SAPIENZA
UNIVERSITÀ DI ROMA

"Sapienza" Università di Roma
Ingegneria dell'Informazione,
Informatica e Statistica
Dipartimento di Informatica

Calcolo delle Probabilità

Appunti integrati con il libro "The Probability Tutoring Book", Carol Ash
e revisionati dalla professoressa Giovanna Nappo

Autore
Simone Bianco

14 ottobre 2025

Indice

Informazioni e Contatti

Appunti e riassunti personali raccolti in ambito del corso di *Calcolo delle Probabilità* offerto dal corso di laurea in Informatica dell'Università degli Studi di Roma "La Sapienza".

Ulteriori informazioni ed appunti possono essere trovati al seguente link: **https://github.com/Exyss/university-notes**. Chiunque si senta libero di segnalare incorrettezze, migliorie o richieste tramite il sistema di Issues fornito da GitHub stesso o contattando in privato l'autore :

- Email: **bianco.simone@outlook.it**
- LinkedIn: **Simone Bianco**

Gli appunti sono in continuo aggiornamento, pertanto, previa segnalazione, si prega di controllare se le modifiche siano già state apportate nella versione più recente.

Prerequisiti consigliati per lo studio:

Apprendimento del materiale relativo al corso *Metodi Matematici per l'Informatica*

Licence:

1

Introduzione alla probabilità

1.1 Probabilità classica

Prima di poter parlare di probabilità, è necessario definire ciò di cui essa si occupa. Il calcolo di una probabilità corrisponde allo studio di un fenomeno osservabile esclusivamente dal punto di vista della possibilità o meno del suo verificarsi. La probabilià è uno strumento molto utile per effettuare ragionamenti e prendere decisioni, ma essa può essere *interpretata* in svariati modi.

Supponiamo ad esempio di trovarci in ospedale. Il nostro amico Marco è molto malato ed ogni trattamento sembra fallire. Un dottore ci notifica dell'esistenza di un farmaco che possa *probabilmente* salvargli la vita. Preoccupati, chiediamo al dottore quale sia la probabilità che il farmaco funzioni. Il medico afferma, tramite le sue conoscenze del campo, che "con buona probabilità" il farmaco funzionerà. Volendo utilizzare un approccio più matematico, chiediamo il rateo di successo del farmaco su 100 pazienti, ossia quanti dei 100 pazienti sono guariti dopo la somministrazione. Dopo aver controllato, il medico afferma che il rateo sia 65 casi su 100, rassicurandoci.

L'affermazione iniziale del medico si basa su una probabilità data dal **pensiero soggettivo**, mentre la nostra domanda finale si basa su una probabilità data dalla **frequenza degli esiti**. Di primo occhio, penseremmo che il nostro approccio sia "più matematico" o "più corretto". Uno scettico potrebbe però contestare anche la nostra affermazione:

- E se le condizioni iniziali degli esiti presi in analisi sono diverse?
- E se il farmaco viene somministrato per la prima volta dal medico?
- E se gli esiti registrati sono stati tutti fortuiti?
- E se aumentando il campione il rateo di successo diminuisce?

Con solo questo caso in analisi, concludiamo facilmente che parlare di probabilità è *molto* difficile. L'obiettivo principale di questo capitolo è quindi quello di introdurre concetti elementari che ci permettano di discutere di probabilità. In particolare, ci concentreremo inizialmente sulla **probabilità classica**. Questa tipologia elementare di probabilità si basa sul concetto di **eventi** ed **esiti**.

Consideriamo ad esempio il lancio di una moneta. Tale *evento* può avere solo due *esiti*: testa o croce. Intuitivamente, ci aspettiamo che ciascuno dei due esiti abbia il 50% di probabilità di accadere, ossia "metà e metà". Nella probabilità classica, questi due esiti costituiscono lo **spazio ambiente**, ossia l'insieme di tutti gli esiti possibili per il fenomeno preso in esame. Lo spazio ambiente viene generalmente indicato con il simbolo Ω, mentre gli esiti vengono generalmente indicati con un numero. Ad esempio, in questo caso abbiamo $\Omega = \{0, 1\}$, dove l'esito 0 corrisponde a "croce" e l'esito 1 corrisponde a "testa".

Nella probabilità classica, la probabilità di un evento A viene (solitamente) calcolata utilizzando la frequenza degli esiti positivi nello spazio ambiente, ossia la quantità di esiti dello spazio ambiente che rende vero l'evento A. In questo caso, dunque, abbiamo che:

$$\Pr[\text{il risultato del lancio è testa}] = \frac{\text{numero di esiti testa}}{\text{numero di esiti totali}}$$

Questa tipologia di probabilità può essere comodamente modellata tramite la teoria degli insiemi. In particolare, dato lo spazio ambiente Ω, possiamo definire un **evento** come un semplice sottoinsieme di $A \subseteq \Omega$, dove gli elementi di A corrispondono a tutti gli esiti che rendono vero l'evento A. Ad esempio, l'insieme $T = \{1\}$ corrisponde all'evento "testa". Questo ci permette di calcolare la probabilità dell'esempio precedente nel seguente modo:

$$\Pr[T] = \frac{|T|}{|\Omega|} = \frac{1}{2} = 50\%$$

ottenendo il 50% di probabilità che ci aspettavamo. Tuttavia, notiamo facilmente che questa interpretazione della probabilità non è sempre un corretto modello:

- E se la moneta è stata truccata al fine di rendere uno dei due esiti più probabile?
- E se lo spazio ambiente contiene un numero infinito di esiti?

Per la prima domanda, il modello può essere "spremuto" in modo da forzare un modello corretto. Ad esempio, se la moneta è stata truccata in modo da ottenere il 75% delle volte l'esito "testa", possiamo considerare lo spazio ambiente $\Omega' = \{0, 1, 11, 111\}$ ed imporre che $T' = \{1, 11, 111\}$ e che $C' = \{0\}$, dove C è l'evento descrivente l'esito "croce". In questo modo, otteniamo che:

$$\Pr[T'] = \frac{|T'|}{|\Omega'|} = \frac{3}{4} = 75\%$$

Per il secondo caso, invece, è chiaramente impossibile adattare il modello in quanto non possiamo dividere per una cardinalità infinita. Vedremo nelle sezioni successive un modello **assiomatico** della probabilità che ci permetterà di risolvere entrambe le problematiche (e molte altre) comodamente.

Definizione 1.1: Spazio ambiente ed Evento (prob. classica)

Dato un insieme finito $\Omega = \{\omega_1, \ldots, \omega_n\}$ detto spazio ambiente, definiamo un sottoinsieme $A \subseteq \Omega$ come evento su Ω. La probabilità classica dell'evento A è definita come:

$$\Pr[A] = \frac{|A|}{|\Omega|}$$

1.2 Modelli principali del Calcolo Combinatorio

Dopo aver enunciato le basi della probabilità classica, risulta evidente la necessità di ripassare le basi del **calcolo combinatorio**, ossia la branca della matematica atta a studiare metodi per computare quantità di elementi ed oggetti. Difatti, domande come "qual è la probabilità di pescare un asso da un mazzo di 52 carte francesi?" nella probabilità classica vengono ridotte a due sotto-problemi: calcolare la quantità di assi in un mazzo di 52 carte francesi (ossia 4, 1 per seme) e la quantità di carte in un mazzo di 52 carte francesi (ossia 52, ovviamente).

In questa sezione verranno riassunte le basi del calcolo combinatorio, elencandone i principali principi e modelli. Per appunti più approfonditi sull'argomento (in particolare spiegazioni ed esempi ulteriori delle formule mostrate in seguito), si rimanda agli appunti del corso *Metodi Matematici per l'Informatica* (reperibili qui).

Partiamo con un problema semplice. Supponiamo che in un urna vi siano 3 palline bianche e 2 palline nere. Ci chiediamo quale sia la probabilità che, estraendo due palline, la seconda sia bianca? Formalizziamo il problema secondo il modello della probabilità classica. Prima di tutto, individuiamo l'**evento elementare**, ossia la forma assunta da un singolo esito, come l'estrazione *ordinata* di due palline $\omega = (\omega_1, \omega_2)$, dove ω_1 è la prima pallina estratta e ω_2 è la seconda pallina estratta.

Possiamo numerare le palline da 1 a 5, dicendo che le palline bianche son quelle da 1 a 3, mentre le restanti sono nere. Detto ciò, possiamo capire che l'insieme degli eventi elementari, ossia lo spazio ambiente, corrisponde a tutte le coppie di numeri naturali compresi fra 1 a 5 senza ripetizioni.

$$\Omega = \{(\omega_1, \omega_2) \mid \omega_1 \in \{1,2,3,4,5\}, \omega_2 \in \{1,2,3,4,5\} - \{\omega_1\}\}$$

È di facile intuizione capire che $|\Omega| = 5 \cdot 4 = 20$ poiché abbiamo 5 possibilità per la prima pallina e 4 possibilità per la seconda pallina (visto che una è stata già pescata), dunque è sufficiente moltiplicare le varie combinazioni tra loro. In questo caso, abbiamo implitamente utilizzato il **principio moltiplicativo**, il quale afferma che dati due insiemi finiti A, B si ha che:

$$|A \times B| = |A| \cdot |B|$$

Tornando all'esempio, definiamo come B_2 l'evento "la seconda pallina estratta senza ripetizioni è bianca". Formalmente, abbiamo che:

$$B_2 = \{(\omega_1, \omega_2) \in \Omega \mid \omega_2 \in \{1,2,3\}\}$$

Ancora una volta, tramite il principio moltiplicativo osserviamo facilmente che $|B_2| = 4 \cdot 3 = 12$ poiché per la seconda estrazione abbiamo 3 possibili palline bianche e per la prima estrazione abbiamo tutte le altre rimanenti (non ci interessa il colore essendo la prima estrazione). Possiamo quindi concludere che:

$$\Pr[B_2] = \frac{|B_2|}{|\Omega|} = \frac{12}{20} = \frac{3}{5} = 0.6$$

Passiamo adesso al riepilogo di alcuni modelli fondamentali nella quale spesso ricadono i problemi di conteggio. L'idea principale dietro ogni modello viene comunemente metaforizzata secondo il problema delle palline nelle urne.

- **Disposizioni senza ripetizione di classe k per n elementi**. Si hanno n palline numerate dentro un urna e si vogliono fare k estrazioni con *senza rimpiazzo* (ossia ogni pallina pescata viene eliminata dall'urna). L'ordine di estrazione delle palline è di nostro interesse, sequenze come $1, 2, 3$ e $1, 3, 2$ vengono considerate come *diverse*. Per problemi di questo tipo, la quantità $D_{n,k}$ di sequenze possibili estraendo k palline su n totali è data da:

  $$D_{n,k} = \frac{n!}{(n-k)!}$$

  Quando $k = n$, questo modello viene detto **permutazione**.

- **Disposizioni con ripetizione di classe k per n elementi**. Si hanno n palline numerate dentro un urna e si vogliono fare k estrazioni *con rimpiazzo* (ossia ogni pallina pescata viene subito reinserita nell'urna). L'ordine di estrazione delle palline è di nostro interesse, dunque sequenze come $1, 2, 3$ e $1, 3, 2$ vengono considerate come *diverse*. Per problemi di questo tipo, la quantità $D'_{n,k}$ di sequenze possibili estraendo k palline su n totali è data da:

  $$D'_{n,k} = n^k$$

- **Combinazioni di classe k per n elementi**. Si hanno n palline numerate dentro un urna e si vogliono fare k estrazioni *senza rimpiazzo* (ossia ogni pallina pescata viene subito reinserita nell'urna). L'ordine di estrazione delle palline è di nostro interesse, dunque sequenze come $1, 2, 3$ e $1, 3, 2$ vengono considerate come *uguali*. Per problemi di questo tipo, la quantità $C_{n,k}$ di sequenze possibili estraendo k palline su n totali è data da:

  $$C_{n,k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

  dove $\binom{n}{k}$ è il **coefficiente binomiale**

- **Anagrammi di classi $k_1, \ldots, k_m$ per n elementi**. Si hanno n palline numerate dentro un urna. Le palline sono di m tipologie e per ogni tipologia i vi sono k_i palline (dunque $k_1 + \ldots + k_m = n$). Si vogliono fare n estrazioni *senza rimpiazzo* (ossia ogni pallina pescata viene subito reinserita nell'urna). L'ordine di estrazione delle palline è di nostro interesse, ma le palline della stessa tipologia vengono considerate come uguali tra loro, dunque sequenze come $1_R, 2_R, 3_B$ e $2_R, 3_B, 1_R$ (dove R e B rappresentano la tipologia della pallina) vengono considerate *uguali* mentre sequenze come $1_R, 2_R, 3_R$ e $1_R, 2_R, 3_B$ vengono considerate *diverse*. Per problemi di questo tipo, la quantità $A_{k_1,\dots,k_m}$ di sequenze possibili estraendo n palline su n totali è data da:

$$A_{k_1,\dots,k_m} = \frac{n!}{k_1! \cdot \ldots \cdot k_m!} = \binom{n}{k_1,\dots,k_m}$$

dove $\binom{n}{k_1,\dots,k_m}$ è il **coefficiente multinomiale**

È opportuno osservare come il coefficiente binomiale sia un caso speciale del coefficiente multinomiale dove vi sono solo due tipologie: quella contenente gli oggetti di nostro interesse e quella che non li contiene. Ricordiamo inoltre l'identità conosciuta come **binomio di Newton**, la quale afferma che:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Vediamo ora alcuni esempi per comprendere come applicare tali modelli combinatori nel calcolo delle probabilitè classiche.

1. Un codice PIN è formato da 4 cifre scelte tra 0 e 9 (quindi 10 cifre totali). Le cifre possono ripetersi (es. 1123 è valido) e l'ordine di digitazione conta (dunque 1234 è diverso da 4321). Supponiamo che un ladro voglia indovinare il PIN di un bancomat inserendo un codice casuale. Vogliamo sapere quale sia la probabilità dell'evento A definito come "il ladro indovina il PIN al primo tentativo".

   Osserviamo come la quantità di esiti totali, ossia la cardinalità dello spazio ambiente, sia data da tutti i PIN possibili, i quali sono modellati da una disposizione con ripetizione di classe 4 per 10 elementi. Concludiamo quindi che $|\Omega| = D'_{10,4} = 10^4$. Inoltre, è evidente che esista un solo esito favorevole all'evento A, ossia il PIN da indovinare. Concludiamo quindi che:

$$\Pr[A] = \frac{1}{10^4} = 0.01\%$$

2. In una classe di 20 studenti, si devono eleggere una delegazione di 3 rappresentanti con ruoli distinti: presidente, vicepresidente e segretario. Ogni studente può essere eletto per un solo ruolo e (ad esempio vi è distinzione tra il caso in cui Marco viene eletto come presidente e il caso in cui viene eletto come vicepresidente). Vogliamo sapere quale sia la probabilità dell'evento B definito come "Anna, Marco e Luca vengono eletti in questo ordine".

   Osserviamo come la quantità di esiti totali sia data da tutte delegazioni da 3 rappresentanti possibili, le quali sono modellati da una disposizione senza ripetizione di classe 3 per 20 elementi. Concludiamo quindi che $|\Omega| = \frac{20!}{17!} = 20 \cdot 19 \cdot 18$. Inoltre, è evidente che esista un solo esito favorevole all'evento B, ossia la delegazione Anna, Marco e Luca. Concludiamo quindi che:

$$\Pr[B] = \frac{1}{\frac{20!}{17!}} = \frac{1}{20 \cdot 19 \cdot 18} = \frac{1}{6840} \approx 0.0001\%$$

3. Considerando la stessa situazione del punto precedente, vogliamo ora sapere la probabilità dell'evento C definito come "Anna, Marco e Luca vengono eletti in qualsiasi ordine".

   La cardinalità dello spazio ambiente rimane chiaramente inalterata, ma la quantità di esiti favorevoli aumenta. In particolare, tali esiti sono modellati da una dispozione senza ripetizione di classe 3 per 3 elementi (ossia una permutazione di 3 elementi), dunque $|C| = \frac{3!}{0!} = 3 \cdot 2 \cdot 1$. Concludiamo quindi che:

$$\Pr[C] = \frac{\frac{3!}{0!}}{\frac{20!}{17!}} = \frac{3 \cdot 2 \cdot 1}{20 \cdot 19 \cdot 18} = \frac{1}{1140} \approx 0.0009\%$$

4. Da un mazzo standard di 52 carte francesi, si estraggono 5 carte casualmente, senza rimetterle nel mazzo (come nel poker classico). Vogliamo sapere la probabilità dell'evento D definito come "la mano pescata è composta da esattamente una carta cuori".

   Osserviamo come la quantità di esiti totali sia data da tutte pescate di 5 carte possibili, le quali sono modellati da una combinazione di classe 5 per 52 elementi, dunque $|\Omega| = \binom{52}{5}$. Per quanto riguarda la quantità di esiti favorevoli, anche essi possono essere modellati dalle combinazioni. In particolare, poiché l'ordine di pescata non conta, gli eventi favorevoli sono modellati dalle pescate in cui viene pescata 1 carta dalle 13 di cuori disponibili e 4 carte dalle 39 rimanenti, senza contare l'ordine di pescata, dunque $|D| = \binom{13}{1}\binom{39}{4}$. Concludiamo quindi che:

$$\Pr[D] = \frac{\binom{13}{1}\binom{39}{4}}{\binom{52}{5}} = \frac{1069263}{2598960} \approx 0.41\%$$

1.3 Operazioni tra eventi

La definizione data di probabilità classica ci permette di prendere in prestito alcuni concetti della teoria degli insiemi ed applicarli nel calcolo delle probabilità. Ad esempio, consideriamo il lancio di un dado (non truccato) a 6 facce, descritto dallo spazio ambiente $\Omega = \{1, 2, 3, 4, 5, 6\}$. Consideriamo ora l'evento P descritto dagli esiti in cui la faccia risultante sia un numero pari, ossia $P = \{2, 4, 6\}$. Sia inoltre $F_{\leq 3}$ l'evento descritto dagli esiti in cui la faccia risultate sia un numero minore o uguale a 3, ossia $F_{\leq 3} = \{1, 2, 3\}$.

Consideriamo quindi l'evento "la faccia risultante è un numero pari ed è un numero minore o uguale a 3". Tramite la teoria degli insiemi, questo evento è facilmente descrivibile tramite l'**intersezione** $P \cap F_{\leq 3} = \{2\}$, ottenendo quindi che:

$$\Pr[P \cap F_{\leq 3}] = \frac{|P \cap F_{\leq 3}|}{|\Omega|} = \frac{1}{6}$$

Similmente, l'evento "la faccia risultante è un numero pari o è un numero minore o uguale a 3" è describile come l'**unione** $P \cup F_{\leq 3} = \{1, 2, 3, 4, 6\}$, ottenendo quindi che:

$$\Pr[P \cup F_{\leq 3}] = \frac{|P \cup F_{\leq 3}|}{|\Omega|} = \frac{5}{6}$$

Altre basilari operazioni insiemistiche sono la **differenza** e il **complemento**. Dati due insiemi $A, B \subseteq \Omega$, la differenza insiemistica di A con B corrisponde all'insieme $A - B$ definito come:

$$A - B = \{x \in A \mid x \notin B\}$$

ossia l'insieme degli elementi di A che non si trovano in B. Osserviamo come tale operazione sia *non commutativa*. Ad esempio, per gli eventi precedenti abbiamo che $P - F_{\leq 3} = \{4, 6\}$ e che $F_{\leq 3} - P = \{1, 3\}$.

Il complemento insiemistico su un insieme $A \subseteq \Omega$, indicato con $\overline{A}$ (oppure A^c), corrisponde al sottoinsieme dello spazio ambiente contenente tutti gli esiti opposti che non rientrano nell'evento A, ossia:

$$\overline{A} = \{x \in \Omega \mid x \notin A\}$$

In altre parole, abbiamo che $\overline{A} = \Omega - A$. Osserviamo come tale operazione sia un'*involuzione*, ossia un'operazione inversa di se stessa. Difatti, abbiamo che:

$$\overline{\overline{A}} = \Omega - (\Omega - A) = \{x \in \Omega \mid x \notin \Omega - A\} = \{x \in \Omega \mid x \in A\} = A$$

In questi casi, risultano fondamentali le principali proprietà insiemistiche di base di seguito elencate (con le quali si dovrebbe già essere familiari):

- *Proprietà disgiuntiva*:

$$A \cap \overline{A} = \varnothing$$

- *Proprietà associativa*:

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

- *Proprietà distributiva*:

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

- *Leggi di De Morgan*:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \qquad \overline{A \cap B} = \overline{A} \cup \overline{B}$$

- *Principio di inclusione-esclusione*:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

L'uso di questi operatori insiemistici risulta fondamentale nella probabilità classica, poiché esso ci permette di "riscrivere" molti eventi complessi in termini di eventi meno complicati. Di seguito, riportiamo due esempi:

1. Consideriamo ancora l'evento $P \cap F_{\leq 3}$ descritto precedentemente. Consideriamo ora gli eventi $F_{=3}$ e $F_{<3}$, definiti analogamente. Osserviamo facilmente che $F_{\leq 3} = F_{=3} \cup F_{<3}$. Per tanto, abbiamo che:

$$P \cap F_{\leq 3} = P \cap (F_{=3} \cup F_{<3}) = (P \cap F_{=3}) \cup (P \cap F_{<3}) = \varnothing \cup \{2\} = \{2\}$$

concludendo quindi che:

$$\Pr\left[P \cap F_{\leq 3}\right] = \frac{|P \cap F_{\leq 3}|}{|\Omega|} = \frac{1}{6}$$

2. Tramite le *leggi di De Morgan* possiamo riscrivere l'evento "la faccia risultate non è né un numero pari né un numero minore o uguale a 3", ossia l'evento $\overline{P} \cap \overline{F_{\leq 3}}$ come l'evento "la faccia risultante non è un numero pari o un numero minore o uguale a 3", ossia l'evento $\overline{P \cup F_{\leq 3}}$, concludendo che:

$$\Pr\left[\overline{P} \cap \overline{F_{\leq 3}}\right] = \Pr\left[\overline{P \cup F_{\leq 3}}\right] = \frac{\left|\overline{P \cup F_{\leq 3}}\right|}{|\Omega|} = \frac{1}{6}$$

dove $\overline{P \cup F_{\leq 3}} = \{5\}$.

Osserviamo inoltre come gli eventi complementari ci permettano di derivare una prima fondamentale proprietà della probabilità classica (che verrà estesa anche agli altri tipi di probabilità). Dato un evento $A \subseteq \Omega$, poiché Ω contiene tutti gli elementi di A abbiamo che:

$$\Pr\left[\overline{A}\right] = \frac{|\Omega - A|}{|\Omega|} = \frac{|\Omega| - |A|}{|\Omega|} = 1 - \frac{|A|}{|\Omega|} = 1 - \Pr[A]$$

In altre parole, la probabilità di un evento complementare è uguale a 1 meno la probabilità dell'evento di partenza.

Teorema 1.1: Probabilità di un evento complementare

Dato un evento $A \subseteq \Omega$, si ha che:

$$\Pr\left[\,\overline{A}\,\right] = 1 - \Pr[A]$$

Similmente, il *principio di inclusione-esclusione* nella probabilità classica ci permette di riscrivere la probabilità di eventi disgiunti (ossia quelli della forma "A o B") tramite somme e differenze. Dati due eventi $A, B \subseteq \Omega$, abbiamo che:

$$\begin{aligned}
\Pr[A \cup B] &= \frac{|A \cup B|}{|\Omega|} \\
&= \frac{|A| + |B| - |A \cap B|}{|\Omega|} \\
&= \frac{|A|}{|\Omega|} + \frac{|B|}{|\Omega|} + \frac{|A \cap B|}{|\Omega|} \\
&= \Pr[A] + \Pr[B] - \Pr[A \cap B]
\end{aligned}$$

Come per il precedente risultato, estenderemo questa proprietà anche alle altre forme probabilità e non solo a quella classica.

Teorema 1.2: Regola della somma di probabilità

Dati due eventi $A, B \subseteq \Omega$, si ha che:

$$\Pr\left[A \cup B\right] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

Questi teoremi risultano estremamente convenienti in molte situazioni. Ad esempio, consideriamo l'evento A definito come "su 2 pescate da un mazzo di 52 carte francesi almeno una è un asso" (assumendo che la carta venga reinserita nel mazzo). Tale evento può essere riscritto come l'unione di due eventi A_1 e A_2, dove ogni A_i rappresenta l'evento "la pescata i-esima è un asso", dunque $A = A_1 \cup A_2$. Per il principio di inclusione-esclusione, abbiamo che:

$$\Pr[A_1 \cup A_2] = \Pr[A_1] + \Pr[A_2] - \Pr[A_1 \cap A_2]$$

Essendoci solo 4 assi (uno per seme) ed essendo la carta pescata reinserita nel mazzo (dunque non vi è differenza tra le due pescate), otteniamo facilmente che:

$$\Pr[A_1] = \Pr[A_2] = \frac{4}{52}$$

mentre, utilizzando il principio moltiplicativo, la probabilità $\Pr[A_1 \cap A_2]$ può essere calcolata come il rapporto tra il numero di coppie di carte di nostro interesse e il numero totale di coppie di carte, ossia:

$$\Pr[A_1 \cap A_2] = \frac{4 \cdot 4}{52 \cdot 52}$$

Concludiamo quindi che:

$$\begin{aligned}
\Pr[A_1 \cup A_2] &= \Pr[A_1] + \Pr[A_2] - \Pr[A_1 \cap A_2] \\
&= \frac{4}{52} + \frac{4}{52} - \frac{4 \cdot 4}{52 \cdot 52} \\
&= \frac{400}{2704} \\
&\approx 0.148
\end{aligned}$$

Il PIE ci permette quindi di "spezzare" il calcolo in più sotto-calcoli molto immediati. E se l'evento A fosse stato definito su 10 pescate? In tal caso osserviamo che:

$$\begin{aligned}
\Pr[A_1 \cup \ldots \cup A_{10}] &= \Pr[A_1] + \Pr[A_2 \cup \ldots \cup A_{10}] - \Pr[A_1 \cap (A_2 \cup \ldots \cup A_{10})] \\
&= \Pr[A_1] + \Pr[A_2 \cup \ldots \cup A_{10}] - \Pr[(A_1 \cap A_2) \cup \ldots \cup (A_1 \cap A_{10})]
\end{aligned}$$

Notiamo quindi che la scomposizione da vita ad altre probabilità formate da unioni, le quali vanno a loro volta ricorsivamente scomposte utilizzando il PIE, rendendo quindi il calcolo estremamente impegnativo. Il primo teorema, invece, ci permette di "rigirare la frittata" calcolando la probabilità dell'evento opposto, ossia l'evento $\overline{A}$ definito come "su 10 pescate da un mazzo di 52 carte nessuna è un asso". Questo trucco viene comunemente detto *passaggio per il complementare*.

$$\begin{aligned}
\Pr[A_1 \cup \ldots \cup A_{10}] &= 1 - \Pr[\,\overline{A_1 \cup \ldots \cup A_{10}}\,] \\
&= 1 - \Pr[\,\overline{A_1} \cap \ldots \cap \overline{A_{10}}\,] \\
&= 1 - \frac{(52-4)^{10}}{52^{10}} \\
&= 1 - \left(\frac{48}{52}\right)^{10} \\
&\approx 0.55
\end{aligned}$$

La regola della somma ha forti implicazioni nel calcolo della probabilità. In particolare, essa ci permette di "spacchettare" eventi complessi in calcoli più semplici suddividendo l'evento originale in casistiche. Immaginiamo di avere un'urna contenente 2 palline bianche e 3 palline nere. Supponiamo inoltre di avere due mazzi di carte. Al primo mazzo sono state levate tutte le carte con seme cuori (dunque contengono un seme rosso e due neri). Al secondo mazzo, invece, sono state levate tutte le carte con seme picche (dunque contengono due semi rossi e un seme nero). I mazzi e l'urna sono legati dalle seguenti regole:

1. Se la pallina estratta è bianca, peschiamo una carta dal primo mazzo
2. Se la pallina estratta è nera, peschiamo una carta dal secondo mazzo

Vogliamo sapere la probabilità (classica) dell'evento R definito come "pesco una carta con seme rosso". Calcolare questa probabilità non risulta facile in quanto vi siano troppi eventi dipendenti tra loro da analizzare. Possiamo quindi sfruttare gli operatori insiemistici per scomporre l'evento R in due eventi più semplici da calcolare. Sia B l'evento "pesco una pallina bianca". Poiché $R \subseteq \Omega$, abbiamo che:

$$R = \Omega \cap R = (B \cup \overline{B}) \cap R = (B \cap R) \cup (\overline{B} \cap R)$$

Infine, usando la regola della somma tra probabilità abbiamo che:

$$\begin{aligned}
\Pr[R] &= \Pr[(B \cap R) \cup (\overline{B} \cap R)] \\
&= \Pr[B \cap R] + \Pr[\overline{B} \cap R] - \Pr[(B \cap R) \cap (\overline{B} \cap R)] \\
&= \Pr[B \cap R] + \Pr[\overline{B} \cap R] - \Pr[\varnothing] \\
&= \Pr[B \cap R] + \Pr[\overline{B} \cap R]
\end{aligned}$$

A questo punto, calcolare $\Pr[R]$ risulta molto semplice in quando le due congiunzioni ci permettono di sapere con certezza da quale mazzo andremo a pescare (osserviamo che si tratta di due eventi non indipendenti):

$$\Pr[B \cap R] = \frac{2}{5} \cdot \frac{13}{39} \qquad \Pr[\overline{B} \cap R] = \frac{3}{5} \cdot \frac{26}{39}$$

dunque:

$$\Pr[R] = \Pr[B \cap R] + \Pr[\overline{B} \cap R] = \frac{2}{5} \cdot \frac{13}{39} + \frac{3}{5} \cdot \frac{26}{39}$$

Osserviamo come la condizione fondamentale in grado di far funzionare tale trucco è la *digiunzione* tra gli eventi B e $\overline{B}$, ossia il fatto che $B \cap \overline{B} = \varnothing$. Difatti, questo trucco può essere generalizzato ad una qualsiasi *partizione* dello spazio ambiente Ω. La generalizzazione di questa techica viene detta **probabilità totale**.

Definizione 1.2: Partizione

Sia X un insieme. Diciamo che $A_1, \dots, A_n$ è una partizione di X quando:

- Per ogni $i \leq n$ vale che $A_i \subseteq X$
- Per ogni coppia di indici $i \neq j$ vale che $A_i \cap A_j = \varnothing$
- $X = \bigcup_{i=1}^{n} A_i$

Proposizione 1.1: Probabilità totale

Sia B un evento. Data una partizione $A_1, \dots, A_n$ di Ω si ha che:

$$\Pr[B] = \sum_{i=1}^{n} \Pr[A_i \cap B]$$

1.4 Indipendenza stocastica

Consideriamo il lancio di due monete. Vogliamo sapere qual è la probabilità che entrambe le monete diano esito "testa". Questo evento è comodamente modellabile dall'evento intersezione $T_1 \cap T_2$ corrispondente a "la prima moneta è testa e la seconda moneta è testa". Usando la probabilità classica,notiamo facilmente che la situazione sia modellabile da una disposizione con ripetizione $D'_{2,2}$, come se volessimo disporre le lettere $\{T, C\}$ su due posizioni. In questo caso, l'unico esito positivo è dato dalla singola disposizione TT. Otteniamo quindi che $\Pr[T_1 \cap T_2] = \frac{1}{4}$. Ponendo maggior attenzione, osserviamo che:

$$\Pr[T_1 \cap T_2] = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \Pr[T_1] \cdot \Pr[T_2]$$

Notiamo quindi che in questo caso la probabilità dell'evento intersezione è uguale al prodotto delle probabilità dei singoli eventi.

Consideriamo ora invece due pescate dallo stesso mazzo di 52 carte. Vogliamo sapere quale sia la probabilità dell'evento "la prima carta pescata è un asso di cuori e la seconda carta pescata è di cuori", dunque la probabilità dell'evento $\Pr[P_1 \text{ è asso di cuori} \cap P_2 \text{ è cuori}]$.

Se l'evento considerato fosse "la prima carta pescata è un asso di cuori", calcoleremmo tale probabilità come $\Pr[P_1 \text{ è asso di cuori}] = \dfrac{1}{52}$ poiché vi è un solo asso di cuori sulle

52 carte totali. Per quanto riguarda il nostro evento in analisi, invece, tramite il principio moltiplicativo otteniamo facilmente che gli esiti possibili siano $52 \cdot 51$ (52 possibilità per la prima pescata e 51 per la seconda) e che quelli positivi siano $1 \cdot 12$ (1 possibilità per la prima pescata e 12 per la seconda), concludendo che:

$$\Pr[P_1 \text{ è asso di cuori} \cap P_2 \text{ è cuori}] = \frac{1 \cdot 12}{52 \cdot 51} = \frac{1}{52} \cdot \frac{12}{51}$$

Tuttavia, osserviamo che la probabilità dell'evento "la seconda carta pescata è di cuori" la probabilità risulterebbe $\Pr[P_2 \text{ è di cuori}] = \frac{13}{52}$ poiché non abbiamo assolutamente alcuna informazione sulla prima carta pescata, dunque siamo in una situazione identica a quella in cui nessuna carta è stata pescata dal mazzo (in altre parole, è come se la carta fosse solo stata spostata dal mazzo). Concludiamo quindi che:

$$\Pr[P_1 \text{ è asso di cuori} \cap P_2 \text{ è cuori}] \neq \Pr[P_1 \text{ è asso di cuori}] \cdot \Pr[P_2 \text{ è cuori}]$$

Qual è dunque la differenza tra le due situazioni analizzate? Nella prima situazione, i due lanci di moneta *non si influenzano* a vicenda, ossia l'esito del primo lancio non ha alcun effetto sul secondo e viceversa. Nella seconda situazione, invece, la seconda pescata sarà sicuramente *influenzata* dalla prima pescata in quando un asso di cuori è stato già estratto dal mazzo. I due eventi della seconda situazione sono quindi **dipendenti** tra loro, mentre quelli della prima sono **indipendenti**.

In generale, dunque, diciamo che due eventi A e B sono *indipendenti* tra loro quando si verifica che:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$$

La nozione di *indipendenza stocastica* viene generalizzata anche al caso tra n eventi: se nessuno degli n eventi si influenza a vicenda indipendentemente da quanti e quali di essi vengano considerati, gli eventi godono di indipendenza stocastica. Nel caso di tre eventi A, B, C, ad esempio, è necessario che tutte e quattro le seguenti condizioni si verifichino:

$$\Pr[A \cap B] = \Pr[A]\Pr[B] \qquad \Pr[A \cap C] = \Pr[A]\Pr[C] \qquad \Pr[B \cap C] = \Pr[B]\Pr[C]$$

$$\Pr[A \cap B \cap C] = \Pr[A]\Pr[B]\Pr[C]$$

Definizione 1.3: Indipendenza stocastica

Diciamo che n eventi $A_1, \dots, A_n$ sono stocasticamente indipendenti quando per ogni sottoinsieme $I \subseteq \{1, \dots, n\}$ si ha che:

$$\Pr\left[\bigcap_{i \in I} A_i\right] = \prod_{i \in I} \Pr[A_i]$$

Quando si trattano probabilità congiunte, dunque, risulta essenziale comprendere se gli eventi coinvolti siano indipendenti o meno.

Osservazione 1.1

Osserviamo che il concetto di indipendenza statistica è del tutto slegato dal concetto di *disgiunzione* tra eventi. Difatti, è possibile che vi siano due eventi A e B tali che $A \cap B = \varnothing$ (dunque $\Pr[A \cap B] = 0$) ma tali che $\Pr[A]\Pr[B] > 0$.

1.5 Probabilità condizionata e teorema di Bayes

Fino ad ora abbiamo ristretto la nostra attenzione ad eventi semplici in cui gli esiti dipendono solo da un istante di tempo iniziale. In molti casi, invece, la probabilità di un evento A di veridicità ignota potrebbe essere influenzata dal verificarsi di un evento precedente B (dunque i due eventi non avvengono in contemporanea). Formalmente, questo tipo di evento viene scritto come $\Pr[A \mid B]$, dove la sbarra in mezzo viene letta come "dato che" o "sapendo che".

Come al solito, diamo un'intuizione usando la probabilità classica. Consideriamo un mazzo da 52 carte. Vogliamo sapere quale sia la probabilità dell'evento $2R \mid 1A$ definito come "pescare una mano di poker da 5 carte con esattamente due re sapendo che nella mano pescata vi è esattamente un asso". Normalmente, per calcolare la probabilità dell'evento "pescare una mano di poker da 5 carte con esattamente due re" procederemmo considerando le combinazioni semplici. In particolare, la candinalità dello spazio ambiente risulta essere $\binom{52}{5}$ poiché dalle 52 totali ne stiamo scegliendo 5 casuali senza ripetizioni e senza ordine, mentre la cardinalità dell'evento favorevole risulta essere $\binom{4}{2}\binom{48}{3}$ poiché stiamo scegliendo 2 re dai 4 presenti nel mazzo e 3 carte dalle 48 carte che non sono un re.

$$\Pr[2R] = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$$

Nel caso dell'evento $2R \mid 1A$, invece, sappiamo già che una delle 5 carte è un asso e che non vi sono altri assi nella mano. Questo restringe chiaramente la cardinalità dei casi possibili considerabili (dunque come se stessimo usando uno spazio ambiente ristretto), portando le scelte totali a $\binom{4}{1}\binom{48}{4}$. Per quanto riguarda gli eventi favorevoli, invece, le scelte diventano $\binom{4}{1}\binom{4}{2}\binom{44}{2}$ poiché stiamo scegliendo 1 asso, 2 re e 2 carte dalle 44 che non sono nè assi nè re.

$$\Pr[2R \mid 1A] = \frac{\binom{4}{1}\binom{4}{2}\binom{44}{2}}{\binom{4}{1}\binom{48}{4}}$$

Con un po' di manipolazione algebrica osserviamo come:

$$\Pr[2R \mid 1A] = \frac{\binom{4}{1}\binom{4}{2}\binom{44}{2}}{\binom{4}{1}\binom{48}{4}} \cdot \frac{\binom{52}{5}}{\binom{52}{5}} = \frac{\Pr[2R \cap 1A]}{\Pr[1A]}$$

In altre parole, una probabilità condizionata $A \mid B$ non è nient'altro che il rapporto tra la probabilità congiunta $A \cap B$ e la probabilità dell'evento B. Questa definizione viene utilizzata per tutte le probabilità (dunque non solo in quella classica).

Definizione 1.4: Probabilità condizionata

Dati due eventi A e B, la probabilità dell'evento "A dato B" è definita come:

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}$$

Sebbene la probabilità condizionata sia definita tramite la probabilità congiunta, in molte situazioni la prima risulta più "immediata" da calcolare. Difatti, spesso non è neanche richiesto passare per la definizione esatta, permettendoci di inferire il risultato direttamente tramite la logica. Questo ci permette di sfruttare la probabilità condizionata per calcolare più facilmente la probabilità congiunta tramite la seguente formula inversa:

$$\Pr[A \cap B] = \Pr[A \mid B] \Pr[B]$$

Consideriamo il seguente esempio (sempre nella probabilità classica). Vengono lanciati due dadi eqquilibrati a sei facce. Vogliamo sapere la probabilità che la somma dei due risultati sia uguale a 6 sapendo che esattamente uno dei due dadi dia 2 come risultato. In questo caso, abbiamo solo 11 esiti totali possibili (la coppia $(2, 2)$ viene contata una sola volta) di cui solo 2 sono positivi (le coppie $(2, 5)$ e $(5, 2)$), concludendo che:

$$\Pr[\text{Somma 6} \mid \text{Almeno un 2}] = \frac{2}{11}$$

Similmente, abbiamo $\Pr[\text{Almeno un 2}] = \frac{11}{36}$ dato dagli 11 esiti positivi sui 36 totali. Concludiamo quindi che:

$$\begin{aligned}\Pr[\text{Somma 6} \;\cap\; \text{Almeno un 2}] &= \Pr[\text{Somma 6} \mid \text{Almeno un 2}] \Pr[\text{Almeno un 2}] \\ &= \frac{2}{11} \cdot \frac{11}{36} \\ &= \frac{22}{396}\end{aligned}$$

Questa formulazione della probabilità congiunta ci permette di stabilire una nuova formulazione del concetto di **eventi indipendenti**. Difatti, osserviamo che se A e B sono due eventi indipendenti allora:

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[A] \Pr[B]}{\Pr[B]} = \Pr[A]$$

Questo risultato non dovrebbe stupire nessuno: il concetto di eventi indipendenti esprime proprio il fatto che l'evento B non influenzi in alcun modo l'evento A, rendendo la probabilità inalterata. Ovviamente, questa implicazione vale anche al contrario, ossia $\Pr[A \mid B] = \Pr[A]$ se e solo se A e B sono indipendenti.

Proposizione 1.2: Indipendenza stocastica (a due eventi)

Due eventi A e B sono indipendenti se e solo se $\Pr[A \mid B] = \Pr[A]$.

Similmente, possiamo definire una nuova forma equivalente di **probabilità totale**.

Proposizione 1.3: Probabilità totale (2° formulazione)

Sia B un evento. Data una partizione $A_1, \dots, A_n$ di Ω si ha che:

$$\Pr[B] = \sum_{i=1}^{n} \Pr[B \mid A_i] \cdot \Pr[A_i]$$

Vediamo ora un esempio. Abbiamo 10 confezioni di lampadine, ciascuna contenente 12 lampadine. Delle 10 confezioni totali, 9 contengono 10 lampadine funzionanti e 2 rotte, mentre la scatola rimanente contiene 2 lampadine funzionanti e 10 rotte. Ci chiediamo quale sia la probabilità di pescare due lampadine guaste da una qualsiasi scatola.

Il problema ci chiede di calcolare la probabilità dell'evento $2R$ definito come "peschiamo due lampadine rotte". Tramite la probabilità totale, sappiamo che:

$$\Pr[2R] = \Pr[2R \mid B] \Pr[B] + \Pr[2R \mid \overline{B}]$$

dove B è l'evento "peschiamo da una confezione buona". Dai dati del problema, ricaviamo immediatamente che $\Pr[B] = \frac{9}{10}$. Per le due probabilità condizionate, invece, l'evento noto ci permette di sapere da quale tipologia di scatola stiamo pescando, ottenendo che:

$$\Pr[2R \mid B] = \frac{\binom{2}{2}}{\binom{12}{2}} = \frac{1}{\binom{12}{2}} = \frac{10! \cdot 2!}{12!} = \frac{2}{12 \cdot 11} = \frac{1}{66}$$

$$\Pr[2R \mid B^c] = \frac{\binom{10}{2}}{\binom{12}{2}} = \frac{10!}{8! \cdot 2!} \cdot \frac{10! \cdot 2!}{12!} = \frac{10 \cdot 9}{12 \cdot 11} = \frac{15}{22}$$

Concludiamo quindi che:

$$\Pr[2R] = \Pr[2R \mid B] \cdot \Pr[B] + \Pr[2R \mid \overline{B}] \cdot \Pr[\overline{B}] = \frac{1}{66} \cdot \frac{9}{10} + \cdot \frac{15}{22} \cdot \frac{1}{9} = \frac{9}{110}$$

Giusti a questo punto, discutiamo di quella che potremmo chiamare "proprietà cardine" della probabilità condizionata, dovuta a T. Bayes. Questa proprietà deriva da una semplice osservazione: essendo l'operatore $\cap$ commutativo, l'evento $A \cap B$ può essere letto anche come $B \cap A$. Per tanto, abbiamo che:

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}$$

Di primo occhio, questa semplice manipolazione algebrica non sembra così grandiosa. Tuttavia, un occhio attento può notare come tramite questa formulazione vengano **invertiti gli eventi condizionali**. Questo ha una potenza incredibile nel calcolo probabilistico: possiamo inferire la probabilità di una causa tramite l'esito della causa stessa.

Teorema 1.3: Teorema di Bayes

Dati due eventi A e B, si ha che:

$$\Pr[A \mid B] = \frac{\Pr[B \mid A]\Pr[A]}{\Pr[B]}$$

Consideriamo gli stessi dati dell'esempio precedente. Questa volta, ci chiediamo quale sia la probabilità che, avendo pescato due lampadine guaste, la scatola da cui abbiamo pescato sia quella contenente le 10 lampadine guaste. Tramite il teorema di Bayes, otteniamo facilmente che:

$$\Pr[\overline{B} \mid 2R] = \frac{\Pr[2R \mid \overline{B}]\Pr[\overline{B}]}{\Pr[2R]} = \frac{15}{22} \cdot \frac{1}{10} \cdot \frac{100}{9} = \frac{5}{6}$$

1.6 Definizione assiomatica della probabilità

Giunti a fine capitolo, dopo aver sperimentato e discusso le principali definizioni, tecniche e proprietà che un modello semplice di probabilità come quella classica rispetta, siamo pronti a dare una definizione assiomatica di probabilità. Ricordiamo che l'obiettivo di tale definizione è risolvere i problemi evidenziati ad inizio capitolo (incompletezza della probabilità basata sulla frequenza, spazi di probabilità infiniti, ...).

Osserviamo come trovare una definizione assiomatica non sia un compito facile. Chiaramente, questa definizione assiomatica deve preservare tutte le proprietà e definizioni che abbiamo dato nelle sezioni precedenti (poiché altrimenti la probabilità classica non rispetterebbe gli assiomi). Nel 1933, il matematico russo Andrey Kolmogorov propose tre assiomi per assolvere questo compito. A quasi 100 anni di distanza dalla sua proposta, questi tre assiomi sono divenuti lo standard alle fondamenta della teoria della probabilità.

Definizione 1.5: Assiomi di Kolmogorov

La tripla $(\Omega, \mathcal{P}(\Omega), \Pr)$ è detta spazio di probabilità se:

1. $\Pr : \mathcal{P}(\Omega) \to [0, 1]$, con $[0, 1] \subset \mathbb{R}$.
2. $\Pr[\Omega] = 1$
3. Per ogni sequenza $A_1, A_2, \ldots \subseteq \Omega$ tali che $A_i \cap A_j = \varnothing$, per ogni $i \neq j$, vale che:

$$\Pr\left[\bigcup_{i=1}^{\varnothing} A_i\right] = \sum_{i=1}^{\varnothing} \Pr[A_i]$$

Poniamo due osservazioni cruciali sugli assiomi appena formulati, in particolare sul primo e il terzo:

- Il primo assioma afferma che la probabilità è una funzione reale dall'insieme di tutti gli eventi (corrispondente all'insieme potenza di Ω) all'intervallo reale $[0, 1] \subset \mathbb{R}$. Implicitamente, questo ci dice non solo che tutti gli eventi assumono una probabilità compresa tra 0 e 1, ma anche che tali probabilità sono "dettate a priori" poiché definite dalla funzione stessa. In altre parole, questo ci slega dal vincolo della probabilità basata su frequenze (come nel caso di quella classica) e insiemi ambiente finiti.

- Il terzo assioma utilizzi una sequenza infinita di eventi ma esso implica in modo diretto una "versione finita" dello stesso assioma assioma: per ogni sequenza finita $A_1, \ldots, A_n$ basta considerare la sequenza infinita $A_1, \ldots, A_n, \varnothing, \varnothing, \ldots$. Difatti, la formulazione "infinita" data dell'assioma è strettamente necessaria solo affinché sia possibile trattare anche spazi di probabilità con insieme ambiente di cardinalità infinita.

Per dimostrare che tali assiomi sono una scelta valida per la definizione di probabilità che ci aspettiamo, è sufficiente dimostrare che le seguenti quattro proprietà valgono.

Proposizione 1.4

Dato uno spazio di probabilità $(\Omega, \mathcal{P}(\Omega), \Pr)$, si ha che:

- $\Pr[\varnothing] = 0$
- Per ogni $A, B \subseteq \Omega$ tali che $A \subseteq B$ vale che $\Pr[A] \leq \Pr[B]$.
- Per ogni $A \subseteq \Omega$ vale che $\Pr[\overline{A}] = 1 - \Pr[A]$
- Per ogni $A, B \subseteq \Omega$ vale che $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$

Dimostrazione. Prima di tutto, è opportuno specificare che nel dimostrare le quattro proprietà dobbiamo dimenticare completamente tutto ciò che abbiamo discusso nelle sezioni precedenti. In altre parole, possiamo utilizzare solo ed esclusivamente i tre assiomi di Kolmogorov.

1. Poiché $\Omega = \Omega \cup \varnothing$ e poiché $\Omega \cap \varnothing = \varnothing$, per il terzo assioma (in particolare la sua versione finita) abbiamo che $\Pr[\Omega] = \Pr[\Omega] + \Pr[\varnothing]$, ma ciò può essere vero solo se $\Pr[\varnothing] = 0$.

2. Fissiamo due eventi $A, B \subseteq \Omega$ tali che $A \subseteq B$. Poiché $B = (B - A) \cup A$ e poiché $(B - A) \cap A = \varnothing$, per il terzo assioma (in particolare la sua versione finita) abbiamo che $\Pr[B] = \Pr[B - A] + \Pr[A]$. Per il primo assioma, sappiamo che ognuna di tali probabilità è non-negativa. Per tanto, concludiamo che $\Pr[B] \geq \Pr[A]$.

3. Fissiamo un evento $A \subseteq \Omega$. Poiché $\Omega = A \cup \overline{A}$ e poiché $A \cap \overline{A} = \varnothing$, per il terzo assioma (in particolare la sua versione finita) abbiamo che $\Pr[\Omega] = \Pr[A] + \Pr[\overline{A}]$. Per il secondo assioma, sappiamo che $\Pr[\Omega] = 1$, dunque $1 = \Pr[A] + \Pr[\overline{A}]$, ma ciò può essere vero solo se $\Pr[\varnothing] = 1 - \Pr[A]$.

4. Fissiamo due eventi $A, B \subseteq \Omega$. Poiché $A \cup B = A \cup (B - A)$ e poiché $A \cap (B - A) = \varnothing$, per il terzo assioma (in particolare la sua versione finita) abbiamo che $\Pr[A \cup B] = \Pr[A] + \Pr[B - A]$. Inoltre, osserviamo come $B - A = B - (A \cap B)$, ottenendo che $\Pr[A \cup B] = \Pr[A] + \Pr[B - (A \cap B)]$. Successivamente, poiché $B = (B - (A \cap B)) \cup (A \cap B)$ e poiché $(B - (A \cap B)) \cap (A \cap B) = \varnothing$, per il terzo assioma (sempre la sua versione finita) abbiamo che $\Pr[B] = \Pr[B - (A \cap B)] + \Pr[A \cap B]$. Unendo le due equazioni ottenute, concludiamo che $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$

$\square$

Una volta ricavate le precedenti quattro proprietà, è possibile ricavare tutte le proprietà, principi e teoremi ricavati nelle sezioni precedenti, a patto che le definizioni date (indipendenza, condizionamento, ...) restino inalterate. D'ora in poi, tratteremo la probabilità solo tramite il concetto di spazio di probabilità – dunque assegnando le singole probabilità tramite la funzione Pr – abbandonando quindi definitivamente la probabilità classica.

In particolare, osserviamo come la definizione assiomatica implichi che affinché sia noti tutti i valori dettati da Pr è sufficiente che siano definite le probabilità degli *eventi atomici*, ossia degli eventi contenenti un solo esito. Consideriamo ad esempio l'insieme ambiente $\Omega = \{a, b, c, d\}$. Impostiamo le seguenti probabilità per gli eventi atomici $\{a\}, \{b\}, \{c\}$ e $\{d\}$:

$$\Pr[\{a\}] = \frac{1}{8} \qquad \Pr[\{b\}] = \frac{1}{4} \qquad \Pr[\{c\}] = \frac{1}{2} \qquad \Pr[\{d\}] = \frac{1}{8}$$

Una volta definite queste quattro probabilità, il terzo assioma impone automaticamente tutte le altre probabilità.

$$\begin{gathered}
\Pr[\{a, b\}] = \Pr[\{a\}] + \Pr[\{b\}] = \frac{1}{8} + \frac{1}{4} = \frac{3}{8} \\
\Pr[\{a, c\}] = \Pr[\{a\}] + \Pr[\{c\}] = \frac{1}{8} + \frac{1}{2} = \frac{5}{8} \\
\Pr[\{a, d\}] = \Pr[\{a\}] + \Pr[\{d\}] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\
\Pr[\{b, c\}] = \Pr[\{b\}] + \Pr[\{c\}] = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \\
\Pr[\{b, d\}] = \Pr[\{b\}] + \Pr[\{d\}] = \frac{1}{4} + \frac{1}{8} = \frac{3}{8} \\
\Pr[\{c, d\}] = \Pr[\{c\}] + \Pr[\{d\}] = \frac{1}{2} + \frac{1}{8} = \frac{5}{8} \\
\Pr[\{a, b, c\}] = \Pr[\{a\}] + \Pr[\{b\}] + \Pr[\{c\}] = \frac{1}{8} + \frac{1}{4} + \frac{1}{2} = \frac{7}{8} \\
\Pr[\{a, b, d\}] = \Pr[\{a\}] + \Pr[\{b\}] + \Pr[\{d\}] = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} = \frac{1}{2} \\
\Pr[\{a, c, d\}] = \Pr[\{a\}] + \Pr[\{c\}] + \Pr[\{d\}] = \frac{1}{8} + \frac{1}{2} + \frac{1}{8} = \frac{3}{4} \\
\Pr[\{b, c, d\}] = \Pr[\{b\}] + \Pr[\{c\}] + \Pr[\{d\}] = \frac{1}{4} + \frac{1}{2} + \frac{1}{8} = \frac{7}{8} \\
\Pr[\{a, b, c, d\}] = \Pr[\{a\}] + \Pr[\{b\}] + \Pr[\{c\}] + \Pr[\{d\}] = \frac{1}{8} + \frac{1}{4} + \frac{1}{2} + \frac{1}{8} = 1
\end{gathered}$$

2
Variabili aleatorie