Nozioni Preliminari per la Probabilità - Sommario

Nozioni preliminari per la ludopatia.

A. INTRODUZIONE MATEMATICA

A1. L'essenza della Probabilità

Essenza della Probabilità

Breve discorso introduttorio al calcolo della probabilità: concezione filosofica del determinismo, limiti del determinismo. Necessità della probabilità per misurare gli eventi in assenza di informazioni: concezione analogica della probabilità con la misurazione.

---

1. Preambolo

DETERMINISMO. Esistono degli eventi già *predeterminati*, dall'inizio?
Secondo la concezione *deterministico-matematica* della realtà, sì. Se lancio un dado, allora dovrebbe essere possibile calcolare con *totale accuratezza* il valore del risultato. Tuttavia, sappiamo che è molto difficile farlo: servirebbero un sacco di informazioni! Come ad esempio l'aerodinamica della forma, la forza del vento, la forza con cui tiro il dado ed eccetera...
Quindi, in assenza di queste informazioni, posso comunque *associare* un numero all'*evento in cui esce un numero fissato*?

LA PROBABILITA'. La risposta è sì, ed è qui che la teoria della probabilità viene ad aiutarci. Nonostante l'*impossibilità di ottenere la predizione precisa dell'output*, possiamo comunque in un modo *"misurare"* questi output.
Vogliamo creare una specie di *bilancia*, che al posto di misurare la *massa* di

un'*oggetto*, misuri la *"probabilità"* che accada un evento.
Quindi mediante gli strumenti della matematica, costruiremo dei blocchi fondamentali per la teoria della probabilità, che a sua volta sarà in grado di misurare la *"probabilità"* degli eventi in qualsiasi situazioni.

A2. Strutture Matematiche Fondamentali

Strutture Matematiche della Probabilità

Definizioni delle strutture matematiche per la probabilità: algebra, probabilità su un'algebra. Estensioni delle nozioni di algebra e probabilità: sigma-algebra, misura di probabilità su sigma-algebra. Sigma-algebra di Borel in R^N.

---

1. Algebra e Probabilità

#Definizione

Definizione (algebra per un insieme).

Supponiamo di avere un insieme Ω finito, una famiglia $\mathcal{A}$ non-vuota di sottoinsiemi di Ω (in genere la famiglia di *tutti i sottoinsiemi* di Ω).
Se valgono le seguente regole:
A1. *L'esistenza dell'evento certo e dell'evento impossibile*

$$\Omega, \emptyset \in \mathcal{A}$$

A2. *Chiusura dell'unione e dell'intersezione*

$$A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A} \wedge A \cap B \in \mathcal{A}$$

A3. *L'esistenza dell'evento opposto (oppure elemento complementare in Ω)*

$$A \in \mathcal{A} \implies \mathcal{C}_{\mathcal{A}} A \in \mathcal{A}$$

Allora la famiglia $\mathcal{A}$ si dice *algebra*.

#Definizione

Definizione (eventi elementari, spazio degli eventi elementari, eventi casuali, eventi mutualmente esclusivi).

Si può dare una *nomenclatura* per elementi degli insiemi appena definiti, in particolare nel contesto della *probabilità*.

Per Ω può essere definita come lo *spazio degli eventi elementari*, per gli elementi $\omega \in \Omega$ possono essere definiti come *eventi elementari*, invece per gli elementi $A \in \mathcal{A}$ possono essere definiti come *eventi casuali*.
Se due elementi $A, B \in \mathcal{A}$ hanno intersezione nulla, si dice che sono *eventi mutualmente esclusivi*.

#Definizione

Definizione (probabilità su un'algebra).

Supponiamo di avere un'*algebra* $\mathcal{A}$ (Definizione 1 (algebra per un insieme)).
Sia definita una funzione P del tipo

$$P : \mathcal{A} \longrightarrow [0,1]$$

(in realtà si può scegliere un altro intervallo chiuso del tipo $[a,b]$, la scelta di $[0,1]$ è una pura convenzione)
Se valgono le seguenti:
P1. *L'evento certo ha la probabilità massima e l'evento nullo ha la probabilità nulla*

$$P(\Omega) = 1; P(\emptyset) = 0$$

P2. *Additività*
Supponendo che A, B siano disgiunti, ovvero tali che $A \cap B = \emptyset$, allora vale che

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

#Osservazione

Osservazione (la necessità di estendere questi concetti).

Fin'ora abbiamo visto che i concetti di *algebra* e di *probabilità* servono per modellizzare i *spazi probabilistici finiti*, dove l'insieme degli eventi campionari Ω è *finito* (Definizione 1 (spazio di probabilità finito)).

Tuttavia, questo non è *sufficiente* per modellizzare *tutti* i *spazi probabilistici*. Infatti, se volessimo modellizzare i *spazi probabilistici discreti*, ovvero con un insieme Ω *infinito* (ma comunque *numerabile*!), ci serve estendere il concetto di *algebra* e di *probabilità*.

2. σ-Algebra e σ-Probabilità

#Definizione

Definizione (σ-Algebra su uno spazio campionario).

Una famiglia $\mathcal{A}$ di parti di un insieme Ω (*che può essere infinito*) si dice *σ-algebra* (dove σ sta per *"numerabile"*) se valgono le seguenti regole.
σA1. *L'esistenza dell'evento certo e impossibile*

$$\Omega, \emptyset \in \mathcal{A}$$

σA2. *L'esistenza dell'evento opposto*

$$A \in \mathcal{A} \implies \mathcal{C}_{\mathcal{A}} A \in \mathcal{A}$$

σA3. *Chiusura dell'intersezione e unione infinita*

$$\{\mathcal{A}_1, \ldots, \mathcal{A}_k\} \subset \mathcal{A} \implies \bigcup_{n=1}^{\infty} \mathcal{A}_n, \bigcap_{n=1}^{\infty} \mathcal{A}_n \in \mathcal{A}$$

#Osservazione

Osservazione (il concetto di sigma-algebra è un'estensione di algebra).

Evidentemente *ogni* σ-algebra è anche un'algebra, dal momento che le *regole assiomatiche* del σ-algebra sono una versione *"più forte"* delle regole dell'algebra.

In particolare le due strutture *coincidono* se Ω ha un *numero finito* di *elementi*.

#Definizione

Definizione (σ-probabilità).

Sia $\mathcal{A}$ un σ-algebra su Ω. Una *misura di probabilità* (o σ-probabilità) è una funzione $p : \mathcal{A} \longrightarrow [0, 1]$ tale che valgano le seguenti regole.
σP1. *Gli eventi estremi hanno i valori estremi*

$$p(\emptyset) = 0, p(\Omega) = 1$$

σP2. *σ-additività*
Se $(A_n)_n \subset \mathcal{A}$ è *una successione di insiemi a due a due disguinti*, allora vale che

$$\sum_{n=1}^{+\infty} p(A_n) = p\left(\bigcup_{n=1}^{\infty} A_n\right)$$

3. σ-Algebra di Borel in R^N

#Osservazione

Osservazione (problema preliminare).

Supponiamo di avere una *figura di forma indeterminata* in una *circonferenza* e di voler calcolare la probabilità che, lanciando una freccia casualmente, si colpisce la *figura*.

Un buon approccio è quello di approssimare l'intero cerchio con *piccoli rettangoli* e di fare una proporzione tra *piccoli rettangoli appartenente alla figura nera* e tra tutti i *rettangoli disponibili*.

Vogliamo quindi approssimare questa figura con una *famiglia di rettangoli*, che faremo con la seguente struttura matematica, e allo stesso tempo di *trovare una struttura matematica* che ci permetta di modellizzare questo *problema di natura probabilistica*.

#Definizione

Definizione (σ-algebra di Borel in 2D).

Definiamo la *σ-algebra di Borel in* $\mathbb{R}^2$ come la *più piccola σ-algebra* che *contiene tutti i rettangoli*, cioè

$$\mathcal{B}(\mathbb{R}^2) = \{(a,b)\times(c,d) : a,b,c,d \in \mathbb{R}\}$$

Se chiara dal contesto, indichiamo la *sigma-algebra di Borel* semplicemente come $\mathcal{B}$.

#Definizione

Definizione (σ-algebra di Borel generalizzata in N).

Si può generalizzare il concetto di *σ-algebra di Borel* in $\mathbb{R}^N$, definendola come la seguente:

$$\mathcal{B}(\mathbb{R}^N) = \left\{\prod_{k=1}^{N}(a_k, b_k) : a_k, b_k \in \mathbb{R}\right\}$$

In particolare sarà utile vedere la *σ-algebra di Borel* in $N=3$ come la *σ-algebra* più piccola contenente *parallelepipedi*, in $N=1$ contenente *segmenti*.
Con dimensioni più grandi, non sarà più (fisicamente) possibile fare delle *analogie geometriche*, si penserà dunque N come il *numero delle variabili*.

#Proposizione

Proposizione (analisi di σ-algebra di Borel in 1D).

Iniziamo ad analizzare $\mathcal{B} = \mathcal{B}(\mathbb{R}^1)$.
Vediamo che $\mathcal{B}$ contiene le *semirette*, dato che possono essere *espressione* di *unione infinita di segmentini finiti*. Infatti,

$$(a,+\infty) = \bigcup_{k=1}^{\infty}(a, a+k)$$

Allora da ciò discende che i *loro complementari* fanno parte di $\mathcal{B}$, ovvero semirette *chiuse*. Ovvero, intervalli del tipo $(-\infty, a]$ o $[a, +\infty)$.
Ma allora, considerando l'intersezione di *semirette chiuse*, abbiamo che gli *intervalli chiusi limitati* del tipo $[a, b]$ sono elementi di $\mathcal{B}$! Ma allora anche $(a, b) \in \mathcal{B}$!

Questa breve analisi vuole dire che *"tutti gli intervalli di ogni tipo fanno parte della σ-algebra di Borel in* $\mathbb{R}$.

A3. Spazio Finito di Probabilità

Spazio di Probabilità Finito

Definizione e proprietà dello spazio di probabilità finito. Regole per la modellizzazione di uno spazio probabilistico finito.

---

0. Voci Correlate

- Strutture Matematiche della Probabilità

1. Definizione di Spazio di Probabilità Finito

#Definizione

Definizione (spazio di probabilità finito).

Sia Ω un *insieme finito*, sia $\mathcal{A}$ un'*algebra* e sia $P : \mathcal{A} \longrightarrow [0, 1]$ una *probabilità*.
Allora la terna $(\Omega, \mathcal{A}, P)$ si dice *spazio di probabilità finito*.

2. Proprietà dei spazi di probabilità finiti

#Proposizione

Proposizione (prime proprietà dei spazi di probabilità finiti).

Uno spazio di probabilità $(\Omega, \mathcal{A}, P)$ gode delle seguente proprietà.
SP1. (*Subadditività*)
Siano $A_1, \ldots, A_n \in \mathcal{A}$ degli eventi *a due a due disgiunti* (ovvero tali che $A_i \cap A_j = \emptyset$ per qualsiasi i, j scegliamo). Allora vale la relazione

$$\sum_{k=1}^{n} P(A_k) = P\left(\bigcup_{k=1}^{n} A_k\right)$$

Se *non* sono disgiunti, allora vale che

$$\sum_{k=1}^{n} P(A_k) \geq P\left(\bigcup_{k=1}^{n} A_k\right)$$

SP2. (*La somma della probabilità di due eventi*)

$$\forall A, B \in \mathcal{A}, P(A) + P(B) = P(A \cup B) + P(A \cap B)$$

SP3. (*Monotonia della probabilità*)
Si ha la probabilità della sottrazione degli eventi $P(A \diagdown B) = P(A) - P(B)$ per $A, B \in \mathcal{A}$ tali che $B \subseteq A$.
Vale che $P(A \diagdown B) \geq 0$, di conseguenza $B \subseteq A \implies P(B) \leq P(A)$; ovvero se un'*evento minore* è *"contenuto"* in un *"evento maggiore"*, allora l'evento maggiore è più probabile (o ugualmente) dell'evento minore.

#Osservazione

Osservazione (interpretazione intuizionistica della SP2).

Vediamo di dare un senso alla SP2, ovvero l'enunciato

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

Qui stiamo *sommando* due eventi che non sono necessariamente *mutualmente esclusivi*; quindi sommando le loro probabilità rischiamo di avere una specie die *"eccesso"*, che in questo caso è rappresentato da $A \cap B$. Infatti, $A \cap B$ sostanzialmente vuol dire l'evento in cui accadono sia A che B.

Questo è perfettamente in *linea* con l'analogia della *probabilità* come la *misurazione degli eventi*: misurando la lunghezza di due blocchi che condividono una parte, si ha un'eccesso.

#Proposizione

Proposizione (il complementare dell'evento è calcolabile in un modo).

Dato che $\mathcal{A}$ è un'*algebra*, se $A \in \mathcal{A}$ allora $\mathcal{C}_\Omega A \in \mathcal{A}$.
In particolare vale la relazione

$$P(\mathcal{C}_\Omega A) = 1 - P(A)$$

3. Esempi

#Esempio

Esempio (il dado).

Vogliamo modellizzare lo *spazio di probabilità finito* per un dado D6 (ovvero in gergo *"geek"*, a sei facce). Assumiamo che il dado non sia truccato. Prendiamo allora

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

e $\mathcal{A}$ la *famiglia di tutti i sottoinsiemi di* Ω.

Prima di tutto sappiamo subito due cose:

1. La probabilità che non esca nessun numero è 0, ovvero *impossibile*.
2. La probabilità che esca un qualsiasi numeri di Ω è 1, ovvero *certo*.
   E invece per gli altri eventi? Dal momento che abbiamo assunto che il dado sia *bilanciato*, ovvero le probabilità che escano un singolo numero abbiamo lo stesso *"peso"*, abbiamo

$$P(1) + P(2) + \ldots + P(6) = 1 \implies P(1) = \ldots = P(6) = \frac{1}{6}$$

Ora abbiamo tutti gli strumenti per calcolare la probabilità di ogni evento casuale.

Numeri pari? Basta prendere $P(\{2,4,6\}) = P(2) + P(4) + P(6)$. Numeri dispari? Ragionamento analogo.

4. Regole per modellizzare uno Spazio di Probabilità

#Osservazione

Osservazione (ci sono regole precise per modellizzare spazi probabilistici).

Notiamo che quando si modellizza uno *spazio probabilistico*, usiamo implicitamente dei *principi fondamentali* che possono essere usati per tutti le situazioni. Ora ne vediamo alcune.

#Proposizione

Proposizione (trovare gli insiemi di riferimento).

Scegliamo l'insieme finito Ω come l'insieme che contiene tutti gli *eventi elementari*, ovvero tutti i *singoli esiti possibili*, del tipo $\Omega = \{\omega_1, \ldots, \omega_n\}$.
Scegliamo adesso una *stringa* Δ come un insieme di *numeri* del tipo $\Delta = \{\alpha_1, \ldots, \alpha_n\}$ tali che la *somma di tutti i numeri della stringa* sia 1.
Ovvero $\sum_{1<i\leq n} \alpha_i = 1$.

Notiamo benissimo che la scelta dei numeri $\alpha_1, \ldots, \alpha_n$ è completamente *arbitraria*! L'unica condizione è che la *somma* di tutti questi numeri faccia 1.

#Proposizione

Proposizione (estensione della probabilità su eventi casuali).

Adesso parliamo della *funzione probabilità* $p : \mathcal{A} \longrightarrow [0,1]$.
Prima di tutto poniamo

$$p(\{\omega_k\}) = \alpha_k, \forall k \in \{1, \ldots, n\}$$

Ovvero associamo la *probabilità dell'evento elementare* k-esimo al valore k-esimo della stringa Δ.

Adesso vogliamo estendere la definizione di *probabilità sugli eventi casuali*, ovvero il restante degli elementi di $\mathcal{A}$.
Poniamo un qualsiasi elemento $A \in \mathcal{A}$ come un *sottoinsieme qualsiasi non-vuoto* di Ω, dunque

$$p(A) = \sum_{\omega \in A} p(\{\omega\})$$

ovvero come la *"somma di tutti i singoli eventi elementari"*.
Poniamo infine $p(\emptyset) = 0$.

#Proposizione

Proposizione (altre conseguenze).

Da queste regole elementari appena poste, seguono altre regole derivate.
Ad esempio si ha $p(\Omega) = 1$, dal momento che stiamo sommando *tutti* gli elementi di Δ, che per costruzione deve essere 1.
Inoltre si verifica che se due eventi A, B sono *mutualmente esclusivi*, allora la probabilità che accada *una di queste* è la loro somma individuale.
Ovvero

$$A \cap B = \emptyset \implies p(A \cup B) = p(A) + p(B)$$

Questa viene verificata dal momento che possiamo *"unire le sommatorie"*

$$\sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega) = \sum_{\omega \in (A \cup B)} p(\omega)$$

A4. Spazio Discreto di Probabilità

Spazio di Probabilità Discreto

Definizione di spazio di probabilità discreto come generalizzazione di spazi di probabilità finiti. Proprietà dei spazi probabilistici discreti. Esempio del lancio della moneta ripetuta infinitamente.

---

1. Definizione di Spazio di Probabilità Discreto

#Definizione

Definizione (spazio di probabilità discreto).

Sia $(\Omega, \mathcal{A}, p)$ un *spazio di probabilità*. Questo *spazio di probabilità* si dice *discreto* se vale che Ω è *al più numerabile* (ovvero esiste una biiezione tra Ω e $\mathbb{N}$) e che $\mathcal{A}$ è la *famiglia di tutti i sottoinsiemi di* Ω. In particolare, $\mathcal{A}$ è una *sigma-algebra*.

#Proposizione

Proposizione (modellizzazione di uno spazio probabilistico discreto).

Adesso vogliamo capire *come definire* questo spazio probabilistico discreto, in particolare per quanto riguarda la probabilità p. Come possiamo associare un *numero* ad una *somma infinita di numeri*? Qui ci serve la *teoria delle serie*. Possiamo porre

$$p(A) = \begin{cases} \sum_{\omega \in A} p(\{\omega\}), \emptyset \neq A \subset \Omega \\ 0, A = \emptyset \end{cases}$$

Questa definizione è *ben posta*, dato che Ω è un *insieme numerabile* e quindi $\mathcal{A}$ *al più numerabile*.

#Osservazione

Osservazione (la probabilità degli eventi casuali è ben posta).

Se A è *numerabile* allora interpreto la probabilità

$$\sum_{\omega \in A} p(\{\omega\})$$

come una *serie*, dal momento che A è un elemento di un sottoinsieme di elementi infiniti di Ω.
Ma allora non si avrebbe il rischio di avere una definizione *mal posta*?
Ovvero che *cambiando l'ordine in cui sommo i termini* $p(\{\omega\})$ posso cambiare il *risultato* della somma?

> La risposta è no, dal momento che stiamo trattando di *serie convergenti a termini positivi*, quindi *assolutamente convergenti* (Teorema 4 (dell'assoluta convergenza)). Dunque per il *teorema di Riemann* (Teorema 7 (di Riemann)), la somma della serie non cambia.

#Proposizione

> **Proposizione (proprietà dei spazi di probabilità).**
>
> Sia $(\Omega, \mathcal{A}, p)$ uno *spazio di probabilità discreto*. Allora valgono le seguenti.
> A. *Continuità dal basso*
> Se $(A_n)_n$ è una *successione crescente di eventi casuali*, ovvero che valga $A_n \subset A_{n+1}$ (*"un evento è sempre incluso nel prossimo"*), allora vale il limite
>
> $$p\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_n p(A_n)$$
>
> B. *Continuità dall'alto*
> Se $(A_n)_n$ è una *successione decrescente di eventi casuali*, ovvero $A_{n+1} \subset A_n$ (ovvero *"un evento include sempre il prossimo evento"*), vale il limite
>
> $$p\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_n p(A_n)$$

#Corollario

> **Corollario (sigma-subadditività dei spazi probabilistici discreti).**
>
> Sia $(A_n)_n$ una *successione di eventi in uno spazio probabilistico discreto* $(\Omega, \mathcal{A}, p)$. Allora vale che
>
> $$p\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{n=1}^{+\infty} p(A_n)$$

#Osservazione

Osservazione (bilanciare la disuguaglianza).

Prendiamo la disuguaglianza del Corollario 5 (sigma-subadditività dei spazi probabilistici discreti), per due *eventi non disgiunti* A, B: si ha

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

Da un punto di vista geometrico, l'elemento $P(A \cap B)$ rappresenta l'*"eccesso"* della somma $A \cup B$; infatti, se $A \cap B = C \neq \emptyset$, si può vedere che vale la relazione

$$A \cup B = (A \diagdown C) \cup (B \diagdown C) \cup C$$

allora

$$P(A \cup B) = P(A) - P(C) + P(B) - P(C) + P(C)$$

ovvero la tesi

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

Dalla stessa idea si ha

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

2. Esempi di Modelli Probabilistici Discreti

#Esempio

Esempio (moneta lanciata infinitamente).

Presa una moneta e *lanciandola ripetutamente*, vogliamo determinare la *probabilità che esca testa al lancio n-esimo ma non prima*.

Per approcciarci a questo problema, possiamo scegliere Ω come *l'insieme delle successioni a valori in* $\{0, 1\}$. Ad esempio, un elemento di Ω potrebbe essere $(0, 0, 1, 0, 1, 0, 1, \ldots)$.
Allora prendiamo un suo *elemento generico* $\omega_n \in \Omega$ come la *famiglia*

delle successioni che hanno 1 nei *primi elementi* e 0 all'n-esimo (dopo non ci importa più come si comporta).
Vediamo che $\Omega = \{\omega_n, n \in \mathbb{N}\}$.
Per rispondere alla *domanda*, possiamo ragionevolmente stabilire che *in ogni singolo lancio la probabilità che esca testa o croce sia la stessa* (ovvero che la moneta non sia truccata). Inoltre, gli eventi elementari sono *indipendenti*.

Allora possiamo assegnare a ω_1 il numero 0.5: infatti la probabilità che esca testa al primo lancio è proprio 0.5. Per quanto riguarda invece ω_2, possiamo fare la *metà della metà*: ovvero 0.25. Andando avanti, assegniamo

$$\omega_n = \frac{1}{2^n}$$

Notiamo che la serie formata dalle probabilità degli eventi elementari è *convergente* con *somma* 1.

$$\sum_{n=1}^{+\infty} \frac{1}{2^n} = 1$$

Quindi questo è *perfettamente in linea col nostro modello probabilistico*.

Adesso, possiamo iniziare a *rispondere delle domande*. Se vogliamo ad esempio calcolare la probabilità che esca testa *per la prima volta tra il terzo e sesto lancio (compresi)*, basta fare la somma

$$\frac{1}{2^3} + \ldots + \frac{1}{2^6}$$

Oppure se vogliamo sapere la probabilità che esca *testa* in un *lancio pari*, basta fare la somma

$$\sum_{n=1}^{+\infty} \frac{1}{2^{2n}} = \sum_{n=1}^{+\infty} \frac{1}{4^n} = \frac{1}{3}$$

#Osservazione

> Osservazione (attenzione nel modellizzare i modelli probabilistici).

> Come possiamo notare, abbiamo preso Ω come un *insieme fatto di famiglie di successioni*, invece di *un insieme fatto di singole successioni*. Perché questa scelta? Non potevamo semplicemente scegliere gli eventi casuali di Ω come le singole successioni?
>
> La risposta è sì, ma questo ci creerebbe problemi. Infatti, in questo caso Ω non diventerebbe più numerabile: se volessimo assegnare ad ogni singolo evento ω un numero tale che la somma di tutti gli elementi rimanga 1, bisognerebbe dare un *numero infinitamente piccolo*.
>
> Per spiegare meglio questo concetto, vediamo un *esempio pratico*. Ci viene chiesto di *modellizzare la risposta elastica* di un *ponte di ferro costruito con delle travi*. Possiamo considerare un modello partendo dai *singoli atomi di ferro*? Sì. Conviene farlo? No, dato che il *concetto di elasticità* viene definito a partire dalle *travi*, non *sugli atomi*.

B. LA PROBABILITA' CONDIZIONALE

B1. Definizione di Probabilità Condizionale

Probabilità Condizionale

Probabilità condizionale. Esempio preliminare, definizione di probabilità condizionale, osservazioni sulla probabilità condizionale, indipendenza di due eventi. Regola della catena. Esempio concreto di modello probabilistico con probabilità condizionali.

---

1. Osservazione preliminare

#Osservazione

> Osservazione (la probabilità può cambiare al variare di informazioni disponibili).

Consideriamo un'urna con *dieci palline* numerate da 1 a 10. Qual è la probabilità che, estraendo una pallina, si ha una pallina con un numero *minore o uguale* di 5? Ovviamente è 0.5, dato che $p(A) = 5/10 = 0.5$. Infatti, $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 10\}$ e $A = \{1, 2, 3, 4, 5\}$.

Adesso cambiamo le cose: supponiamo ora di sapere a priori (tramite un insider) che il numero sarà *sicuramente pari*. Ma allora quale sarà la probabilità di avere un numero *minore o uguale di 5*? Qui le cose cambiano, dato che dobbiamo *cambiare lo spazio campionario* a $\Omega' = \{2, 4, 6, 8, 10\}$ e prendiamo l'insieme $A' = \{\omega' \in \Omega' : \omega \leq 5\}$. Quindi la probabilità *è scesa* a $\tilde{p}(A') = 2/5$.

Si osserva in particolare che $A' = A \cap \Omega'$ e vale la relazione

$$\tilde{p}(C) = \frac{|A \cap \Omega'|}{|\Omega'|} = \frac{p(A \cap \Omega')}{p(\Omega')}$$

Vedremo di *definire rigorosamente* questo concetto.

2. Definizione di Probabilità Condizionale

#Definizione

Definizione (probabilità condizionale).

Siano A, B due *eventi* in uno *spazio di probabilità* $(\Omega, \mathcal{A}, p)$ con $p(B) > 0$. Chiamiamo la *"probabilità condizionale di A dato B"* la quantità definibile come

$$\boxed{p(A|B) := \frac{p(A \cap B)}{p(B)}}$$

Notare che $A|B$ *non ha nessun significato* ed è solo un modo per esprimere *"evento A dato B"*.

#Osservazione

Osservazione (i casi estremi).

Notiamo che se $P(B) = 0$, la definizione perde senso. Infatti, se B non *ha nessuna probabilità di accadere*, allora non ha senso chiedere *se un certo evento A accadrebbe o meno a seconda dell'evento B*, dato che tanto l'evento B è impossibile.

Parimenti, se $P(B) = 1$ non *avrebbe senso considerare* la probabilità condizionale, dato che rimane la stessa. Infatti, $P(B) = 1 \implies P(A|B) = P(A)$.

#Proposizione

Proposizione (la probabilità condizionale costituisce una probabilità).

Sia $p(\cdot|B)$ la funzione del tipo

$$p(\cdot|B) : \mathcal{A} \longrightarrow [0, 1]$$

cioè $\forall A \in \mathcal{A}$, $p(A|B)$, allora questa funzione è una probabilità su $\mathcal{A}$.

#Proposizione

Proposizione (la regola della catena).

Siano $A_1, \ldots, A_n$ degli *eventi* in uno *spazio di probabilità* $(\Omega, \mathcal{A}, p)$ tali che la *probabilità di tutti gli eventi fino all'$n-1$-esimo non sia nulla*, ovvero

$$p\left(\bigcap_{k=1}^{n-1} A_k\right) > 0$$

allora vale la seguente regola:

$$p\left(\bigcap_{k=1}^{n} A_k\right) = p(A_1) \cdot p(A_2|A_1) \cdot p(A_3|A_2 \cap A_1) \cdot \ldots \cdot p\left(A_n | \bigcap_{k=1}^{n-1} A\right.$$

3. Esempi di Problemi con la Probabilità Condizionale

#Esempio

Esempio (l'urna).

Supponiamo di avere un'urna con 90 palline, da cui ne estriamo 5. Vogliamo sapere la probabilità che escano i numeri 1 e 7, e sapendo che dei 5 numeri estratti 3 saranno sicuramente dispari.

Modellizziamo questo problema come il seguente:

$$\Omega = \{\omega \subset \{1, \dots, 90\} : |\omega| = 5\}$$

Dal calcolo combinatorio è noto che $|\Omega| = \binom{90}{5}$.

Adesso prendiamo A come l'evento in cui *"i numeri estratti contengono 1 e 7"*, l'evento B come *"tre dei numeri estratti sono dispari"*. Per calcolare $|A|$, $|B|$ e $|A \cap B|$, sarà necessario usare i *strumenti del calcolo combinatorio*, da cui discende che:

$$|A| = \binom{88}{3}$$
$$|B| = \binom{45}{2}\binom{45}{3}$$
$$|A \cap B| = 43\binom{45}{2}$$

Adesso possiamo calcolare $P(A)$ e $P(A|B)$. Abbiamo dunque

$$P(A) = \frac{\binom{88}{3}}{\binom{90}{5}} = \dots \approx 0.0025$$

e

$$P(A|B) = \frac{43\binom{45}{2}}{\binom{90}{5}} \cdot \frac{\binom{90}{5}}{\binom{45}{2}\binom{45}{3}} = \dots \approx 0.003$$

In questo caso la probabilità di A è *aumentata*.

#Osservazione

Osservazione (avvolte si parte dalle probabilità condizionali).

Vediamo adesso un esempio in cui si *costruisce un modello probabilistico* a partire dalle *informazioni sulle probabilità condizionali*. Questo si rivelerà utile, spesso nei *casi reali* in cui si hanno *quasi esclusivamente* probabilità condizionali.

Ad esempio, spesso si hanno informazioni del tipo "la probabilità che un incidente accada è del 20% se piove".

#Esempio

Esempio (le urne).

Supponiamo di avere *due urne*, etichettate come α e β.
L'urna α contiene *tre palline nere* e *una pallina bianca*; l'urna β contiene invece *una pallina nera* e *una pallina bianca*.
Adesso estraiamo una palla da una delle urne. Si suppone che la scelta dell'urna sia equiprobabile. Qual è la probabilità di pescare una pallina nera?

Prima di tutto costruiamo il modello probabilistico.

$$\Omega = \{\alpha_n, \alpha_b, \beta_n, \beta_b\}$$

dove il singolo elemento rappresenta l'esito *"si è scelta l'urna ... e la pallina ..."*.
Adesso, prendiamo $A = \{\alpha_n, \alpha_b\}$ e $N = \{\alpha_n, \beta_n\}$ come gli *eventi casuali* in cui *"pesco una qualsiasi palla dall'urna alfa"* e *"ho pescato una palla nera da una qualsiasi urna"*.
Grazie alle ipotesi iniziali, conosciamo le seguenti probabilità condizionali.

$$p(A) = \frac{1}{2}, P(N|A) = \frac{3}{4}, P(N|\mathcal{C}A) = \frac{1}{2}$$

e usando la definizione di probabilità condizionale si ha

$$P(N \cap A) = \frac{3}{4 \cdot 2} = \frac{3}{8}$$

(il calcolo è analogo per $P(N \cap \mathcal{C}A)$).
Adesso si osserva che

$$N := \{\alpha_n, \beta_n\} = (N \cap A) \cup (N \cap \mathcal{C}A)$$

che ci dà il risultato finale

$$P(N) = P(N \cap A) + P(N \cap \mathcal{C}A) = \frac{5}{8}$$

#Osservazione

Osservazione (avvolte la probabilità condizionale è ininfluente).

Osserviamo che non sempre vale $P(\cdot) \neq P(\cdot|B)$, dato un'evento B tale che $P(B) > 0$. Infatti, vediamo un esempio.

Supponiamo di lanciare due volte una moneta. Se prendiamo A come le volte in cui *esce la testa al secondo lancio* e B come le volte in cui *esce la testa al primo lancio*, notiamo che $P(A) = P(A|B)$ e che $P(B) = P(B|A)$.
Infatti, $|A| = |B| = 2$ e $|A \cap B| = 1$.
Formalizzeremo questo concetto sotto forma di *eventi indipendenti* (1).

B2. Formule relative alla Probabilità Condizionale

Formula di Disintegrazione e delle Probabilità Totali

Formule utili per calcolare probabilità senza dover ricondurre al modello probabilistico: formula di disintegrazione, formula delle probabilità totali per casi generali e per dicotomie.

---

0. Voci correlate

- Probabilità Condizionale

1. Formula di disintegrazione e delle probabilità totali

#Osservazione

Osservazione (lo scopo di queste formule).

Supponiamo di conoscere delle *probabilità condizionali*; se voglio calcolare un qualsiasi altro dato, come ad esempio la probabilità di un evento dalle probabilità condizionali, di solito bisognava ricostruire l'intero *modello probabilistico* $(\Omega, \mathcal{A}, p)$ (1).

Con le formule che vedremo, ci sarà un metodo per evitare tutto ciò.

#Proposizione

Proposizione (formule di disintegrazione e delle probabilità totali).

Sia $(\Omega, \mathcal{A}, p)$ un *spazio di probabilità*, sia $(B_n)_{n \in N} \subset A$ una *partizione finita o numerabile* di Ω (ovvero una successione finita o infinita di sottoinsiemi disgiunti di Ω tali che la loro intersezione riconduca a Ω stesso; ovvero $\bigcap_n B_n = \Omega$).

Allora valgono le seguenti formule.
i. *Formula di disintegrazione*

$$\forall A \in \mathcal{A}, p(A) = \sum_{n \in N} p(A \cap B_n)$$

ii. *Formula delle probabilità totali*

$$p(B_k) > 0, \forall n \in N \implies p(A) = \sum_{n \in N} p(A|B_n) \cdot p(B_n)$$

#Dimostrazione

DIMOSTRAZIONE della Proposizione 2 (formule di disintegrazione e delle probabilità totali)
Non daremo una dimostrazione completa, bensì un'idea della dimostrazione.
i. Ricordando che $(B_n)_n$ è una partizione di Ω, possiamo scrivere

$$A = A \cap \Omega = A \cap \left(\bigcup_{n \in N} B_n \right) \overset{\text{prop. distributiva}}{\Longrightarrow} A = \bigcup_{n \in N} (A \cap B_n)$$

Adesso, ricordando che $(B_n)_n$ sono tutti disgiunti, posso passare alla probabilità e asserire l'uguaglianza

$$p(A) = p\left(\bigcup_{n \in N} (A \cap B_n) \right) = \sum_{n \in N} p(A \cap B_n)$$

che è la tesi.
ii. Basta ricordarsi della definizione di probabilità condizionale, per cui

$$p(A|B_n) := \frac{p(A \cap B_n)}{P(B_n)} \overset{p(B_n)>0}{\Longrightarrow} p(A \cap B_n) = P(A|B_n)P(B_n)$$

e dato che $(B_n)_n$ formano una partizione per Ω, posso sfruttare la formula di disintegrazione per cui posso passare alla probabilità

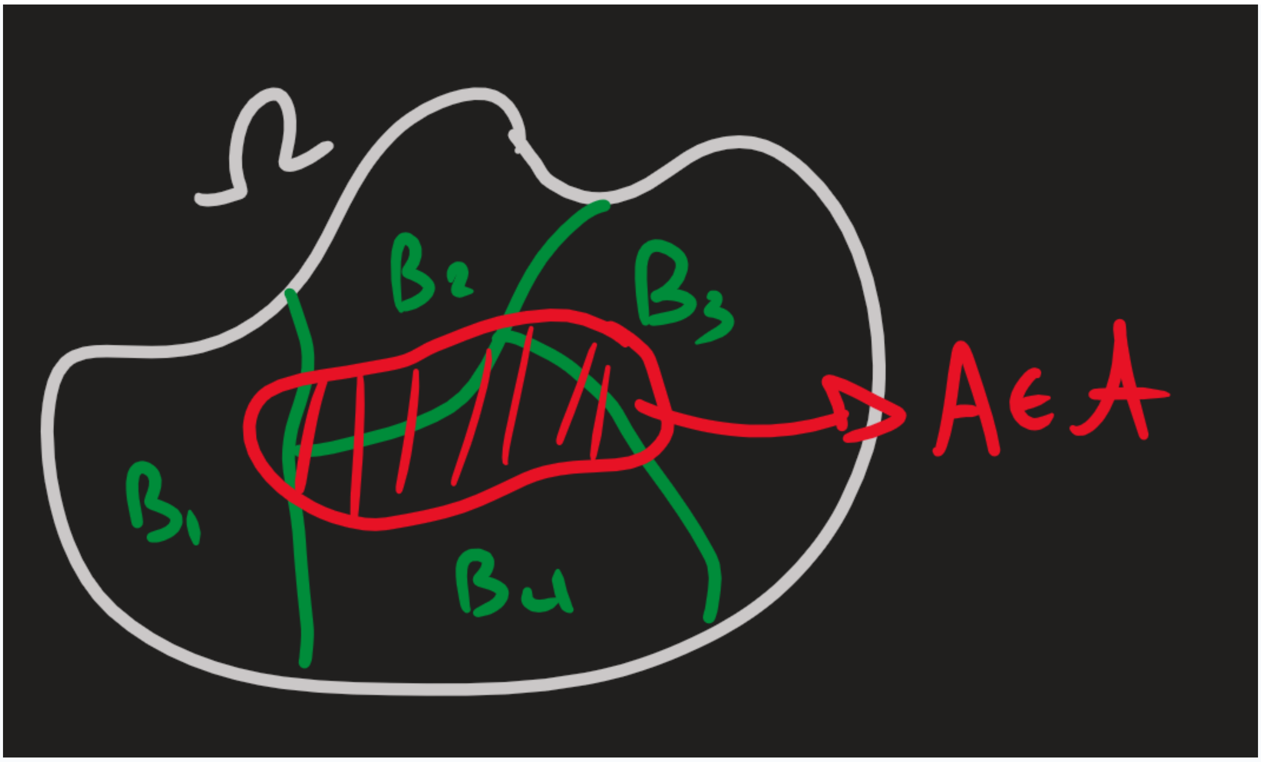
$$p(A) = \sum_{n \in N} p(A \cap B_n) = \sum_{n \in N} p(A|B_n) \cdot p(B_n)$$

che è la tesi. ■

#Osservazione

> Osservazione (interpretazione geometrica della prima formula).
>
> Possiamo dare un'interpretazione geometrica per la *formula di disintegrazione*.
>
> Supponiamo di avere una forma Ω, e di tagliarli in n parti, che poi formeranno B_1, B_2, eccetera...
> Allora per misurare l'area di un segmento A (che rappresenterebbe un evento complesso in $\mathcal{A}$), possiamo semplicemente prendere *tutte le intersezioni* tra A e i B_n, poi per sommarli.

FIGURA 1.1. (*Interpretazione geometrica*)



2. Caso particolare per le dicotomie

#Corollario

Corollario (formula delle probabilità totali per le dicotomie).

Siano $A, B \in \mathcal{A}$ degli eventi che hanno probabilità non-estreme ($A, B \in (0,1)$). Allora vale che

$$p(A) = p(A|B) \cdot p(B) + p(A|\mathcal{C}_\Omega B) \cdot (1 - p(B))$$

B3. Definizione di Eventi Indipendenti

Eventi Indipendenti

Definizione di indipendenza di eventi; prima definizione per due eventi; definizione generalizzata.

---

1. Definizione di due eventi indipendenti

#Definizione

Definizione (due eventi indipendenti).

Siano A, B degli *eventi* in uno *spazio di probabilità*. A, B si dicono *indipendenti* se vale che

$$A, B \text{ indipendenti} : P(A \cap B) = P(A) \cdot P(B)$$

In particolare se $P(B) > 0$, vale che

$$P(A) = P(A|B)$$

oppure se $P(A) > 0$, vale che

$$P(B) = P(B|A)$$

cioè *"la probabilità di A non cambia conoscendo B e viceversa"*.

2. Definizione generalizzata di eventi indipendenti

#Definizione

Definizione (tre eventi indipendenti).

Tre eventi A, B, C in uno spazio di probabilità si dicono *indipendenti* se sono *due a due indipendenti* e se vale

$$p(A)p(B)p(C) = p(A \cap B \cap C)$$

#Definizione

Definizione (n eventi indipendenti).

Una successione di eventi $(B_n)_{n \in N}$ si dicono *indipendenti* se per *ogni sottoinsieme finito* $J \subset N$, con $|J| \geq 2$ abbiamo che

$$\forall J \subset N, p\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} p(B_j)$$

#Osservazione

Osservazione (la necessità di dover aggiungere un'altra condizione).

Notiamo che, quando abbiamo effettuato il passaggio da $n = 2$ a $n = 3$, abbiamo dovuto aggiungere un'ulteriore condizione, ovvero che la *produttoria degli eventi devono valere la probabilità dell'intersezione degli eventi*. Giustifichiamo questa necessità col seguente *esempio*.

Supponiamo che nel dipartimento di Matematica, Informatica (e Geoscienze) di Trieste *tre* matematici, in particolare un analista, un algebrista e uno statistico, stanno giocando a dadi, *ognuno lanciando un dado* (trascurando dettagli inutili, stiamo lanciando *tre dadi*).

Contrassegno l'evento A come *"il primo e il secondo dado hanno lo stesso risultato"* e l'evento B come *"il secondo e il terzo dado hanno lo stesso risultato"*. Gli eventi A, B sono indipendenti, dato che $p(A \cap B) = p(A)p(B) = \frac{1}{6} \cdot \frac{1}{6}$.
Adesso contrassegniamo l'evento C come *"il primo e il terzo dado hanno lo stesso risultato"*: in questo caso, vediamo che A, B, C rimangono *a due a due indipendenti*.

Tuttavia, A non è più indipendente da $B \cap C$ (ovvero che capitino assieme); se $B \cap C$ accade, allora per forza A dev'essere vero! Ovvero $p(A|B \cap C) = 1$.

Di conseguenza, possiamo dire che A, B, C *non sono indipendenti*, nonostante il fatto che siano comunque *a due a due indipendenti*.

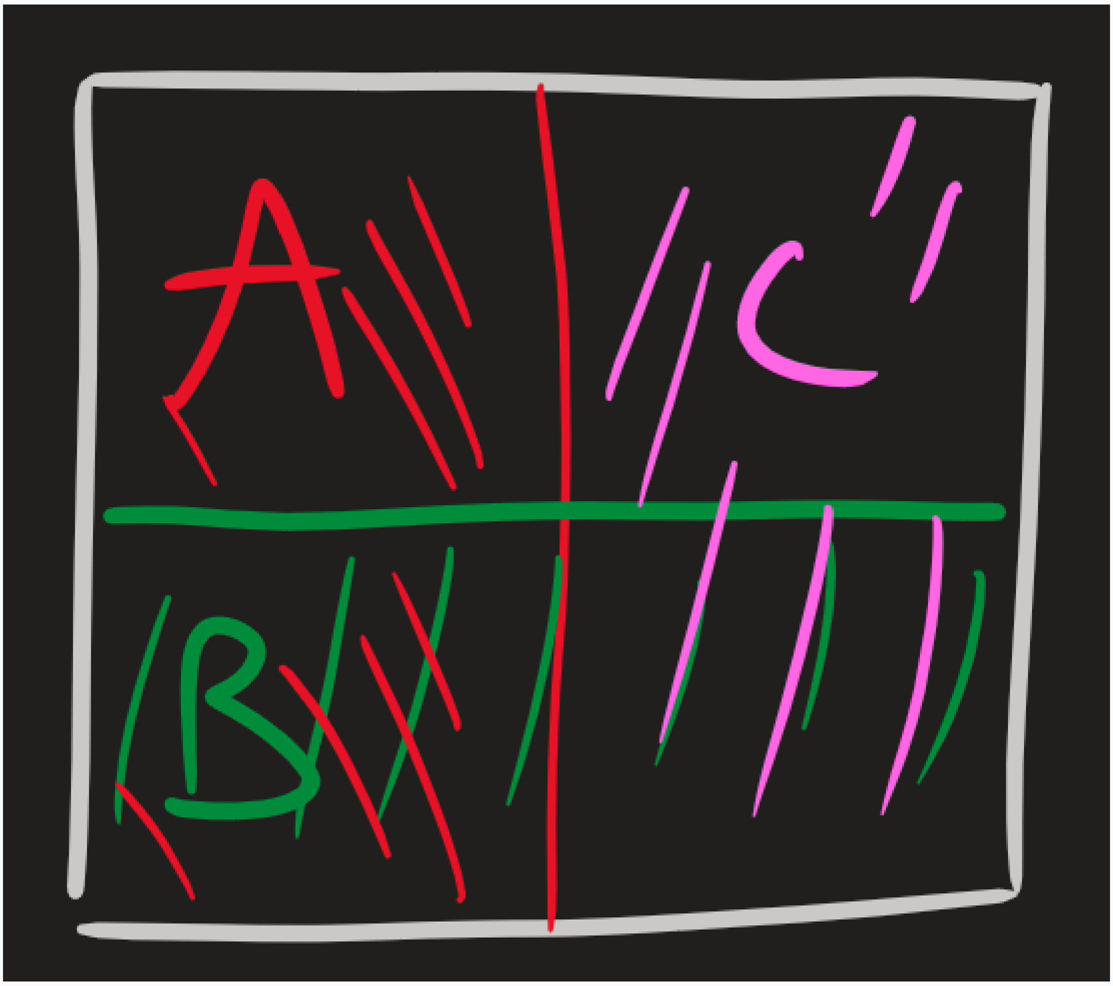
3. Le Non-Proprietà degli Eventi

#Osservazione

Osservazione (la indipendenza non gode della proprietà transitiva).

> Notiamo che l'*indipendenza tra eventi* non gode della *proprietà transitiva*. Ad esempio, se A è *indipendente* da B e B è *indipendente* da C, non deve vuol dire che A è *indipendente* da C.
>
> Interpretiamo questa cosa un una visione geometrica: consideriamo i tre eventi come dei *"segmenti del quadrato"* con area 0.5. Li poniamo in una maniera tale che $p(A \cap B) = p(B \cap C) = \frac{1}{4}$ (ovvero A, B e B, C sono *indipendenti*) ma $p(A \cap C) = 0$, che rende A, B non-indipendenti.

FIGURA 3.1. (*La non transitività dell'indipendenza tra eventi*)



#Proposizione

Proposizione (la condizione aggiuntiva fornisce una condizione sufficiente).

> Notiamo che la *"condizione generale"* per n eventi non dice nulla per l'*indipendenza a j a j* di questi eventi.
> Ovvero,

$$p\left(\bigcap_{n\in N} B_n\right) = \prod_{n\in N} p(B_n)$$

non deve *necessariamente implicare*

$$\forall J \subset N, p\left(\bigcap_{j\in J} B_j\right) = \prod_{j\in J} p(B_j)$$

#Dimostrazione

DIMOSTRAZIONE della Proposizione 6 (la condizione aggiuntiva fornisce una condizione sufficiente)

Si può dimostrare il caso $n = 3$ con un controesempio: tenendo conto del *lancio dei tre dadi* con gli eventi

$$\begin{gathered}A = \{(x, y) \in \Omega : y = 1, 2, 5\} \\ B = \{(x, y) \in \Omega : y = 4, 5, 6\} \\ C = \{(x, y) \in \Omega : x + y = 9\}\end{gathered}$$

provando che A, B, C *non sono a due a due indipendenti*, ma vale che

$$p(A \cap B \cap C) = p(A)p(B)p(C)$$

confutando la contronominale. ■

4. Il complementare degli Eventi Indipendenti

#Osservazione

Osservazione (i complementari di due eventi indipendenti rimangono indipendenti).

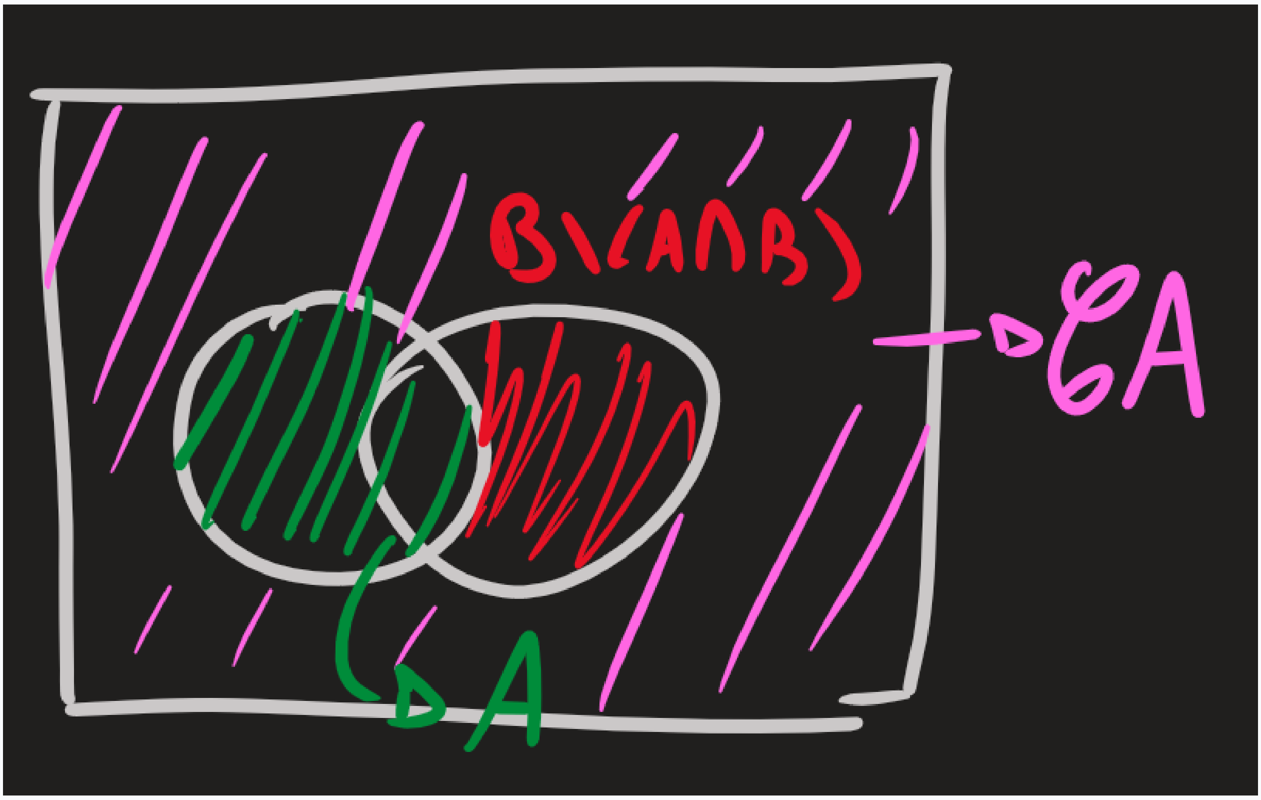
Suppongo che due eventi A, B sono *indipendenti*. Cosa succede se passo al complementare? Rimangono indipendenti o no?

Spoiler: rimangono indipendenti, dal momento che possiamo pensare $\mathcal{C}A \cap B$ (o viceversa) come $B \diagdown (A \cap B)$. (ovvero rimuoviamo da B la parte in comune con A, *figura 4.1.*). Da qui seguono i seguenti passaggi:

$$
\begin{aligned}
p(\complement A \cap B) = p(B \setminus (A \cap B)) &= p(B) - p(A \cap B) \\
&= p(B) - p(A)p(B) \\
&= p(B)(1 - p(A)) \\
&= p(B)p(\complement A)
\end{aligned}
$$

che sarebbe la tesi. Generalizzando questa osservazione in più piani, abbiamo le seguenti proposizioni.

FIGURA 4.1. (*Idea del primo passaggio*)



#Proposizione

Proposizione (condizioni equivalenti tra eventi indipendenti).

Siano A, B degli *eventi indipendenti*. Allora sono equivalenti

$$
A \text{ ind. } B \iff \complement A \text{ ind. } B \iff A \text{ ind. } \complement B \iff \complement A \text{ ind. } \complement B
$$

(*N.B. La scritta* ind. *sta per "... indipendente da ..."*)

#Proposizione

Proposizione (scegliendo i complementari di alcuni eventi indipendenti rimangono indipendenti).

Sia $(A_n)_{n \in N}$ una famiglia di *eventi indipendenti*. Sia $J \subset N$ un *sottoinsieme di indici* qualsiasi.
Definiamo la nuova famiglia di eventi $(B_n)_{n \in N}$ come

$$B_i = \begin{cases} A_i, i \in J \\ \complement A_i, i \notin J \end{cases}$$

(*a parole stiamo scegliendo alcuni elementi e li stiamo convertendo in loro complementari*)
Segue che anche $(B_n)_{n \in N}$ è una *famiglia di eventi indipendenti*.

B4. Formula di Bayes

Formula di Bayes

Formula di Bayes (tre versioni): enunciato, dimostrazione. Interpretazione della terza versione. Osservazione sul test del COVID-19: quello che conta è il rapporto di probabilità condizionali su probabilità generali.

---

0. Voci correlate

- Probabilità Condizionale

1. Formula di Bayes

#Teorema

Teorema (formula di Bayes).

Sia $(\Omega, \mathcal{A}, p)$ uno *spazio di probabilità*, e siano A, B degli eventi.
Allora valgono le seguenti formule.
i. *Formula di Bayes*

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

ii. *Formula di Bayes per le dicotomie*

$$p(A), p(B) > 0 \implies p(B|A) = \frac{p(A|B)p(B)}{p(A|B)p(B) + p(A|\mathcal{C}B)p(\mathcal{C}B)}$$

#Corollario

Corollario (formula delle probabilità delle cause).

Sia $(\Omega, \mathcal{A}, p)$ uno *spazio di probabilità*, e sia $(B_n)_{n \in N}$ una *partizione* di Ω, tale che $p(B_n) > 0, \forall n$.
Allora vale che

$$p(B_i|A) = \frac{p(A|B_i)p(B_i)}{p(A)} = \frac{p(A|B_i)p(B_i)}{\sum_{n \in N} p(A|B_n)p(B_n)}$$

#Dimostrazione

DIMOSTRAZIONE del Teorema 1 (formula di Bayes).
Basta osservare che, per definizione, ho

$$p(B|A)p(A) = p(A \cap B) = p(B \cap A) = p(A|B)p(B)$$

da cui posso ricavare la tesi. ■

#Osservazione

Osservazione (interpretazione della formula di Bayes).

Prendendo la *terza versione della formula di Bayes* (ovvero Corollario 2 (formula delle probabilità delle cause)), posso interpretare gli eventi della partizione $(B_n)_n$ come le *"cause"* al verificarsi dell'evento A. Quindi, la formula enunciata ci permette di calcolare che, quando ho l'evento A, è stato proprio l'evento B_i a causare.

2. Osservazione sulla formula di Bayes

#Osservazione

Osservazione (osservazione sull'importanza dei valori).

Osserviamo una caratteristica dedotta dalla *formula di Bayes*.

Dalla formula abbiamo

$$p(B|A) = \frac{p(A|B)p(B)}{p(A|B)p(B) + p(A|\mathcal{C}B)p(\mathcal{C}B)}$$

da cui discende, per valori di $p(B)$ non nulli,

$$p(B|A) = \frac{p(A|B)}{p(A|B) + \frac{p(A|\mathcal{C}B)}{p(B)}p(\mathcal{C}B)}$$

Notiamo che il valore $p(B|A)$ tende a 1 per valori più piccoli di

$$\rho = \frac{p(A|\mathcal{C}B)}{p(B)}$$

che sarebbe il rapporto tra la probabilità di avere *un'evento A che non* sia causato da B e la probabilità di B.
Al contrario, il valore $p(B|A)$ tende a scendere per valori di ρ più grandi.

Infine, possiamo interpretare questo valore ρ come *"una costante di proporzionalità per la conversione da $p(A|B)$ a $p(B|A)$"*

#Osservazione

Osservazione (esempio del test di SARS-COVID19).

Applichiamo quest'osservazione con un esempio.

Supponiamo di avere un test clinico per determinare la presenza della presenza del *coronavirus*. Etichettiamo A come gli eventi per cui ho un *paziente positivo al test* e B come gli eventi per cui ho *un paziente effettivamente malato*.

Facciamo finta di sapere che il *test risulta positivo* al 99% col *virus presente*, e il *test risulta positivo* al 2% col *virus assente* (ovvero i falsi positivi). Inoltre, la probabilità di *beccarsi il virus* è di 0.2%
Questo sembra un buon test, no? Vediamo

Traducendoli nel nostro linguaggio, abbiamo

$$p(A|B) = 0.99, p(A|\mathcal{C}B) = 0.02, p(B) = 0.002$$

Adesso, per ricavare la *probabilità di avere un malato effettivo da un test positivo*, usiamo l'osservazione appena ricavata:

$$p(B|A) = \frac{0.99}{0.99 + \frac{0.02}{0.002}0.998} = 0.09 \approx 9\%$$

Accipicchia! Ma cos'è successo? Abbiamo solo la probabilità del 9% per determinare se un paziente è *effettivamente malato* da un *test positivo*?

Infatti, analizzando la situazione, vediamo che il valore ρ è *molto grande*, ovvero

$$\rho = 10$$

che *"porta giù"* il valore $p(B|A)$.

Quindi, se volessimo un *buon risultato*, dovremmo avere dei valori $p(A|\mathcal{C}B)$ (ovvero la probabilità dei falsi positivi) *molto più piccoli* rispetto a $p(B)$.

$$p(A|\mathcal{C}B) \ll p(B) \sim \rho < 1$$

C. LO SCHEMA DELLE PROVE RIPETUTE

C1. Definizione dello Schema di Bernoulli

Definizione di Schema delle Prove Indipendenti

Definizione di Schema delle Prove Indipendenti.

---

0. Voci correlate

- Eventi Indipendenti

1. Esempio Preliminare

#Esempio

Esempio (il lancio della moneta).

Facciamo il seguente esperimento: lanciamo n volte una moneta, per la quale ha una probabilità $0 < q < 1$ di ottenere *testa* (quindi stiamo rinunciando all'*equiprobabilità* dell'esperimento; inoltre supponiamo che esistano *solo due esiti*). Supponiamo che *tutti i risultati dei singoli lanci siano indipendenti tra di loro*.

Consideriamo quindi lo spazio campionario degli eventi elementari come le n-uple di *zeri e uni*, ovvero

$$\Omega = \{0, 1\}^n$$

Adesso definiamo la famiglia degli eventi complessi $(A_k)_{k \leq n}$ definendo il singolo evento complesso A_i come *"la probabilità di uscire* 1 *(o testa) al lancio* i*-esimo"*.
Visto che la probabilità di ottenere testa in un singolo lancio è q, possiamo porre

$$p(A_i) = q, \forall i \leq n$$

Adesso prendiamo un $\omega \in \Omega$, espresso come $\omega = \{\omega_1, \ldots, \omega_n\}$ con ω_i la componente i-esima di questa n-upla.
Poi prendiamo l'insieme degli indici $\{1, \ldots, n\}$ e lo dividiamo in due sottoinsiemi I_0, I_1 definiti come

$$I_0 = \{1 \leq i \leq n : \omega_i = 0\}, I_1 = \{1 \leq i \leq n : \omega_i = 1\}$$

Possiamo quindi vedere ogni elemento $\omega \in \Omega$ come l'intersezione

$$\{\omega\} = \bigcap_{i \in I_1} A_i \cap \bigcap_{i \in I_2} \mathcal{C}A_i$$

Per le ipotesi iniziali, sappiamo che gli eventi delle intersecatorie sono *indipendenti*, dunque posso passare alla probabilità

$$p(\{\omega\}) = \prod_{i \in I_1} p(A_i) \cdot \prod_{i \in I_2} p(\mathcal{C}A_i) = q^{|I_1|}(1-q)^{|I_2|}$$

Notando che per ottenere $|I_1|$ basta sommare tutti gli elementi di ω, possiamo scrivere

$$|I_1| = \sum_{\omega_i \in \omega} \omega_i; \, |I_0| = n - |I_1|$$

il ché ci porta al risultato finale

$$\boxed{p(\{\omega\}) = q^{\sum_{\omega_i \in \omega} \omega_i} (1-q)^{n - \sum_{\omega_i \in \omega} \omega_i}}$$

2. Definizione Formale dello schema di Bernoulli

#Definizione

Definizione (schema delle prove indipendenti / ripetute / di Bernoulli).

Sia $(\Omega, \mathcal{A}, p)$ uno *spazio di probabilità*, formato da n prove *ripetute ed indipendenti*, che possono avere *solo due esiti possibili* (etichettati con $1, 0$ e detti tradizionalmente come *"successo"* e *"insuccesso"*).

Questo spazio di probabilità si dice *"schema delle prove indipendenti* (o *ripetute* o *di Bernoulli*).

C2. Proprietà dello Schema di Bernoulli

Proprietà dello Schema delle Prove Indipendenti

Risultati relativi allo schema delle prove indipendenti: formula per calcolare la probabilità di un evento elementare; formula per calcolare la probabilità del primo successo al lancio j; formula per calcolare k probabilità; la scimmia di Borel

---

0. Voci correlate

- Definizione di Schema delle Prove Indipendenti

1. Risultati Relativi allo Schema di Bernoulli

#Proposizione

Proposizione (risultati relativi allo schema delle prove indipendenti).

Sia $(\Omega, \mathcal{A}, p)$ uno *schema delle prove indipendenti*. Si ha che
i. *formula per calcolare la probabilità ogni elemento campionario*

$$p(\{\omega\}) = q^{\sum_{\omega_i \in \omega} \omega_i} (1-q)^{n - \sum_{\omega_i \in \omega} \omega_i}$$

ii. *formula per calcolare il primo successo al j-esimo esperimento*

$$P(C_j) = (1-q)^{j-1} q$$

iii. *formula per calcolare k successi*

$$p(C_k) = \binom{n}{k} q^k (1-q)^{n-k}$$

#Dimostrazione

DIMOSTRAZIONE del Proposizione 1 (risultati relativi allo schema delle prove indipendenti)
N.B. L'importanza di questo teorema è quello prima di capire il ragionamento, poi la formula. Quindi leggere bene!
i. Già dimostrata in Esempio 1 (il lancio della moneta)
ii. Formalizziamo l'evento C_j come

$$C_j = \{\omega \in \Omega : \omega_1 = \ldots = \omega_{j-1} = 0 \land \omega_j = 1\}$$

allora prendo l'intersecatoria e passo alla probabilità

$$C_j = \bigcap_{i \leq j-1} \complement A \cap A_j \implies p(C_j) = (1-q)^{j-1} q$$

che è la tesi. Notiamo la *non-dipendenza* da n di questa formula.
iii. Prendiamo C_k come l'evento in cui ho k componenti di ω per cui *valgono* 1. Associo ad un dato $\omega \in C_k$ con la *famiglia degli indici* H tali per cui $\omega_{h \in H} = 1$. Chiaramente $|H| = k$ (ovvero il numero degli indici per cui ho k successi è

proprio k stesso). Quindi posso creare una corrispondenza *biunivoca* tra $\omega \in C_k$ e H. Posso provare in questo modo che

$$|C_k| = \binom{n}{k}$$

(*a parole, sto prendendo la combinazione di k elementi su n elementi*)
Inoltre, notiamo che per la *formula i.* sappiamo che per un $\omega \in C_k$

$$p(\{\omega\}) = q^k(1-q)^{n-k}$$

allora ho

$$p(C_k) = \sum_{\omega \in C_k} p(\{\omega\}) = \binom{n}{k} q^k (1-q)^{n-k}$$

che è la tesi. ■

#Osservazione

> Osservazione (la somma della probabilità degli eventi elementari è 1).
>
> Osserviamo che per induzione è possibile provare la somma
>
> $$\sum_{\omega \in \Omega} p(\{\omega\}) = 1$$

2. La Scimmia di Borel

#Osservazione

> Osservazione (la probabilità di avere 0 successi diventa sempre più piccolo).
>
> Notiamo che la probabilità A di avere 0 successi (o n insuccessi) si può scrivere come segue.
>
> $$p(A) = (1-q)^n$$
>
> (per una dimostrazione dettagliata vedere 1, *formula 3*)
> Allora posso prendere il limite

$$\lim_n p(A) = \lim_n (1-q)^n = 0$$

Ma allora per la definizione del limite, qualunque valore piccolo $\varepsilon > 0$ fissato, *deve esistere* un numero di esperimenti $n_0(\varepsilon)$ tale che $1 - q^n < \varepsilon$, quindi ho la probabilità di *avere almeno un successo* del $p(\mathcal{C}A) > 1 - \varepsilon$! Quindi prima o poi c'è sempre la speranza di aver un successo, qualunque valore di $q \in (0,1)$ ho.

Illustriamo questa osservazione col seguente teorema/esempio.

#Esempio

Esempio (la scimmia instancabile di Borel).

Supponiamo di mettere una *scimmia* davanti ad un *laptop*. Riuscirà, pigiando i tasti a caso, a *comporre il primo capitolo di Harry Potter* (che ha circa 5000 caratteri)?

Se prendiamo che ci sono 25 *tasti* sulla tastiera, vediamo che ci sono 25^{5000} disposizioni, che ha la grandezza di 10^{7000}. Quindi, ho la probabilità di avere *almeno un successo* del

$$q = 25^{-5000}$$

Prendiamo $\varepsilon = \frac{1}{10^4}$; per l'osservazione appena vista, abbiamo che esiste un numero di n_0 tentativi tali che ho la probabilità di avere almeno un successo

$$p(\mathcal{C}A) = 1 - \frac{1}{10^4} = 0.9999$$

Quindi ho la probabilità del 99.99% che la scimmia possa comporre il primo capitolo di Harry Potter! Wow! Ma qual è il valore di questo n_0? Vediamo che

$$(25^{-5000})^{n_0} \leq \frac{1}{10^4} \implies n_0 \geq 4 \cdot 25^{5000}$$

che è un numero decisamente molto grande.