1 Introduzione

Per un gran numero di fenomeni non è possibile, o non è conveniente, dare una descrizione deterministica, ovvero dare un insieme di leggi che, note certe premesse (condizioni iniziali) permettono di descrivere il fenomeno con certezza e precisione. Questo avviene ad esempio quando non si ha una conoscenza completa di come si sviluppa il fenomeno, ad esempio a causa di un numero eccessivo di variabili da tenere in considerazione oppure a causa dell'incertezza introdotta dagli apparati di misura delle grandezze in gioco. Per descrivere tali problemi è dunque necessario introdurre un approccio probabilistico, cercando di stimare la probabilità con cui si verifica un dato evento. La branca della matematica che si occupa di modellizzare problemi di predizione e derivare delle "regole di calcolo" che, nota la probabilità di eventi elementari, permettono di calcolare la probabilità di eventi complessi è detta *calcolo delle probabilità* .

1.1 Cosa si intende per probabilità di un evento?

L'interesse matematico per questo tipo di problemi ha origine nel 1600 (con lo studio della teoria dei giochi), ma per avere una formalizzazione matematicamente rigorosa del concetto di probabilità di un evento dobbiamo aspettare fino all'inizio del 1900 con l'assiomatizzazione di Kolmogorov.
Da un punto di vista intuitivo possiamo pensare alla probabilità di un evento come il grado di fiducia nel suo verificarsi. Per trattare questo concetto in maniera matematicamente rigorosa il matematico russo A.N. Kolmororov ha introdotto la definizione di probabilità in termini di una *misura* non negativa sull'insieme dei possibili eventi.

2 Spazio delle prove ed eventi

Supponiamo di effettuare un esperimento il cui risultato non è prevedibile con certezza, però è noto l'insieme dei possibili risultati. Tale insieme viene detto **spazio delle prove** e viene indicato con Ω. Ad esempio se l'esperimento consiste nel lancio di un dado lo spazio Ω è l'insieme

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

Se invece l'esperimento consiste nella misura del tempo impiegato dal un nucleo radioattivo a decadere lo spazio Ω sará l'insieme di tutti i numeri reali positivi, ovvero $\Omega = [0, +\infty)$.

Definiamo **evento** ogni sottoinsieme dello spazio Ω. Ad esempio nel caso del lancio di un dado, l'evento "esce il numero 2" è rappresentato dall'insieme $A = \{2\}$, mentre l'evento "esce un numero dispari" è rappresentato dall'insieme $B = \{1, 3, 5\}$. Nell'esperimento sul decadimento del nucleo radioattivo l'evento "il nucleo decade dopo l'istante T_1 e prima dell'istante T_2" è rappresentato dall'insieme $A = [T_1, T_2]$.

Tale modellizzazione degli eventi come insiemi dello spazio delle prove è utile perchè le operazioni di unione, di intersezione insiemistica e di passaggio al complementare possono venire applicate il questo contesto.
Dati due eventi $A, B \subseteq \Omega$, l'evento **unione** $A \cup B$ di A e B rappresenta l'evento in cui "si verifica l'evento A oppure l'evento B". L'evento **intersezione** $A \cap B$ di A e B rappresenta l'evento in cui "si verificano sia l'evento A sia l'evento B", mentre l'evento complementare di A, indicato con $A^c = \Omega \setminus A$ rappresenta l'evento in cui "non si verifica l'evento A".

Ad esempio, nel caso del lancio di un dado, se $A = \{1, 3, 5\}$ rappresenta l'evento "esce un numero dispari" e $B = \{1, 2, 3, 4\}$ rappresenta l'evento "esce un numero minore o uguale a 4" allora $A \cup B = \{1, 2, 3, 4, 5\}$ rappresenta l'evento "esce un numero dispari o un numero minore o uguale a 4", $A \cap B = \{1, 3\}$ rappresenta l'evento "esce un numero dispari minore di 4", mentre $A^c = \{2, 4, 6\}$ rappresenta l'evento "esce un numero pari".

Se due eventi A, B sono tali che $A \cap B = \emptyset$ allora A e B sono detti *disgiunti* o *incompatibili*.

3 Probabilità di eventi

Modellizziamo ora la probabilità di un evento. Ad ogni evento assegnamo un numero, compreso tra 0 e 1, che misura quanto è verosimile che tale evento si verifichi.

Definizione 1. *Dato uno spazio delle prove Ω, una probabilità P è un'applicazione che assegna ad ogni sottoinsieme $A \subseteq \Omega$ un numero reale $P(A)$, che soddisfa le seguenti proprietà :*

1. $0 \leq P(A) \leq 1$,
2. $P(\Omega) = 1$,
3. *data una successone $\{A_n\}_n$ di eventi a due a due disgiunti si ha che $P(\cup_n A_n) = \sum_n P(A_n)$.*

Dall'ultima proprietà segue che $P(\emptyset) = 0$.

Ad esempio nell'esperimento del lancio di un dado, se questo non è truccato è verosimile supporre che tutti i 6 possibili risultati siano equiprobabili, assegnamo quindi

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

La probabilità dei restanti eventi viene calcolata utilizzando la proprietà 3 di P, ad esempio rappresentando l'evento $\{1, 2, 4\}$ come unione di eventi disgiunti $\{1, 2, 4\} = \{1\} \cup \{2\} \cup \{4\}$ abbiamo:

$$P(\{1, 2, 4\}) = P(\{1\}) + P(\{2\}) + P(\{4\}) = \frac{1}{2}.$$

Questo esempio è un caso particolare della *probabilità uniforme*. Dato uno spazio delle prove Ω con un numero finito di elementi n allora la probabilità uniforme su Ω è quella che assegna ad ogni evento semplice di Ω, ovvero ad ogni sottoinsieme $\{\omega\}$ di Ω con un unico elemento, la stessa probabilità p. Il valore p può quindi venire calcolato notando che

$$1 = P(\Omega) = P(\cup_{\omega\in\Omega}\{\omega\}) = \sum_{\omega\in\Omega} P(\{\omega\}) = np,$$

da cui $p = 1/n$. Quindi la probabiltà di un sottoinsieme $A \subseteq \Omega$ con m elementi è data da

$$P(A) = m/n = \frac{\#A}{\#\Omega}.$$

In tutti i casi in cui si suppone che su un insieme Ω sia definita una probabilità uniforme, la probabilità di un evento A viene calcolata come rapporto tra il numero di elementi di A ed il numero di elementi di Ω. Per calcolare queste quantità vengono applicate le regole del *calcolo combinatorio*.

3.1 Alcune regole di calcolo per P

- $P(A^c) = 1 - P(A)$, infatti A e A^c sono disgiunti e $P(A \cup A^c) = P(\Omega) = 1$.
- Dati due eventi A, B, non necessariamente disgiunti, la probabilità della loro unione è data da:

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

Ad esempio consideriamo l'esperimento in cui si lancia due volte un dado. Lo spazio delle prove è l'insieme Ω delle coppie (n_1, n_2), dove $n_1, n_2 =$

$1, 2, 3, 4, 5, 6$. In tutto ha $6 * 6 = 36$ elementi. Se il dado non è truccato, possiamo supporre che ogni singolo evento sia equiprobabile, ovvero che su Ω sia definita la probabilità uniforme
Supponiamo di voler calcolare la probabilità dell'evento "esce almeno una volta il numero 6". Tale evento è rappresentato dall'insieme

$$E = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6),(1,6),(2,6),(3,6),(4,6),(5,6)\}.$$

Tale evento può essere rappresentato anche come unione dell'evento A "al primo lancio esce il numero 6" e dell'evento B "al secondo lancio esce il numero 6". Abbiamo che $P(A) = 1/6$, $P(B) = 1/6$, mentre $P(A \cap B) = P(\{(6,6)\}) = 1/36$. Si ha dunque

$$P(A \cup B) = 1/6 + 1/6 - 1/36 = 11/36.$$

3.2 Probabilità condizionata

Sia dato uno spazio delle prove Ω e una probabilità P sugli eventi di Ω. Fissiamo inoltre un evento $E \subseteq \Omega$, tale che $P(E) > 0$. Si definisce **probabilità condizionata** di un evento A dato l'evento E, e si indica con $P(A|E)$, una nuova probabilità che esprime un giudizio su quanto è verosimile che l' evento A si verifichi *sapendo che si è verificato l'evento* E.
Viene calcolata tramite la formula seguente

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

che risulta più intuitiva se scritta come $P(A \cap E) = P(E)P(A|E)$. $P(\ |E)$ è una nuova misura di probabilità a tutti gli effetti, infatti verifica le proprietà 1,2,3 della definizione 1.

Consideriamo l'esperimento del lancio di due dadi (non truccati). Supponiamo che l'evento E sia "il risultato del primo lancio è 4" e l'evento A "la somma dei risultati dei due lanci è 6". Vogliamo calcolare la probabilità che la somma dei risultati dei due lanci dia 6 sapendo che il risultato del primo lancio è 4, ovvero la probabilità condizionata di A dato E.
Per calcolare tale valore possiamo ragionare nel modo seguente. Dato che sappiamo che il risultato del primo lancio è 4, allora per il lancio dei due dadi si presentano le seguenti possibilità $\{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$. Nell'ipotesi che il dado non sia truccato allora questi 6 coppie di risultati sono equiprobabili e la loro probabilità è $1/6$. L'evento "la somma dei risultati dei due lanci è 6" è rappresentato dall'insieme $(4,2)$ la cui probabilità è appunto $1/6$.

Potevamo però ragionare anche nel modo seguente. La probabilità condizionata di A dato E è data da

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

dove $P(E) = 1/6$ e $P(A \cap E) = P((4,2)) = 1/36$, quindi

$$P(A|E) = \frac{1/36}{1/6} = 1/6.$$

3.3 Eventi indipendenti

Dato uno spazio delle prove Ω e una probabilità P sui sottoinsiemi di Ω, due eventi $A, B \subseteq \Omega$ sono detti indipendenti se

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

In altre parole A e B sono indipendenti se $P(A|B) = P(A)$: il verificarsi dell'evento B non ha influenza sulla possibilità di verificarsi dell'evento A.

Ad esempio, se consideriamo il lancio di due dadi, indichiamo con A l'evento "la somma dei risultati è 6" e con B l'evento "il risultato del lancio del primo dado è 4", allora si ha che

$$P(A \cap B) = P(\{(4,2)\}) = 1/36$$

mentre $P(B) = 1/6$, $P(A) = P\left(\{(1,5),(2,4),(3,3),(4,2),(5,1)\}\right) = 5/36$, quindi

$$P(A)P(B) = \frac{1}{6}\frac{5}{36} \neq \frac{1}{36} = P(A \cap B)$$

quindi i due eventi non sono indipendenti.
Consideriamo ora l'evento C "la somma dei risultati è 7" e con B l'evento "il risultato del lancio del primo dado è 4", allora si ha che

$$P(C \cap B) = P(\{(4,3)\}) = 1/36$$

mentre $P(B) = 1/6$, $P(C) = P\left(\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}\right) = 6/36 = 1/6$, quindi

$$P(C)P(B) = \frac{1}{6}\frac{1}{6} = \frac{1}{36} = P(C \cap B)$$

quindi i due eventi sono indipendenti.

La definizione di indipendenza può essere estesa ad un numero arbitrario n di eventi. Dati n eventi $A_1, A_2, ..., A_n$, questi sono detti indipendenti se per ogni $m \leq n$, scelto un sottoinsieme $A_{1'}, A_{2'}, ..., A_{m'}$ di $A_1, A_2, ..., A_n$ si ha che

$$P(A_{1'} \cap A_{2'} \cap ... \cap A_{m'}) = P(A_{1'})P(A_{2'}) \cdots P(A_{m'}).$$

Il seguente esempio mostra come in certe situazioni tre eventi non sono indipendenti se considerati insieme, anche se sono a due a due indipendenti. Consideriamo l'esperimento dell'estrazione di 4 palline, numerate da 1 a 4, da un'urna. Consideriamo i tre eventi

$$A_1 = \{1, 2\}, \quad A_2 = \{1, 3\}, \quad A_3 = \{1, 4\}.$$

Assumendo una probabilità uniforme su $\Omega = \{1, 2, 3, 4\}$ abbiamo che

$$P(A_1) = 1/2 \quad P(A_2) = 1/2, \quad P(A_3) = 1/2,$$

$$\begin{aligned} P(A_1 \cap A_2) &= P(\{1\}) = 1/4 = P(A_1)P(A_2), \\ P(A_1 \cap A_3) &= P(\{1\}) = 1/4 = P(A_1)P(A_3), \\ P(A_2 \cap A_3) &= P(\{1\}) = 1/4 = P(A_2)P(A_3), \end{aligned}$$

però abbiamo che

$$P(A_1 \cap A_2 \cap A_3) = P(\{1\} = 1/4 \neq P(A_1)P(A_2)P(A_3) = 1/8.$$

3.4 La formula di Bayes

Dato uno spazio delle prove Ω, probabilità P sui sottoinsiemi di Ω e due eventi $A, B \subset \Omega$ tali che $P(B) \neq 0$ e $P(A) \neq 0$, la formula di Bayes mette in relazione la probabilità condizionata di A dato B con la probabilità condizionata di B dato A:

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

Tale formula si dimostra facilmente notando che la probabilità dell'evento intersezione di A e B può essere ottenuta sia come $P(A \cap B) = P(B|A)P(A)$, sia come $P(A \cap B) = P(A|B)P(B)$, da cui

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

Il seguente esempio mostra un'applicazione interessante della formula di Bayes.

Una popolazione è composta per il 40% da fumatori (F) e per il 60% da non fumatori (N). Il 25% dei fumatori e il 7% dei non fumatori siono affetti da una malattia respiratoria cronica. Qual'è la probabilità che una persona affetta dalla malattia sia fumatore ? Qual'è la probabilità che una persona affetta dalla malattia sia non fumatore?
Indichiamo con Ω l'insieme di tutti gli individui, con F l'insieme degli individui che fumano, con N l'insieme degli individui che non fumano e con M l'insieme degli individui affetti dalla malattia. I dati in nostro possesso sono

$$P(F) = 0.4, \qquad P(N) = 0,6, \qquad P(M|F) = 0.25, \qquad P(M|N) = 0.07.$$

Vogliamo calcolare $P(F|M)$ e $P(N|M)$. Per utilizzare la formula di Bayes $P(F|M) = \frac{P(M|F)P(F)}{P(M)}$ dobbiamo prima calcolare $P(M)$. Rappresentiamo l'insieme M come l'unione disgiunta di $M \cap F$ e di $M \cap N$, quindi

$$\begin{aligned} P(M) &= P(M \cap F) + P(M \cap N) \\ &= P(M|F)P(F) + P(M|N)P(N) = \\ &= 0.25 * 0.4 + 0.07 * 0.6 = 0.142 \end{aligned}$$

quindi

$$P(F|M) = \frac{P(M|F)P(F)}{P(M)} = \frac{0.25 * 0.4}{0.142} = 0.704$$

$$P(N|M) = \frac{P(M|N)P(N)}{P(M)} = \frac{0.07 * 0.6}{0.142} = 0.296$$

4 Elementi di calcolo combinatorio

Il calcolo combinatorio fornisce una serie di regole di calcolo per calcolare il numero di elementi di un insieme.
É basato sul *principio di fattorizzazione*:
Se si devono compiere k scelte e per la $i-$ scelta abbiamo N_i possibilità , allora il numero totale di alternative possibili è

$$n_1 * n_2 * .... * n_k$$

Esempio Se abbiamo 6 camice, 2 pantaloni, 5 paia di calzini e 3 paia di scarpe, il numero totale di modi in cui ci si può vestire è $6 * 2 * 5 * 3 = 180$.
Esempio Considerando le 21 lettere dell'alfabeto italiano il numero totale di parole (sensate o meno) composte da 4 lettere è $21 * 21 * 21 * 21 = (21)^4$
Esempio Un'urna contiene 90 palline, numerate da 1 a 90. Si effettuano 4 estrazioni e dopo ogni estrazione la pallina estratta viene reinserita nell'urna. Allora il numero totale di sequenze di numeri estratti è $90*90*90*90 = (90)^4$.

4.1 Disposizioni

Definizione 2. *Sia N un insieme con n elementi $N = \{x_1, x_2, ..., x_n\}$. Una disposizione di k elementi di N è una k-upla ordinata $(x_{i_1}, x_{i_2}, ..., x_{i_k})$ di elementi di N tutti distinti fra loro.*

In numero totale delle possibili disposizioni di k elementi di un insieme con n elementi è

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Esempio Considerando le 21 lettere dell'alfabeto italiano il numero totale di parole (sensate o meno) composte da 4 lettere distinte è $21 * 20 * 19 * 18 = \frac{21!}{17!}$
Esempio Un'urna contiene 90 palline, numerate da 1 a 90. Si effettuano 4 estrazioni e dopo ogni estrazione la pallina estratta non viene reinserita nell'urna. Allora il numero totale di sequenze di numeri estratti è $90 * 89 * 88 * 87 = 90!/86!$.

4.2 Permutazioni

Definizione 3. *Sia N un insieme con n elementi $N = \{x_1, x_2, ..., x_n\}$. Una disposizione di n elementi di N, ovvero una n-upla ordinata $(x_{i_1}, x_{i_2}, ..., x_{i_nk})$ di elementi distinti di N, è detta permutazione di N.*

Il numero totale di permutazioni di un insieme con n elementi è $n!$
Esempio Una classe è composta da 20 studenti. Il numero totale di modi in cui possono venire ordinati (es ordine alfabetico...) è 20!

4.3 Combinazioni

Definizione 4. *Sia N un insieme con n elementi $N = \{x_1, x_2, ..., x_n\}$. Una combinazione di k elementi di N è un sottoinsieme di N composto da k elementi $\{x_{i_1}, x_{i_2}, ..., x_{i_k}\}$*

Il numero totale delle possibili combinazioni di n elementi a k a k è

$$\frac{n!}{(n-k)!k!}$$

Esempio Una classe è composta da 20 studenti. Il numero totale di gruppi formati da tre studenti è $\frac{20!}{3!17!}$.

4.4 Un esempio

Calcoliamo la probabilità che in un'aula con k studenti, almeno 2 compiano gli anni lo stesso giorno.

Indichiamo con N l'insieme dei 365 giorni giorni dell'anno (per semplicità escludiamo il 29 febbraio). Numeriamo i giorni dell'anno da 1 a 365:

$$N = \{1, 2, 3, ..., 365\}$$

Rappresentiamo i compleanni dei k studenti della classe con un $k-$pla ordinata di elementi di N e indichiamo con Ω l'insieme di tali $k-$ple ordinate:

$$\Omega = \{\omega = (\omega_1, \omega_2, ..., \omega_k),\ \omega_i \in N\}.$$

L'insieme Ω ha $(365)^k$ elementi distinti.

Se supponiamo per semplicità che tutti i giorni dell'anno siano equiprobabili, ovvero che ci sia una distibuzione uniforme delle nascite (cosa di fatto non completamente vera), allora ogni elemento di Ω è equiprobabile e su Ω possiamo considerare la probabilità uniforme P. La probabilità di un generico sottoinsieme $A \subset \Omega$ è quindi

$$P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{(365)^k}$$

Vogliamo ora calcolare la probabilità dell'evento "almeno due persone compiono gli anni lo stesso giorno", rappresentato dall'insieme $A \subset \Omega$ delle $k-$ple ordinate $(\omega_1, \omega_2, ..., \omega_k)$ in cui almeno due elementi sono uguali. Di fatto è più semplice calcolare la probabilità dell'evento complementare A^c e poi trovare la probabilità di A tramite la formula

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

L'evento complementare di A è l'evento "tutte le persone compiono gli anni in giorni diversi", rappresentato dall'insieme $A^C \subset \Omega$ delle $k-$ple ordinate $(\omega_1, \omega_2, ..., \omega_k)$ in cui tutti gli elementi elementi sono distinti. A^C è quindi l'insieme delle diposizioni di k elementi dell'insieme N ed il numero totale di tali disposizioni è $n!/k! = 365!/k!$. La probabilità di A^c è dunque

$$P(A^c) = \frac{365!}{k!(365)^k}$$

e la probabilità di A è quindi

$$P(A) = 1 - P(A^C) = 1 - \frac{365!}{k!(365)^k}$$

Ad esempio, per k=23 otteniamo $P(A) = 0.507 > 1/2$.

5 Variabili aleatorie

Spesso in un esperimento non si è interessati solamente ai possibili risultati, bensì ad una funzione del risultato. Immaginiamo ad esempio il lancio di due dadi e supponiamo che si vinca se la somma dei risultati sia maggiore o uguale a 7, mentre si perda altrimenti. In questo caso la variabile di interesse non è la coppia dei valori ottenuti nei due lanci quanto la loro somma.
Un altro esempio è il gioco in cui lanciando un dado se il risultato è strettamente maggiore di 3 allora si vince una somma pari a S, mentre se il risultato è minore o uguale di 3 si perde una somma pari a 2S. Supponiamo che ci interessi sapere qual'è la somma che ci aspettiamo di vincere, o perdere, dopo 5 partite.
Questa classe di problemi può essere formalizzata introducendo il concetto di variabile aleatoria.

Definizione 5. *Sia Ω uno spazio delle prove e P una probabilità sui sottoinsiemi di Ω. Si chiama variabile aleatoria una funzione $X : \Omega \to \mathbb{R}$.*

Si usa distinguere tra *variabili aleatori discrete*, ovvero variabili aleatorie X che possono assumere solo un numero discreto di valori, e *variabili aleatorie continue* ovvero variabili aleatorie che assumono valori all'interno di un insieme continuo.
Ad esempio nell'esperimento del lancio di un due dadi, scegliendo come variabile aleatoria X la somma dei risultati dei due lanci, abbiamo che X puó assumere i valori 2,3,4,5,6,7,8,9,10,11,12 e dunque è una variabile aleatoria discreta.
Nell'esperimento della misura del decadimento di un nucleo radiaottivo, scegliendo come variabile aleatoria X l'instante in cui il nucleo decade, abbiamo che, in linea di principio, X può assumere un qualsiasi valore all'interno dell'intervallo $[0, +\infty)$. X è dunque una variabile aleatoria continua.

Data una variabile aleatoria X, possiamo introdurre, dato un insieme $I \subset \mathbb{R}$, la probabilità che la variabile X assuma valori contenuti all'interno di A, ovvero

$$P(\{\omega \in \Omega, : X(\omega) \in I\})$$

L'applicazione, che ad un insieme $I \subset \mathbb{R}$ associa $P(\{\omega \in \Omega, : X(\omega) \in I\})$ è detta *legge* o *distribuzione* della variabile aleatoria X.

Si definisce inoltre la *funzione di distribuzione* $F_X : \mathbb{R} \to [0, 1]$, data da

$$F_X(t) := P(\{\omega \in \Omega : X(\omega) \leq t\}) = P(X \leq t)$$

per il suo significato, F è monotona crescente (in generale non strettamente). Si ha che

$$\lim_{t\to-\infty} F_X(t) = 0, \qquad \lim_{t\to+\infty} F_X(t) = 1$$

inoltre

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

5.1 Densità di probabilità

Consideriamo una variabile aleatoria discreta X, che può quindi assumere solo un insieme discreto di valori $\{x_1, x_2, ...\}$. La distribuzione di X è completamente specificata se conosciamo la probabilità degli insiemi

$$\{\omega \in \omega \,:\, X(\omega) = x_i\}, \qquad i = 1, 2, ...$$

Indichiamo con $p(x_i) = P(\{\omega \in \omega \,:\, X(\omega) = x_i\})$. la funzione p è detta *densità di probabilità discreta* e gode delle seguenti proprietà :

- $p(x_i) \geq 0$ per ogni i,
- $\sum_i p(x_i) = 1$.

Nota la densità discreta p, allora per ogni $I \subseteq \mathbb{R}$, si ha che

$$P(X \in I) = P(\{\omega \in \Omega \,:\, X(\omega) \in I\}) = \sum_{x_i \in I} p(x_i).$$

Ad esempio, nel caso dell'esperimento del lancio di un due dadi, scegliendo come variabile e aleatoria X la somma dei risultati dei due lanci, abbiamo che la distribuzione di X è data da:

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

Quindi, ad esempio,

$$P(4 \leq X \leq 6) = p(4) + p(5) + p(6) = \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{12}{36}$$

Nel caso di una variabile aleatoria continua X, si ha che $P(X = x) = 0$ e si definisce la *densità di probabilità* f, come quella funzione $f : \mathbb{R} \to \mathbb{R}^+$ tale che

$$F_X(t) = \int_{-\infty}^{t} f(u)du$$

$$P(X \in [a,b]) = \int_{a}^{b} f(u)du$$

Notiamo che, per il teorema fondamentale del calcolo integrale, si ha che:

$$f(t) = F'_X(t)$$

Nel caso dell'esperimento della misura del tempo di decadimento di un nucleo radioattivo si ha che

$$P(X \leq t) = 1 - e^{-\lambda t}, t > 0 \qquad P(X \in [T_1, T_2]) = e^{-\lambda T_1} - e^{-\lambda T_2}.$$

dove λ è una costante positiva che caratterizza l'elemento.
La densità di probabilità f è data da

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \qquad t \geq 0 \\ f(t) &= 0 \qquad\quad\; t < 0 \end{aligned}$$

5.2 Media e varianza

Nei problemi concreti, spesso non è conveniente e neppure necessario conoscere l'intera distribuzione della variabile aleatoria X, ma è sufficiente conoscere un numero finito di parametri, che indicano il modo in cui X è distribuita.

5.2.1 Media

Si definisce *media* o *speranza matematica* o *valore atteso* di una variabile aleatoria X, e si indica col simbolo $\mathbb{E}[X]$ o μ_X il numero

$$\mu_X = \mathbb{E}[X] := \begin{cases} \sum_i x_i p(x_i) & \text{per variabili aleatorie discrete} \\ \int_{-\infty}^{+\infty} x f(x) dx & \text{per variabili aleatorie continue} \end{cases}$$

nell'ipotesi la sommatoria (estesa a tutti i possibili valori che possono essere assunti dalla variabile aleatoria discreta) o l'integrale siano convergenti. In caso contrario la media non esiste.

Ad esempio, nel caso dell'esperimento del lancio di un due dadi, scegliendo come variabile e aleatoria X la somma dei risultati dei due lanci, la media di X è data da

$$\begin{aligned} \mu_X = \mathbb{E}[X] &= 2p(2) + 3p(3) + 4p(4) + 5p(5) + 6p(6) + 7p(7) + 8p(8) + 9p(9) + \\ &\qquad +10p(10) + 11p(11) + 12p(12) \\ &= 2\frac{1}{36} + 3\frac{2}{36} + 4\frac{3}{36} + 5\frac{4}{36} + 6\frac{5}{36} + 7\frac{6}{36} + 8\frac{5}{36} + 9\frac{4}{36} + 10\frac{3}{36} + \\ &\qquad +11\frac{2}{36} + 12\frac{1}{36} \\ &= \frac{252}{36} = 7 \end{aligned}$$

Nel caso dell'esperimento della misura del tempo di decadimento di un nucleo radioattivo, dove la densità di probabilità f è data da

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \qquad t \geq 0 \\ f(t) &= 0 \qquad\quad\; t < 0 \end{aligned}$$

abbiamo che il valore medio è dato da

$$\mu_X = \mathbb{E}[X] = \int_0^{+\infty} t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

5.2.2 Varianza

La varianza di una variabile aleatoria X, indicata con il simbolo σ_X^2, è definita come il valore atteso del quadrato dello scarto tra il valore della variabile X e al sua media μ_X:

$$\sigma_X^2 = \mathbb{E}[(X - \mu_X)^2].$$

La varianza fornisce quindi una misura della "dispersione" della variabile X rispetto alla sua media μ_X.
Per una variabile aleatoria discreta si calcola come:

$$\sigma_X^2 = \sum_i (x_i - \mu_X)^2 p(x_i),$$

mentre per una variabile aleatoria continua si calcola come

$$\sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx,$$

(nell'ipotesi che la sommatoria o l'integrale convergono, in caso contrario la varianza non esiste).
Si dimostra facilmente una formula alternativa per il calcolo della varianza:

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mu_X)^2$$

Per una variabile aleatoria discreta si calcola quindi come:

$$\sigma_X^2 = \sum_i x_i^2 p(x_i) - (\mu_X)^2,$$

mentre per una variabile aleatoria continua si calcola come

$$\sigma_X^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - (\mu_X)^2.$$

Ad esempio, nel caso dell'esperimento del lancio di un due dadi, scegliendo come variabile e aleatoria X la somma dei risultati dei due lanci, abbiamo visto che la media μ_X vale $\mu_X = 7$. Per calcolare la varianza, calcoliamo prima $\mathbb{E}[X^2]$

$$\begin{aligned}
\mathbb{E}[X^2] &= 2^2p(2) + 3^2p(3) + 4^2p(4) + 5^2p(5) + 6^2p(6) + 7^2p(7) + 8^2p(8) + 9^2p(9) + \\
&\qquad +10^2p(10) + 11^2p(11) + 12^2p(12) \\
&= 4\frac{1}{36} + 9\frac{2}{36} + 16\frac{3}{36} + 25\frac{4}{36} + 36\frac{5}{36} + 49\frac{6}{36} + 64\frac{5}{36} + 81\frac{4}{36} + 100\frac{3}{36} + \\
&\qquad +121\frac{2}{36} + 144\frac{1}{36} \\
&= \frac{1974}{36} = 54.8
\end{aligned}$$

Quindi

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mu_X)^2 = 54.8 - (7)^2 = 5.8$$

Nel caso dell'esperimento della misura del tempo di decadimento di un nucleo radioattivo, dove la densità di probabilità f è data da

$$\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \qquad t \geq 0 \\
f(t) &= 0 \qquad\qquad t < 0
\end{aligned}$$

abbiamo che il valore medio è dato da $\mu_X = 1/\lambda$, mentre

$$\mathbb{E}[X^2] = \int_0^{+\infty} t^2 \lambda e^{-\lambda t} dt = \frac{2}{\lambda^2}.$$

Quindi la varianza è data da

$$\sigma_X^2 = \mathbb{E}[X^2] - (\mu_X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

5.3 Variabili aleatorie congiunte

Due variabili aleatorie $X, Y : \Omega \to \mathbb{R}$ definite sullo stesso spazio Ω vengono dette variabili aleatorie congiunte.

Ad esempio consideriamo l'esperimento del lancio di due dati (non truccati) e indichiamo con X la somma dei risultati e con Y il valore assoluto della differenza.

COnsideriamo l' esperimento che consiste nello scegliere a caso un individuo dalla popolazione di una città e nel misurarne l'altezza X e il peso Y. In tal caso X e Y sono due variabili aleatorie congiunte in quanto definite sullo stesso insieme Ω, l'insieme degli individui della città .

Se le due variabili X e Y sono entrambe discrete, allora ha senso calcolare la probabilità che assumano rispettivamente dei particolari valori x_i e y_j. Si definisce *densità di probabilità congiunta* di X, Y la probabilità

$$P(X = x_i, Y = y_j) \equiv p_{X,Y}(x_i, y_j)$$

che gode delle seguenti proprietà :

- $\sum_{x_i, y_j} p_{X,Y}(x_i, y_j)$.
- Sommando la densità congiunta su tutti i possibili valori di una variabile casuale, si ottiene la densità di probabilità dell'altra variabile, detta anche *densità marginale*:

$$\sum_{y_j} p_{X,Y}(x_i, y_j) = p_X(x_i),$$

$$\sum_{x_i} p_{X,Y}(x_i, y_j) = p_Y(y_j).$$

  $p_X(x_i)$ ha il significato di probabilità che la variabile X assuma il valore x_i indipendentemente dal valore assunto dalla variabile Y. Analogamente $p_Y(y_j)$ ha il significato di probabilità che la variabile Y assuma il valore y_j indipendentemente dal valore assunto dalla variabile X.

Nel caso in cui le due variabili aleatorie congiunte sono entrambe continue, allora si definisce la funzione *densità di probabilità congiunta* $f_{X,Y}$ come quella funzione $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ tale che

$$P(a \leq X \leq b, c \leq Y \leq d) = \int\int_{[a,b]\times[c,d]} f_{X,Y}(x, y)dxdy$$

Come nel caso di variabili aleatorie discrete, la funzione densità di probabilità congiunta soddisfa le seguenti proprietà

- $\int\int_{\mathbb{R}^2} f_{X,Y}(x, y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x, y)dxdy = 1.$
- Integrando la densità congiunta su tutti i possibili valori di una variabile casuale, si ottiene la densità di probabilità dell'altra variabile, detta anche *densità marginale*:

$$\int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy = f_X(x),$$

$$\int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx = f_Y(y).$$

$f_X(x)$ ha il significato di densità di probabilità della variabile X indipendentemente dal valore assunto dalla variabile Y, cioé:

$$P(a \leq X \leq b, Y \in \mathbb{R}) = \int_a^b f_X(x)dx.$$

Analogamente $f_Y(y)$ ha il significato di densità di probabilità della variabile Y indipendentemente dal valore assunto dalla variabile X, cioé:

$$P(X \in \mathbb{R}, Y \in [c,d]) = \int_c^d f_Y(y)dy.$$

Due variabili aleatorie congiunte X, Y sono dette *indipendenti* se:

$$p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j), \qquad \text{se } X \text{ e } Y \text{ sono variabili aleatorie discrete}$$

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \qquad \text{se } X \text{ e } Y \text{ sono variabili aleatorie continue}$$