da $P(A) = 0.933$. $\lozenge$

Sommario

Premessa

Si considera la situazione in cui, noto il risultato di un qualche esperimento, si vuole determinare la probabilità che esso sia dovuto ad una certa causa, o condizione sperimentale.

Ciò accade, ad esempio, quando l'esperimento avviene in due stadi e, pur essendo noto il risultato finale, non si è a conoscenza del risultato ottenuto al primo stadio.

Esempio. *Due urne*. Si considerino due urne indistinguibili. La prima contiene quattro palline bianche e sei nere, la seconda tre palline bianche e cinque nere.

Si sceglie a caso un'urna, senza sapere quale delle due, e si estrae da essa una pallina. Se la pallina èbianca, ci si chiede quale è la probabilità che essa provenga dalla prima urna.

Sia $B =$ "la pallina estratta bianca" e $A_i =$ "si sceglie l'urna i", $i = 1, 2$, si cerca $P(A_1 \mid B)$.

Utilizzando la formula di moltiplicazione, si si ottiene

$$P(A_1 \cap B) = P(A_1)P(B \mid A_1) = (1/2)(4/10) = 1/5.$$

Per la formula delle probabilità totali ha che

$$\begin{aligned} P(B) &= P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) \\ &= (1/2)(4/10) + (1/2)(3/8) = 31/80. \end{aligned}$$

Quindi, per la definizione di probabilità condizionata,

$$P(A_1 \mid B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{1/5}{31/80} \doteq 0.516.$$

Si noti che $P(A_1 \mid B) > P(A_1)$ e questo trova una giustificazione nel fatto che la prima urna contiene una porzione maggiore di palline bianche ed inoltre si suppone di avere estratto pallina bianca. $\Diamond$

Teorema di Bayes

Il **teorema di Bayes** afferma che:

dato un evento B non trascurabile e una partizione A_i, $i \in I \subseteq \mathbf{N}$, di Ω costituita da eventi non trascurabili, si ha che, per ogni $i \in I$,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}.$$

Se $P(B)$ non è nota, si può utilizzare la formula delle probabilità totali $P(B) = \sum_{j \in I} P(A_j)P(B|A_j)$.

Infatti, per la definizione di probabilità condizionata, applicando la formula di moltiplicazione, si ha che, per ogni $i \in I$,

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{P(B)},$$

che è il risultato cercato.

Thomas Bayes è stato un matematico e ministro presbiteriano britannico. Il teorema che prende il suo nome è stato pubblicato postumo nel 1763.

Si possono fare le seguenti considerazioni:

- le probabilità $P(A_i)$, $i \in I$, vengono chiamate **probabilità a priori** (**iniziali**) della condizione sperimentale i-esima e riflettono quelle che sono le conoscenze disponibili prima della realizzazione dell'esperimento;
- le probabilità $P(A_i|B)$, $i \in I$, vengono chiamate **probabilità a posteriori** (**finali**) e tengono conto del fatto che l'esperimento si è concluso e l'evento B si è realizzato;
- il teorema di Bayes esprime formalmente una procedura coerente di apprendimento dall'esperienza;
- $P(A_i|B)$ risulta proporzionale a $P(A_i)P(B|A_i)$, mentre la quantità $P(B)$, presente al denominatore, è un fattore di normalizzazione;
- $P(B|A_i)$ è chiamata la **verosimiglianza** di A_i; si può interpretare concettualmente come la verosimiglianza che il verificarsi di B ha attribuito alla condizione sperimentale A_i.

Esempio. *Linee di produzione.* Un'azienda produce il 30% dei suoi articoli con una prima linea di produzione, che fornisce 8 pezzi difettosi su 100, mentre il restante 70% con una seconda linea, che fornisce 5 pezzi difettosi su 100.

Si sceglie a caso un articolo, senza sapere da quale linea provenga, e viene scartato perché difettoso. Si vuole calcolare la probabilità che provenga dalla prima linea di produzione.

Indicato con $B =$ "l'articolo selezionato è difettoso" e con $A_i =$ "l'articolo selezionato proviene dall'i-esima linea", $i = 1, 2$, si ha che $P(B|A_1) = 8/100$, $P(B|A_2) = 5/100$, $P(A_1) = 3/10$, $P(A_2) = 7/10$.

Per il teorema di Bayes, la probabilità cercata è

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} = \frac{24}{59} \doteq 0.407.$$

$\Diamond$

Esempio. *Fumatori*. Una popolazione presenta il 32% di fumatori. È noto che il 25% dei fumatori e il 5% dei non fumatori è affetto da una patologia respiratoria cronica.

Si sceglie a caso un individuo dalla popolazione. Quale è la probabilità che sia affetto dalla patologia? Se l'individuo risulta ammalato, quale è la probabilità che sia fumatore?

Indicato con $B =$ "l'individuo scelto è ammalato" e con $A =$ "l'individuo scelto è fumatore", si ha che $P(B|A) = 0.25$, $P(B|A^c) = 0.05$, $P(A) = 0.32$, $P(A^c) = 0.68$.

Per la formula della probabilità totale e per il teorema di Bayes si ottiene, rispettivamente,

$$P(B) = 0.32 \cdot 0.25 + 0.68 \cdot 0.05 \doteq 0.114.$$

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.32 \cdot 0.25}{0.114} \doteq 0.407.$$

$\Diamond$

Esempio. *Spam* (continua). Si consideri l'esempio delle tre caselle di posta elettronica.

Nell'ipotesi che si riceva un messaggio catalogato come *spam*, si vuole calcolare la probabilità che provenga dalla seconda casella di posta.

Avendo già calcolato in precedenza $P(B) = 0.016$, per il teorema di Bayes la probabilità cercata è

$$P(A_2|B) = \frac{P(A_2)P(B|A_2)}{P(B)} = \frac{0.2 \cdot 0.02}{0.016} = 0.25.$$

$\Diamond$