

Statistica Biomedica

Davide Corso, Alessandra Nardi

5 novembre 2024

Indice

I	Probabilità	3
1	Definizioni e prime leggi	1
1.1	Il concetto di evento	2
1.1.1	Eventi composti	3
1.2	Sigma Algebra e probabilità assiomatica	5
1.2.1	Probabilità Assiomatica	6
1.2.2	Conseguenze dell'impostazione assiomatica	7
1.3	La Probabilità Totale	8
1.4	Un piccolo spazio di probabilità	9
1.5	La Probabilità Condizionata	11
1.6	Legge delle probabilità composte	14
1.7	Teorema di Bayes	17
2	Variabili aleatorie	27
2.1	Variabili aleatorie discrete	27
2.1.1	La variabile aleatoria Bernoulliana	28
2.1.2	Valore atteso e varianza nel discreto	30
2.1.3	Il modello Binomiale	35
2.1.4	La famiglia Binomiale Negativa	40
2.1.5	Il modello Geometrico	41
2.1.6	Il modello di Poisson	42
2.2	Variabili aleatorie continue	43
2.2.1	Dal discreto al continuo, un passaggio delicato	43
2.2.2	La variabile aleatoria Uniforme	46
2.2.3	Il modello Normale	47
2.2.4	Il modello Esponenziale	50
2.3	Ancora sull'indipendenza	51
2.4	Due leggi importanti	52
2.4.1	La legge dei Grandi Numeri	52
2.4.2	Il Teorema Centrale del Limite	53
3	Esercizi	54

Parte I

Probabilità

Capitolo 1

Definizioni e prime leggi

Iniziamo il nostro percorso parlando di probabilità. Diverse sono le ragioni di questa scelta. Soprattutto nell'ambito della biomedicina si è spesso chiamati a prendere decisioni in condizioni di incertezza. Incerta è spesso la risposta ad una terapia o la sopravvivenza di un paziente a lungo termine. Nei contesti medici, caratteristica non eliminabile degli esperimenti è la presenza di un margine di errore nel risultato finale così come qualsiasi marcatore di una data patologia non è in grado di portare ad una diagnosi certa. In tutti questi casi diventa essenziale essere capaci di quantificare l'incertezza nel modo corretto e solo il linguaggio della probabilità può aiutarci a farlo. Sebbene il concetto di probabilità sia intuitivo, la sua definizione è complessa al punto che nel tempo si sono succedute diverse scuole di pensiero.

La nascita del concetto di probabilità è legata all'ambito dei giochi e delle scommesse. Tra le prime definizioni vogliamo ricordare quella di Pascal secondo cui “la probabilità di un evento è il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili, *purché questi ultimi siano tutti ugualmente probabili*”. Questa definizione, generalmente nota come definizione classica, sebbene tautologica richiamando al suo interno il concetto stesso di probabilità, è immediatamente operativa ed importante logicamente. Sancisce infatti come condizione essenziale per la sua applicabilità che gli eventi osservabili abbiano uguale probabilità di accadere, siano cioè in qualche senso scambiabili. Benché del tutto intuitiva questa condizione viene spesso dimenticata nel linguaggio corrente. Immaginiamo semplicemente un campione di pazienti a cui viene somministrato un farmaco. Possiamo valutare la probabilità di risposta come rapporto tra il numero di pazienti che rispondono ed il totale dei pazienti trattati? È evidente come la condizione di equiprobabilità, naturalmente verificata nei giochi, non è generalmente rispettata nell'ambito sperimentale.

Un primo superamento di questo limite è offerto dalla definizione frequentista. Questa si fonda sull'idea che, in una successione di prove fatte nelle medesime condizioni, la frequenza (relativa) di un evento si avvicina

alla sua probabilità e che l'approssimazione tende a migliorare con l'aumentare del numero delle prove (legge empirica del caso).

Si stabilisce così uno stretto legame tra frequenze relative e probabilità al punto che quest'ultima viene definita come il limite a cui tende la frequenza relativa dell'evento al tendere ad infinito del numero delle prove. Ritroveremo questo legame più volte nell'ambito dell'inferenza statistica dove dovremo imparare a valutare la qualità di questa approssimazione al variare del numero delle prove. Il concetto di limite stabilisce infatti che per qualsiasi numero finito di prove, frequenza relativa e probabilità non coincideranno.

Citiamo infine per la sua rilevanza nell'ambito della teoria della probabilità la definizione soggettiva secondo cui la probabilità è il grado di fiducia che una persona ha nel verificarsi dell'evento. Definizione bella, ampia ma non operativa ed estremamente soggettiva in questa semplice formulazione. Per ovviare a questi limiti sono state introdotte delle condizioni nel cui merito non entreremo. Vedremo infatti come tutte le definizioni proposte sono state superate e riconciliate dalla definizione assiomatica proposta da Kolmogorov nel 1933. Prima di enunciarla dobbiamo però definire il terreno su cui andrà operare la probabilità, ovvero il suo dominio.

1.1 Il concetto di evento

Il calcolo delle probabilità fa riferimento ad un prova o esperimento il cui esito è incerto ma di cui siamo in grado di elencare i possibili risultati. Questo punto è essenziale per circoscrivere i confini della nostra incertezza. Indichiamo con Ω (*Omega*) l'insieme degli *eventi elementari*, cioè di quegli eventi, non scindibili, che osserviamo come diretto risultato del nostro esperimento. La definizione di evento è molto ampia; evento è la risposta positiva ad un farmaco o il semplice fatto che domani piova. Esiste tuttavia una condizione che deve essere rispettata: ci deve essere un momento a posteriori in cui si è in grado di verificare se quell'evento si sia realizzato o meno. In qualche misura l'evento deve mantenere una sua oggettività. Ad esempio il fatto che una persona sia simpatica non può essere considerato un evento poiché mantiene un grado di soggettività. Non avrebbe senso parlare di probabilità rispetto a qualcosa che sarà vero per alcuni e falso per altri allo stesso tempo. Attenzione anche a non rimanere nel generico: il lancio di un dado non è un evento perché non è verificabile, mentre lo è il fatto che, lanciando un dado, esca il numero 3.

Facciamo qualche esempio restando nell'ambito dei giochi che, pur essendo lontano dalla realtà sperimentale, ci aiuterà a capire i primi concetti fondamentali:

- nel lancio di un comune dado, gli eventi osservabili sono i numeri da 1 a 6.

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

- se lanciamo in sequenza due monete

$$\Omega = \{ TT, CC, TC, CT \}$$

- nel lancio successivo di due dadi, così come nel caso precedente, gli eventi osservabili sono coppie di valori. Quindi il risultato del lancio di due dadi è un insieme costituito da vettori di due elementi, dove ogni elemento è un numero compreso tra 1 e 6.

È lasciata allo studente la scrittura dell'insieme Omega. Come cambierebbe Ω se lanciassimo simultaneamente due dadi?

Ω diventerà in futuro l'insieme di tutti i possibili risultati del nostro esperimento, è quindi importante capirne la natura. Come evidenzia l'ultimo esempio, la sua struttura dipende da cosa siamo effettivamente in grado di osservare. Ma non solo. Può accadere che il ricercatore scelga di semplificare la struttura di Ω : ai fini di un dato studio può essere sufficiente riportare se un dato gene sia espresso o meno invece che il livello esatto di espressione misurato. In questo modo Ω passa dall'insieme dei numeri reali positivi ad un insieme di soli due elementi. Una semplificazione importante ma accompagnata da una importante perdita di informazione che deve sempre essere accuratamente valutata.

1.1.1 Eventi composti

Oltre agli eventi semplici potremmo essere interessati ad eventi composti, come per esempio l'unione degli eventi singoli. Immaginiamo di lanciare un dado a 6 facce, e di essere interessati all'uscita di un numero pari; tale interesse è quindi rivolto a uno specifico sottoinsieme di Ω . Ma quando dico “*esce un numero pari*” sto dicendo: *esce 2 oppure 4 oppure 6*. Per comprendere meglio quali operazioni possiamo fare tra eventi osserviamo che questi ultimi sono assimilabili ad insiemi: saranno quindi valide tutte le operazioni definite sugli insiemi. Vediamo in dettaglio le principali

- **Unione** ($A \cup B$) $\rightarrow$ si verifica quando *almeno uno* degli eventi che lo costituiscono, si verifica.

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ oppure } \omega \in B\}$$

Si osservi che formalmente ω è unione di tutti i possibili eventi semplici e poiché uno di essi dovrà necessariamente verificarsi ω viene definito come *l'evento certo*.

- **Intersezione** $(A \cap B) \rightarrow$ si verifica quando si verificano **entrambi** gli eventi.

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ e } \omega \in B\}.$$

- **Complemento** $(A^c) \rightarrow$ contiene tutti gli eventi elementari in Ω tranne gli eventi contenuti nell'insieme A

È evidente che per poter definire il complemento di un evento dovremmo aver preliminarmente definito Ω ($A^c = \Omega \setminus A$) $A^c = \overline{A} = \{\omega \in \Omega : \omega \ni A\}$

Il complemento di Ω è l'insieme vuoto $\emptyset$ (non lo zero!) noto anche come *evento impossibile*.

Esistono poi delle proprietà che saranno importanti nel prosieguo.

- **Implicazione:** $(A \subset B)$ l'evento A è contenuto nell'evento B (gli elementi di A si trovano anche in B).

A è contenuto in B significa che **A implica B** , mentre non è vero che B implica A perché

- se A si verifica, si verifica anche B
- se B si verifica, non necessariamente si verifica anche A .

- **Eventi incompatibili:** due eventi sono incompatibili quando la loro intersezione è l'insieme vuoto ($A \cap B = \emptyset$).

Si tratta di eventi che, non avendo elementi in comune, non possono verificarsi contemporaneamente. Si noti che più eventi sono incompatibili se lo sono a due a due.

- **Eventi necessari:** due eventi sono necessari quando la loro unione è Ω ($A \cup B = \Omega$).

La necessità implica che uno dei due eventi dovrà verificarsi per forza. La definizione si estende naturalmente a più eventi che sono necessari se la loro unione è Ω .

Osserviamo che Ω è l'evento necessario per definizione essendo unione degli tutti gli eventi elementari osservabili (*singoletti*).

Inoltre un insieme di eventi necessari e incompatibili è chiamato *partizione di Ω* .

Ad esempio nel lancio del dado gli eventi pari e dispari costituiscono una partizione di Ω . Il concetto di partizione riveste particolare importanza sia nei metodi di classificazione sia nell'inferenza statistica. In entrambi i casi è utile creare all'interno di Ω dei sottoinsiemi di eventi che siano al loro interno, in qualche senso, simili.

- **Legge di De Morgan:** $(A \cup B)^c = A^c \cap B^c$

Questa legge stabilisce l'esistenza di una relazione, seppure poco intuitiva, tra unione ed intersezione di eventi.

- **Proprietà associativa:** valida sia per l'unione sia per l'intersezione:

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

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

- **Proprietà distributiva:** dell'unione rispetto all'intersezione e viceversa

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

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

- Da queste proprietà segue che

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$$

Questa relazione consente di riscrivere qualsiasi evento A come la parte di A contenuta in B unita alla parte di A non contenuta in B (e quindi contenuta nel suo complemento). Si tratta di una proprietà molto utile nei casi in cui siamo in grado di calcolare la probabilità di un evento solo mettendo questo evento in relazione con un altro. Ad esempio la probabilità di guarigione di un paziente può dipendere in modo sostanziale dal fatto che risponda o meno ad un farmaco.

1.2 Sigma Algebra e probabilità assiomatica

Come descritto in precedenza, il primo passo nel calcolo della probabilità è definire Ω cioè l'insieme degli eventi elementari. Tuttavia potremmo essere interessati a calcolare la probabilità di eventi diversi da questi, noti come *eventi composti*. Abbiamo già citato nel lancio del dado gli eventi *esce un numero pari* e *esce un numero dispari* che non elementi di *Omega*.

Abbiamo allora la necessità di affiancare ad Ω , una sua Algebra A_Ω che contenga tutti gli eventi di interesse, sia semplici sia composti. Per costruirla ricordiamo che la definizione di evento richiede semplicemente che sia a posteriori possibile stabilire con certezza se si sia verificato o meno. Da questa definizione segue che:

- *Omega* è un evento, quindi $\Omega \in A_\Omega$
- se A è un evento, lo è anche il suo complemento $A \in A_\Omega \Rightarrow A^c \in A_\Omega$
- se A e B sono eventi, anche la loro unione è un evento $A, B \in A_\Omega \Rightarrow A \cup B \in A_\Omega$

Queste proprietà definiscono la nostra algebra di insiemi che sarà chiusa sia rispetto all'unione (se due eventi sono in A ci sarà anche la loro unione), sia rispetto alla negazione (se un evento è in A ci sarà anche il suo complemento).

Nel semplice caso del lancio di una moneta:

$$\Omega = \{T, C\}$$

$$A_\Omega = \{\{T\}, \{C\}, \{T \cup C\} = \Omega, \emptyset\}.$$

Appena il numero di eventi elementari in Ω cresce, la corrispondente algebra diventa molto complessa. Si provi per esercizio a costruire l'algebra relativa al lancio di un dado dove $\Omega = \{1, 2, 3, 4, 5, 6\}$.

In alcuni casi la struttura di Ω può non corrispondere agli eventi osservabili ma ad una loro partizione. Per esempio sempre nel caso del lancio del dado potremmo essere interessati semplicemente agli eventi: *minore o uguale a 2* o *maggiore di 2* ($\Omega = \{\text{evento} \leq 2, \text{evento} > 2\}$). Questa scelta se porta da un lato ad una importante semplificazione di Ω e dell'algebra ad esso collegata, implica dall'altro una importante perdita d'informazione: sapremo che è uscito un numero maggiore di 2 ma non più quale valore con esattezza. È chiaro che un livello informativo con $\Omega = \{1, 2, 3, 4, 5, 6\}$ è più preciso ma il suo prezzo è la presenza di strutture più complesse. Questa situazione corrisponde ad un principio generale dell'inferenza statistica dove semplificazioni nella struttura dei dati sono sempre *pagate* da una perdita di parte dell'informazione disponibile. La decisione circa quale livello d'informazione mantenere nella rilevazione dei risultati è un punto fondamentale nel disegno di qualsiasi esperimento e la scelta di rinunciare a parte dell'informazione deve essere sempre valutata con attenzione. Immaginate di misurare il livello di espressione di un dato gene: è preferibile riportare il livello esatto oppure semplicemente se il gene risulta espresso o meno (rispetto ad una prefissata soglia)?

Definito *Omega* e la sua algebra possiamo finalmente occuparci della probabilità degli eventi oggetto di studio. In altri termini possiamo completare la costruzione del nostro **spazio di probabilità**.

1.2.1 Probabilità Assiomatica

Abbiamo già visto diverse possibili definizioni di probabilità, tuttavia quella a cui faremo riferimento nel prosieguo fu introdotta nel 1933 da Kolmogorov. Si tratta di una definizione assiomatica dove la probabilità viene considerata un ente matematico astratto senza chiarirne il suo significato. La probabilità viene infatti definita come una funzione che associa ad ogni evento della classe A_Ω un numero reale e che soddisfa i seguenti assiomi:

- C1: $P(A) \geq 0$ (per ogni elemento $A \in A_\Omega$)

- C2: $P(\Omega) = 1$
- C3: Se $A \cap B = \emptyset$,
allora

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

Esiste in realtà un quarto assioma che tratta situazioni al limite e che omettiamo per semplicità.

Da questa definizione segue che ogni funzione che rispetti queste tre condizioni può essere considerata una misura di probabilità. In particolare le tre definizioni che abbiamo discusso nell'introduzione - classica, frequentista e soggettiva - soddisfano i tre assiomi e possono essere considerate misure di probabilità secondo Kolmogorov.

Dagli assiomi C1-C3 derivano tutte le proprietà dello spazio di probabilità (Ω, A_Ω, P) . Vediamo le prime elementari conseguenze che valgono ovviamente solo per quelle misure di probabilità che rispettano questi assiomi.

In generale nel definire una misura di probabilità si parte assegnando le probabilità agli eventi singoli in Ω . I risultati che vedremo sono essenziali per poter calcolare le probabilità di eventi composti a partire da queste.

1.2.2 Conseguenze dell'impostazione assiomatica

- per ogni evento $A \in A_\Omega$
 $P(\bar{A}) = 1 - P(A)$ **Dim.:** $\bar{A} \cup A = \Omega \implies P(\bar{A}) + P(A) = 1$ per il secondo e terzo assioma
- $P(\emptyset) = 0$
Dim.: $P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0$
- $P(A) \leq 1$ per ogni evento $A \subseteq A_\Omega$
Dim.: $P(A) = 1 - P(\bar{A}) \leq 1$ dove $P(\bar{A}) \geq 0$
- Se $A \subset B$ allora $P(A) \leq P(B)$
Dim.: $B = (B \cap A) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A})$
Dal momento che
 $A \cap (B \cap \bar{A}) = B \cap (A \cap \bar{A}) = B \cap \emptyset = \emptyset$
possiamo applicare il terzo assioma e scrivere
 $P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$
dove l'ultimo passaggio è spiegato dal fatto che, in base al primo assioma, $P(B \cap \bar{A}) \geq 0$.

1.3 La Probabilità Totale

Il terzo assioma può essere esteso al caso di $A_1, \dots, A_n$ eventi a due a due incompatibili (additività finita). Si dimostra infatti che $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Vediamo ora come fare se si è interessati a calcolare la probabilità dell'unione tra due insiemi che non sono incompatibili poiché la loro intersezione non è vuota.

$$\begin{aligned} &\text{Per ogni coppia di eventi } A \text{ e } B \in \mathcal{A}_\Omega \\ &P(A \cup B) = P(A) + P(B) - P(A \cap B) \end{aligned}$$

In questa legge l'ultima sottrazione è necessaria poiché se sommassimo semplicemente le due probabilità l'intersezione verrebbe considerata due volte.

Questo risultato è vero anche nel caso di insiemi disgiunti: basta osservare che $P(A \cap B) = 0$.

Quindi il terzo assioma rientra in questo risultato anche se dal punto di vista formale lo precede. Vedremo infatti come la dimostrazione di questo teorema richieda anche il terzo assioma.

Dim.: Iniziamo scrivendo

$$(A \cup B) = A \cup (B \cap A^c)$$

Importante: *l'evento $(B \cap A^c)$ si verifica solo se si verifica A^c . Ma se si verifica A^c non si verifica l'evento A . Quindi $(B \cap A^c)$ è **incompatibile** con A .*

Notiamo che gli elementi che decidono questa incompatibilità sono due: l'intersezione che richiede il verificarsi di entrambi gli eventi e la presenza di A^c . Questi due elementi insieme rendono l'evento $(B \cap A^c)$ incompatibile con l'evento A .

Quindi data la loro incompatibilità per il terzo assioma

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

Per arrivare alla formula che vogliamo dimostrare, ovvero $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, dobbiamo ancora dimostrare che:

$$P(B \cap A^c) = P(B) - P(B \cap A) \text{ (per sostituzione)}$$

Partiamo dal fatto che

$$B = (B \cap A) \cup (B \cap \bar{A})$$

però per verificarsi l'evento $(B \cap A^c)$ bisogna che si verifichi A^c , ma se si verifica A^c non si verifica A e quindi non si verifica l'evento $(B \cap A)$. Quindi $(B \cap A)$ è **incompatibile** con $(B \cap A^c)$.

Ora, trasformando tutto in probabilità e esplicitando $P(B \cap A^c)$ possiamo scrivere

$$P(B) = P(B \cap A) + P(B \cap A^c)$$

$$P(B \cap A^c) = P(B) - P(B \cap A)$$

quindi

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

□

Questo risultato può essere esteso al caso di tre eventi $A, B, C \in A_\Omega$
 $P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(A \cap B) + P(A \cap C) + P(B \cap C)] + P(A \cap B \cap C)$

Lasciamo alla vostra intuizione l'estensione a più eventi.

Dovremo ricordarci di questo teorema tutte le volte che parliamo della probabilità che si verifichi almeno uno di due (o più) eventi. La probabilità che almeno uno di due geni sia un marcatore di una data patologia non è la somma delle singole probabilità a meno che i due geni non possano esserlo contemporaneamente. La probabilità che un paziente risponda ad almeno uno di due trattamenti non è la somma delle singole probabilità a meno che non siamo in grado di escludere che il paziente possa rispondere ad entrambi.

Abbiamo visto come il calcolo della probabilità di una unione di eventi coinvolga la loro intersezione. Tuttavia calcolare la probabilità di una intersezione può essere complesso e dobbiamo iniziare ad occuparcene.

1.4 Un piccolo spazio di probabilità

Riprendiamo l'esempio del lancio di un dado e cominciamo a costruire uno spazio di probabilità.

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

A partire da Ω costruiamo la sua algebra A_Ω che comprende tutti gli eventi singoli (*singoletti*), tutte le unioni di due eventi elementari, tutte le unioni di tre eventi elementari, tutte le unioni di quattro eventi elementari, tutte le unioni di cinque eventi elementari, Ω ed il suo complemento cioè l'insieme vuoto (è lasciata allo studente la formalizzazione di tale insieme).

Per definire la misura di probabilità P , e quindi la probabilità di ciascun elemento di A_Ω , partiamo dai singoletti. Questo in genere è sufficiente perché tutte le altre probabilità degli eventi composti si potranno ricavare applicando le leggi viste finora. Se il dado è bilanciato tutti gli eventi elementari avranno la stessa probabilità di essere osservati. Potremo quindi applicare la definizione classica (o definizione di Pascal) di probabilità secondo cui la probabilità di ciascuno dei singoletti sarà $1/6$.

Consideriamo ora i seguenti eventi:

$$\mathbf{A} = \{2 \cup 3\}$$

$$\mathbf{B} = \{1 \cup 3 \cup 5\}$$

Per calcolare la probabilità della loro unione possiamo seguire due strade.

Possiamo osservare che $(A \cup B) = \{1 \cup 2 \cup 3 \cup 5\}$ ed applicare la definizione classica $\frac{\text{casi favorevoli}}{\text{casi totali possibili}} = \frac{4 \text{ casi favorevoli}}{6 \text{ casi totali possibili}}$. Osserviamo tuttavia che *se i casi possibili non fossero equiprobabili, la definizione classica non sarebbe più applicabile.*

In generale la via maestra è quella di ricondurci alle probabilità degli eventi elementari (che ricordiamo sono per definizione incompatibili) applicando le leggi viste. Avremo

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = [P(2) + P(3)] + [P(1) + P(3) + P(5)] - P(3) = [1/6 + 1/6] + [1/6 + 1/6 + 1/6] - 1/6 = 4/6.$$

Questa formula può apparire ridondante ma è l'unica che potremo utilizzare se gli eventi elementari non fossero equiprobabili. Immaginiamo che il dado sia stato truccato per favorire l'uscita del numero 3. Sia Ω sia la sua algebra A_Ω restano immutate perché gli eventi osservabili sono gli stessi. Cambia invece la definizione di P .

Supponiamo sia: $P(3) = 1/2, P(1) = P(2) = P(4) = P(5) = P(6) = 1/10$.

Nel definire queste probabilità dobbiamo ricordare che la loro somma dovrà essere 1 (perché?).

Adesso utilizzare la definizione classica sarebbe sbagliato. Avremo infatti

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = [P(2) + P(3)] + [P(1) +$$

$$P(3) + P(5)] - P(3) = [1/10 + 1/2] + [1/10 + 1/2 + 1/10] - 1/2 = 6/10 + 7/10 - 5/10 = 8/10 \neq 4/6$$

1.5 La Probabilità Condizionata

Il concetto di probabilità non ha una natura statica ma varia al variare delle nostre conoscenze. Durante un esperimento possono intervenire elementi nuovi che modificano la nostra incertezza sul risultato finale. La definizione che andiamo ad illustrare ci insegna come rivedere la nostra misura di probabilità alla luce di un evento B che si è verificato.

Sia (Ω, A_Ω, P) il nostro spazio di probabilità e sia B un evento in $A_{\Omega'}$ con $P(B) > 0$. Definiamo probabilità dell'evento A dato B , cioè probabilità dell'evento A sapendo che l'evento B si è verificato, come:

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

Stiamo di fatto definendo una nuova misura di probabilità sulla base di quella originaria.

Dobbiamo quindi verificare che la definizione sia corretta, cioè che questa nuova misura rispetti i tre assiomi di Kolmogorov. Vediamo in dettaglio:

- Presi un insieme qualsiasi $A \in A_\Omega$ $P(A|B) \geq 0$
Segue dalla definizione poiché $P(B) > 0$ per ipotesi e $P(A \cap B) \geq 0$ dal momento che la misura originaria P è una probabilità e come tale rispetta il primo assioma.

- $P(\Omega|B) = 1$

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

- Presi due insiemi A e $C \in A_\Omega$ con $(A \cap C) = \emptyset$
 $P((A \cup C)|B) = P(A|B) + P(C|B)$

$$\begin{aligned} P((A \cup C)|B) &= \frac{P((A \cup C) \cap B)}{P(B)} = \frac{P((A \cap B) \cup (C \cap B))}{P(B)} = \\ &= \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} = P(A|B) + P(C|B) \end{aligned}$$

Dalla definizione di probabilità condizionata segue in modo naturale il concetto di indipendenza (in probabilità).

Siano A e B due insiemi $\in A_\Omega$, se

$$P(A|B) = P(A)$$
 allora gli eventi A e B sono *indipendenti*.

Due eventi sono indipendenti se sapere che B si è verificato non modifica la probabilità che si verifichi A o, in altri termini, ***la conoscenza di un evento che non modifica l'incertezza sull'altro***. Vedremo come, da un punto di vista *statistico*, l'indipendenza implica che l'evento B non è informativo su A .

Poiché *il concetto di indipendenza è un concetto simmetrico*, allora è vero anche che

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

Ricordando la definizione di probabilità condizionata, se A e B sono eventi indipendenti potremo scrivere $P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$.

Da cui

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

Questa uguaglianza, molto utile quando si vuole verificare l'indipendenza di due eventi, è preferita da alcuni come definizione di indipendenza per la sua naturale simmetria. Torneremo a breve sul concetto di indipendenza.

Per comprendere l'idea di probabilità condizionata torniamo a giocare con i dadi ma questa volta immaginiamo di lanciarne due. Cominciamo con l'osservare che il risultato del lancio di due dadi è una coppia di numeri o analogamente un vettore di due elementi dove ciascuno dei due elementi è un numero intero da 1 a 6.

Per capire quali vettori contenga Ω dobbiamo riflettere su come vengono lanciati i dadi.

Il vettore $(1,2)$ sarà infatti distinguibile dal vettore $(2,1)$ soltanto se i dadi vengono lanciati in sequenza (o simultaneamente ma sono di colori diversi). Se si lanciano contemporaneamente due dadi *identici*, allora i due eventi $(1,2)$ e $(2,1)$ non saranno distinguibili e quindi coincideranno. Ovviamente anche nel caso di due lanci successivi possiamo decidere di non essere interessati all'ordine e considerare equivalenti le coppie $(1,2)$ e $(2,1)$ anche se potenzialmente saremmo in grado di distinguerle.

Notiamo che in generale il modo in cui viene condotto un esperimento contribuisce a determinare lo spazio dei possibili risultati. Tuttavia è sempre il ricercatore a decidere il grado di dettaglio nell'osservazione degli eventi.

Ipotizziamo di lanciare i due dadi in sequenza e di tener conto dell'ordine in cui vengono osservati i numeri. Avremo

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

Se indichiamo con $\Omega' = \{1, 2, 3, 4, 5, 6\}$ si è soliti esprimere Ω come lo spazio prodotto $\Omega = \Omega' \times \Omega'$ che conterrà 36 possibili eventi elementari. La classe A_Ω si complica (è lasciata allo studente la sua costruzione), proviamo ad accennarla

$$A_\Omega = \{(1, 1), (1, 2), \dots, \{(1, 1) \cup (1, 2)\}, \dots\}$$

Per definire una opportuna misura di probabilità osserviamo come tutte le coppie di valori condividano la stessa probabilità di essere osservate dal momento che i due lanci sono indipendenti ed il dado non è truccato.

Possiamo allora applicare la definizione classica di probabilità in base alla quale ogni singolo evento avrà probabilità $1/36$ di verificarsi.

Costruito lo spazio di probabilità relativo alla nostra prova ipotizziamo di sapere che al primo lancio del dado è uscito il numero 2. Questo nuovo elemento modifica lo spazio dei possibili risultati che si ridurrà a

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

con la relativa sigma algebra

$$A_{\Omega''} = \{(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), \{(2, 1) \cup (2, 2)\}, \{(2, 1) \cup (2, 3)\}, \dots, \Omega'', \emptyset\}$$

Ora gli eventi osservabili sono diventati 6. Se indichiamo con B l'evento "esce 2 al primo lancio" notiamo come Ω'' coincida proprio con B, l'evento noto. Sapere che B si è verificato ha modificato lo spazio di probabilità originario, limitando la nostra incertezza agli elementi contenuti in B. Per poter utilizzare la definizione classica di probabilità dovremmo dimostrare preliminarmente che gli eventi in Ω'' sono equiprobabili. La definizione di probabilità condizionata ci viene in aiuto consentendoci di modificare la probabilità di qualsiasi evento senza dover costruire formalmente un nuovo spazio di probabilità. Sia $A = (2, 6)$. La sua probabilità *a priori* è $1/36$. Noto l'evento B avremo

$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{6}{36}} = 1/6$. La probabilità a posteriori di A sapendo che B si è verificato.

1.6 Legge delle probabilità composte

Dalla definizione di probabilità condizionata possiamo ricavare quella che è nota come legge delle probabilità composte:

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

Nel caso di indipendenza, essendo $P(A|B) = P(A)$ ritroviamo l'uguaglianza $P(A \cap B) = P(A) \cdot P(B)$. Questa legge ci consente di calcolare la probabilità dell'intersezione di due eventi anche nel caso che siano dipendenti ed ha delle implicazioni importanti, che proviamo a comprendere tornando alle nostre urne.

Lo schema classico prevede un'estrazione casuale di palline colorate (con colori diversi) da un'urna. Gli elementi fondamentali, che variano da esperimento ad esperimento, sono la composizione dell'urna e le modalità di estrazione. Resta invece costante il principio che tutte le palline hanno la stessa probabilità di essere estratte e quindi la possibilità di utilizzare la definizione classica di probabilità. Ritroveremo una situazione simile nell'inferenza statistica quando andremo ad estrarre in modo casuale un campione di individui da una data popolazione in modo da garantire che ciascun individuo abbia la stessa probabilità di essere estratto. La differenza sostanziale è che nell'inferenza statistica la composizione della popolazione non è nota e conoscerne alcune caratteristiche attraverso l'osservazione del campione è l'obiettivo.

Per ora immaginiamo di avere un'urna composta come segue:

- 10 palline in totale, tutte equiprobabili ad ogni estrazione, di cui
 - 4 bianche
 - 6 azzurre

La nostra prova consiste nell'estrazione successiva di due palline senza che la prima venga rimessa nell'urna. *Quale è la probabilità che la seconda pallina estratta sia azzurra?* A volte si considera impossibile valutare questa probabilità senza sapere quale colore aveva la prima pallina estratta. In verità le leggi della probabilità ci consentono di affrontare il problema semplicemente considerando tutto ciò che poteva

accadere alla prima estrazione. Indichiamo con $B1$ l'evento *la prima pallina estratta è bianca*, con $B2$ l'evento *la seconda pallina estratta è bianca*; $A1$ e $A2$ indicheranno gli eventi corrispondenti nel caso la pallina estratta sia azzurra con $A1 = B1^c$ e $A2 = B2^c$. L'insieme dei possibili risultati della nostra prova sarà $\Omega = \{A1 \cap A2, A1 \cap B2, B1 \cap A2, B1 \cap B2\}$. Riscriviamo quindi l'evento $A2$ mettendolo in relazione con quello che può accadere nella prima estrazione

$$A2 = A2 \cap \Omega = (A2 \cap A1) \cup (A2 \cap A1^c) = (A2 \cap A1) \cup (A2 \cap B1)$$

Ad eventi uguali corrisponderanno uguale probabilità

$$P(A2) = P((A2 \cap A1) \cup (A2 \cap B1))$$

Osserviamo che $(A2 \cap A1)$ è incompatibile con $(A2 \cap B1)$, infatti l'evento $(A2 \cap A1)$ si verifica solo se si verifica $A1$ che è incompatibile con $B1$ e di conseguenza con l'evento $(A2 \cap B1)$. Per il terzo assioma avremo quindi

$$P(A2) = P(A2 \cap A1) + P(A2 \cap B1)$$

Utilizzando la *legge delle probabilità composte* potremo scrivere:

$$P(A2) = P(A1)P(A2|A1) + P(B1)P(A2|B1)$$

Ora siamo in grado di calcolare tutte le probabilità coinvolte

$$P(A2) = 6/10 \cdot 5/9 + 4/10 \cdot 6/9 = 1/3 + 4/15 = 9/15 = 3/5$$

Questa logica è importante in tutti i casi in cui un evento aleatorio *finale* dipenda dal verificarsi o meno di un evento aleatorio *intermedio*. Immaginiamo che la sopravvivenza di un paziente dipenda in modo cruciale dal fatto che risponda ad un certo farmaco. Siamo in grado di valutare la sua probabilità di sopravvivenza anche prima di sapere se risponderà o meno? Potremo farlo se siamo in grado di stimare tre quantità: la probabilità che il paziente risponda al trattamento e la probabilità di sopravvivenza in caso positivo e in caso negativo.

Torniamo alla nostra urna per capire meglio come possiamo verificare se due eventi sono indipendenti. Consideriamo ad esempio gli eventi $A1$ e $A2$ definiti in precedenza. Per verificare la loro indipendenza possiamo prendere due strade:

1. Utilizzare la condizione $P(A2|A1) = P(A2)$

calcolare separatamente $P(A2)$ e $P(A2|A1)$ e quindi verificare se $P(A2|A1) = P(A2)$. Se l'uguaglianza è vera allora questi due eventi sono indipendenti.

2. Utilizzare la condizione $P(A1 \cap A2) = P(A1) \cdot P(A2)$

calcolare separatamente la probabilità dei due eventi e quella della loro intersezione e verificare se vale l'uguaglianza $P(A1 \cap A2) = P(A1) \cdot P(A2)$. Ovviamente la probabilità dell'intersezione non può essere calcolata come prodotto delle probabilità dei due eventi visto che non possiamo assumere l'indipendenza a priori.

Proviamo a seguire la prima strada. Abbiamo già verificato che $P(A2) = 3/5$. Avremo $P(A2|A1) = 5/9$. Poiché $5/9 \neq 3/5$ i due eventi sono dipendenti.

Allo stesso risultato si giunge osservando che $P(A1 \cap A2) = P(A1) \cdot P(A2|A1) = 6/10 \cdot 5/9 = 1/3$. Poiché $P(A1) = 6/10$ e $P(A2) = 3/5$ avremo $P(A1) \cdot P(A2) = 6/10 \cdot 3/5 = 9/25$. Da $1/3 \neq 9/25$ segue che i due eventi sono dipendenti.

Ovviamente le cose cambiano se la prima pallina estratta viene rimessa nell'urna. Si lascia al lettore dimostrare che in questo caso i due eventi $A1$ e $A2$ sono indipendenti.

L'indipendenza in probabilità è un concetto complesso che sfugge almeno in parte all'intuizione. Per comprenderlo meglio vediamo i suoi legami con l'idea di incompatibilità. Capita che gli studenti confondano questi due termini (forse perché entrambi iniziano con "in"). In verità sono molto diversi. Ricordiamo che due eventi sono **incompatibili** se la loro intersezione è vuota, cioè se non si possono verificare contemporaneamente. In questa definizione non compare la probabilità, e questo sottolinea il fatto che questa proprietà resta tale anche se modifichiamo la nostra misura di probabilità. Al contrario l'indipendenza è strettamente legata a quest'ultima al punto che due eventi indipendenti possono diventare dipendenti se la nostra misura di probabilità cambia. Vediamo un semplice esempio. Torniamo a lanciare un dado e definiamo:

$$A = \{5 \cup 6\}$$

$$B = \{\text{esce un numero pari}\}$$

Se il dado è bilanciato possiamo utilizzare la definizione classica di probabilità da cui

$$P(A) = 2/6$$

$$P(B) = 3/6$$

La loro intersezione contiene solo il numero 6 con $P(A \cap B) = 1/6$. Osserviamo che $P(A) \cdot P(B) = 6/36 = 1/6$. Essendo la probabilità

dell'intersezione uguale al prodotto delle probabilità dei due eventi possiamo concludere che A e B sono indipendenti.

Ora immaginiamo che il dado sia truccato in modo da favorire l'uscita del numero 6. Nel nuovo spazio di probabilità Ω e la sua algebra A_Ω resteranno immutati ma cambierà la misura di probabilità che sarà adesso

- $P(6) = 1/2$;
- $P(k) = 1/10$ per $k = 1, \dots, 5$.

Verifichiamo se in questo nuovo spazio, i due eventi A e B rimangono indipendenti.

$$P(A) = P(5 \cup 6) = P(5) + P(6) = 1/10 + 1/2 = 6/10$$

$$P(B) = P(2 \cup 4 \cup 6) = P(2) + P(4) + P(6) = 1/10 + 1/10 + 1/2 = 7/10$$

$$\text{con } P(A) \cdot P(B) = 6/10 \cdot 7/10$$

L'intersezione resta il solo numero 6 con $P(A \cap B) = 1/2$

Da $P(A \cap B) \neq P(A) \cdot P(B)$ segue che adesso i nostri eventi sono dipendenti. Favorire l'uscita del 6 ha creato un legame tra A e B anche se gli eventi sono rimasti uguali a se stessi.

Abbiamo visto come due eventi compatibili possano risultare sia dipendenti sia indipendenti. Cosa accade se i due eventi A e B sono incompatibili? Dalla definizione di incompatibilità segue che $A \cap B = \emptyset$ da cui $P(A \cap B) = 0$. Se i due eventi hanno probabilità diversa da zero di verificarsi non potrà mai essere $P(A \cap B) = P(A) \cdot P(B)$. Quindi due eventi incompatibili sono necessariamente dipendenti se escludiamo il caso in cui almeno uno dei due abbia probabilità 0 di verificarsi. Almeno in questo caso l'intuizione ci conforta: se due eventi sono incompatibili il verificarsi dell'uno escluderà il verificarsi dell'altro.

1.7 Teorema di Bayes

Dalla definizione di probabilità condizionata si può derivare il **Teorema di Bayes**, semplice nella dimostrazione ma le cui conseguenze sono state enormi. Lo proponiamo in una forma elementare.

Dato uno spazio di probabilità (Ω, A_Ω, P) consideriamo due eventi A e B con probabilità diversa da zero.

Ora applichiamo la definizione di probabilità condizionata dando per noto l'evento B.

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

Questo implica che

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

Ora applichiamo la stessa definizione considerando noto l'evento A, cioè *invertendo* il ruolo dei due eventi.

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

da cui

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

Entrambe queste relazioni sono vere e poiché i membri di sinistra sono uguali potremo uguagliare i membri di destra

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

Da cui

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

Un semplice passaggio algebrico ci porta al Teorema di Bayes

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

Questo teorema stabilisce una relazione logica tra le probabilità condizionate quando invertiamo il ruolo degli eventi. Per capirne il senso proviamo ad applicarlo nell'ambito dei test diagnostici.

Ogni test diagnostico ha associate due misure che ne definiscono la qualità: sensibilità e specificità. Definiamo i seguenti eventi

- H^+
il soggetto è effettivamente malato di HIV;
- H^-
il soggetto *non* ha contratto l'HIV;
- T^+
il test è risultato positivo;
- T^-
il test è risultato negativo.

Ovviamente valgono le relazioni $H^- = (H^+)^c$ e $T^- = (T^+)^c$. Possiamo allora definire come

- **Sensibilità** $P(T^+|H^+)$ di un test: è la probabilità che il test sia positivo dato che il paziente è malato.
- **Specificità** $P(T^-|H^-)$ di un test: è la probabilità che il test risulti negativo, dato che il paziente è sano.

Notiamo come entrambe le misure siano probabilità condizionate. La sensibilità ci indica quanto il test è sensibile nel riconoscere la malattia quando è presente, la specificità misura invece quanto il test è specifico per quella patologia, risultando negativo quando la patologia è assente.

La prima viene in genere stimata su un campione di individui malati riportando la frequenza relativa dei soggetti positivi al test, la seconda su un campione di individui non affetti dalla malattia come frequenza relativa dei soggetti negativi al test. Osserviamo che il condizionamento definisce la popolazione di riferimento di cui il campione deve essere *rappresentativo*. Inoltre si ribadisce il legame tra frequenza relativa e probabilità; in questo caso la prima va a stimare la seconda.

Il test E.L.I.S.A. per l'HIV è caratterizzato da:

- Sensibilità $P(T^+|H^+) = 0.99$
- Specificità $P(T^-|H^-) = 0.99$

Immaginiamo che un individuo si sottoponga a questo test con esito positivo: è certamente malato ?

A volte nel rispondere a questa domanda si fa erroneamente riferimento alla sensibilità 0.99 del test. Tuttavia quest'ultima è la probabilità che il test sia positivo dato che il soggetto è affetto dalla patologia mentre quella che vorremmo conoscere è la probabilità che l'individuo sia effettivamente malato, noto che il test è risultato positivo. Il teorema di Bayes ci insegna che le due probabilità sono diverse e ci indica la retta via per il calcolo.

$$P(H^+|T^+) = P(T^+|H^+) \cdot \frac{P(H^+)}{P(T^+)}$$

In questa espressione $P(H^+)$ è la *probabilità a priori* che il soggetto sia malato prima di conoscere l'esito del test. Se l'individuo non appartiene a particolari categorie a rischio possiamo stimare questa

probabilità utilizzando il dato sulla prevalenza della malattia nel paese dove risiede.

Apriamo una piccola parentesi per ricordare che due sono le misure della diffusione di una data patologia:

- **la prevalenza** è data dal *rapporto del numero dei soggetti malati ad una certa data sulla popolazione di riferimento*. Tale definizione ha un'accezione statica poiché ci fornisce una fotografia della situazione in quel momento.
- **l'incidenza** è data dal *rapporto del numero di nuovi casi che si sono sviluppati nell'arco di un prefissato periodo di tempo diviso il numero dei soggetti esposti a rischio all'inizio del periodo (cioè il numero di soggetti che non avevano ancora la malattia)*. Ha un'accezione dinamica poiché misura l'evoluzione della patologia.

I dati sull'incidenza dell'HIV nell'anno solare in Italia sono facilmente reperibili e ci dicono per l'anno 2018 di circa 5.7 nuovi casi su 100000 residenti. Meno noti sono quelli sulla prevalenza; immaginiamo che in Italia sia dell'ordine di 1 caso su 500.

Applichiamo il teorema di Bayes

$$P(H^+|T^+) = P(T^+|H^+) \cdot \frac{P(H^+)}{P(T^+)}$$

Per calcolare $P(T^+)$ osserviamo che

$$\Omega = \{T^+ \cap H^+, T^+ \cap H^-, T^- \cap H^+, T^- \cap H^-\}$$

Possiamo quindi scrivere

$$T^+ = T^+ \cap \Omega = (T^+ \cap H^+) \cup (T^+ \cap H^-)$$

Ad uguali eventi corrisponderanno uguali probabilità

$$P(T^+) = P((T^+ \cap H^+) \cup (T^+ \cap H^-))$$

Osserviamo che $(T^+ \cap H^+) \cap (T^+ \cap H^-) = (T^+ \cap (H^+ \cap H^-)) = T^+ \cap \emptyset = \emptyset$

dove nel primo passaggio abbiamo applicato la proprietà associativa. Essendo gli eventi incompatibili possiamo applicare il terzo assioma

$$P(T^+) = P(T^+ \cap H^+) + P(T^+ \cap H^-)$$

Applicando la legge delle probabilità composte

$$P(T^+ \cap H^+) = P(T^+|H^+) \cdot P(H^+)$$

$$P(T^+ \cap H^-) = P(T^+|H^-) \cdot P(H^-) = (1 - P(T^-|H^-))(1 - P(H^+))$$

Notiamo che $P(T^+|H^-)$ è il **complemento** della specificità. Ricordiamo infatti che quando ricaviamo un evento complementare, agiamo sull'**evento aleatorio** e non sul condizionamento che è noto. Sostituendo otteniamo

$$P(H^+|T^+) = P(T^+|H^+) \cdot \frac{P(H^+)}{P(T^+|H^+) \cdot P(H^+) + (1 - P(T^-|H^-))(1 - P(H^+))}$$

$$P(H^+|T^+) = 0.99 \cdot \frac{P(H^+)}{0.99 \cdot P(H^+) + 0.01 \cdot (1 - P(H^+))}$$

Ipotizzando $P(H^+) = 1/500$ avremo

$$P(H^+|T^+) = 0.99 \cdot \frac{1/500}{0.99 \cdot 1/500 + 0.01 \cdot (1 - 1/500)} = 0.165$$

La probabilità che l'individuo abbia effettivamente contratto l'HIV dato l'esito positivo del test è solo del 16.5% (confortante per chi ha fatto il test).

Questo risultato può apparire sconcertante soprattutto tenendo conto dell'altissimo livello di sensibilità e specificità del test utilizzato. Il punto critico di questa valutazione è la bassissima probabilità a priori che ha il nostro paziente di essere malato: $1/500 = 0.002$. Il risultato positivo agisce modificando questa probabilità a priori, per arrivare una probabilità a posteriori di 0.165.

Ovviamente possiamo supporre che, se il soggetto si è sottoposto ad un test, la sua probabilità iniziale fosse più alta rispetto a quella di un individuo qualsiasi scelto a caso dalla popolazione italiana. Immaginiamo allora che il soggetto appartenga ad una categoria a rischio dove adesso la prevalenza dell'HIV è di $1/50$. Come cambia la sua probabilità a posteriori?

Abbiamo

$$P(H^+|T^+) = 0.99 \cdot \frac{1/50}{0.99 \cdot 1/50 + 0.01 \cdot (1 - 1/50)} = 0.669$$

Come vedete adesso la sua probabilità di aver contratto il virus è decisamente più alta. È altresì evidente l'importanza che la probabilità a priori gioca nel calcolo.

Questo esempio ci aiuta a capire la difficoltà nel portare avanti politiche di screening sull'intera popolazione dove la maggior parte degli individui è sana. Se infatti i soggetti sani diventano la maggioranza, noi magari sbagliamo poco ma sbagliamo poco su tantissima gente. La specificità del test diagnostico diventa un elemento critico quando applicata in una condizione in cui la prevalenza di una patologia è bassissima. Anche se il test è affidabile, noi lo stiamo mettendo in una condizione di difficile applicazione perché stiamo continuando ad operare il test su soggetti sani e quindi inevitabilmente otterremo un elevato numero di falsi positivi.

Torniamo sul percorso logico che abbiamo seguito applicando il Teorema di Bayes. Siamo partiti da una probabilità a priori $P(H^+)$ e, noto il risultato del test, attraverso il teorema di Bayes, siamo arrivati ad una probabilità a posteriori $P(H^+|T^+)$.

Questa stessa logica è alla base di un approccio alternativo all'inferenza statistica classica noto come *inferenza Bayesiana*. Senza aver la pretesa di discutere in questa sede un argomento così vasto e complesso, proviamo ad intuirne alcuni principi sulla base di un semplice esempio.

Immaginiamo di avere tre scatole identiche. Ogni scatola contiene 2 monete: la prima scatola due monete da 50 centesimi, la seconda una moneta da 50 centesimi e una da un euro, la terza due monete da un euro.

Estraiamo casualmente una scatola e, da questa scatola una moneta che risulta essere da 50 centesimi.

Qual è la probabilità che la moneta provenga dalla prima, dalla seconda o dalla terza scatola?

Definiamo gli eventi di interesse e le relative probabilità a priori:

$$A = (\text{La moneta proviene dalla prima scatola}) \longrightarrow P(A) = \frac{1}{3}$$

$$B = (\text{La moneta proviene dalla seconda scatola}) \longrightarrow P(B) = \frac{1}{3}$$

$$C = (\text{La moneta proviene dalla terza scatola}) \longrightarrow P(C) = \frac{1}{3}$$

L'estrazione casuale della scatola implica che le probabilità a priori siano le stesse. Definiamo $D = \{\text{la moneta estratta è da 50 cent.}\}$.

Vogliamo calcolare la probabilità che la moneta provenga dalle tre scatole, noto si tratta di una moneta da 50 centesimi; in simboli $P(A|D)$, $P(B|D)$, $P(C|D)$.

Anche la moneta viene estratta casualmente dalla scatola scelta, avremo quindi

$$P(D|A) = 1$$

$$P(D|B) = \frac{1}{2}$$

$$P(D|C) = 0$$

Dal teorema di Bayes

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D)}$$

dove

$$P(D) = P((D \cap A) \cup (D \cap B) \cup (D \cap C))$$

Essendo eventi incompatibili, avremo

$$P(D) = P(D \cap A) + P(D \cap B) + P(D \cap C)$$

Per la legge delle probabilità composte

$$P(D \cap A) = P(D|A) \cdot P(A)$$

$$P(D \cap B) = P(D|B) \cdot P(B)$$

$$P(D \cap C) = P(D|C) \cdot P(C)$$

Sostituendo

$$P(A|D) = \frac{P(D|A) \cdot P(A)}{P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C)}$$

Da cui

$$P(A|D) = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}$$

Analogamente

$$P(B|D) = \frac{P(D|B) \cdot P(B)}{P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C)}$$

Da cui

$$P(B|D) = \frac{\frac{1}{2} \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{3}{6}} = \frac{1}{3}$$

Infine

$$P(C|D) = \frac{P(D|C) \cdot P(C)}{P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C)} = 0$$

Questo semplice esempio può aiutarci ad intuire la logica alla base dell'approccio *Bayesiano* dell'inferenza. Immaginiamo che le tre scatole siano ipotesi scientifiche da valutare e che $P(A)$, $P(B)$, $P(C)$ siano le loro *probabilità a priori*, tutte uguali nel nostro esempio. Eseguiamo il nostro esperimento e indichiamo con D l'evidenza sperimentale. Il Teorema di Bayes ci consente di rivalutare le probabilità delle nostre ipotesi alla luce dei dati sperimentali calcolando le loro *probabilità a posteriori*, $P(A|D)$, $P(B|D)$, $P(C|D)$. È evidente che, dopo l'estrazione della moneta da 50 cent., l'ipotesi più probabile diviene la prima scatola.

Il rapporto $\frac{P(D|A)}{P(D|B)}$, noto come *rapporto di verosimiglianza* delle due ipotesi sulla base dei nostri dati sperimentali, è quello che trasforma il rapporto delle probabilità a priori nel rapporto delle probabilità a posteriori.

$$\frac{P(A)}{P(B)} \cdot \frac{P(D|A)}{P(D|B)} = \frac{P(A|D)}{P(B|D)}$$

$$\frac{1/3}{1/3} \cdot \frac{1}{1/2} = \frac{2/3}{1/3}$$

Le due scatole avevano inizialmente la stessa probabilità ma, osservata la moneta estratta, la prima scatola diventa due volte più probabile della seconda scatola.

La definizione delle probabilità a priori, che caratterizza l'impostazione *bayesiana*, è un punto delicato che ha suscitato e suscita molte

discussioni. L'idea che nessuna ricerca parta da zero è largamente condivisa ma resta la difficoltà di tradurre le conoscenze a priori in probabilità. Inoltre il fatto che le conclusioni finali dipendano dalle probabilità a priori è a volte letto come un punto di debolezza dell'impostazione *bayesiana*.

Concludiamo osservando che questa logica può essere iterata come dimostra l'esempio che segue. Le probabilità a posteriori dopo il primo esperimento diventeranno le probabilità a priori per l'esperimento successivo, riflettendo le nostre conoscenze all'inizio del secondo esperimento.

Rimettiamo la moneta nella scatola da cui l'abbiamo estratta, ripetiamo l'estrazione dalla stessa scatola ed immaginiamo che esca nuovamente una moneta da 50 cent.

Quali sono adesso le probabilità che la scatola che abbiamo estratto sia la prima, la seconda o la terza.

Sia $E = \{\text{esce di nuovo una moneta da 50 cent.}\}$

Le probabilità a posteriori dopo la prima estrazione possono essere considerate probabilità a priori prima della seconda estrazione: $\rightarrow$

$$P(A) = \frac{2}{3}, \quad P(B) = \frac{1}{3}, \quad P(C) = 0$$

Dobbiamo calcolare

$$P(A|E), \quad P(B|E), \quad P(C|E)$$

$$P(A|E) = \frac{P(E|A) \cdot P(A)}{P(E)}$$

$$P(E) = P(E \cap A) \cup P(E \cap B) \cup P(E \cap C)$$

Essendo eventi incompatibili avremo

$$P(E) = P(E \cap A) + P(E \cap B) + P(E \cap C)$$

$$P(E \cap A) = P(E|A) \cdot P(A)$$

$$P(E \cap B) = P(E|B) \cdot P(B)$$

$$P(E \cap C) = P(E|C) \cdot P(C)$$

quindi:

$$\begin{aligned}
 P(A|E) &= \frac{P(E|A) \cdot P(A)}{P(E|A) \cdot P(A) + P(E|B) \cdot P(B) + P(E|C) \cdot P(C)} \\
 &= \frac{1 \cdot \frac{2}{3}}{1 \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{2} + 0} = \frac{4}{5}
 \end{aligned}$$

$$P(B|E) = \frac{P(E|B) \cdot P(B)}{P(E)}$$

$$\begin{aligned}
 P(B|E) &= \frac{P(E|B) \cdot P(B)}{P(E|A) \cdot P(A) + P(E|B) \cdot P(B) + P(E|C) \cdot P(C)} \\
 &= \frac{\frac{1}{2} \cdot \frac{1}{3}}{1 \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0} = \frac{1}{5}
 \end{aligned}$$

$$P(C|E) = \frac{P(E|C) \cdot P(C)}{P(C)}$$

$$P(C|E) = \frac{P(E|C) \cdot P(C)}{P(E|A) \cdot P(A) + P(E|B) \cdot P(B) + P(E|C) \cdot P(C)} = 0$$

La seconda estrazione di una moneta da 50 cent ha evidentemente rafforzato l'ipotesi che la scatola inizialmente estratta fosse la prima.

Capitolo 2

Variabili aleatorie

Il punto di partenza è ancora un esperimento il cui risultato è incerto. Quanto visto finora ci ha permesso di formalizzare quest'incertezza rendendola trattabile tramite la definizione di uno spazio di probabilità. Il passo successivo va nella direzione di costruire dei *modelli statistico-probabilistici* che sintetizzino le caratteristiche dell'esperimento e le nostre conoscenze a priori sul fenomeno. Tuttavia per costruire modelli abbiamo bisogno di lavorare sull'asse reale (o su un suo sottoinsieme) che ci consente definire un'unità di misura e di calcolare distanze tra eventi. Partiamo da uno spazio di probabilità (Ω, A_Ω, P) e definiamo allora come ***variabile aleatoria un'applicazione che ci porta da Ω in $\mathbb{R}$, cioè che ci consente di spostarci dal nostro Ω sui Reali.*** Formalmente si tratta di una funzione $X : \Omega \mapsto \mathbb{R}$ che associa a ogni risultato elementare del nostro esperimento $\omega \in \Omega$ un valore reale $X(\omega)$.

2.1 Variabili aleatorie discrete

La definizione di probabilità cambia a seconda che i valori osservabili siano discreti oppure continui. Ma qual è la differenza fra discreto e continuo? Il discreto viene spesso associato ad un numero finito di valori osservabili che diventano infiniti nel continuo. Questa idea intuitiva richiede tuttavia una precisazione. Partiamo dalla definizione di continuo. In ambito scientifico un continuo è un insieme totalmente ordinato che è *densamente ordinato*, cioè tra due valori dell'insieme c'è sempre un terzo valore compreso fra i primi due. Semplicemente il continuo non ha *buchi* e può essere associato alla retta reale. Il discreto non soddisfa questa condizione e può essere messo in corrispondenza con i numeri naturali (incluso lo 0). Iniziamo considerando variabili aleatorie discrete. Vedremo in dettaglio alcuni modelli discreti

impo rtanti ma altri ne esistono e questa trattazione non vuole essere esaustiva.

2.1.1 La variabile aleatoria Bernoulliana

Consideriamo l'esperimento più semplice costituito da una sola prova il cui esito è *dicotomico*, testa o croce oppure, più in generale, successo o insuccesso.

Ora associamo al successo e all'insuccesso due valori reali. *In genere si sceglie di assegnare al successo il valore 1 e all'insuccesso il valore 0* anche se altre scelte sono possibili. Resta così definita una variabile aleatoria *Bernoulliana* i cui valori osservabili sono 0 e 1.

Tornando all'esempio del lancio di una moneta abbiamo uno spazio di probabilità dove $\Omega = \{T, C\}$, definiamo X , $A_\Omega = \{\{T\}, \{C\}, \{T \cup C\} = \Omega, \emptyset\}$. Assumendo che la moneta sia bilanciata, potremo utilizzare per P la definizione classica di probabilità con $P(T) = P(C) = 1/2$

Ci trasferiamo da Ω ad un sottoinsieme dei numeri reali definendo X come $X(\text{Testa}) = 1$ e $X(\text{Croce}) = 0$ Il nostro interesse si è adesso spostato da Ω ad $\Omega_X = \{1, 0\} \subset \mathbb{R}$ con $A_{\Omega_X} = \{\{1\}, \{0\}, \{1 \cup 0\} = \Omega_X, \emptyset\}$. Per completare la costruzione del nuovo spazio di probabilità definiamo P_X a partire dalla probabilità P definita nello spazio originario come

$$P_X(X = 0) = P(\omega \in \Omega : X(\omega) = 0) = P(\text{Croce}) = 1/2$$

$$P_X(X = 1) = P(\omega \in \Omega : X(\omega) = 1) = P(\text{Testa}) = 1/2$$

Stiamo di fatto assegnando al possibile risultato 0 la probabilità dell'evento originario croce a cui la funzione X aveva associato 0. Analogamente per 1.

Osserviamo che in generale, perchè questo passaggio sia possibile, è necessario che ad ogni evento in A_{Ω_X} corrisponda un evento nell'originaria A_Ω cioè che la funzione X sia *misurabile*.

La variabile aleatoria Bernoulliana costituisce un primo semplice modello a descrivere un esperimento il cui esito può essere successo o insuccesso. Pur nella sua semplicità si tratta di un modello alla base di modelli più complessi che trovano applicazione reale nella sperimentazione. Tuttavia in condizioni sperimentali in genere la probabilità di successo non è nota anzi è proprio per la sua *stima* che viene condotto un esperimento. Pensiamo ad esempio ai famosi esperimenti condotti da Mendel per stimare la probabilità che un carattere recessivo ricompaia in seconda generazione, evento che potremo definire *successo*.

Se indichiamo con π tale probabilità di successo, è chiaro che la probabilità di insuccesso sarà $1 - \pi$.

$$P(X = 0) = 1 - \pi$$

$$P(X = 1) = \pi$$

In una notazione più generale

$$P(X = x; \pi) = \pi^x(1 - \pi)^{1-x}$$

Quella che abbiamo scritto è la *distribuzione di X* cioè una funzione che dipende da π e che associa ad ogni risultato osservabile x la sua probabilità. In generale le lettere maiuscole in probabilità indicano le variabili aleatorie mentre con le lettere minuscole vengono indicati in modo generico i valori che possono assumere.

Infatti se si sostituisce 0 a x avremo

$$P(X = 0) = \pi^0(1 - \pi)^{1-0} = 1 - \pi$$

Se invece si sostituisce 1

$$P(X = 1) = \pi^1(1 - \pi)^{1-1} = \pi$$

Inoltre nei modelli è convenzione indicare le quantità non note con delle lettere greche che in statistica prenderanno il nome di *parametri*

Per iniziare a comprendere il rapporto tra modello e dati osservati e come il modello preceda l'osservazione, immaginiamo di partecipare ad una scommessa sul risultato del lancio di una moneta. Ovviamente è importante per noi sapere se la moneta sia bilanciata o meno. Poiché l'esito del lancio è dicotomico potremo ragionevolmente assumere un modello Bernoulliano a guidare il nostro esperimento. Si configurano tre diversi possibili obiettivi :

- vogliamo sapere quanto vale la probabilità di testa π (stima puntuale);
- vogliamo verificare l'ipotesi che la moneta sia bilanciata (verifica d'ipotesi);
- vogliamo prevedere il risultato del prossimo lancio, in quanto è nostro interesse scommettere sul risultato più verosimile (previsione).

Stima, verifica d'ipotesi e previsione sono i tre capitoli dell'inferenza statistica che vedremo nel proseguo del corso. In tutti e tre i casi abbiamo bisogno di dati sperimentali che nel nostro caso possiamo

ottenere lanciando ripetutamente la nostra moneta. Ora immaginiamo di aver eseguito 1000 lanci osservando 491 volte croce e 509 volte testa.

Possiamo sintetizzare i risultati **osservati** attraverso quella che chiamiamo **variabile statistica** $\tilde{X}$

Valori Osservati	Frequenze Assolute n_i	Frequenze Relative $f(x_i) = n_i/n$
0 (croce)	491	0.491
1 (testa)	509	0.509
Totale	1000	1

Quanto descritto in questa tabella è una distribuzione di frequenze dove ritroviamo i valori 0 e 1 che adesso corrispondono però a valori realmente **osservati** e non più a valori **osservabili** come nel caso della variabile aleatoria Bernoulliana. Accanto ad ogni valore, non abbiamo più la sua probabilità di essere osservato, ma la frequenza (relativa) con cui quel valore è stato osservato.

Una variabile aleatoria costituisce un modello per il nostro esperimento ed appartiene ad un universo aleatorio dove non c'è nulla di osservato ma piuttosto un insieme di valori osservabili con le relative probabilità. In questo mondo aleatorio, che possiamo immaginare pre-sperimentale si costruisce tutta la teoria inferenziale.

Una variabile statistica è invece un modo di descrivere dei dati reali e si colloca nel universo post-sperimentale dell'osservato

Il modello contiene traduce in parametri gli elementi incogniti del nostro studio, l'evidenza empirica contenuta nei dati osservati sarà lo strumento per arrivare a conoscerli.

Tornando al nostro semplice esperimento come possiamo stimare π ? Dimostreremo che lo stimatore migliore è proprio la frequenza relativa dei successi che risulta pari a 0.509. Restano aperti gli altri quesiti: questo valore è in grado di confermare che la moneta è bilanciata? Quale valore è più probabile tra 0 e 1?

Per poter rispondere dobbiamo ancora capire molte cose.

2.1.2 Valore atteso e varianza nel discreto

Iniziamo ricordando la definizione di media aritmetica nel caso di variabili statistiche come somma dei valori osservati diviso il loro numero.

Quale valore assume la media dei dati osservati nel nostro piccolo esperimento? Attenzione perché i dati sono descritti attraverso una distribuzione di frequenze: abbiamo osservato il valore 0 491 volte ...

Dalla tabella, la media risulta essere 0.509.

$$Media[\tilde{X}] = \frac{(509 \cdot 1) + (409 \cdot 0)}{1000} = 0.509$$

oppure analogamente

$$Media[\tilde{X}] = (0 \cdot 0.409) + (1 \cdot 0.509) = 0.509$$

In generale la media di una **variabile statistica** sarà

$$Media[\tilde{X}] = \sum_{i=1}^c x_i \cdot f(x_i)$$

dove x_i sono i valori osservati e $f(x_i)$ le corrispondenti frequenze relative

Possiamo chiederci se esiste una quantità analoga nel mondo aleatorio.

Ora, se per una variabile statistica la media era la somma dei valori osservati moltiplicati per le loro frequenze, il **valore atteso** di una variabile aleatoria discreta si definisce come la somma, non più dei valori osservati ma dei valori osservabili moltiplicati, non più per le loro frequenze, ma per le relative probabilità di essere osservati

$$E[X] = \sum_{i=1}^c x_i \cdot p(x_i)$$

Le analogie nella definizione non devono far dimenticare che il significato di queste quantità è diverso: la media è sintesi di dati osservati e ci indica la posizione della nostra distribuzione sull'asse reale; il valore atteso non corrisponde a nessuna reale osservazione, descrivendo piuttosto quello che ci attendiamo di osservare in media. Proviamo a calcolare il valore atteso di una variabile aleatoria bernoulliana. Ricordando che $P(X = 0) = 1 - \pi$ e $P(X = 1) = \pi$ avremo

$$E[X] = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi$$

È importante sottolineare che il valore atteso coincide in questo caso con la probabilità di successo valore in realtà non osservabile se non negli estremi 0 e 1. Questa peculiarità del modello bernoulliano dipende in parte dalla struttura Ω "povero in quanto" di valori osservabili contenendo solo 0 e 1.

Notiamo inoltre come il risultato $E[X] = \pi$ dipenda dalla definizione di variabile aleatoria Bernoulliana. Potremmo definire una diversa variabile aleatoria Z associando al successo sempre il valore 1 ma penalizzando l'insuccesso associandogli il valore -1. Come cambierebbe il valore atteso?

Torniamo al mondo dei dati osservati per ricordare la definizione di varianza come media degli scarti di ogni singola osservazione rispetto alla media aritmetica, elevati al quadrato. La varianza dovrebbe sempre accompagnare il valore medio nel descrivere i dati poichè la media è un indice estremamente sintetico (solo un numero a sintesi di n valori) e ricordiamo che *ad ogni sintesi corrisponde una perdita di informazione*. La media nasconde infatti la variabilità dei dati osservati: la stessa media può corrispondere ad una situazione con bassa variabilità in cui i valori osservati sono tutti molto simili e quindi molto simili al valor medio, oppure può mascherare la presenza di un'alta variabilità con osservazioni molto più piccole e molto più elevate del valor medio. Notiamo come nel primo caso la media ben descrive i dati mentre nel secondo caso la sua rappresentatività può essere molto ridotta. La varianza è chiamata pertanto a descrivere la variabilità delle osservazioni e, indirettamente, la capacità della media aritmetica a rappresentare. Ora se i dati si presentano in una distribuzione di frequenze, come nel nostro esempio sul lancio di una moneta, possiamo definire la **Varianza di una variabile statistica** come

$$V[\tilde{X}] = \sum_{i=1}^c (x_i - \text{Media}[\tilde{X}])^2 \cdot f(x_i)$$

Nel nostro esempio avremo

$$V[\tilde{X}] = (0 - 0.509)^2 \cdot 0.491 + (1 - 0.509)^2 \cdot 0.509 = 124.92$$

Possiamo chiederci se, come per la media, è possibile definire una quantità equivalente nel mondo aleatorio.

La risposta è affermativa, si definisce infatti la **Varianza di una variabile aleatoria discreta** come la sommatoria dei valori *osservabili* meno il valore atteso, al quadrato, dove ciascuno scarto è moltiplicato per la corrispondente probabilità di essere osservato

$$V[X] = \sum_{i=1}^c (x_i - E[X])^2 \cdot p(x_i)$$

Nel caso di una variabile aleatoria Bernoulliana avremo

$$V[X] = (1 - \pi)^2 \cdot \pi + (0 - \pi)^2 \cdot (1 - \pi) = \pi \cdot (1 - \pi)$$

Abbiamo visto come, nel caso di dati osservati, la varianza misura la variabilità delle osservazioni intorno alla media. Ora cosa descrive la varianza nel caso di una variabile aleatoria?

Osserviamo che il valore minimo di $V[X]$ è 0 e si ottiene nei due casi estremi in cui $\pi = 0$ o $\pi = 1$. Entrambi i casi corrispondono a situazioni in cui il risultato del lancio è certo: 0 nel primo caso, 1 nel

secondo. Quindi la varianza è nulla se il risultato del nostro esperimento è prevedibile cioè l'incertezza è nulla.

Osserviamo ancora che $V[X]$ raggiunge il massimo quando $\pi = 0.5$, condizione di massima incertezza sul risultato del lancio (la moneta è perfettamente bilanciata) che non è in alcun modo prevedibile.

Il caso di una variabile aleatoria Bernoulliana ci aiuta a comprendere il significato di varianza nel mondo aleatorio, non più una misura di variabilità ma una misura di incertezza.

Abbiamo visto come i concetti di media e varianza, definiti per dati osservati, trovano un loro equivalente nelle variabili aleatorie con un significato tuttavia diverso.

Riepilogo:• **Variabile Statistica $\tilde{X}$**

- **Media:** sommatoria dei valori osservati moltiplicati per le loro frequenze

$$Media[\tilde{X}] = \sum_{i=1}^c x_i \cdot f(x_i)$$

- **Varianza:** sommatoria degli scarti di ogni singola osservazione dalla media (quindi $x_i - media$), al quadrato, e poi moltiplicati per le corrispondenti frequenze relative

$$V[\tilde{X}] = \sum_{i=1}^c (x_i - Media[\tilde{X}])^2 \cdot f(x_i)$$

• **Variabile Aleatoria X**

- **Valore Atteso:** sommatoria (somma) dei valori *osservabili* moltiplicati per le relative probabilità di essere osservati

$$E[X] = \sum_{i=1}^c x_i \cdot p(x_i)$$

- **Varianza:** sommatoria dei valori *osservabili* meno il valore atteso, al quadrato, e moltiplicati per le corrispondenti probabilità di essere osservati

$$V[X] = \sum_{i=1}^c (x_i - E[X])^2 \cdot p(x_i)$$

• **Osservazioni**

- Valore Atteso **Variabile Aleatoria Bernoulliana**

$$E[X] = \pi$$

- Varianza **Variabile Aleatoria Bernoulliana**

$$V[X] = \pi \cdot (1 - \pi)$$

2.1.3 Il modello Binomiale

Abbiamo visto come la variabile aleatoria Bernoulliana descriva il risultato di una prova il cui esito è dicotomico. Tuttavia, in generale un esperimento prevede l'esecuzione di prove o osservazioni. Immaginiamo quindi di ripetere la prova ipotizzando che questa avvenga sempre nelle medesime condizioni e che le prove siano tra loro indipendenti. Fare n prove significa formalmente che il risultato dell'esperimento non è più una singola variabile aleatoria ma un **vettore di n variabili aleatorie**

$$(X_1, \dots, X_n)$$

tra loro indipendenti e tutte ugualmente distribuite secondo una legge Bernoulliana. Osserviamo che il fatto che la prova venga ripetuta nelle medesime condizioni implica che le n variabili aleatorie condivideranno lo stesso parametro π cioè che *la probabilità di successo sarà la stessa in tutte le osservazioni*.

Ogni possibile risultato dell'esperimento sarà un vettore di n elementi, dove ciascun elemento potrà assumere il valore 0 oppure 1. Avremo in tutto 2^n vettori osservabili che costituiranno il nostro Ω , cioè l'insieme dei possibili risultati del nostro esperimento. Questi vettori saranno equiprobabili? Proviamo a calcolare la probabilità di osservare un generico vettore

$$P\{X_1 = x_1, \dots, X_n = x_n\}$$

dove il valore osservato x_i potrà assumere valore 0 oppure 1. Notiamo come il maiuscolo è generalmente utilizzato per indicare una variabile aleatoria mentre in corsivo vengono indicati, seppure in modo generico, i valori osservati. La notazione svolge, sia nella probabilità sia nella statistica, un ruolo importante. Sebbene può inizialmente risultare ostica, il mio suggerimento è fare amicizia, diventerà una vostra fedele alleata.

Osserviamo inoltre che la virgola corrisponde alla condizione *and* quindi formalmente la probabilità di osservare un certo vettore corrisponde alla probabilità di n intersezioni.

Poichè abbiamo ipotizzato prove indipendenti, questa probabilità diventerà il prodotto delle probabilità dei singoli eventi

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \cdot \dots \cdot P\{X_n = x_n\} = \prod_{i=1}^n Pr\{X_i = x_i\} = \prod_{i=1}^n \pi^{x_i} \cdot (1 - \pi)^{1-x_i}$$

Quest'ultimo passaggio richiede la condizione che abbiamo ipotizzato inizialmente di un valore di π costante nelle diverse prove. Questa condizione implica che *l'esperimento sia ripetibile in condizioni analoghe*.

È chiaro che se operiamo ad esempio in ambito medico, tipicamente l'esperimento non è ripetibile nelle medesime condizioni; al contrario, in esperimenti di laboratorio, su cavie o, più in generale, in esperimenti fortemente controllati π è approssimabile a una costante.

La probabilità di osservare un generico vettore assumerà valori diversi a seconda del valore di π e di quanti successi (e insuccessi) osserveremo. Solo in un caso tutti i possibili risultati saranno equiprobabili, quale?

Immaginiamo ora di non essere interessati al risultato di ogni singola prova ma soltanto al numero complessivo di successi osservati. Introduciamo allora una nuova variabile aleatoria, che chiamiamo Y , definita come il numero di successi nelle n prove indipendenti, eseguite nelle medesime condizioni, ovvero con π costante.

Siamo di fronte ad una variabile aleatoria discreta della quale dovremo capire quali valori può assumere e con che probabilità.

Per semplicità ipotizziamo di eseguire soltanto 3 prove; è evidente che potremo osservare da 0 a tre successi. Con che probabilità osserverò questi valori ?

Proviamo a riassumere i possibili risultati nella tabella seguente

Y	X_1	X_2	X_3	Prob
0	0	0	0	$(1 - \pi)^3$
1	1	0	0	$\pi(1 - \pi)^2$
1	0	1	0	$\pi(1 - \pi)^2$
1	0	0	1	$\pi(1 - \pi)^2$
2	1	1	0	$\pi^2(1 - \pi)$
2	1	0	1	$\pi^2(1 - \pi)$
2	0	1	1	$\pi^2(1 - \pi)$
3	1	1	1	π^3

L'evento 0 successi ovvero ($Y = 0$) corrisponde al solo vettore (0, 0, 0) ovvero ($X_1 = 0 \cap X_2 = 0 \cap X_3 = 0$).

Con che probabilità osserverò questo vettore?

$$\begin{aligned}
 P(Y = 0) &= P(X_1 = 0 \cap X_2 = 0 \cap X_3 = 0) \\
 &= P(X_1 = 0) \cdot P(X_2 = 0) \cdot P(X_3 = 0) \\
 &= \pi^0(1 - \pi)^{1-0} \cdot \pi^0(1 - \pi)^{1-0} \cdot \pi^0(1 - \pi)^{1-0} \\
 &= (1 - \pi)^1 \cdot (1 - \pi)^1 \cdot (1 - \pi)^1 = (1 - \pi)^3
 \end{aligned}$$

Se siamo interessati alla probabilità che $Y = 1$ seguiamo la stessa logica osservando però che questo evento corrisponde ad una di tre possibili terne.

$$\begin{aligned}
 P(Y = 1) &= P((X_1 = 1 \cap X_2 = 0 \cap X_3 = 0) \cup \\
 &\quad \cup (X_1 = 0 \cap X_2 = 1 \cap X_3 = 0) \cup \\
 &\quad \cup (X_1 = 0 \cap X_2 = 0 \cap X_3 = 1)) \\
 &= P(X_1 = 1 \cap X_2 = 0 \cap X_3 = 0) + \\
 &\quad + P(X_1 = 0 \cap X_2 = 1 \cap X_3 = 0) + \\
 &\quad + P(X_1 = 0 \cap X_2 = 0 \cap X_3 = 1) \\
 &= P(X_1 = 1) \cdot P(X_2 = 0) \cdot P(X_3 = 0) + \\
 &\quad + P(X_1 = 0) \cdot P(X_2 = 1) \cdot P(X_3 = 0) + \\
 &\quad + P(X_1 = 0) \cdot P(X_2 = 0) \cdot P(X_3 = 1) \\
 &= \pi \cdot (1 - \pi) \cdot (1 - \pi) + \\
 &\quad + (1 - \pi) \cdot \pi \cdot (1 - \pi) + \\
 &\quad + (1 - \pi) \cdot (1 - \pi) \cdot \pi \\
 &= 3 \cdot \pi \cdot (1 - \pi)^2
 \end{aligned}$$

Notiamo che la probabilità delle unioni diventa la somma delle probabilità perché gli eventi considerati sono incompatibili. Potete ricavare le probabilità di osservare 2 o 3 successi in modo analogo.

In generale se ipotizziamo n prove Bernoulliane ($X_i, i = 1, \dots, n$) condotte nelle medesime condizioni (π costante) e indipendenti, e indichiamo con $Y = \sum_{i=1}^n X_i$ il numero totale dei successi avremo

$$\Omega_Y = (0, \dots, n)$$

$$P(Y = k; n, \pi) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$$

Il coefficiente binomiale $\binom{n}{k} (n \text{ su } k) = \frac{n!}{k!(n-k)!}$ conta tutti i modi in cui in n prove si possono osservare k successi ($\binom{3}{1} = \frac{3!}{1!(2)!} = 3$ nel caso descritto in precedenza di 1 successo su 3 prove). Per convenzione $0! = 1$.

Anche la famiglia Binomiale contiene al suo interno diverse distribuzioni di probabilità, una per ogni scelta della coppia di parametri n e π . Vedremo tuttavia come questi due parametri svolgeranno in ambito sperimentale ruoli diversi: n è generalmente fissato nella fase di disegno dell'esperimento mentre sarà π il parametro di interesse inferenziale.

Come per la variabile aleatoria Bernoulliana anche nel caso Binomiale possiamo calcolarne il valore atteso e la varianza. Questa volta non utilizzeremo le definizioni ma piuttosto le loro proprietà. Sappiamo infatti che il valore atteso di una somma di variabili aleatorie è uguale alla somma dei loro valori attesi. Ricordando che il valore di π resta

costante nelle n prove avremo

$$E[Y] = E[\sum_{i=1}^n X_i] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \pi = n \cdot \pi$$

Una proprietà analoga vale per la varianza ma a condizione che le variabili aleatorie siano indipendenti. Poiché nel nostro caso abbiamo ipotizzato prove indipendenti potremo scrivere

$$V[Y] = V[\sum_{i=1}^n X_i] = \sum_{i=1}^n V[X_i] = \sum_{i=1}^n \pi(1 - \pi) = n\pi(1 - \pi)$$

Vediamo una semplice applicazione del modello Binomiale per meglio comprenderne le caratteristiche.

Immaginiamo di avere due sequenze nucleotidiche allineate:

Seq 1	A	G	T	A	G	C	T	T	A	G
		•	•		•	•		•		•
Seq 2	G	G	T	C	G	C	A	T	G	G

Capire se e quanto due sequenze siano simili è un problema che introntriamo ad esempio quando si osservano sequeze di DNA di due individui e l'ipotesi da verificare è se questi due individui abbiano un progenitore comune.

Per misurare la somiglianza di queste due sequenze iniziamo col chiederci con che probabilità in corrispondenza della stessa posizione si presenterà la stessa base (match) e quanti match ci attendiamo sull'intera sequenza. Una domanda di questo tipo non è banale: un match infatti potrebbe verificarsi del tutto casualmente oppure potrebbe essere influenzato dalla presenza di un progenitore comune.

Possiamo riformulare il quesito chiedendoci se il numero di match osservati è troppo elevato rispetto a quello che ci attenderemmo casualmente, fornendo in tal mod una evidenza a favore della presenza di un comune progenitore. Distinguere tra ciò che è casuale e ciò che corrisponde invece ad una relazione di causa effetto è un problema che incitreremo ancora in ambito inferenziale.

Se agisse soltanto il caso, cosa ci attenderemmo? Iniziamo con il formalizzare il problema utilizzando il linguaggio della probabilità. Indichiamo con X_i la presenza (1) o assenza (0) di un match in corrispondenza dell'iesimo sito. Si tratta ovviamente di una variabile aleatoria Bernoulliana dove π indica la probabilità di ossevare un match in quella posizione. Siamo interessati al numero totale di match osservati

nell'intera sequenza (Y) ma il passaggio ad un modello Binomiale richiede il verificarsi di precise assunzioni. La prima riguarda il fatto che il valore di π resti costante nelle diverse posizioni. La seconda assunzione è quella di indipendenza delle prove che, nel nostro caso, si traduce nell'ipotesi che osservare ad esempio un match in posizione 1 non modifichi la probabilità di osservare un match (o un mismatch) nella seconda posizione o in quelle successive. L'ipotesi di indipendenza tra nucleotidi è discussa in letteratura e spesso viene accettata al fine di semplificare i metodi anche se un certo grado di dipendenza tra siti, soprattutto se vicini, probabilmente non può essere escluso a priori.

Se accettiamo come vera l'indipendenza e ipotizziamo che le 4 basi compaiano in ciascun nucleotide in modo casuale allora la probabilità di osservare un match sarà costante e pari ad $1/4$ (provate a dimostrarlo formalmente).

Il numero osservato di match nel nostro esempio è pari a 6, maggiore del numero atteso per pura casualità pari a $E[Y] = n\pi = 10 \cdot 0.25 = 2.5$. Questa evidenza a favore dell'ipotesi di un progenitore comune è sufficiente a farci escludere la casualità? Torneremo sull'argomento più avanti.

Un aspetto interessante del modello Binomiale può essere illustrato tornando al lancio di un dado bilanciato. Immaginiamo di voler scommettere sul fatto che esca il numero 1. Pur essendo 6 i risultati possibili nulla osta a considerare come *successo* l'evento esce 1 e come *insuccesso* l'evento esce un numero diverso da 1. Se il dado è bilanciato la nostra probabilità di vincere sarà $1/6$. Ora immaginiamo di lanciare due volte il dado. L'intuizione ci porterebbe a credere che la nostra probabilità di vittoria raddoppi visto che abbiamo raddoppiato il numero dei lanci. E invece si dimostra che la probabilità che nei due lanci esca 1 almeno una volta è pari a $2\pi - \pi^2 = 2/6 - (1/6)^2 = 0.31$. Provate a dimostrarlo utilizzando proprio il modello Binomiale. Se poi lanciamo il dado 3 volte questa probabilità diventa $3\pi - 3\pi^2 + \pi^3 = 3/6 - 3(1/6)^2 + (1/6)^3 = 0.42$ e non $3\pi = 0.5$. Tuttavia l'intuito non è completamente fuori strada, si dimostra che la probabilità che esca 1 in n lanci di un dado bilanciato può essere approssimata ad $n\pi$ ma soltanto se π è molto molto piccolo, diciamo un dado bilanciato con 100 faccie.

Vediamo come dare una rappresentazione grafica della distribuzione Binomiale (e più in generale di una distribuzione nel discreto). Ricordiamo che il modello Binomiale è una famiglia di distribuzioni quindi dovremo fissare i valori di n e π per poi riportare sull'asse X i valori osservabili (Ω_Y). A ciascun valore corrisponderà una barra verticale la cui lunghezza sarà pari alla sua probabilità di essere osservato.

Nel grafico 2.1 sono sovrapposte 3 diverse distribuzioni Binomiali con $n=20$ e rispettivamente $\pi = 0.1$, $\pi = 0.5$ e $\pi = 0.9$. Solo nel caso $\pi = 0.5$ la distribuzione risulta simmetrica rispetto al suo valore atteso, essendo uguale la probabilità di successo o insuccesso nella singola prova. Se l'evento è raro sarà ovviamente più probabile osservare pochi successi mentre se la probabilità di successo è molto alta sarà probabile osservare sul totale delle prove un numero elevato di successi.

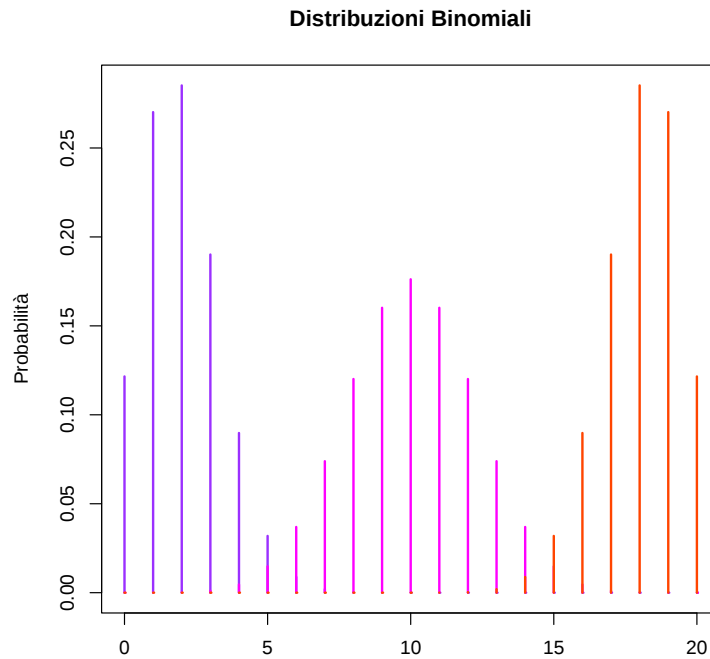


Figura 2.1: Distribuzioni Binomiali al variare di π

2.1.4 La famiglia Binomiale Negativa

Come nel modello Binomiale immaginiamo prove indipendenti tutte ugualmente distribuite secondo una legge Bernoulliana con π costante. Cambiamo però la regola di arresto: invece di fermarci dopo n prove terminiamo il nostro esperimento quando abbiamo ottenuto il k -esimo successo. Ne segue che adesso, essendo fissato il numero di k successi, sarà aleatorio il numero N di prove che dovremo eseguire per ottenerli. L'insieme Ω_N sarà

$$\Omega_N = \{k, \dots\}$$

Da notare che per ottenere k successi è necessario fare almeno k prove.

La distribuzione di probabilità di N può essere ricavata osservando che, per ottenere k successi in n prove dobbiamo ottenere $(k-1)$ successi nelle prime $(n-1)$ prove ed un successo nell'ultima prova eseguita

$$\begin{aligned} Pr\{N = n; \pi, k\} &= \binom{n-1}{k-1} \pi^{k-1} (1 - \pi)^{(n-1)-(k-1)} \pi = \\ &= \binom{n-1}{k-1} \pi^k (1 - \pi)^{n-k} \end{aligned}$$

Come si può notare la distribuzione Binomiale Negativa differisce dalla distribuzione Binomiale soltanto per un termine: mentre nella prima è presente $\binom{n-1}{k-1}$ nella seconda è presente $\binom{n}{k}$. Questa piccola differenza ha generato molte discussioni in ambito inferenziale su cui torneremo nel prosieguo.

2.1.5 Il modello Geometrico

Restiamo ancora nel nostro esperimento fatto di prove Bernoulliane indipendenti con probabilità di successo costante π . La variabile aleatoria Y che descrive la distribuzione Geometrica raccoglie il numero di prove che necessarie **prima** di osservare il primo fallimento (che resta escluso dal conteggio). Intuitivamente abbiamo che

$$\Omega_Y = \{0, 1, 2, \dots\}$$

Si noti come questo insieme contenga anche lo 0 che corrisponde al caso in cui già nella prima prova si osserva un fallimento.

La probabilità di eseguire y prove prima di osservare il primo fallimento sarà

$$Pr\{Y = y; \pi\} = (1 - \pi)^y \pi$$

Y viene a volte indicata come la lunghezza di un *success run* e trova applicazione negli studi di affidabilità ad esempio per componenti meccaniche. Possiamo ritrovarla anche nel confronto tra sequenze allineate di nucleotidi: nell'esempio visto in precedenza, se assumiamo come successo la presenza di un match, la lunghezza massima osservata per un *success run* è stata 2.

2.1.6 Il modello di Poisson

Tra le famiglie discrete non possiamo non menzionare il modello di Poisson che descrive la probabilità che un evento aleatorio accada un certo numero di volte in un determinato intervallo di tempo. Una variabile aleatoria di Poisson X è caratterizzata dalla distribuzione

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

con $\Omega_X = [0, 1, \dots, +\infty)$

Perchè questo modello sia applicabile devo tuttavia verificarsi alcune condizioni: il verificarsi di un primo evento non deve influenzare la probabilità che si verifichi un secondo evento e due eventi non possono verificarsi simultaneamente. Inoltre deve essere compatibile con i dati sperimentali la caratteristica vincolante del modello che $E(X) = V(X) = \lambda$.

Il parametro λ , rappresentando il valore atteso di X nell'intervallo di tempo considerato, è strettamente legato alla sua lunghezza. Per questo motivo può essere conveniente riscrivere $\lambda = \alpha t$, dove α è il tasso medio con cui si presenta l'evento in un intervallo unitario di tempo. Questa diversa parametrizzazione ci consente di chiarire l'ulteriore condizione alla base del modello che vuole α indipendente dal verificarsi di un qualsiasi evento. Spesso questa assunzione è riportata per semplicità come α costante anche se non è strettamente necessario.

Possiamo ad esempio utilizzare il modello di Poisson per descrivere in probabilità il numero di asteroidi con una prefissata dimensione minima che impattano la terra in un anno solare o il numero di difetti che presenta una certa superficie. Notiamo che, in questo ultimo caso, dovremo specificare la dimensione della superficie a cui facciamo riferimento ed il tempo verrà sostituito dallo spazio.

L'utilizzo del modello di Poisson, come di qualsiasi altro modello parametrico, dovrebbe sempre essere accompagnato da una attenta verifica della coerenza tra le caratteristiche dell'esperimento e dei dati osservati e i vincoli e le proprietà del modello. Ad esempio il modello di Poisson potrebbe non essere adatto a descrivere il numero di terremoti con una magnitudo superiore ad una data soglia, registrati in certa area in un prefissato intervallo di tempo (si noti in questo caso la presenza di tempo e spazio). Infatti il verificarsi di un terremoto in genere aumenta la probabilità di nuovi eventi nei giorni successivi.

Anche le caratteristiche del modello possono non adattarsi ai dati osservati: può accadere che la variabilità osservata sia superiore al valore atteso oppure che la frequenza di 0 eventi sia più elevata rispetto

alla probabilità prevista in base al modello di Poisson. Si parla in questi casi di *overdispersion* e di *zero inflated Poisson model*.

2.2 Variabili aleatorie continue

2.2.1 Dal discreto al continuo, un passaggio delicato

Abbiamo visto come nel caso di variabili aleatorie discrete la probabilità unitaria totale venga distribuita sui singoli eventi. Ad ogni singolo valore in Ω verrà assegnata la sua probabilità di essere osservato ricordando che, nel rispetto dei primi due assiomi di Kolmogorov, queste probabilità dovranno essere sempre maggiori o uguali a 0 e la loro somma dovrà essere 1. Il grafico 2.2 mostra sovrapposte 3 distribuzioni Binomiali $\pi = 0.15$ e rispettivamente $n = 10$, $n = 50$, $n = 100$ e $n = 150$.

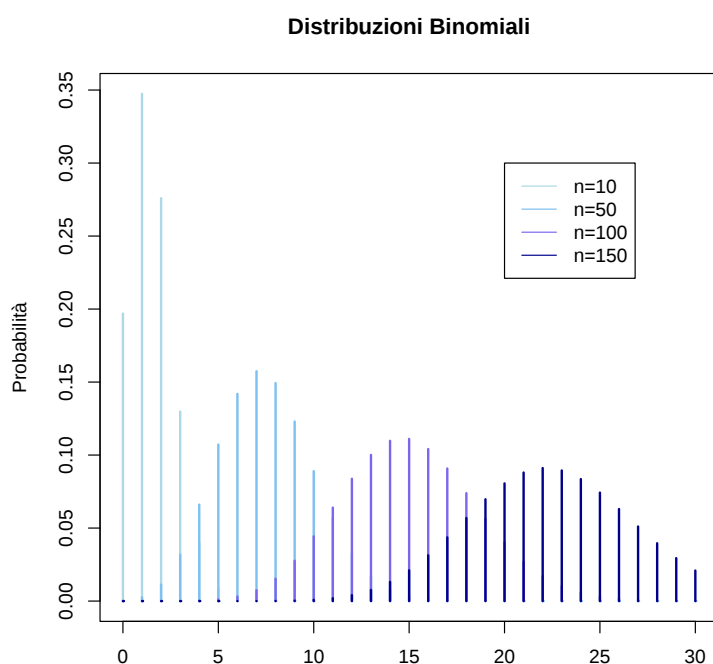


Figura 2.2: Distribuzioni Binomiali al crescere di n

Osserviamo come, al crescere di n , le distribuzioni tendano a *schiacciarsi* sull'asse X , con il valore massimo della probabilità che passa da 0.35 a circa 0.1. Cosa accade? Al crescere di n cresce il numero degli eventi osservabili in Ω e, di conseguenza, dovendo distribuire una

massa unitaria di probabilità su un numero crescente di valori, ciò che spetta a ciascun valore sarà sempre di meno. Se proviamo a spingere n ad infinito la probabilità di ciascun punto diventerà prossima a 0. Ora cosa accade se passiamo da una infinità numerabile ad una infinità continua di punti? La probabilità del singolo punto diventa necessariamente zero, in quanto non esiste un modo di assegnare una probabilità diversa da 0 a un'infinità continua di punti mantenendo il vincolo che la somma sia 1. I valori osservabili diventano davvero troppi, basti pensare al fatto che nell'intervallo aperto $(0, 1]$ nel discreto non c'è nessun valore osservabile, mentre nel continuo i valori osservabili sono già infiniti.

Immaginiamo di essere ad una festa di compleanno di dividere una torta tra 10 bambini: ad ogni bambino daremo una fetta, qualcuna più piccola, qualcuna più grande. Ora immaginiamo di avere la stessa torta da dividere questa volta tra 50 bambini. Per quanto ci sforziamo di tagliare fette sottili non riusciremo mai a ricavarne 50. Se vogliamo evitare il peggio dobbiamo cambiare strategia. Continuiamo a fare 10 fette ma adesso diano una fetta a 5 bambini. E speriamo che le conseguenze non siano troppo gravi!

In probabilità facciamo qualcosa di simile. Nle continuo, non riuscendo ad assegnare una probabilità a ciascun punto, andiamo ad assegnarla ad un intervallo, cioè ad un insieme di infiniti punti. Tuttavia, per definire formalmente la probabilità, dobbiamo passare per la definizione di *densità di probabilità* che sarà una qualsiasi funzione $f(x)$ da Ω in $\mathbb{R}^+$ che soddisfi due condizioni:

- $f(x) \geq 0$
- $\int_{\Omega} f(t)dt = 1$

Ricordiamo che adesso *Omega* coincide con $\mathbb{R}$ o con un suo sottoinsieme. La probabilità di osservare valori compresi in un dato intervallo sarà l'area corrispondente sotto questa funzione. Più formalmente, se X è una variabile aleatoria continua

$$P(x \leq X < x + dx) = \int_x^{x+dx} f_X(t)dt \approx f(x)dx$$

La figura 2.3 mostra un esempio di densità di probabilità dove l'area in rosso indica la probabilità di osservare valori tra 20 e 30.

Sottolineiamo alcuni punti importanti.

- Quando parliamo di probabilità con riferimento ad una variabile aleatoria continua dobbiamo sempre riferirci ad intervalli e la probabilità corrisponderà sempre ad un'area sotto una densità.

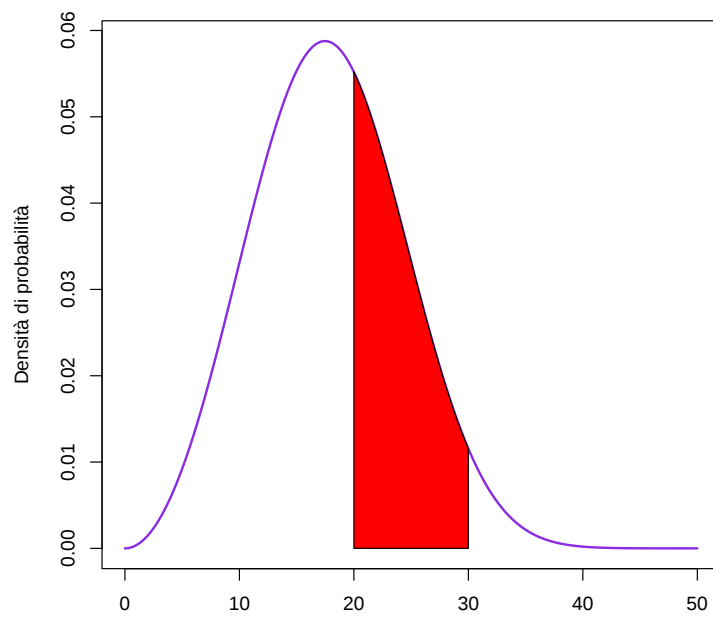


Figura 2.3: Un esempio di densità

- Sull'asse delle ordinate non abbiamo, come nel caso discreto, delle probabilità se semplicemente i valori che può assumere la funzione $f_X(x)$, cioè *densità di probabilità*.
- La forma di $f_X(x)$ decide come le probabilità viene distribuita sull'asse reale. Dove la curva è più elevata avremo intervalli con probabilità maggiori, dove la curva è più bassa con probabilità minori, ovviamente a parità di ampiezza. Dove la curva è vicina a 0, avremo eventi improbabili.
- La condizione $f(x) \geq 0$ assicura che venga rispettato il primo assioma di Kolmogorov.
- La condizione $\int_{\Omega} f(t)dt = 1$ garantisce il rispetto del secondo assioma $P(\Omega) = 1$.
- La probabilità di osservare valori in un intervallo infinitesimo $(x, x + dx)$ può essere approssimata con l'area del rettangolo di base dx ed altezza $f(x)$, confermando che i valori di $f(x)$ non sono delle probabilità.
- Quando Ω è l'intera retta reale la condizione $\int_{\Omega} f(t)dt = 1$ non è semplice da realizzare. Infatti l'integrale può facilmente divergere se la funzione non tende velocemente per x che tende a $\pm\infty$. Le densità che vedremo hanno tutte *forme chiuse* o perché Ω è un sottoinsieme dei reali o perché la curva converge rapidamente verso zero in entrambe le code.

2.2.2 La variabile aleatoria Uniforme

Abbiamo già incontrato la distribuzione uniforme nel discreto, ad esempio nel lancio di un dado. In questa semplice prova i risultati osservabili sono 6 e, se il dado è bilanciato, ognuno di essi ha la stessa probabilità di essere osservato. Proprio quest'ultima condizione di equiprobabilità degli eventi elementari definisce il modello uniforme.

È possibile adattare questa definizione al continuo? Abbiamo visto come la probabilità debba adesso riferirsi ad intervalli e non più a singoli valori osservabili. Ne segue che l'idea di equiprobabilità riguarderà adesso tutti gli intervalli osservabili *a parità di ampiezza* (quest'ultima condizione è essenziale altrimenti la condizione di uguale probabilità perderebbe senso).

Perché questo sia vero la densità di probabilità di una variabile aleatoria uniforme Y dovrà essere costante ($f_Y(x) = c$) o, in altri termini, una retta parallela all'asse X . Osserviamo tuttavia che, se Ω_Y è l'intera

retta reale, viene violata la seconda condizione richiesta ad una densità dal momento che l'integrale di una costante su $\mathbb{R}$ diverge. Si parla in questo caso di densità *impropria*.

La distribuzione uniforme è definibile in modo proprio soltanto se Ω_Y è un intervallo finito $[a, b]$. Allora la sua densità uniforme diventerà

$$f_Y(x) = c = \frac{1}{b-a} \quad x \in [a, b]$$

Provate ora ad immaginare un esperimento i cui possibili risultati, ovvero intervalli osservabili nel continuo a parità di ampiezza, abbiano la stessa probabilità.

2.2.3 Il modello Normale

La Normale o Gaussiana è una famiglia di densità che, svolge un ruolo centrale nell'inferenza statistica. Formalmente parliamo di una variabile aleatoria Y definita su tutta la retta reale, $\Omega = \mathbb{R}$, e la cui densità ha l'espressione

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

Si tratta, come dicevamo, di una famiglia di densità, una per ciascuna coppia di valori di μ e σ^2 che hanno caratteristiche comuni.

Prima di descriverle in dettaglio apriamo una parentesi per tornare sulla definizione di valore atteso e varianza di una variabile aleatoria. Abbiamo visto la loro definizione nel discreto e il passaggio al continuo richiede qualche modifica. Infatti, nel caso di variabili aleatorie continue il valore atteso diventa

$$E[X] = \int_{\Omega} x f_X(x) dx$$

mentre la varianza

$$V[X] = \int_{\Omega} (x - E[X])^2 f_X(x) dx$$

Osserviamo come, rispetto al discreto, l'integrale ha preso il posto della sommatoria e le probabilità dei valori osservabili sono adesso sostituite da $f_X(x)dx$, rimanendo invariato il senso delle due quantità.

Nel caso di una variabile aleatoria Normale i due parametri μ e σ^2 corrispondono proprio al valore atteso e alla varianza

$$E[Y] = \int_{\mathbb{R}} y \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy = \mu$$

$$V[Y] = \int_{\mathbb{R}} (y-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy = \sigma^2$$

Ai più coraggiosi il compito di dimostrare queste uguaglianze.

Torniamo alle caratteristiche della famiglia Normale per osservare che tutte le sue densità hanno un solo punto di massimo proprio in corrispondenza di $x = \mu$ per poi decrescere in modo simmetrico tendendo a 0 al tendere di x a $\pm\infty$.

Ne segue che in un modello Normale i valori centrali intorno a μ sono i più probabili, intervalli a destra e sinistra di μ , a parità di ampiezza e di distanza da μ , avranno la stessa probabilità, mentre diventano molto improbabili valori molto distanti da quello atteso.

Queste caratteristiche del modello sono vincolanti e dovranno essere coerenti con il fenomeno oggetto di studio tutte le volte che andremo ad *assumere un modello Normale*. Al contrario μ e σ^2 sono elementi di flessibilità che andranno ad adattarsi ai dati sperimentali. In particolare μ decide la posizione della densità sull'asse reale mentre σ^2 controlla quanto la curva è concentrata intorno a μ . Se la varianza è bassa la curva si concentra intorno al suo valore atteso e più prevedibile è il risultato della nostra osservazione; se la varianza è alta la curva è più dispersa e più imprevedibile è il valore che osserveremo. In figura 2.4 sono riportati alcuni esempi di densità Normali al variare del valore atteso e della varianza.

Tra le densità Normali la Normale Standard $Z \sim N(0, 1)$, caratterizzata da $\mu = 0$ e $\sigma^2 = 1$, ha un ruolo particolare. Ricordiamo che la probabilità di osservare valori in un certo intervallo nel continuo è definita come l'area sotto la densità e il suo calcolo richiede una integrazione. Come per molte altre funzioni la densità Normale non è integrabile e quindi questo calcolo richiede metodi numerici. Per la Normale Standard sono riportati in tavole gli integrali riferiti ad intervalli del tipo $(-\infty, x)$. Ad esempio la probabilità di osservare valori nell'intervallo $(-\infty, -1.96)$ è pari a 0.025 e, dalla simmetria di questa funzione, segue che $P(-1.96 < Z < 1.96) = 0.95$. Ritroveremo questi valori più avanti nell'ambito della statistica inferenziale. Dalle probabilità di intervalli del tipo $(-\infty, x)$ è possibile ricavare le probabilità di qualsiasi intervallo (come?). Tuttavia possiamo chiederci cosa accade se la nostra densità Normale non è quella standard. Sia

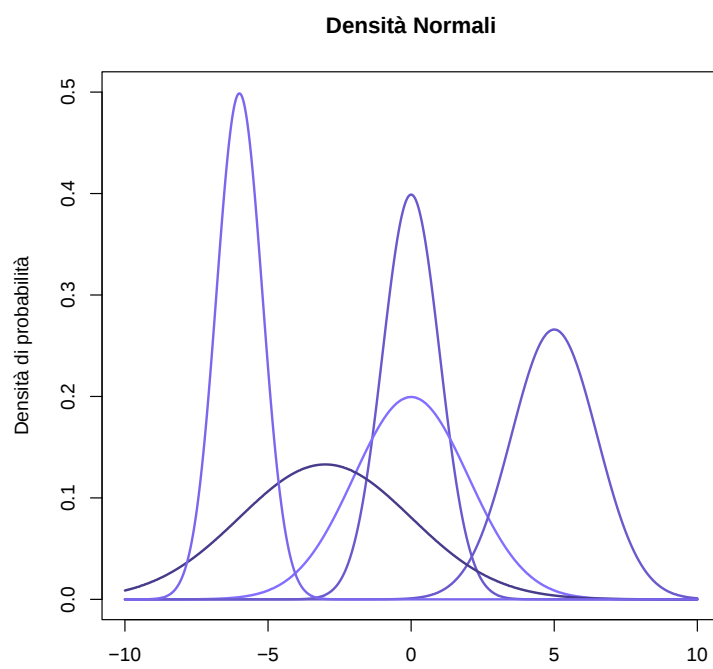


Figura 2.4: Esempi di densità Normali

$Y \sim N(3, 4)$ e ipotizziamo di voler calcolare $P(Y < 2)$. Con semplici passaggi possiamo ricondurci a $Z \sim N(0, 1)$

$$P(Y < 2) = P(Y - 3 < 2 - 3) = P\left(\frac{Y - 3}{\sqrt{4}} < \frac{2 - 3}{\sqrt{4}}\right) = P(Z < -1/2)$$

Questa trasformazione, nota come *standardizzazione*, prevede prima una traslazione: sottraendo ad Y (cioè a tutti i valori che può assumere) il suo valore atteso portiamo quest'ultimo a 0. Per ricondurre ad 1 la varianza operiamo un cambiamento di scala dividendo per la deviazione standard. Vale la pena sottolineare che in questo secondo passaggio è proprio la deviazione standard a diventare la nostra nuova unità di misura. Spesso, erroneamente si attribuisce alla *standardizzazione* la capacità di *normalizzare* la densità di partenza. In realtà standardizzando semplicemente portiamo a 0 ed 1 rispettivamente il valore atteso e la varianza della nostra variabile aleatoria che seguirà un modello Normale soltanto se lo era in origine, come nel nostro caso.

Nel precedente esempio, la standardizzazione è stata eseguita su una variabile aleatoria. La stessa trasformazione può essere fatta su una variabile statistica. Ipotizziamo, ad esempio, di aver rilevato l'altezza di 100 individui e di avere calcolato media e varianza. Se ora sottraiamo a ciascun valore osservato la media e lo dividiamo per la deviazione standard otterremo nuovi dati con media 0 e varianza 1.

2.2.4 Il modello Esponenziale

Una variabile aleatoria X appartenente a questo modello è caratterizzata dalla seguente densità di probabilità

$$f_X(x) = \lambda e^{-\lambda x}$$

dove $\Omega_X = [0, +\infty)$. In questa distribuzione, la variabile aleatoria X che si distribuisce secondo il modello esponenziale ($X \sim \text{Exp}(\lambda)$) possiede $E[X] = 1/\lambda$ e $E[X^2] = 1/\lambda^2$ rispettivamente come valore atteso e varianza.

Si tratta di una famiglia di densità spesso utilizzata per modellare durate di vita e può essere considerata una variante continua della distribuzione Geometrica poichè calcola il tempo fino al verificarsi del primo fallimento. Non è un caso che sia la distribuzione Esponenziale sia quella Geometrica soddisfino una proprietà nota come *assenza di memoria*, in simboli

$P(X > s + t | X > s) = P(X > t)$ Se immaginiamo che X descriva il

tempo di attesa per il verificarsi di un evento aleatorio, questa proprietà implica che la probabilità di attendere un tempo t dopo che abbiamo già atteso un tempo s senza che l'evento si sia verificato è la stessa che avevamo inizialmente. Il fatto che il tempo che passa non incida sulla probabilità di fallimento rende questo modello adatto a descrivere ad esempio la durata di vita di componenti elettroniche. Potremmo utilizzare questo modello per descrivere il tempo di attesa per il prossimo autobus? Da una diversa prospettiva l'assenza di memoria può essere letta come la presenza di un rischio di fallimento che rimane costante nel tempo. Se questa condizione risulta troppo restrittiva è possibile passare a modelli più flessibili come la famiglia Weibull o Gamma che ammettono un rischio monotono e quindi crescente o decrescente nel tempo e conetngono il modello Esponenziale come caso particolare.

2.3 Ancora sull'indipendenza

Ci siamo già trovati, e accadrà ancora nel trattare i metodi inferenziali, a considerare simultaneamente più variabili aleatorie descritte da un vettore $(X_1, X_2, \dots, X_n)$. Nonostante la trattazione delle variabili aleatorie multiple vada oltre gli scopi di questo testo, diventa importante definire alcuni concetti. Per semplicità tratteremo il caso di un vettore di due variabili, essendo immediata l'estensione a dimensioni maggiori. In questo caso per descrivere $\Omega_{(X_1, X_2)}$ avremo bisogno di due dimensioni; il lancio di due dadi è l'esempio più elementare. Se indichiamo con X_1 e X_2 il risultato del primo e del secondo lancio, l'insieme $\Omega_{(X_1, X_2)}$ sarà

	1	2	3	4	5	6
1	1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2	2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3	3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5	5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6	6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

dove abbiamo riportato a margine Ω_{X_1} e Ω_{X_2} . Assumeremo quasi sempre che le variabili aleatorie considerate siano tra loro indipendenti. Abbiamo definito l'indipendenza nel caso di eventi e, sebbene sia intuitivo nel caso dei dadi, diventa necessario estendere questo concetto alle variabili aleatorie. In generale, nel caso discreto, la distribuzione congiunta di X_1 ed X_2 descrive la probabilità di osservare ciascuna delle coppie in $\Omega_{(X_1, X_2)}$ e due variabili aleatorie sono indipendenti se l'accadere di un qualsiasi evento che riguarda X_1 non modifica la probabilità

di un qualsiasi evento relativo ad X_2 . Formalmente:

$$P(X_1 = i, X_2 = j) = P(X_1 = i) \cdot P(X_2 = j) \text{ per ogni } (i, j) \in \Omega_{(X_1, X_2)}.$$

Questa definizione si estende al caso continuo: se indichiamo con X e Y due variabili aleatorie continue e con $f_{(X,Y)}(x, y)$ la loro densità congiunta la condizione di indipendenza diventa

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y) \text{ per ogni } (x, y) \in \Omega_{(X,Y)}.$$

2.4 Due leggi importanti

La teoria della probabilità è vasta e contiene molti importanti teoremi. Accenniamo soltanto a due di essi nella loro versione più semplice perchè ad essi faremo riferimento nel prosieguo quanto tratteremo le basi dell'inferenza statistica.

2.4.1 La legge dei Grandi Numeri

La definizione frequentista della probabilità aveva già evidenziato i legami tra probabilità e frequenze relative. Questo teorema torna su questo legame ma con il valore di un risultato dimostrato. Sia $\{X_n\}$ una successione di variabili aleatorie indipendenti ed identicamente distribuite (diremo da ora in poi i.i.d.). con valore atteso $E(X_n) = \mu$ e varianza $V(X_n) = \sigma^2$ entrambi finiti. Allora si dimostra che

$$\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n} \rightarrow_{n \rightarrow \infty} \mu$$

Nel caso particolare in cui le variabili aleatorie seguono una distribuzione Bernoulliana con $E(X_n) = \pi$ la media coinciderà con la frequenza relativa dei successi che tenderà a $\mu = \pi$ al tendere di n ad infinito. Più in generale, se siamo interessati alla probabilità di un evento del tipo $(Y_i \in C)$ dove Y_i è una variabile aleatoria continua e $C = (a, b)$ potremo ricondurci al caso Bernoulliano definendo una variabile aleatoria indicatrice dell'evento $X_i = 1$ se $Y_i \in C$ e $X_i = 0$ altrimenti.

Purtoppo questa legge è citata spesso a sproposito o in modo eccessivamente ottimistico; alcune persone, accanite giocatrici al lotto, scommettono sui numeri che ritardano confidando sul fatto che necessariamente dovranno uscire ma dimenticando che la convergenza è all'infinito.

2.4.2 Il Teorema Centrale del Limite

Come per il teorema precedente sia $\{X_n\}$ una successione di variabili aleatorie i.i.d. con valore atteso $E(X_n) = \mu$ e varianza $V(X_n) = \sigma^2$ entrambe finite. Allora la loro media standardizzata convergerà in distribuzione ad una variabile aleatoria Normale standard. In simboli

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \xrightarrow{n \rightarrow \infty} Z \sim N(0, 1)$$

Si tratta di un risultato che svolge un ruolo importante in ambito inferenziale. Purtroppo anch'esso viene a volte interpretato male: ricordate che è la media a normalizzarsi e non le singole variabili aleatorie che continueranno a seguire la loro distribuzione originaria. Se non potremo assumere nel nostro studio un modello normale questo teorema potrà aiutarci ma non cambiare la situazione.

Capitolo 3

Esercizi

Sugli eventi

Dati tre eventi A , B , e C , esprimere, utilizzando unione ed intersezione, gli eventi che seguono:

1. si verifica almeno uno dei tre eventi;
2. si verifica al massimo uno dei tre eventi;
3. nessuno dei tre eventi si verifica;
4. tutti gli eventi si verificano;
5. si verifica A e B ma non C ;
6. si verifica solo uno dei tre eventi.

1. $A \cup B \cup C$;
2. $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \cup (A^c \cap B^c \cap C^c)$;
3. $(A^c \cap B^c \cap C^c)$;
4. $(A \cap B \cap C)$;
5. $(A \cap B \cap C^c)$;
6. $(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)$.

Ancora sugli eventi

Dimostrare che, per ogni coppia di eventi A e B, $A \cup B = A \cup (A^c \cap B)$
 Per la proprietà distributiva e la definizione di complemento possiamo scrivere

$$\begin{aligned} A \cup (A^c \cap B) &= (A \cup A^c) \cap (A \cup B) \\ &= \Omega \cap (A \cup B) \\ &= (A \cup B) \end{aligned}$$

Eventi indipendenti

Siano A e B due eventi indipendenti tali che

$$P(A) = 0.8 \text{ e } P(B^c) = 0.4$$

1) Quanto vale $P(B)$?

$$P(B) = 1 - P(B^c) = 1 - 0.4 = 0.6$$

2) Quanto vale $P(A \cap B)$?

$$P(A \cap B) = (\text{per indipendenza}) P(A) \cdot P(B) = 0.8 \cdot 0.6 = 0.48$$

3) Quanto vale $P(A \cup B)$?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.6 - 0.48 = 0.92$$

3) Quanto vale $P(A^c \cap B)$?

$$\begin{aligned} P(A^c \cap B) &= (\text{per indipendenza}) P(A^c) \cdot P(B) = \\ &= (1 - P(A)) \cdot P(B) = (1 - 0.8) \cdot 0.6 = 0.2 \cdot 0.6 = 0.12 \end{aligned}$$

Strani eventi

Esistono in probabilità degli eventi particolari che prendono il nome di eventi *quasi certi* e *quasi impossibili*. Questi eventi, pur essendo diversi dall'insieme vuoto e da Ω hanno rispettivamente probabilità 0 ed 1 di verificarsi e si comportano (in probabilità) come se fossero l'insieme vuoto ed Ω .

Infatti se B è quasi impossibile ($P(B)=0$) allora $P(A \cap B) = 0$ e $P(A \cup B) = P(A)$. Proviamo a dimostrarlo

$$P(A \cap B) \leq P(B) = 0 \text{ poich\`e } (A \cap B) \subseteq B$$

In base al primo assioma la probabilità non può essere minore di 0 da cui $P(A \cap B) = 0$.

Scriviamo $(A \cup B) = A \cup (B \cap A^c)$ dove questi ultimi due eventi sono incompatibili.

Infatti, utilizzando la proprietà associativa e la definizione di evento complementare, abbiamo $A \cap (B \cap A^c) = B \cap (A \cap A^c) = B \cap \emptyset = \emptyset$.

Possiamo quindi applicare il terzo assioma e sfruttare quanto appena dimostrato

$$P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) = P(A) + 0 = P(A)$$

Se invece B è un evento quasi certo ($P(B)=1$) allora $P(A \cap B) = P(A)$ e $P(A \cup B) = 1$. Lasciamo al lettore la dimostrazione.

L'urna colorata

In un'urna si trovano 10 palline, 5 viola e 5 arancioni. Calcolare la probabilità che, in due estrazioni successive, escano due palline dello stesso colore sia reinserendo sia non reinserendo la pallina nell'urna dopo la prima estrazione.

Prima di tutto costruiamo lo spazio dei possibili eventi Ω

$$\Omega = \{(A1, A2), (V1, V2), (A1, V2), (V1, A2)\}$$

Riscriviamo l'evento che escano due palline dello stesso colore come unione di eventi elementari e quindi incompatibili

$$\{(A1 \cap A2) \cup (V1 \cap V2)\}$$

In base al 3° assioma, possiamo scrivere:

$$P\{(A1 \cap A2) \cup (V1 \cap V2)\} = P(A1 \cap A2) + P(V1 \cap V2)$$

Se reinseriamo nell'urna la pallina dopo la prima estrazione, gli eventi nelle intersezioni sono Indipendenti.

Avremo pertanto

$$P\{(A1 \cap A2) \cup (V1 \cap V2)\} = P(A1) \cdot P(A2) + P(V1) \cdot P(V2)$$

$$P\{(A1 \cap A2) \cup (V1 \cap V2)\} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

Se non reinseriamo la pallina estratta nell'urna avremo ancora

$$P\{(A1 \cap A2) \cup (V1 \cap V2)\} = P(A1 \cap A2) + P(V1 \cap V2)$$

In questo caso gli eventi nelle intersezioni sono dipendenti, applichiamo quindi la legge delle Probabilità Composte $P(A \cap B) = P(B) \cdot P(A|B)$

$$P(A1 \cap A2) + P(V1 \cap V2) = P(A1) \cdot P(A2|A1) + P(V1) \cdot P(V2|V1)$$

$$P\{(A1 \cap A2) \cup (V1 \cap V2)\} = \frac{5}{10} \cdot \frac{4}{9} + \frac{5}{10} \cdot \frac{4}{9} = \frac{4}{9}$$

Il medico indeciso

Un medico si trova a dover decidere se sia preferibile somministrare al paziente immediatamente una nuova terapia antivirale sperimentale o provare inizialmente una terapia standard e passare al farmaco sperimentale solo in caso di fallimento. Teniamo conto che la probabilità di risposta della terapia standard è 0.4, mentre se somministriamo subito il nuovo antivirale la probabilità di guarigione $P(A)$ è 0.8. Tuttavia la terapia standard, in uso da molti anni, a fronte di una minore probabilità di risposta, garantisce una maggiore affidabilità rispetto agli effetti secondari. Il dubbio riguarda il rischio che l'efficacia del nuovo farmaco possa diminuire se somministrato dopo un fallimento. Qual è la probabilità di risposta minima, dopo fallimento della prima terapia, perchè sia attuabile la terapia in due fasi? Definiamo gli eventi $A1 = \{\text{Il paziente risponde al primo farmaco}\}$ con $P(A1) = 0.4$ e $A2 = \{\text{Il paziente risponde al secondo farmaco}\}$. Avremo

$$\Omega = \{(A1 \cap A2), (A1^c \cap A2), (A1^c \cap A2^c), (A1 \cap A2^c)\}$$

Calcoliamo $P(A2)$ nel caso di doppia somministrazione

$$P(A2) = P(A1) \cdot P(A2|A1) + P(A1^c) \cdot P(A2|A1^c)$$

Assumiamo

$$P(A2|A1) = 1$$

poiché se ha avuto effetto il 1° farmaco, il paziente è già guarito.

$$P(A1^C) = 1 - P(A1)$$

La probabilità di risposta al secondo farmaco dopo fallimento del primo rappresenta la mia incognita. Per semplicità scriviamo $P(A2|A1^C) = x$ e andiamo a sostituire i valori

$$P(A2) = 0.4 \cdot 1 + (1 - 0.4) \cdot x$$

$P(A2)$ deve essere almeno 0,8

$$0.8 \leq 0.4 \cdot 1 + (1 - 0.4) \cdot x$$

$$0.8 \leq 0.4 + 0.6x$$

$$0.6x \geq 0.8 - 0.4$$

$$x \geq \frac{0.4}{0.6} = 0.7$$

La terapia in due fasi è accettabile solo se la probabilità di risposta del farmaco sperimentale a seguito di fallimento del farmaco standard resta di almeno 0.7

Le monete di Pinocchio

Lanciamo 3 monete bilanciate una dopo l'altra.

1. Calcolare la probabilità dei seguenti eventi

- $A = \{\text{La prima testa esce al terzo lancio}\}$
- $B = \{\text{Esce due volte testa}\}$
- $C = \{\text{Esce croce al secondo lancio}\}$

Avremo

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

dove, per semplicità, abbiamo omissso il simbolo di intersezione all'interno di ciascuna terna.

Poichè i risultati dei singoli lanci sono indipendenti e la moneta bilanciata, gli eventi in Ω sono equiprobabili avendo ciascuno probabilità $1/8$ di verificarsi. Se infatti indichiamo con $(ijk) = (icapj \cap k)$ un generico evento in Ω , dove i, j e k possono essere T o C, avremo, per l'indipendenza dei lanci $P(i, j, k) = P(i) \cdot P(j) \cdot P(k)$. Essendo la moneta bilanciata ciascuna probabilità sarà pari ad $1/2$ da cui $P(i, j, k) = P(i) \cdot P(j) \cdot P(k) = 1/2 \cdot 1/2 \cdot 1/2 = 1/8$.

La condizione di equiprobabilità è essenziale per poter utilizzare la definizione classica di probabilità (casi favorevoli su casi possibili). Essendo verificata potremo applicare la definizione classica nel calcolo delle probabilità a seguire.

$$A = \{(CCT)\}$$

$$P(A) = \frac{1}{8}$$

$$B = \{(TTC), (TCT), (CTT)\}$$

Come in Ω la virgola indica l'unione degli eventi che sono tra loro incompatibili.

$$P(B) = \frac{3}{8}$$

$$C = \{(TCT), (TCC), (CCT), (CCC)\}$$

$$P(C) = \frac{4}{8} = \frac{1}{2}$$

2. Analizzare le possibili coppie di eventi e chiarire quali sono incompatibili. Ricordiamo che affinché due eventi siano incompatibili la loro intersezione deve essere $= \emptyset$

$$1) (A \cap B) = \emptyset \rightarrow \text{sono eventi incompatibili}$$

$$2) (A \cap C) = \{(CCT)\} \rightarrow \text{non sono eventi incompatibili}$$

$$3) (B \cap C) = \{(TCT)\} \rightarrow \text{non sono eventi incompatibili}$$

3. Gli eventi A, B e C costituiscono una partizione di Ω

No, gli eventi indicati non sono una partizione di Ω poiché non sono disgiunti a due a due e la loro unione

$(A \cup B \cup C) = \{(CCT), (TTC), (TCT), (CTT), (TCC), (CCC)\}$
è diversa da Ω .

3. Analizzare gli eventi a due a due e valutare quali sono indipendenti

Iniziamo considerando gli eventi A e B; se fossero indipendenti dovrebbe valere la seguente uguaglianza

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

Abbiamo

$$P(A \cap B) = 0 \neq \frac{1}{8} \cdot \frac{3}{8} = P(A) \cdot P(B)$$

Dal momento che l'uguaglianza non è verificata gli eventi sono dipendenti

$$P(A \cap C) = \frac{1}{8} \neq \frac{1}{8} \cdot \frac{1}{2} = P(A) \cdot P(C)$$

Anche in questo caso si tratta di eventi dipendenti

$$P(B \cap C) = \frac{1}{8} \neq \frac{3}{8} \cdot \frac{1}{2} = P(B) \cdot P(C)$$

Ancora una volta si tratta di eventi dipendenti

Lanciamo i dadi

Un dado viene lanciato due volte in successione. Indichiamo con X_1 e X_2 il risultato del primo e del secondo lancio

Calcolare la distribuzione della somma dei due risultati $Y = X_1 + X_2$

$X_1 = \text{"Risultato 1 lancio"} \longrightarrow \Omega_{X_1} = \{1, 2, 3, 4, 5, 6\}$, $P(X_1 = k =$

$$\frac{1}{6}) \quad \forall k \in \Omega_{X_1}$$

$$X_2 = \text{“Risultato 2 lancio”} \longrightarrow \Omega_{X_2} = \{1, 2, 3, 4, 5, 6\}, \quad P(X_2 = j = \frac{1}{6}) \quad \forall j \in \Omega_{X_2}$$

$$Y = X_1 + X_2 \longrightarrow \Omega_Y = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

Riportiamo nella matrice a seguire tutti i risultati osservabili per il vettore aleatorio (X_1, X_2) ($\Omega_{(X_1, X_2)}$)

	1	2	3	4	5	6
1	1, 1	1, 2	1, 3	1, 4	1, 5	1, 6
2	2, 1	2, 2	2, 3	2, 4	2, 5	2, 6
3	3, 1	3, 2	3, 3	3, 4	3, 5	3, 6
4	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5	5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6	6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

Osserviamo che

$$P\{(X_1 = k) \cap (X_2 = j)\} = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \quad \forall (k, j) \in \Omega_{(X_1, X_2)}$$

poiché i risultati dei due lanci sono indipendenti e il dado è bilanciato.

Possiamo allora applicare la definizione classica di probabilità come rapporto di casi favorevoli all'evento su casi possibili

$$P(Y = 2) = P(X_1 = 1 \cap X_2 = 1) = \frac{1}{36}$$

Si noti come il primo passaggio ci consente di passare da Ω_Y , dove gli eventi osservabili non sono equiprobabili, ad $\Omega_{(X_1, X_2)}$ dove abbiamo appena dimostrato l'equiprobabilità.

Analogamente

$$P(Y = 3) = P((X_1 = 1 \cap X_2 = 2) \cup (X_1 = 2 \cap X_2 = 1)) = \frac{2}{36}$$

per arrivare alla distribuzione di probabilità che segue

$$\begin{aligned}
P(Y = 2) &= \frac{1}{36} \\
P(Y = 3) &= \frac{2}{36} \\
P(Y = 4) &= \frac{3}{36} \\
P(Y = 5) &= \frac{4}{36} \\
P(Y = 6) &= \frac{5}{36} \\
P(Y = 7) &= \frac{6}{36} \\
P(Y = 8) &= \frac{5}{36} \\
P(Y = 9) &= \frac{4}{36} \\
P(Y = 10) &= \frac{3}{36} \\
P(Y = 11) &= \frac{2}{36} \\
P(Y = 12) &= \frac{1}{36}
\end{aligned}$$

Allo stesso risultato si può giungere utilizzando le leggi della probabilità invece della definizione classica

$$P(Y = 4) = P((X_1 = 1 \cap X_2 = 3) \cup (X_1 = 2 \cap X_2 = 2) \cup (X_1 = 3 \cap X_2 = 1))$$

Dato che gli eventi sono incompatibili possiamo riscrivere la probabilità della loro unione come somma delle probabilità dei singoli eventi

$$P(Y = 4) = P(X_1 = 1 \cap X_2 = 3) + P(X_1 = 2 \cap X_2 = 2) + P(X_1 = 3 \cap X_2 = 1)$$

da cui, sempre per l'indipendenza dei lanci

$$P(Y = 4) = \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{3}{36}$$

Questo secondo metodo è l'unico applicabile nel caso in cui uno o entrambi i dadi siano truccati. Si determini la distribuzione di Y nel caso in cui il primo dado sia truccato con $P(X_1 = 1) = \frac{1}{2}$ e $P(X_1 = i) = \frac{1}{10}, i = 2 \dots 6$ mentre il secondo dado resti bilanciato.

Ancora sui dadi

Immagiamo di sapere che il risultato del primo lancio è stato un numero maggiore di 3. Calcolare la distribuzione di $Y|X_1 > 3$

Adesso l'insieme dei valori osservabili per Y sarà

$$\Omega_{Y|X_1 > 3} = \{5, 6, 7, 8, 9, 10, 11, 12\}$$

Utilizzando la definizione di probabilità condizionata abbiamo

$$P(Y = 5|X_1 > 3) = \frac{P(Y = 5 \cap X_1 > 3)}{P(X_1 > 3)} = \frac{P(X_1 = 4 \cap X_2 = 1)}{P(X_1 > 3)} = \frac{1/36}{18/36} = 1/18$$

Si noti che per applicare la definizione classica siamo tornati all'insieme $\Omega_{(X_1, X_2)}$. Otteniamo analogamente le probabilità degli altri valori osservabili di Y

$$\begin{aligned} P(Y = 5) &= \frac{1}{18} \\ P(Y = 6) &= \frac{2}{18} \\ P(Y = 7) &= \frac{3}{18} \\ P(Y = 8) &= \frac{3}{18} \\ P(Y = 9) &= \frac{3}{18} \\ P(Y = 10) &= \frac{3}{18} \\ P(Y = 11) &= \frac{2}{18} \\ P(Y = 12) &= \frac{1}{18} \end{aligned}$$

Ad un risultato analogo saremmo arrivati osservando che sapere che il risultato del primo lancio è stato un numero maggiore di 3 riduce il numero delle coppie osservabili a

	1	2	3	4	5	6
4	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6
5	5, 1	5, 2	5, 3	5, 4	5, 5	5, 6
6	6, 1	6, 2	6, 3	6, 4	6, 5	6, 6

Avremmo potuto utilizzare la definizione classica di probabilità con

riferimento a questo nuovo Omega ma soltanto dopo aver dimostrato che, anche in questo caso, gli eventi osservabili sono equiprobabili.

Keine Panik

Un paziente risulta positivo al test per l'epatite B; quale è la probabilità che il paziente sia effettivamente malato?

Definiamo gli eventi :

$$\begin{aligned} B^+ &= \{ \text{paziente malato} \} \\ B^- &= \{ \text{paziente sano} \} \\ T^+ &= \{ \text{test positivo} \} \\ T^- &= \{ \text{test negativo} \} \end{aligned}$$

Ipotizziamo che il test utilizzato sia caratterizzato dai seguenti valori:

- Sensibilità $P(T^+|B^+) = 0.95$
- Specificità $P(T^-|B^-) = 0.90$

Vogliamo sapere $P(B^+|T^+)$

Applichiamo il teorema di Bayes.

$$P(B^+|T^+) = \frac{P(T^+|B^+) \cdot P(B^+)}{P(T^+)}$$

La $P(T^+)$, ovvero la probabilità che il test risulti positivo, non è nota in generale ma può essere ricavata come segue :

$$P(T^+) = P(T^+ \cap B^+) + P(T^+ \cap B^-)$$

Utilizzando la probabilità condizionata, riscriviamo le intersezioni come:

$$\begin{aligned} P(T^+ \cap B^+) &= P(T^+|B^+) \cdot P(B^+) \\ P(T^+ \cap B^-) &= P(T^+|B^-) \cdot P(B^-) \end{aligned}$$

Per ricondurci ai valori di sensibilità e specificità facciamo i complementi opportuni ricordando che, in caso di eventi condizionati, l'evento noto non deve essere modificato (essendo appunto noto):

$$P(T^+ \cap B^-) = P(T^+|B^-) \cdot P(B^-) = (1 - P(T^-|B^-)) \cdot (1 - P(B^+))$$

Riscriviamo il teorema di Bayes andando a sostituire:

$$P(B^+|T^+) = \frac{P(T^+|B^+) \cdot P(B^+)}{P(T^+|B^+) \cdot P(B^+) + (1 - P(T^-|B^-)) \cdot (1 - P(B^+))}$$

In questa espressione $P(B^+)$ indica la probabilità di essere malati a priori; per stimarla possiamo utilizzare i dati di prevalenza dell'epatite B che in Italia sono intorno a $\frac{1}{100000}$ affetti $\rightarrow P(B^+) = \frac{1}{10}$

Sostituiamo i valori:

$$P(B^+|T^+) = \frac{0,95 \cdot \frac{1}{100000}}{0,95 \cdot \frac{1}{100000} + (1 - 0,90) \cdot (1 - \frac{1}{100000})} = 0,000094$$

Strategie di vendita

Un commerciante di automobili usate venderà un'automobile al prossimo cliente con probabilità 0.3. Se questo accade, allora la macchina venduta ha la stessa probabilità di costare \$4000 o \$6000.

Sia X la cifra spesa dal cliente

1) Quale è la distribuzione di X?

Iniziamo osservando che $\Omega_X = \{0, 4000, 6000\}$. Per calcolare la probabilità di questi valori dobbiamo tuttavia introdurre una seconda variabile aleatoria Bernoulliana Y che assumerà valore 1 se il cliente decide di acquistare la macchina e 0 altrimenti, con

$P(Y = 1) = 0.3$. Avremo

$$\begin{aligned} P(X = 0) &= P(X = 0 \cap Y = 0) + P(X = 0 \cap Y = 1) \\ &= P(X = 0|Y = 0) \cdot P(Y = 0) + P(X = 0|Y = 1) \cdot P(Y = 1) \\ &= 1 \cdot 0.7 + 0 \cdot 0.3 \\ &= 0.7 \end{aligned}$$

Con ragionamento analogo potete ottenere

$$P(X = k) = \begin{cases} 0.7 & \text{se } k = 0 \\ 0.15 & \text{se } k = 4000 \\ 0.15 & \text{se } k = 6000 \end{cases}$$

2) Quanto vale $E[X]$?

$$E[X] = \sum x_i P(X = x_i) = 0 \cdot 0.7 + 4000 \cdot 0.15 + 6000 \cdot 0.15 = 600 + 900 = 1500$$

3) Quanto vale $Var(X)$?

Poiché $Var(cX) = c^2 Var(X)$ (con c una costante) per semplificare i calcoli troviamo prima $Var(Z)$, dove Z ha distribuzione:

$$P(Z = k) = \begin{cases} 0.7 & \text{se } k = 0 \\ 0.15 & \text{se } k = 4 \\ 0.15 & \text{se } k = 6 \end{cases}$$

Sappiamo anche che $E[cX] = cE[X]$, quindi $E[Z] = 1.5$

$$Var(Z) = E[Z^2] - E[Z]^2 = 0.07 + 4^2 \cdot 0.15 + 6^2 \cdot 0.15 - 1.5^2 = 2.4 + 5.4 - 2.25 = 5.55$$

$$Var(X) = 1000000 Var(Z) = 1000000 \cdot 5.55 = 5550000$$

4) Quanto vale $SD(X)$?

$$SD(X) = \sqrt{Var(X)} = \sqrt{5550000} = 2355.84$$

Le piantine di soia

Un ricercatore semina 10 piantine di soia geneticamente modificate in 10 diversi spazi non comunicanti, per analizzare la resistenza ad una

specifica malattia. Sapendo che la probabilità che la singola piantina si ammali è 0.03, calcolare:

- 1) la probabilità che nessuna pianta si ammali;
- 2) il numero atteso di piante malate;
- 3) la probabilità che si ammali almeno 1 piantina;
- 4) la probabilità che si ammali al massimo 1 piantina;
- 5) la probabilità che almeno 9 piantine siano sane;

Sia X il numero delle piantine ammalate. Poiché la probabilità di ammalarsi delle diverse piantine è costante e siamo in presenza di eventi indipendenti X seguirà una legge binomiale di parametri $n = 10$ e $\pi = 0.03$.

- 1) La prima richiesta equivale quindi a:

$$P(X = 0) = \frac{10!}{0!(10-0)!} \cdot 0.03^0 \cdot (1 - 0.03)^{10-0} = 0.97^{10} = 0.737$$

- 2) Il valore atteso della distribuzione binomiale è:

$$E[X] = 10 \cdot 0.03 = 0.3$$

- 3) Per quanto riguarda la probabilità che si ammali almeno una piantina conviene esprimerla attraverso l'evento complementare che nessuna piantina si ammali, cioè:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.737 = 0.263$$

$$\begin{aligned} 4) \quad P(X \leq 1) &= P(X = 0 \cup X = 1) = P(X = 0) + P(X = 1) = \\ &= \frac{10!}{0!(10-0)!} \cdot 0.03^0 \cdot (1 - 0.03)^{10-0} + \frac{10!}{1!(10-1)!} \cdot \\ &= 0.03^1 \cdot (1 - 0.03)^{10-1} = 0.737 + 0.228 = 0.965 \end{aligned}$$

- 5) Sia Y una nuova variabile aleatoria che descrive il numero di piantine sane. Osservando che $Y + X = 10$ avremo
 $P(Y \geq 9) = P(10 - X \geq 9) = P(-X \geq -1) = P(X \leq 1) = 0.965$

Spese di manutenzione

Le spese di manutenzione sostenute ogni anno da ogni inquilino di un certo condominio hanno distribuzione normale con media \$3000 e deviazione standard \$600. Nell'ipotesi che le spese sostenute in anni diversi siano indipendenti, calcolare la probabilità che nei prossimi due anni le spese di un certo inquilino siano:

- 1) Maggiori di \$5000
- 2) Minori di \$7000
- 3) Comprese tra \$5000 e \$7000

Cominciamo analizzando il primo punto:

- 1) Spese maggiori di \$5000

Se chiamiamo X la spesa annua dell'inquilino e Y la spesa ogni due anni dello stesso inquilino, abbiamo che $Y = X + X$, quindi : $E[Y] = E[X] + E[X] = 3000 + 3000 = 6000$ e, grazie all'ipotesi di indipendenza, $Var[Y] = Var[X] + Var[X] = 600^2 + 600^2 = 720000$. Poiché la somma di variabili aleatorie normali segue ancora un modello normale, Y seguirà una legge normale di media 6000 e varianza 720000.

Quello che dobbiamo trovare è $P(Y > 5000)$. Per farlo utilizzando le tavole della Normale dobbiamo ricondurci al caso di una variabile aleatoria normale standard (valore atteso 0 e varianza 1) convenzionalmente indicata con Z :

$$P(Y > 5000) = P\left(\frac{Y-6000}{\sqrt{720000}} > \frac{5000-6000}{\sqrt{720000}}\right) = P(Z > -1.18)$$

Le tavole riportano in genere l'area da $-\infty$ a -1.18 pari a 0.12. Ricordando che l'area totale sotto ogni densità di probabilità è pari ad 1 otteniamo $P(Y > 5000) = 1 - 0.12 = 0.88$.

- 2) Spese minori di \$7000

$$P(Y < 7000) = P\left(\frac{Y-6000}{\sqrt{720000}} < \frac{7000-6000}{\sqrt{720000}}\right) = P(Z < 1.18)$$

Utilizzando le tavole troviamo il valore richiesto: 0.88

Nota: Avremmo anche potuto dare direttamente la risposta alla seconda domanda, conoscendo quella alla prima: per la simmetria della densità Normale infatti le due aree sono uguali.

3) Spese comprese tra \$5000 e \$7000

$$\begin{aligned} P(5000 < Y < 7000) &= P(X < 7000) - P(X < 5000) = P(X < \\ &\quad 7000) - (1 - P(X \geq 5000)) = \\ &= 0.88 - (1 - 0.88) = 0.76 \end{aligned}$$

Scommesse

Nel 1600 in Francia il gioco d'azzardo era di gran moda e le bische clandestine diffusissime. Un gioco estremamente alla moda allora era il seguente: la "casa" scommetteva alla pari con un giocatore che quest'ultimo, lanciando per 4 volte un dado, avrebbe ottenuto almeno una volta il numero 6. Come vedremo successivamente questo gioco è leggermente favorevole alla casa che "in media" vince il 52% delle volte. Un distinto ed intelligente francese dell'epoca, Antoine Gombauld Cavalier de Méré, frequentatore delle case da gioco, conosceva bene quel gioco ed era incuriosito da una sua possibile variante: la casa scommetteva alla pari con un giocatore che quest'ultimo, lanciando per 24 volte una coppia di dadi, avrebbe ottenuto almeno una volta il doppio 6. Anche questo gioco, secondo il Cavalier de Méré, avrebbe dovuto essere leggermente favorevole alla casa per questo motivo: quando si lancia un dado vi sono 6 possibili risultati, quindi la probabilità che esca il 6 sarà $1/6$. Invece la probabilità che esca un doppio lanciando una coppia di dadi 6 sarà di $1/36$, essendo 36 i risultati distinti, quindi 6 volte più bassa. Lanciando la coppia di dadi 6 volte di più ($24 = 6 \cdot 4$) si dovrebbe controbilanciare l'effetto di considerare un evento meno probabile di un fattore 6 e si dovrebbe avere quindi la stessa probabilità. Al contrario, quest'ultimo gioco non risulta favorevole alla casa, ma al giocatore confermando come spesso le leggi della probabilità non seguono le nostre intuizioni. Anche il Cavalier de Méré, doveva nutrire dei dubbi non sappiamo se per esperienza diretta o per qualche intuizione teorica. In ogni caso decise di parlarne con un altro brillante francese, Blaise Pascal, che si diletta di lettere, di filosofia, di teologia, ma anche di matematica. Pascal risolse il problema postogli dal de Méré provando anche che con 25 lanci il gioco sarebbe tornato favorevole alla casa.

Calcoliamo anche noi la probabilità degli eventi:

$$A = \{ \text{lanciando 4 volte un dado si ottiene almeno un 6} \}$$

$$B = \{ \text{lanciando 24 volte 2 dadi si ottiene almeno un doppio 6} \}.$$

Iniziamo calcolando la probabilità dell'evento complementare di A :

$A^c = \{ \text{lanciando 4 volte un dado non si ottiene alcun 6} \}$

poi sfrutteremo la proprietà: $P(A) = 1 - P(A^c)$.

Indichiamo con T_1, T_2, T_3, T_4 gli eventi: nessun 6 al primo lancio, nessun 6 al secondo lancio, $\dots$, nessun 6 al quarto lancio.

Ora si noti che

$$T_1 \cap T_2 \cap T_3 \cap T_4 = A^c$$

T_1, T_2, T_3, T_4 sono indipendenti.

Allora: $P(A^c) = P(T_1 \cap T_2 \cap T_3 \cap T_4) = P(T_1)P(T_2)P(T_3)P(T_4) = (5/6)(5/6)(5/6)(5/6) = (5/6)^4$

Quindi $P(A) = 1 - (5/6)^4 = 0.518$

Analogamente per $P(B)$ indichiamo con $T_1, T_2, \dots, T_{24}$ gli eventi: nessun doppio 6 al primo (doppio) lancio, nessun doppio 6 al secondo (doppio) lancio, $\dots$, nessun doppio 6 al ventiquattresimo (doppio) lancio.

$P(T_i) = 35/36$ (per ogni $i = 1, \dots, 24$)

$P(B^c) = P(T_1 \cap T_2 \cap \dots \cap T_{24}) = P(T_1)P(T_2) \dots P(T_{24}) = (35/36)^{24}$

$P(B) = 1 - P(B^c) = 1 - (35/36)^{24} = 0.491$

Cosa succede con 25 lanci di due dadi? La situazione si ribalta, e la casa è nuovamente in vantaggio ($P=0.5055$)! Lasciamo al lettore la verifica.

Il paradosso dei compleanni

Immaginiamo che 25 persone partecipino ad una festa di compleanno: ci chiediamo qual è la probabilità che almeno **almeno** 2 persone festeggino il compleanno lo stesso giorno (evento A)?

Proviamo a calcolare la probabilità dell'evento complementare A^c che i compleanni siano tutti in giorni diversi. Consideriamo un primo invitato qualsiasi, ci sono 364 casi favorevoli al fatto che un secondo invitato nasca in un giorno diverso su 365 giorni possibili, 363 casi favorevoli al fatto che un terzo invitato nasca in un giorno diverso sempre su 365 giorni possibili e così via. Avremo quindi $P(A^c) = (364/365)(363/365)(362/365) \dots (341/365) = 0.432$ $P(A^c) = 0.432$, $P(A) = 1 - 0.432 = 0.568$

Il paradosso risiede nel fatto che questa probabilità è largamente superiore a quanto potrebbe dire l'intuito: infatti già in un gruppo di 25 persone la probabilità è circa 0.57, con 30 persone essa supera 0.70, con 50 persone tocca addirittura 0.97, anche se per arrivare all'evento

certo occorre considerare un gruppo di almeno 366 persone (367 se si considera l'anno bisestile).