Appunti di Probabilità

Matematica II - Corso di Laurea Informatica

Indice

CAPITOLO 1

Calcolo combinatorio

Immaginiamo di avere una collezione finita di oggetti (cioè un insieme finito) e di sceglierne o estrarne a sorte alcuni, cioè di costruire un nuovo insieme raggruppando alcuni degli elementi dati. In primo luogo, dovremo fissare le regole del gioco:

- quali e quanti oggetti abbiamo a disposizione? (inquadriamo il problema nell'ambito di uno o più insiemi matematici)
- quali e quanti oggetti scegliamo? (da quanti elementi sarà costituita la nostra collezione?)
- l'ordinamento è importante? (due collezioni che contengono gli stessi oggetti in diverso ordine sono uguali o diverse?)
- si può usare ripetutamente lo stesso oggetto? (la nostra collezione sarà fatta di elementi distinti o si possono avere più ripetizioni di uno stesso oggetto?)

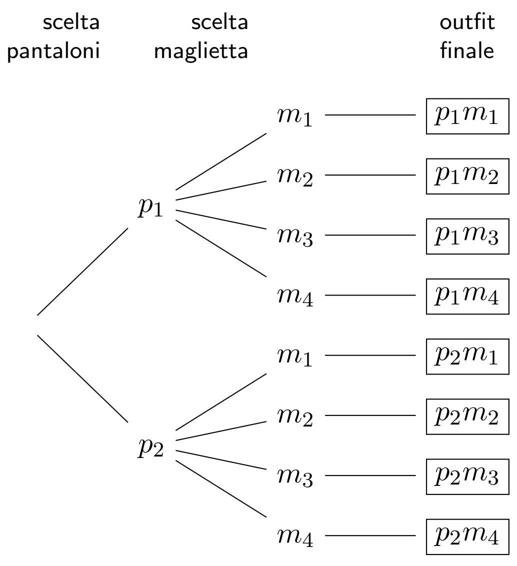
Fatto ciò, possiamo chiederci quanti sono i possibili risultati della nostra scelta, cioè in quanti modi possiamo raggruppare e/o ordinare gli oggetti assegnati. Il **calcolo combinatorio** si occupa di contare tali modi, ovvero le configurazioni, e solitamente risponde a domande quali "Quanti sono...", "In quanti modi...", "Quante possibili combinazioni..." eccetera.

Per poter entrare nel dettaglio della discussione, dobbiamo dare alcune (troppe) definizioni, che illustriamo via via con degli esempi.

1. Raggruppamenti

Se ho due pantaloni e quattro magliette, in quanti modi mi posso vestire? Il mio outfit è un *raggruppamento* in cui posso scegliere un oggetto nell'insieme "pantaloni" e uno nell'insieme "magliette".

Per rispondere alla domanda, costruiamo un diagramma ad albero in cui indichiamo i pantaloni con le lettere p_1, p_2 e le magliette con m_1, m_2, m_3, m_4.



Possiamo scegliere i pantaloni in due modi diversi.
Per ogni scelta di pantaloni, possiamo poi scegliere la maglietta in 4 modi diversi.
In totale abbiamo quindi $2 \cdot 4 = 8$ possibili outfit, come si vede anche contando le foglie dell'albero (steso).

Si può anche inquadrare la questione in modo insiemistico. Indichiamo con $P \coloneqq \{p_1, p_2\}$ l'insieme dei pantaloni e con $M \coloneqq \{m_1, m_2, m_3, m_4\}$ quello delle magliette. Un outfit è una coppia costituita da un elemento di P e uno di M, ovvero un elemento del prodotto cartesiano $P \times M = \{(x,y) : x \in P,\ y \in M\}$, quindi i possibili outfit sono tanti quanti gli elementi del prodotto cartesiano. D'altra parte, in teoria degli insiemi vale il seguente Teorema

TEOREMA 1.1 (Cardinalità dell'insieme prodotto). *Se A e B sono due insiemi finiti, costituiti rispettivamente da n ed m elementi, allora il loro prodotto cartesiano $A \times B$ è formato esattamente da $n \cdot m$ elementi.*
Analogamente, se $A_1, A_2, A_3 \dots A_k$ sono k insiemi finiti, formati rispettivamente da $n_1, n_2, n_3 \dots n_k$ elementi, allora il loro prodotto cartesiano $A_1 \times A_2 \times \cdots A_k$ è formato esattamente da $n_1 \cdot n_2 \cdot n_3 \cdots n_k$ elementi.

In definitiva, possiamo rispondere alla domanda iniziale anche contando il numero degli elementi dell'insieme prodotto $P \times M$. Poiché P ha 2 elementi ed M ne ha 4, rispondiamo che mi posso vestire in $2 \cdot 4 = 8$ modi diversi.

Diamo una definizione generale di raggruppamento e una formula per calcolare il numero di tutti i possibili raggruppamenti.

Definizione 1.1: Raggruppamento

Sia k un numero naturale e siano $A_1, A_2, A_3 \dots A_k$ degli insiemi non vuoti. Un raggruppamento di k elementi è una collezione di k oggetti scelti rispettivamente in $A_1, A_2, \dots A_k$.

Sottolineamo che due raggruppamenti si considerano *diversi* se almeno uno degli oggetti che li compongono differisce. Nell'esempio ininziale, due outfit formati dallo stesso pantalone combinato con magliette differenti sono considerati diversi.

L'applicazione del Teorema 1.1 al calcolo combinatorio fornisce il seguente

Teorema 1.2: Principio di moltiplicazione delle scelte

Sia k un numero naturale e siano $A_1, A_2, A_3 \dots A_k$ degli insiemi finiti, non vuoti. Per contare in quanti modi posso raggruppare un oggetto di A_1, uno di A_2 e così via fino ad A_k bisogna moltiplicare il numero di elementi di A_1 per quello di A_2 e così via fino ad A_k.

Il Teorema 1.2 è anche noto come Teorema fondamentale del calcolo combinatorio.

Esercizio 1.1:

In mensa offrono 3 scelte per il primo, 2 per il secondo e 4 per il contorno. Quanti pasti completi si possono comporre?

Svolgimento. Bisogna contare tutti i raggruppamenti formati da un primo piatto, un secondo e un contorno, che in base al Teorema 1.2 sono $3 \cdot 2 \cdot 4 = 24$. □

Esercizio 1.2:

Se termina un secondo, come cambia il numero dei pasti completi che si possono scegliere?

Esercizio 1.3:

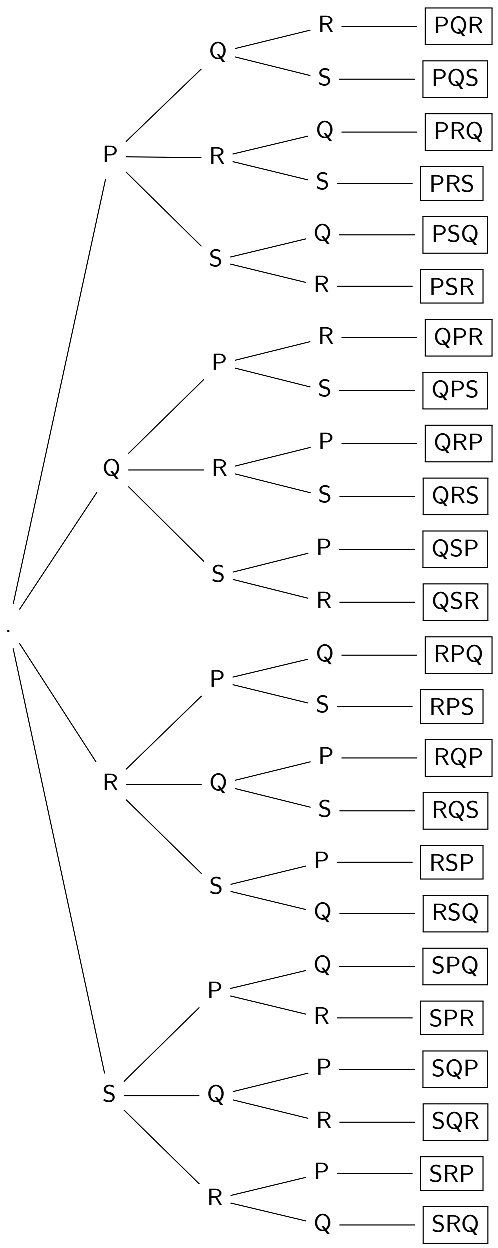
a) *In una compagnia di 8 ragazze e 11 ragazzi, quante coppie formate da un ragazzo e una ragazza si possono formare?*

b) *Ho 5 biglie nere numerate da 1 a 5 e 3 biglie bianche numerate da 1 a 3. Quante coppie formate da una biglia nera e una bianca si possono avere?*
Quante di queste coppie contengono solo numeri dispari? E quante contengono un numero pari e uno dispari?

2. Raggruppamenti ordinati di oggetti distinti

2.1. Disposizioni semplici. Ad una gara con quattro partecipanti (Paolo, Quirino, Roberto e Simone) vengono premiati i primi tre classificati. Quante sono le possibili classifiche dei premiati? Precisiamo che due terne di premiati si intendono diverse se almeno uno dei premiati è diverso, ma anche se i premiati sono gli stessi, ma l'ordine d'arrivo è diverso. La classifica dei premiati è un esempio di *disposizione semplice* di tre oggetti su quattro.

Per rispondere alla domanda, costruiamo un diagramma ad albero in cui indichiamo i partecipanti con le loro iniziali:



Procedendo da sinistra verso destra, nella prima colonna troviamo i possibili vincitori, che sono tanti quanti i partecipanti, cioè 4.
Una volta che il vincitore è noto, passando alla colonna successiva troviamo i $4 - 1 = 3$ partecipanti che possono arrivare secondi, e per il terzo classificato ne restano solo $4 - 2 = 2$.
Complessivamente, le possibili disposizioni sono quindi $4 \cdot 3 \cdot 2 = 24$.

Diamo ora la definizione e la formula generale.

Definizione 2.1: Disposizione semplice

Dati n oggetti distinti, sia $k \leq n$. Una disposizione semplice di k oggetti su n è una collezione ordinata di k oggetti distinti scelti tra gli n disponibili.
La si chiama anche disposizione semplice di n elementi di classe k.
Indichiamo col simbolo $D_{n,k}$ il numero di tutte le disposizioni semplici di k oggetti su n.

Nella definizione abbiamo sottolineato per ben due volte la parola "distinti", per enfatizzare che i k oggetti che formano una disposizione semplice devono essere diversi tra loro. Nel nostro esempio, se Paolo arriva primo allora non potrà essere né secondo né terzo. É per questa caratteristica che la disposizione viene detta *semplice*.
Abbiamo anche sottolineato l'aggettivo "ordinata", a sottolineare che due disposizioni formate dagli stessi oggetti in differente ordine sono da considerarsi diverse. Nel nostro esempio, la classifica PQR (1^o Paolo, 2^o Quirino e 3^o Roberto) è diversa da PRQ (1^o Paolo, 2^o Roberto e 3^o Quirino).

Generalizzando il ragionamento, si può dimostrare la formula che segue.

Proposizione 2.2: Formula per le disposizioni semplici

Siano n e k due numeri naturali con $n \geq k > 0$. Le possibili disposizioni semplici di k elementi su n sono complessivamente $n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)$, in simboli

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

Esercizio 2.1:

a) Quante parole di tre lettere tutte diverse tra loro possiamo formare (contando anche le "parole" che non hanno un significato)?
b) Quanti numeri interi di quattro cifre hanno tutte le cifre diverse tra loro?
c) Quanti sono i possibili podii della finale dei 100m alle Olimpiadi?
d) Una ditta deve assumere tre dipendenti, da collocare rispettivamente in amministrazione, in contabilità e nell'ufficio commerciale. Sapendo che sono arrivati 20 curriculum, in quanti modi diversi si può assumere?

Esercizio 2.2: Estrazione ripetuta senza reintegro (ordinata)

Ho dieci palline numerate da 1 a 10. Ne estraggo tre, una di seguito all'altra, dando importanza all'ordine in cui le estraggo. Quanti sono gli esiti possibili della mia estrazione?

2.2. Permutazioni semplici. Abbiamo visto che per fare una disposizione semplice di k oggetti su n è necessario che k sia minore o uguale a n. Quando k coincide con n, non sto *scegliendo* gli oggetti (li prendo tutti!), piuttosto li sto *ordinando*. In questo caso particolare, parlo di *permutazione*. È questo il caso se guardo l'intero ordine di arrivo in una gara, oppure se compongo gli anagrammi di una parola. In un anagramma, infatti, devo usare tutte le lettere una volta sola.

Definizione 2.3: permutazione semplice

Dati n oggetti distinti, una loro permutazione o riordinamento è una presentazione ordinata di tutti gli n oggetti, ognuno dei quali viene preso una sola volta.
Col simbolo P_n indichiamo il numero totale delle permutazioni di n oggetti distinti.

Due permutazioni diverse differiscono solo per l'*ordine*, ma presentano gli stessi oggetti. Di fatto, $P_n = D_{n,n}$, cioè una permutazione di n oggetti è una disposizione semplice di n oggetti su n. Di conseguenza la Proposizione 2.2 fornisce anche una formula per calcolare il numero delle possibili permutazioni.

Proposizione 2.4: Formula per le permutazioni semplici

Sia n un numero naturale, $n \geq 1$. Le possibili permutazioni semplici di n oggetti sono complessivamente $n \cdot (n-1) \cdot (n-2) \cdots 1$, in simboli

$$P_n = n \cdot (n-1) \cdot (n-2) \cdots 1 = n!$$

Ricordiamo alcune immediate proprietà del fattoriale:

$$0! = 1, \qquad 1! = 1, \qquad (n+1)! = (n+1) \cdot n!, \qquad n! = n \cdot (n-1)!$$

Esercizio 2.3

Calcolare quanti sono tutti i possibili anagrammi della parola PORTA.

Svolgimento. Se accettiamo tutti i possibili risultati (cioè non richiediamo che la parola ottenuta sia di senso compiuto nella nostra lingua), gli anagrammi della parola PORTA corrispondono alle permutazioni di 5 elementi distinti, e sono dunque $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. □

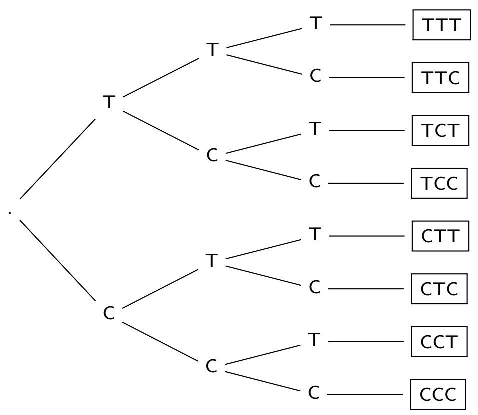
Esercizio 2.4:

Quanti sono i possibili ordini di arrivo della finale dei 100m alle Olimpiadi?
Se sappiamo quali atleti sono saliti sul podio, quanti sono i possibili ordini di arrivo dal quarto all'ottavo posto? Per rispondere a quest'ultima domanda, vi serve sapere chi ha vinto l'oro, l'argento e il bronzo?

3. Raggruppamenti ordinati con ripetizioni

3.1. Disposizioni con ripetizioni. Immaginiamo ora di lanciare una moneta per tre volte. Quanti sono i risultati che possiamo ottenere? Poiché ad ogni lancio possiamo ottenere testa oppure croce, indichiamo l'insieme dei possibili risultati di un lancio con $A \coloneqq \{T, C\}$. L'esito complessivo dopo tre lanci sarà una terna di lettere T e C: poichè ad ogni lancio i possibili risultati sono testa o croce indipendentemente da quanto successo al lancio precedente, nella successione degli esiti può comparire la stessa lettera ripetuta più volte. Inoltre consideriamo diverse due terne costituite dagli stessi esiti ottenuti in ordine diverso. Siamo di fronte ad una disposizione con ripetizioni di tre oggetti su due.

Costruiamo un diagramma ad albero:



Procedendo da sinistra verso destra, nelle prime tre colonne troviamo gli esiti di primo, secondo e terzo lancio, mentre nell'ultima colonna abbiamo elencato la successione degli esiti. Per ogni lancio, i possibili risultati sono due indipendentemente da quanto successo al lancio precedente, quindi nella successione degli esiti può comparire la stessa lettera ripetuta più volte. Poichè ci interessa anche l'ordine in cui sono usciti testa o croce, due terne che hanno le stesse lettere in ordine differente sono da considerarsi diverse.

L'insieme dei possibili esiti è descritto esattamente dal prodotto cartesiano $A \times A \times A$. A ben pensarci, stiamo proprio facendo un raggruppamento di 3 oggetti scelti in tre copie identiche dello stesso insieme A. Ne deduciamo che i possibili esiti sono complessivamente $2 \cdot 2 \cdot 2 = 2^3 = 8$.

Definizione 3.1: Disposizione con ripetizioni

Dati n oggetti distinti, sia k un qualunque numero naturale non nullo. Una disposizione con ripetizioni di k oggetti su n (è una collezione ordinata di k oggetti scelti tra gli n disponibili. La si chiama anche disposizione con ripetizioni di n elementi di classe k.
Col simbolo $D'_{n,k}$ indichiamo il numero di tutte le disposizioni con ripetizioni di k oggetti su n.

Osserviamo che, a differenza delle disposizioni semplici, nelle disposizioni con ripetizioni il numero di elementi scelti (k) puó essere maggiore del numero di oggetti tra cui scegliamo (n). Va poi sottolineato che una disposizione è un insieme ordinato, quindi due disposizioni che contengono gli stessi oggetti in diverso ordine sono differenti.

Il Principio di moltiplicazione delle scelte Teorema 1.2 ci fornisce una formula per calcolare il numero di disposizioni possibili.

Proposizione 3.2: Formula per le disposizioni con ripetizioni

Siano n e k due numeri naturali con $n, k > 0$. Le possibili disposizioni con ripetizioni di k elementi su n sono complessivamente n^k, in formule

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

Esercizio 3.1

Quanti codici di 6 cifre si possono creare utilizzando solo le cifre dispari?

Svolgimento. Le cifre a disposizione sono 5, precisamente $\{1, 3, 5, 7, 9\}$. Un codice è una disposizione di 6 cifre, che certamente si possono ripetere e si differenziano sia per le cifre che appaiono, sia per il loro ordine. Si tratta quindi di una disposizione (con ripetizioni) di classe 6 su 5 oggetti. Le possibili configurazioni sono dunque $D'_{5,6} = 5^6 = 15625$. □

Esercizio 3.2

Quante sono le targhe automobilistiche?

Svolgimento. Una targa automobilistica è una sequenza in cui appare una coppia di lettere, poi una terna di cifre, poi di nuovo una coppia di lettere.
Cominciamo contando quante sono le possibili configurazioni della coppia di lettere iniziali: dobbiamo prendere due lettere e ne abbiamo a disposizione 22 (le lettere O, Q, U, I non sono utilizzate per evitare errori di lettura). Poiché c'interessa l'ordine e possiamo ripetere la stessa lettera, usiamo la formula delle disposizioni non ripetizione $D'_{22,2} = 22^2 = 484$. Chiaramente la stessa cosa vale anche per la coppia finale.
Per quanto riguarda la terna di cifre, invece, le configurazioni possibili sono $D'_{10,3} = 10^3 = 1\,000$.
Infine, per comporre la targa dobbiamo creare un raggruppamento in cui scegliamo una coppia di lettere, una terna di cifre, una coppia di lettere. Il Principio di moltiplicazione delle scelte Teorema 1.2 ci dice allora che le targhe possibili sono $484 \cdot 1\,000 \cdot 484 = 234\,256\,000$. □

3.2. Permutazioni con ripetizioni. Proviamo ora a contare tutti i possibili anagrammi della parola `porto`, compresi quelli che non hanno senso compiuto. In un anagramma, dobbiamo usare tutte le lettere a disposizione (5) una e una sola volta. A differenza dell'esercizio Esercizio 2.3 (gli anagrammi di `porta`), però, non dobbiamo contare tutte le permutazioni semplici, perché ora la lettera o è ripetuta. Ciò fa si che se scambio tra loro la seconda e la quinta lettera non ottengo un nuovo anagramma, ma ripeto la parola porto. Siamo di fronte ad un esempio di permutazione con ripetizioni, e abbiamo già intuito che le possibili configurazioni sono di meno rispetto alle permutazioni semplici. Precisamente sono la metà, perché a ogni coppia di anagrammi della parola porta che differiscono solo perché le lettere o ed a appaiono invertite, corrisponde un solo anagramma della parola porto. Concludo quindi che i possibili anagrammi sono $5!/2 = 60$.

Cosa succede se anagrammo la parola poppa? Gli anagrammi di una parola di 5 lettere distinte sono 5!, ma ora identifico tutti gli anagrammi che differiscono solo per la posizione delle 3 lettere p. Suggerimento: non basta dividere per 3. Per convincercene, indicizziamo le lettere della parola e scriviamo $\mathsf{p}_1\mathsf{o}\mathsf{p}_2\mathsf{p}_3\mathsf{a}$. Ci sono $3! = 6$ permutazioni che cambiano solo le posizioni di $\mathsf{p}_1, \mathsf{p}_2$ e p_3, precisamente

$$\mathsf{p}_1\mathsf{o}\mathsf{p}_2\mathsf{p}_3\mathsf{a},\ \mathsf{p}_1\mathsf{o}\mathsf{p}_3\mathsf{p}_2\mathsf{a},\ \mathsf{p}_2\mathsf{o}\mathsf{p}_1\mathsf{p}_3\mathsf{a},\ \mathsf{p}_2\mathsf{o}\mathsf{p}_3\mathsf{p}_1\mathsf{a},\ \mathsf{p}_3\mathsf{o}\mathsf{p}_1\mathsf{p}_2\mathsf{a},\ \mathsf{p}_3\mathsf{o}\mathsf{p}_2\mathsf{p}_1\mathsf{a}.$$

A questi 3! riordinamenti corrisponde una sola permutazione con ripetizioni. Ne deduco quindi che i possibili anagrammi sono $5!/3! = 20$.

Siamo pronti per dare una definizione ed una formula generale.

Definizione 3.3: Permutazione con ripetizioni

Dati j oggetti distinti e j numeri naturali $n_1, n_2, \dots n_j$, consideriamo l'insieme che contiene n_1 copie del primo oggetto, n_2 copie del secondo, e cosí via. Tale insieme ha $n := n_1 + n_2 + \cdots n_j$ elementi. Una permutazione con ripetizioni è una presentazione ordinata di tutti gli n oggetti dell'insieme.
Indichiamo il numero di tutte le permutazioni con ripetizioni con il simbolo $P'_{n_1,n_2,\dots n_j}$.

Alla luce dei ragionamenti precedenti possiamo affermare che

Proposizione 3.4: Formula per le permutazioni con ripetizioni

Sono dati j oggetti distinti e j numeri naturali $n_1, n_2, \dots n_j$, poniamo $n := n_1 + n_2 + \cdots n_j$. Il numero complessivo delle permutazioni di un insieme che contiene esattamente n_1 copie di un oggetto, n_2 copie di un altro e cosí via è

$$P'_{n_1,n_2,\dots n_j} = \frac{n!}{n_1!n_2!\cdots n_j!}.$$

Si noti che, poiché $1! = 1$, se un oggetto non è ripetuto non dà alcun contributo al denominatore.

Esercizio 3.3

Si calcoli quanti sono tutti gli anagrammi della parola serratura.

Svolgimento. Abbiamo a disposizione complessivamente 9 lettere, di cui però la r e la a sono ripetute, rispettivamente, 3 e 2 volte. La formula precedente ci dice dunque che i possibili anagrammi sono $9!/3!2! = 30\,240$. □

4. Raggruppamenti non ordinati

4.1. Combinazioni semplici. Nelle combinazioni, a differenza delle disposizioni, non si dà importanza all'ordine in cui vengono scelti gli oggetti.

Andrea, Beatrice, Carlo, Davide ed Emma vorrebbero andare ad un concerto, ma hanno a disposizione solo tre biglietti. In quanti modi possono scegliere chi andrà al concerto? Ci interessano solo i nomi dei 3 ragazzi che avranno il biglietto, indipendentemente dall'ordine in cui li elenchiamo, pertanto stiamo operando una combinazione di 3 elementi su 5. Il gruppo composto da Andrea, Beatrice ed Emma corrisponde alla terna ABE, ma anche a tutte le sue permutazioni (o anagrammi) AEB, BAE, BEA, EAB, EBA. Abbiamo complessivamente $P_3 = 6$ terne che rappresentano disposizioni diverse, ma danno luogo ad una unica combinazione. È chiaro che qualunque altra combinazione può provenire da 6 diverse disposizioni. Ne deduciamo che il numero delle combinazioni semplici di 3 oggetti su 5 si ottiene prendendo il numero di disposizioni semplici $D_{5,3}$ e dividendolo per il numero di permutazioni P_3. Otteniamo cosí

$$\frac{D_{5,3}}{P_3} = \frac{\frac{5!}{2!}}{3!} = \frac{5!}{2!3!} = 10.$$

Possiamo darci anche un'altra spiegazione. Elenchiamo i ragazzi in ordine alfabetico, e accanto al loro nome scriviamo la lettera S se hanno ottenuto il biglietto, la lettera N se non l'hanno avuto. Se, ad esempio, hanno avuto il biglietto Andrea, Beatrice ed Emma scriviamo

S	Andrea
S	Beatrice
N	Carlo
N	Davide
S	Emma

A ben pensare, ogni combinazione corrisponde ad una parola di 5 lettere (tante quanti sono i ragazzi) di cui 3 sono S (tante quanti i biglietti) e le restanti sono N. Dunque il numero complessivo di combinazioni semplici è dato da

$$P'_{3,2} = \frac{5!}{3!2!} = 10.$$

Enunciamo la definizione e la formula generale.

Definizione 4.1: Combinazione semplice

Dati n oggetti distinti, sia $k \leq n$. Una combinazione semplice di k oggetti su n è una collezione di k oggetti distinti scelti tra gli n disponibili. La si chiama anche combinazione semplice di n elementi di classe k.
Indichiamo con $C_{n,k}$ il numero complessivo delle combinazioni semplici di k oggetti su n.

Notate che l'unica differenza fra questa nuova definizione e la Definizione 2.1 è che qui manca l'aggettivo "ordinata".

Proposizione 4.2: Formula per le combinazioni semplici

Siano n e k due numeri naturali con $n \geq k > 0$. Le possibili combinazioni semplici di k elementi su n sono complessivamente

$$C_{n,k} = \frac{D_{n,k}}{P_k} = P'_{k,n-k} = \frac{n!}{(n-k)!k!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!}.$$

La formula che appare nella Proposizione 4.2 è il cosiddetto "coefficiente binomiale", che appare tra l'altro nello sviluppo della potenza di un binomio.

Definizione: coefficiente binomiale

Se n e k sono numeri naturali con $n \geq k \geq 0$. Il simbolo $\binom{n}{k}$ si legge coefficiente binomiale n, k e indica il numero

$$\binom{n}{k} \coloneqq \frac{n!}{(n-k)!k!}.$$

Possiamo quindi sintetizzare la Proposizione Proposizione 4.2 nella formula

$$C_{n,k} = \binom{n}{k}.$$

Riportiamo alcune proprietà dei coefficienti binomiali che possono risultare utili.

Proposizione 4.3: Proprietà dei coefficienti binomiali

Siano n, k numeri interi con $n \geq k \geq 0$. Dalla definizione di coefficiente binomiale discende che

$$\binom{n}{n} = \binom{n}{0} = 1,$$

$$\binom{n}{n-k} = \binom{n}{k} \qquad \textit{legge delle classi complementari,}$$

$$\binom{n}{k+1} = \binom{n}{k}\frac{n-k}{k+1} \qquad \textit{formula di ricorrenza.}$$

Esercizio 4.1

a) Un concorso con 20 concorrenti si svolge in due fasi: una scritta ed una orale. Al termine della prova scritta, solo i primi 8 vengono ammessi alla prova orale. Quante sono le possibili configurazioni di concorrenti all'orale? Se gli ammessi all'orale fossero 12 anziché 8, il numero di configurazioni aumenterebbe o diminuirebbe? Perché?

b) Abbiamo 6 scatole numerate da 1 a 6 e 4 anelli identici. Sapendo che ogni anello è dentro una scatola e che nessuna scatola contiene più di un anello, quante sono le possibili configurazioni?

4.2. Combinazioni con ripetizioni. Per dare un esempio di combinazione con ripetizioni, immaginiamo di lanciare per tre volte una moneta e ci chiediamo quanti risultati diversi possiamo avere. A differenza della Sezione 3, però, non distinguiamo tra i tre lanci consecutivi, dunque ci interessa solo sapere quante volte è uscito testa e quante volte croce. Stiamo ora facendo una combinazione di tre oggetti (i risultati dei lanci) su due (testa e croce), che ovviamente presenterà delle ripetizioni.

Definizione 4.4: Combinazione con ripetizioni

Sono dati n e k numeri interi con $n, k > 0$. Dati n oggetti <u>distinti</u>, una combinazione con ripetizioni di k oggetti su n è una collezione di k oggetti (non necessariamente distinti) scelti tra gli n disponibili. La si chiama anche combinazione con ripetizioni di n elementi di classe k.
Il numero di tutte le combinazioni con ripetizioni viene indicato con $C'_{n,k}$.

Due combinazioni, dunque, possono differire perché contengono oggetti diversi, oppure perché contengono gli stessi oggetti ma in diverso <u>numero</u>.

Calcolare il numero delle possibili combinazioni con ripetizioni è un po' più complesso dei casi precedenti. Per farci un'idea, torniamo al nostro esempio, elenchiamo le possibili combinazioni e di fianco scriviamo a quali disposizioni corrispondono:

combinazione	disposizioni corrispondenti	numero di disposizioni
TTT	TTT	$1 = P'_{3,0}$
TTC	TTC , TCT , CTT	$3 = P'_{2,1}$
TCC	TCC , CTC , CCT	$3 = P'_{1,2}$
CCC	CCC	$1 = P'_{0,3}$

È ancora vero che ad ogni combinazione corrispondono tutte le disposizioni che si ottengono per permutazione, ma ora il numero di permutazioni è variabile. In questo caso è facile elencare tutte le possibili combinazioni, e vediamo cosí che sono complessivamente 4.

Per ricavare una formula generale costruiamo un problema equivalente. Immaginiamo di avere 3 caramelle (tante quanti i lanci di moneta) da distribuire tra 2 bambini che si chiamano "Testa" e "Croce". La combinazione TTC, ad esempio, corrisponde a dare 2 caramelle a Testa e 1 a Croce, e la schematizziamo come segue:

Testa		Croce
• •		•
A A	B	A

Scriviamo una A per ogni caramella che diamo e B quando passiamo da un bambino al successivo

In modo analogo, la parola BAAA corrisponde a nessuna Testa e 3 Croce, cioè alla combinazione CCC. Con questo schema, ogni combinazione corrisponde ad una parola formata da 3 lettere A e 1 lettera B, quindi per la Proposizione 3.4 ne abbiamo $P'_{3,1} = \dfrac{(3+1)!}{3!1!} = 4$, come già calcolato in precedenza.

In generale, ogni combinazione con ripetizioni di k oggetti scelti tra n corrisponde ad una parola formata da k lettere A (tante quante sono le scelte che dobbiamo fare) e $n-1$ lettere B (una in meno rispetto agli oggetti tra cui possiamo scegliere, perché corrisponde al "passaggio" da un oggetto al successivo). Cosí, la Proposizione 3.4 fornisce la formula $P'_{k,n-1} = \dfrac{(k+n-1)!}{k!(n-1)!} = \binom{k+n-1}{k}$.

In sintesi

Proposizione 4.5: Formula per le combinazioni con ripetizioni

Siano n e k due numeri naturali con $n, k > 0$. Le possibili combinazioni con ripetizioni di k elementi su n sono complessivamente

$$C'_{n,k} = P'_{k,n-1} = C_{n+k-1,k} = \binom{n+k-1}{k} = \frac{(n+k-1)\cdot(n+k-2)\cdots n}{k!}.$$

Esercizio 4.2

(1) Ho 6 caramelle da distribuire tra 4 bambini, in quanti modi diversi posso farlo?
(2) Ho 4 caramelle da distribuire tra 6 bambini, in quanti modi diversi posso farlo?

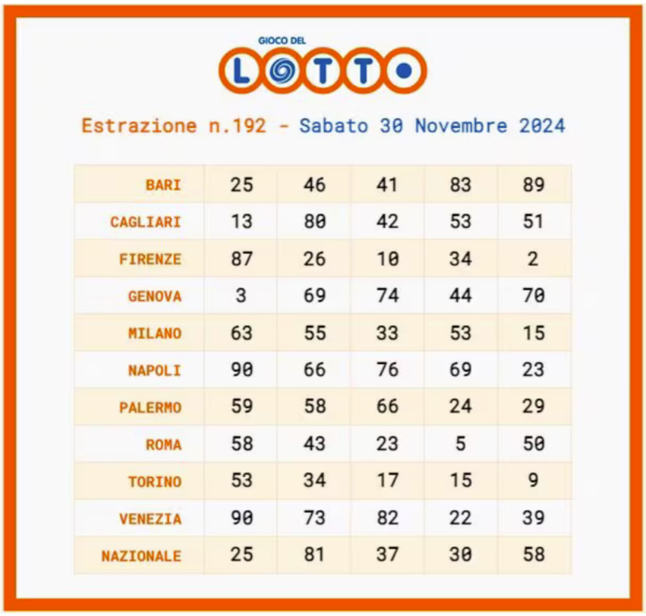
5. Esempi notevoli ed ancora esercizi

Una applicazione tipica dei raggruppamenti si ritrova nella modellizzazione delle estrazioni ripetute. Se abbiamo un'urna contenente delle biglie numerate (e dunque distinte) e facciamo due estrazioni, possiamo procedere in tre modi diversi:

- Estraiamo una biglia, la mettiamo da parte e poi procediamo con la seconda estrazione: si parla di estrazione <u>successiva senza reintegro</u>. Il risultato è una disposizione (perché conosciamo l'ordine di estrazione) semplice (perché il primo numero estratto non potrà essere selezionato nell'estrazione successiva).
- Estraiamo due biglie contemporaneamente: si parla di estrazione <u>simultanea</u>. Il risultato è una combinazione (perché i due numeri estratti non sono ordinati) semplice (perché i due numeri sono necessariamente distinti).
- Estraiamo una biglia, poi la reinseriamo nell'urna prima di procedere con la seconda estrazione: si parla di estrazione <u>successiva con reintegro</u>. Il risultato è una disposizione con ripetizioni (se ci interessa l'ordine di estrazione) o una combinazione con ripetizioni (se non ci interessa l'ordine di estrazione).

Se effettuiamo il reintegro, l'ambiente in cui effettuiamo la seconda estrazione è esattamente identico a quello in cui abbiamo effettuato la prima. Al contrario, se procediamo senza reintegro l'estrazione della seconda biglia avviene in un'urna diversa da quella da cui abbiamo estratto la prima biglia (poiché contiene un oggetto in meno).

Vediamo ora qualche esempio tratto dai “giochi d’azzardo”. Il risultato dell’ultima estrazione del Lotto è:



GIOCO DEL LOTTO

Estrazione n.192 - Sabato 30 Novembre 2024

BARI	25	46	41	83	89
CAGLIARI	13	80	42	53	51
FIRENZE	87	26	10	34	2
GENOVA	3	69	74	44	70
MILANO	63	55	33	53	15
NAPOLI	90	66	76	69	23
PALERMO	59	58	66	24	29
ROMA	58	43	23	5	50
TORINO	53	34	17	15	9
VENEZIA	90	73	82	22	39
NAZIONALE	25	81	37	30	58

Sulla ruota di Napoli abbiamo un raggruppamento di 5 numeri scelti su 90 disponibili. Poiché i numeri vengono estratti in successione dall’urna, non sono ammesse ripetizioni; pertanto il raggruppamento è semplice. Si tratta di una disposizione o di una combinazione? Poiché viene specificato l’ordine dell’estrazione, abbiamo a che fare con una disposizione. D’altra parte, l’ordine con cui vengono estratti i numeri non influisce sulla possibilità di vincere, pertanto a chi gioca al lotto interessano le combinazioni semplici di 5 numeri su 90.
Alla luce di quanto studiato, sappiamo che i possibili risultati dell’estrazione sulla ruota di Napoli sono

$$D_{90,5} = \frac{90!}{85!} = 5\,273\,912\,160.$$

Questo significa che se giochiamo 5 numeri abbiamo una possibilità su 5 273 912 160 di fare cinquina? No! In realtà ne abbiamo *molte* di più, perché vinciamo indovinando la combinazione vincente, indipendentemente dall’ordine. Le combinazioni possibili sono

$$C_{90,5} = \binom{90}{5} = 43\,949\,268.$$

In ogni caso le probabilità a me sembrano molto basse!

Quando giochiamo una colonna al Totocalcio, abbiamo a disposizione 3 oggetti distinti: 1 X e 2. Dobbiamo sceglierli 13 volte (una volta per ogni partita) e ci interessa l’ordine in cui li disponiamo nella colonna (ogni riga corrisponde ad una determinata partita). Si tratta dunque di una disposizione (con ripetizioni) di 13 oggetti su 3.
Il numero di colonne diverse che si possono giocare è

$$D'_{3,13} = 3^{13} = 1\,595\,323.$$

Se lanciamo ripetutamente per tre volte un dado e annotiamo i risultati ottenuti, possiamo farlo in due modi diversi.
Se ricordiamo l’ordine in cui sono usciti i tre numeri, otteniamo una disposizione (con ripetizioni) di 3 elementi su 6. In questo caso abbiamo

$$D'_{6,3} = 6^3 = 216$$

configurazioni diverse. Se invece ci interessano solo i tre numeri usciti, indipendentemente dall'ordine, allora abbiamo una combinazione (con ripetizioni). I risultati possibili sono di meno, precisamente

$$C'_{6,3} = \binom{8}{3} = 56.$$

Se, a parità di premio finale, doveste scegliere tra scommettere sulla cinquina al lotto, sul 13 alla schedina o sul lancio di 3 dadi, quale scegliereste?

Esercizio 5.1

Quante sono le possibili combinazioni di una cassaforte a 5 cifre?

Svolgimento. Abbiamo a disposizione 36 oggetti distinti: le 26 lettere più le 10 cifre da 0 a 9. La combinazione di una cassaforte è una disposizione con ripetizioni di 5 oggetti su 36 (scusate il gioco di parole). Dunque il numero di tutte le combinazioni possibili è $\mathcal{D}'_{36,5} = 36^5 = 60\,466\,176$. □

Esercizio 5.2

Nell'estrazione del lotto (su un'unica ruota)

a) *Quanti sono i terni che si possono giocare?*
b) *Conoscendo il risultato dell'estrazione, quanti sono i terni vincenti?*
c) *Immaginiamo di aver scelto una specifica terna di numeri. Quante sono le estrazioni (senza contare l'ordine di estrazione) che la rendono vincente?*

Esercizio 5.3

a) *Ho deciso di giocare una schedina di Totocalcio di un'unica colonna in cui ci siano esattamente cinque 1, sette X e un 2. In quanti modi diversi posso farlo?*
b) *Tra tutte le possibili colonne, quante contengono esattamente un 2? (con l'aggettivo "esattamente" si precisa che il 2 deve comparire una e una sola volta)*
c) *Quante, invece, contengono almeno un 2? (con l'avverbio "almeno" si precisa che il 2 deve comparire una o più volte)*

Esercizio 5.4

a) *In un'urna ci sono dieci palline numerate da 1 a 10; tre sono bianche e le altre nere. Quante sono le cinquine che contengono esattamente una pallina bianca?*
b) *Un'urna contiene 3 palline di colore diverso: una bianca, una rossa e una nera. Se faccio 4 estrazioni consecutive (rimettendo dentro all'urna la pallina dopo ogni estrazione), quante sequenze di colori diverse posso ottenere?*
c) *Un'urna contiene 12 palline bianche e 3 nere. Se pesco 4 palline simultaneamente, quante quaterne posso ottenere che contengano almeno una pallina nera?*

Esercizio 5.5

Ho un mazzo di carte da poker, composto da 52 carte (senza jolly), ed estraggo simultaneamente due carte.

a) *Quante sono le possibili coppie estratte?*
b) *Quante, tra le possibili coppie estratte, contengono almeno una carta di cuori?*
c) *Quante contengono due carte di quadri?*
d) *Quante contengono una carta di cuori e una di quadri?*

Esercizio 5.6

a) *In un gruppo di 25 persone, devono essere scelti due delegati. In quanti modi diversi è possibile farlo?*
b) *Sapendo che il gruppo è composto da 14 donne e 11 uomini, quante di queste delegazioni rappresentano entrambi i generi?*
c) *Rispondere alle domande precedenti considerando una delegazione composta da tre persone.*

CAPITOLO 2

Probabilità

In probabilità si considera un fenomeno osservabile esclusivamente dal punto di vista della possibilità o meno del suo verificarsi, prescindendo dalla sua natura. Tra due estremi, detti evento certo (ad esempio: lanciando un dado si ottiene un numero compreso tra 1 e 6) ed evento impossibile (ottenere 1 come somma dal lancio di due dadi), si collocano eventi più o meno probabili.

Gli oggetti della probabilità sono gli esperimenti casuali o aleatori. Ciò che contraddistingue gli esperimenti casuali, rispetto agli esperimenti scientifici, è che per loro natura non sono ripetibili. Se lascio cadere un corpo sferico di massa nota da un'altezza assegnata in ambiente neutro, so che esso cadrà per effetto della accelerazione di gravità e dunque posso prevedere con esattezza con quale velocità toccherà il suolo. Ogni volta che questo esperimento scientifico verrà effettuato, si otterrà lo stesso risultato. Se invece lascio cadere un dado, non posso prevedere con esattezza quale faccia mostrerà, e in generale lanci diversi (ma effettuati nelle stesse condizioni) daranno risultati diversi. Si parla in questo caso di esperimento casuale: non possiamo conoscere in anticipo il risultato dell'esperimento, ma sappiamo che, se effettuiamo l'esperimento molte volte, tutti e sei i possibili risultati si presenteranno con la stessa frequenza. Sappiamo prevedere l'andamento medio dell'esperimento, ma non il singolo esito.

1. Definizioni

Precisiamo il vocabolario che useremo d'ora in poi.

Definizione 1.1

Useremo indistintamente le parole **esperimento casuale**, **esperimento aleatorio** *e* **prova**. *Il risultato di un tale esperimento è detto* **esito** *o* **punto campione**. *L'insieme di tutti i possibili esiti di un esperimento aleatorio si dice* **spazio campione** *e viene solitamente indicato con la lettera greca* Ω.
Un **evento** *è un qualunque sottoinsieme dello spazio campione. Dopo avere effettuato una prova, diremo che l'evento si è* **verificato** *se l'esito appartiene all'evento.*

Esempio 1.1: Lancio di un dado

Il lancio di un dado a sei facce è un esperimento aleatorio. Gli esiti sono i numeri da 1 a 6, dunque $\Omega = \{1, 2, 3, 4, 5, 6\}$. *Sono esempi di eventi*

A: uscita di un numero pari. $A = \{2, 4, 6\}$.
B: uscita del numero 6. $B = \{6\}$.
C: uscita di un numero inferiore a 3. $C = \{1, 2\}$.
D: uscita di un numero negativo. $D = \emptyset$.
E: uscita di un numero compreso fra 1 e 6. $E = \{1, 2, 3, 4, 5, 6\}$.

Se lanciamo il dado ed esce 2, allora gli eventi A, C *ed* E *sono verificati, mentre gli eventi* B *e* D *non sono verificati.*

Esempio 1.2: Lancio di una moneta

Anche il lancio di una moneta è un esperimento aleatorio, i cui possibili esiti sono $T =$ *"testa" e* $C =$ *"croce". Dunque lo spazio campione è l'insieme* $\{T, C\}$.

Esempio 1.3: Numero di iscritti

Possiamo interpretare il numero delle immatricolazioni ad informatica come un esperimento aleatorio il cui esito è un numero naturale, cioè $\Omega = \mathbb{N}$. Sono esempi di eventi

A: nessun nuovo iscritto. $A = \{0\}$.
B: meno di 100 iscritti. $B = \{0, 1, 2 \dots 99\}$.
C: almeno 200 iscritti. $C = \{200, 201, \dots\}$.

Esempio 1.4: Altezza di un individuo

Possiamo interpretare come esperimento aleatorio la misura dell'altezza di un individuo scelto a caso, espressa in centimetri e con qualunque livello di precisione (cioè con un numero di decimali potenzialmente infinito). Ciò fa sí che ogni numero reale positivo è un possibile esito, sicché lo spazio campione è l'intervallo $(0, +\infty)$. Sono esempi di eventi

A: altezza compresa fra 170 e 180 cm. $A = [170, 180]$.
B: altezza inferiore a 165 cm. $B = (0, 165)$.
C: altezza superiore a 2 metri. $C = [200, +\infty)$.

Possiamo effettuare operazioni fra eventi facendo le corrispondenti operazioni fra insiemi.

Definizione 1.2: Operazioni tra eventi

$A \wedge B$	$= A \cap B$	*evento intersezione o prodotto logico, si legge "A e B"* *l'evento che si verifica quando sia A che B si verificano,*
$A \vee B$	$= A \cup B$	*evento unione o somma logica, si legge "A o B"* *l'evento che si verifica quando almeno uno tra A e B si verifica,*
$\neg A$	$= \Omega \setminus A$	*evento negazione, si legge "non A"* *l'evento che si verifica quando A non è verificato.*

- *L'evento che coincide con tutto lo spazio campione si dice evento certo.*
- *L'insieme vuoto è detto evento impossibile,*
- *Due eventi A e B sono mutuamente esclusivi o incompatibili se non si possono verificare contemporaneamente, cioè se l'evento $A \wedge B$ è impossibile, ovvero se $A \cap B = \emptyset$.*
- *L'evento A implica l'evento B se $A \subset B$. In questo caso, infatti, se l'esito dell'esperimento appartiene ad A, allora appartiene anche a B.*

Esempio 1.5: Mazzo di carte

Scegliamo casualmente una carta da un mazzo di carte francesi (senza Jolly). Il nostro spazio campione è costituito dalle 52 carte del mazzo. Consideriamo gli eventi

A: un asso,
B: una carta di cuori,
C: una carta di seme rosso,
D: una carta con un numero pari.

Si ha

$A \wedge B$	= *asso di cuori,*	$A \wedge D$	= *evento impossibile ,*
$B \wedge C$	= *una carta di cuori,*	$B \vee C$	= *una carta di seme rosso ,*
$\neg D$	= *una carta con un numero dispari oppure una figura.*		

Gli eventi A e D sono mutuamente esclusivi. L'evento B implica l'evento C.

1.1. La misura di probabilità. Ci vogliamo ora occupare di <u>misurare</u> la probabilità di un dato evento. Precisamente, ad ogni evento vogliamo associare un numero che cresca all'aumentare della probabilità che tale evento si verifichi. Dal punto di vista matematico, dunque, la misura di probabilità sarà una funzione che ha come *dominio* l'insieme degli eventi e fornisce come output un *numero*. Nel seguito la indicheremo con la lettera P.

Definire la legge di questa funzione non è un problema di facile soluzione, e in effetti il concetto di misura di probabilità si è evoluto nel tempo. La prima idea, attribuita a Laplace, è nota come definizione classica di probabilità.

Definizione 1.3: Probabilità - definizione classica

La probabilità di un evento è il rapporto tra il numero dei casi favorevoli all'evento e il numero dei casi possibili.

Per scrivere una formula, introduciamo qualche notazione. Indichiamo con Ω lo spazio campione, con $A \subset \Omega$ un evento e con il simbolo $|\cdot|$ il *numero* di elementi distinti contenuti in un dato insieme. Abbiamo allora

$$P(A) \coloneqq \frac{|A|}{|\Omega|}.$$

Ad esempio, se consideriamo lo spazio campione associato al lancio di un dado e l'evento A : numero pari (Esempio 1.1), abbiamo Ora $|\Omega| = 6$ e $|A| = 3$, dunque $\mathrm{P}(A) = 3/6 = 1/2$.

Osservazione 1.4

Nella vita di tutti i giorni capita spesso di parlare di probabilità in termini di percentuali. In questo caso, ad esempio, diremmo che c'è una probabilità del 50% che esca un numero pari. Non dimentichiamo che le percentuali altro non sono che frazioni con denominatore 100, e infatti $50\% = \dfrac{50}{100} = \dfrac{1}{2}$.

Dalla definizione seguono tre regole:

Proposizione 1.5: Regole di calcolo della probabilità

i) la probabilità di un evento è un numero compreso tra 0 e 1;
ii) la probabilità dell'evento certo è pari a 1;
iii) la probabilità del verificarsi di almeno uno tra due eventi incompatibili è pari alla somma delle probabilità dei due eventi, cioè

$$A \cap B = \emptyset \;\Rightarrow\; \mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B).$$

Esercizio 1.6

Un'urna contiene 7 biglie rosse, 4 bianche e 9 verdi. Si consideri l'esperimento aleatorio di estrarre una biglia dall'urna.

1. *Qual è lo spazio campione corrispondente?*
2. *Si descrivano insiemisticamente gli eventi*
   A: estrazione di una biglia rossa,
   B: estrazione di una biglia verde.
3. *Si calcolino le probabilità di* A *e di* $\neg B$.

Esercizio 1.7

Calcolare la probabilità di fare terno giocando 3 numeri al lotto sulla ruota di Napoli.

Svolgimento. In base alla nozione classica di probabilità, devo calcolare quanti sono gli esiti possibili e quanti quelli favorevoli all'evento. I possibili esiti dell'estrazione del lotto su un'unica ruota sono tutte le combinazioni semplici di 5 numeri su 90, dunque il nostro spazio campione ha $\mathcal{C}_{90,5} = \binom{90}{5}$ punti. Gli eventi a noi favorevoli sono tutte le possibili combinazioni di 5 numeri che contengono i

3 che abbiamo giocato. Esse sono tante quante le combinazioni di 2 numeri sui 97 che non abbiamo giocato, ovvero $\mathcal{C}_{87,2} = \binom{87}{2}$. Di conseguenza la probabilità di fare terno è

$$\frac{\mathcal{C}_{87,2}}{\mathcal{C}_{90,5}} = \frac{\binom{87}{2}}{\binom{90}{5}} = \frac{87!\,5!\,85!}{2!\,85!\,90!} = \frac{5\cdot 4\cdot 3}{88\cdot 89\cdot 90} = \frac{60}{704880} = 0.000085121 \approx 8\cdot 10^{-5} \quad \text{o } 0.008\%.$$

□

Nel seguito (Esercizio 2.11) vedremo tencniche più efficaci per calcolare la probabilità in situazioni di questo tipo, e anche più complesse. Esse si poggiano sulla teoria astratta della probabilità.

La nozione classica di probabilità risulta utile nel caso di spazi campione che contengono un numero finito di esiti equiprobabili (cioè tutti con uguale probabilità di verificarsi), come ad esempio nel lancio di un dado o di una moneta, o nell'estrazione da un mazzo di carte. Se però prendiamo un dado truccato, il risultato non è più corretto poiché a esiti diversi corrisponderanno diverse probabilità. Inoltre, se i possibili esiti sono infiniti (come nell'Esempio 1.2) o addirittura continui (come nell'Esempio 1.4) la formula della probabilità classica (Definizione 1.3) non può essere applicata.

Per valutare la probabilità in queste situazioni ci è utile appellarci alla statistica: dalla analisi delle immatricolazioni negli anni passati possiamo dedurre previsioni sulle immatricolazioni future. In questo caso, stiamo usando una nozione di probabilità mutuata dalla statistica, la cosiddetta nozione frequentista attribuita a von Mises.

Definizione 1.6: Probabilità - definizione frequentista

La probabilità di un evento è il limite cui tende la frequenza relativa dell'evento al crescere del numero degli esperimenti.

Per scrivere una formula, introduciamo qualche notazione. Indichiamo con n il numero di esperimenti effettuati, con A un evento e con n_A il numero degli esperimenti in cui A si è verificato.
Il rapporto n_A/n rappresenta la frequenza di A.
Abbiamo allora

$$P(A) \coloneqq \lim_{n\to\infty} \frac{n_A}{n}.$$

Anche da questa definizione si ricavano le tre regole elencate nella Proposizione 1.5.

La nozione frequentista si applica ad esperimenti casuali i cui esiti non siano ritenuti ugualmente possibili, e siano ipoteticamente infiniti. Tuttavia, essa assume che l'esperimento sia ripetibile più volte, idealmente infinite, sotto le stesse condizioni. Inoltre il limite che appare nella formula non è "calcolabile": chi conosce la legge con cui varia la frequenza (al variare del numero di esperimenti)?

Per ovviare a queste problematiche, De Finetti e Savage hanno proposto una definizione di probabilità applicabile ad esperimenti casuali che non siano necessariamente ripetibili più volte sotto le stesse condizioni

Definizione 1.7: Probabilità - definizione soggettiva

La probabilità di un evento è il prezzo che un individuo razionale ritiene equo pagare per ricevere 1 se l'evento si verifica, 0 se l'evento non si verifica.

L'aggettivo "razionale" si usa per ribadire un criterio di coerenza: la misura di probabilità deve essere attribuita in modo tale che non sia possibile ottenere una vincita o una perdita certa, cioè il prezzo pagato deve essere compreso tra 0 e 1.

Grazie al criterio di coerenza si ricavano dalla definizione soggettiva le stesse tre regole di calcolo della Proposizione 1.5.

La definizione soggettiva consente quindi di calcolare la probabilità di eventi anche quando gli eventi elementari non sono equiprobabili e quando l'esperimento non può essere ripetuto. Rimane fondata, tuttavia, sull'opinione di singoli individui, che potrebbero presentare diverse propensioni al rischio. Basta pensare che molti sarebbero disposti a giocare 1 euro per vincerne 1000, ma pochi giocherebbero un milione di euro per vincerne un miliardo.

Un decisivo cambio di prospettiva viene introdotto negli anni '30 del secolo scorso da Kolmogorov. Egli non si occupa di dare una definizione operativa e fornire indicazioni su come calcolare la probabilità, bensì di dare un assetto formale alla "teoria della probabilità", nell'ambito del quale sia possibile dimostrare teoremi che poi potranno essere utilizzati sia nell'ambito di un approccio frequentista che nell'ambito di un approccio soggettivista.

La definizione assiomatica dà alla probabilità lo status di una branca della matematica ed è quella cui faremo riferimento d'ora in avanti. Per la sua importanza, le dedichiamo una sezione distinta.

1.2. Definizione assiomatica di probabilità e proprietà. Volendo dare una nozione di probabilità che si adatti ad ogni ambito applicativo (in particolare, a spazi campione infiniti e possibilmente non numerabili), dobbiamo innanzitutto chiarire *quali eventi sono misurabili.* Nel caso di spazi campione finiti o numerabili, non ci sono grossi problemi nel misurare la probabilità di un qualunque evento (cioè di un qualunque sottoinsieme dello spazio campione), ma ciò non è più vero se trattiamo con spazi campione che abbiano la potenza del continuo (come $\mathbb{R}$ o un intervallo).

Se Ω è un insieme qualunque, indichiamo con 2^Ω il suo *insieme delle parti*, ovvero la famiglia di tutti i suoi sottoinsiemi. Definiamo poi

Definizione 1.8: σ-algebra

Sia $\mathcal{E} \subset 2^\Omega$ una famiglia non vuota di sottoinsiemi di Ω. $\mathcal{E}$ si dice σ-algebra *se soddisfa le proprietà*

i) se un evento E appartiene a $\mathcal{E}$, vi appartiene anche la sua negazione:

$$E \in \mathcal{E} \ \Rightarrow \ \neg E \in \mathcal{E};$$

ii) se un'infinità numerabile di eventi, $E_1, E_2, ...E_n, ...$ appartiene a $\mathcal{E}$, vi appartiene anche l'evento costituito dalla loro unione:

$$E_n \in \mathcal{E} \ \forall n \in \mathbb{N} \ \Rightarrow \ \bigcup_{n\in\mathbb{N}} E_n \in \mathcal{E}.$$

Dalla definizione di σ-algebra discende che

Proposizione 1.9

Sia $\mathcal{E} \subset 2^\Omega$ una σ-algebra, allora

- *l'evento impossibile e l'evento certo appartengono alla σ-algebra, cioè $\emptyset, \Omega \in \mathcal{E}$;*
- *se un'infinità numerabile di eventi, $E_1, E_2, ...E_n, ...$ appartiene a $\mathcal{E}$, vi appartiene anche l'evento costituito dalla loro intersezione, cioè*

$$E_n \in \mathcal{E} \ \forall n \in \mathbb{N} \ \Rightarrow \ \bigcap_{n\in\mathbb{N}} E_n \in \mathcal{E}.$$

Dato un qualunque insieme Ω, la totalità del suo insieme delle parti è ovviamente una σ-algebra. Se prendiamo $\mathbb{R}$ come spazio campione, la famiglia degli intervalli non è una σ-algebra poiché, ad esempio, l'unione di due intervalli disgiunti non è un intervallo.

Esempio 1.8: algebra di Borel reale

Si dice algebra di Borel la più piccola σ-algebra che contiene tutti gli intervalli. Essa consiste, oltre che di tutti gli intervalli (siano essi aperti o chiusi), di tutte le possibili unioni e intersezioni (finite o numerabili) di intervalli.
Ogni insieme costituito da un singolo punto fa parte dell'algebra di Borel, perché l'insieme $\{x\}$ coincide con l'intersezione della famiglia numerabile di intervalli $A_n = (x - 1/n, x + 1/n)$.

Ora possiamo precisare il significato di **misura di probabilità**.

Definizione 1.10: probabilità - definizione assiomatica

Siano Ω uno spazio campione e $\mathcal{E} \subset 2^\Omega$ una σ-algebra di eventi. Una funzione $\mathrm{P} : \mathcal{E} \mapsto \mathbb{R}$ si dice misura di probabilità se soddisfa le proprietà

(i) la probabilità di ogni evento è nonnegativa, cioè
$$\mathrm{P}(A) \geq 0 \quad \forall A \in \mathcal{E};$$
(ii) l'evento certo ha probabilità uno, cioè
$$\mathrm{P}(\Omega) = 1;$$
(iii) se $A_1, \cdots A_n, \cdots \in \mathcal{E}$ è una famiglia finita o numerabile di eventi mutualmente esclusivi (ovvero tali che $A_n \wedge A_m = \emptyset$ se $n \neq m$), la probabilità della somma logica è la somma delle probabilità, cioè
$$\mathrm{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mathrm{P}(A_n).$$

La terna $(\Omega, \mathcal{B}, \mathrm{P})$ viene detta **spazio di probabilità**.

In particolare l'assioma iii) si applica anche ad una coppia di eventi incompatibili, sicché
$$A \wedge B = \emptyset \implies \mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B).$$

Esempio 1.9: Dado classico

Se $\Omega = \{1, 2, 3, 4, 5, 6\}$ è lo spazio campione associato al lancio di un dado a sei facce, $\mathcal{E}$ la σ-algebra di tutti i suoi possibili sottoinsiemi e P la probabilità classica definita in 1.3, allora la terna $(\Omega, \mathcal{B}, \mathrm{P})$ è uno spazio di probabilità secondo la Definizione 1.10

Esempio 1.10: Dado truccato

Sia $\Omega = \{1, 2, 3, 4, 5, 6\}$ lo spazio campione associato al lancio di un dado a sei facce e $\mathcal{E}$ la σ-algebra di tutti i suoi possibili sottoinsiemi. Si assegni poi la misura di probabilità P come segue

- $\mathrm{P}(\{1\}) = 1/12$, $\mathrm{P}(\{6\}) = 1/3$,
- $\mathrm{P}(\{2\}) = \mathrm{P}(\{3\}) = \mathrm{P}(\{4\}) = \mathrm{P}(\{5\}) = 7/48$,
- *se $A \subset \{1, 2, 3, 4, 5, 6\}$ è un qualunque evento, $\mathrm{P}(A)$ si ottiene sommando le probabilità di tutti gli esiti che lo compongono.*

È facile verificare che la terna $(\Omega, \mathcal{B}, \mathrm{P})$ è uno spazio di probabilità. Essa corrisponde ad un dado "truccato", in cui l'esito 6 è più probabile.

Esempio 1.11: Integrale e probabilità

Siano $\Omega = \mathbb{R}$ *e* $\mathcal{E}$ *l'algebra di Borel (Esempio 1.8). Se* A *è un evento boreliano, definiamo*

$$\mathrm{P}(A) = \int_A \frac{1}{\pi(1+x^2)} dx.$$

Essa è una misura di probabilità in quanto

i) La monotonia dell'integrale (rispetto all'integrando) implica che $P(A) \geq 0$*, poichè*

$$\frac{1}{\pi(1+x^2)} \geq 0.$$

ii) $\mathrm{P}(\mathbb{R}) = \int_{-\infty}^{+\infty} \frac{1}{\pi(1+x^2)} dx = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b \frac{1}{\pi(1+x^2)} dx = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \frac{1}{\pi}(\arctan b - \arctan a) = \frac{1}{\pi}(\frac{\pi}{2} - (-\frac{\pi}{2})) = 1.$

iii) La verifica del terzo assioma è un po' più complicata, ma fattibile... Nel caso di una famiglia finita di intervalli, discende dall'additività dell'integrale rispetto all'insieme di integrazione.

Si noti che <u>*ogni singolo esito ha probabilità nulla*</u>*, cioè* $\mathrm{P}(\{t\}) = 0$ *per ogni* $t \in \mathbb{R}$*, infatti*

$$\mathrm{P}(\{t\}) = \int_t^t \frac{1}{\pi(1+x^2)} dx = 0.$$

Dagli assiomi si ricavano immediatamente alcune proprietà elementari.

Proposizione 1.11: proprietà elementari della probabilità

Sia dato uno spazio di misura $(\Omega, \mathcal{E}, \mathrm{P})$ *e siano* $A, B \in \mathcal{E}$ *due eventi. Allora*

(1.1) $\quad 0 \leq \mathrm{P}(A) \leq 1, \quad \mathrm{P}(\neg A) = 1 - \mathrm{P}(A), \quad \mathrm{P}(\emptyset) = 0,$

(1.2) $\quad \mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \wedge B)$

(1.3) $\quad \mathrm{P}(A \wedge B) \leq \mathrm{P}(A) \leq \mathrm{P}(A \vee B).$

Inoltre se A *implica* B*, allora*

(1.4) $\quad \mathrm{P}(A) \leq \mathrm{P}(B).$

Dimostrazione. Dimostriamo solo la proprietà (1.2), nota anche come legge della probabilità totale.

Risulta utile decomporre gli eventi A, B e $A \wedge B$ nella somma logica di altri eventi fra loro incompatibili, come segue

$$A = (A \setminus B) \vee (A \wedge B), \qquad B = (B \setminus A) \vee (A \wedge B),$$
$$A \vee B = (A \setminus B) \vee (B \setminus A) \vee (A \wedge B).$$

Applicando ora la proprietà assiomatica iii) ad ognuna di queste decomposizioni si ottiene

$$\mathrm{P}(A) = \mathrm{P}(A \setminus B) + \mathrm{P}(A \wedge B), \qquad \mathrm{P}(B) = \mathrm{P}(B \setminus A) + \mathrm{P}(A \wedge B),$$
$$\mathrm{P}(A \vee B) = \mathrm{P}(A \setminus B) + \mathrm{P}(B \setminus A) + \mathrm{P}(A \wedge B).$$

Sommando le prime due uguaglianze si ottiene

$$\mathrm{P}(A) + \mathrm{P}(B) = \mathrm{P}(A \setminus B) + \mathrm{P}(B \setminus A) + 2\mathrm{P}(A \wedge B)$$

e ricordando la terza uguaglianza si ha

$$= \mathrm{P}(A \vee B) + \mathrm{P}(A \wedge B).$$

Sottraendo la quantità $\mathrm{P}(A \wedge B)$ al primo e all'ultimo membro dell'uguaglianza si ha la tesi. $\square$

Gli assiomi (ii) e (iii), assieme alla proprietà (1.1), altro non sono che le regole di calcolo che abbiamo riconosciuto per la probabibilità classica, frequentista e soggettiva. Di fatto, una misura di probabilità è una qualunque funzione degli eventi che soddisfa le tre regole di calcolo empiriche.

Esercizio 1.12

A partire dalla definizione assiomatica di probabilità (Definizione 1.10), dimostrare le proprietà (1.1), (1.3) *e* (1.4).

Esercizio 1.13: lancio di due dadi

Nel lancio due dadi perfetti e distinguibili (immaginiamo che uno sia di colore rosso e l'altro verde), calcolare la probabilità degli eventi

A: somma pari a 8,
B: uscita di almeno un 1.

Svolgimento. Dobbiamo innanzitutto capire quali sono i possibili esiti dell'esperimento aleatorio, cioè individuare lo spazio campione. In linea di principio, lo spazio campione associato è il prodotto cartesiano dei due spazi campione associati ad ogni singolo dado, cioè

$$\Omega = \{1, \ldots, 6\} \times \{1, \ldots, 6\} = \{(i, j) \,:\, i = 1, \ldots, 6;\ j = 1, \ldots, 6\},$$

che possiamo equipaggiare con la probabilità classica. Usando le nostre conoscenze di calcolo combinatorio, possiamo dire che lo spazio campione è formato da tutti i possibili raggruppamenti di due oggetti scelti nell'insieme $\{1, 2, \ldots 6\}$.

Per calcolare la probabilità dell'evento A dovremo prima valutare quali sono gli esiti favorevoli (cioè che danno somma 8): $(2, 6)$, $(6, 2)$, $(3, 5)$, $(5, 3)$, $(4, 4)$ e poi fare il rapporto fra il numero dei casi favorevoli (5) e il numero dei casi possibili ($6 \times 6 = 36$). Ne ricaviamo che la probabilità che la somma sia 8 è $5/36$.

Possiamo calcolare la probabilità dell'evento B in modo analogo: ora gli eventi favorevoli sono 11, precisamente $(1, 1), (1, 2), \cdots (1, 6), (2, 1), (3, 1), \cdots (6, 1)$ e dunque la probabilità è $11/36$. Questo calcolo è semplice ma laborioso: immaginate cosa succederebbe se dovessimo gestire contemporaneamente 10 dadi!

In alternativa, possiamo considerare i due dadi come due copie dello stesso spazio campione $\Omega = \{1, \cdots, 6\}$. In quest'ottica, il nostro evento B si verifica se almeno uno dei due dadi verifica l'evento "uscita dell'1". In simboli abbiamo

B_1 : uscita dell'1 sul dado rosso, $\mathrm{P}(B_1) = 1/6$,
B_2 : uscita dell'1 sul dado verde, $\mathrm{P}(B_2) = \mathrm{P}(B_1) = 1/6$,

e $B = B_1 \vee B_2$. Dunque la legge della probailità totale (1.2) afferma che

$$\mathrm{P}(B) = \mathrm{P}(B_1 \vee B_2) = \mathrm{P}(B_1) + \mathrm{P}(B_2) - \mathrm{P}(B_1 \wedge B_2) = 1/3 - \mathrm{P}(B_1 \wedge B_2).$$

D'altra parte $B_1 \wedge B_2$ corrisponde al solo esito $(1, 1)$, dunque $\mathrm{P}(B_1 \wedge B_2) = 1/36$ e infine $\mathrm{P}(B_1 \vee B_2) = 1/3 - 1/36 = 11/36$. $\square$

2. Probabilità condizionata

Immaginate di giocare a dadi, scommettendo che uscirà il numero 3. Prima di effettuare il lancio, la vostra probabilità di vittoria è la probabilità dell'evento $A = \{3\}$, ovvero $P(A) = 1/6$. Poi il lancio viene effettuato: voi non vedete l'esito (che dunque per voi è ancora aleatorio), ma ricevete l'informazione "È uscito un numero dispari". Sapete dunque che si è verificato l'evento $B = \{1, 3, 5\}$.

A questo punto, la vostra confidenza nella vittoria cambia perché sono cambiate le informazioni in vostro possesso. Se prima valutavate uno spazio campione composto da 6 esiti possibili, ora sapete che in effetti si è verificato uno tra i 3 esiti in B. Dunque la probabilità di vittoria che vi attribuite non è più $1/6$, bensì $1/3$, ovvero il rapporto fra 1 (l'unico esito a voi favorevole) e 3 (gli esiti possibili, alla luce della nuova informazione).
Di fatto, avete calcolato la probabilità dell'evento A, condizionata al verificarsi dell'evento B.

Definizione 2.1: Probabilità condizionata

In uno spazio di misura $(\Omega, \mathcal{E}, \mathrm{P})$, siano $A, B \in \mathcal{E}$ due eventi. Si dice probabilità condizionata di A, dato B (o probabilità di A condizionata a B), la probabilità che l'evento A ha di verificarsi quando si sappia che B si è verificato.
Essa si indica con il simbolo $\mathrm{P}(A|B)$ e si definisce mediante la formula

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

Osservazione 2.2: Probabilità condizionata in senso classico

Quando il nostro spazio di misura è formato da un numero finito di esiti equiprobabili (cioè stiamo lavorando con la definizione classica di probabilità), abbiamo

$$P(A \wedge B) = \frac{\#\text{ esiti favorevoli sia ad } A \text{ che a } B}{\#\text{ esiti possibili}}, \qquad P(B) = \frac{\#\text{ esiti favorevoli a } B}{\#\text{ esiti possibili}},$$

quindi la Definizione 2.1 si semplifica in

$$P(A|B) := \frac{\#\text{ esiti favorevoli sia ad } A \text{ che a } B}{\#\text{ esiti favorevoli a } B}.$$

In questo caso, calcolare la probabilità condizionata a B è equivalente a calcolare la probabilità classica nello spazio campionario B.

Osserviamo esplicitamente che i due eventi che entrano in gioco nella definizione di probabilità condizionata (Definizione 2.1) hanno due ruoli diversi. Quando calcoliamo la probabilità di A condizionata a B, l'evento B è dato per certo (indipendentemente da quale fosse la sua probabilità a priori), lo chiamiamo *evento condizionante*. L'evento A, invece, si ritiene incerto e ci interessa proprio valutare la sua probabilità. Diremo che A è l'*evento condizionato*.

Attenzione dunque a non scambiare i ruoli di condizionante e condizionato.

Esercizio 2.1

Un gruppo di amici è formato da 8 uomini e 6 donne. Degli uomini, 4 sono castani, 3 biondi e uno calvo, mentre delle donne 2 sono castane e 4 bionde.

(i) Qual è la probabilità che una donna sia bionda?
(ii) Qual è la probabilità che una persona coi capelli biondi sia donna?

Svolgimento. Indichiamo

B: capelli biondi,
D: donna.

Nel gruppo di 14 amici, quelli coi capelli biondi sono in totale 7, mentre le donne sono 6. Quindi a priori ho $P(B) = 7/14 = 1/2$ e $P(D) = 6/14 = 3/7$.
Posso rispondere direttamente alla domanda *(i)* osservando che su 6 donne, 4 sono bionde e dunque la probabilità è $4/6 = 2/3$ (qui ho usato la formula semplificata dell'Osservazione 2.2). Oppure posso calcolare la probabilità di B condizionata a D secondo la Definizione 2.1. Ora $B \wedge D$ è il

gruppo delle donne bionde, che comprende 4 persone, dunque $P(B \wedge D) = 4/14 = 2/7$ e

$$P(B|D) = \frac{P(B \wedge D)}{P(D)} = \frac{\frac{2}{7}}{\frac{3}{7}} = \frac{2}{3}.$$

Per quanto riguarda la domanda *(ii)*, posso osservare che su 7 persone bionde, le donne sono 4 e dunque la probabilità è 4/7. Oppure calcolo la probabilità di D condizionata a B, che da definizione è

$$P(D|B) = \frac{P(B \wedge D)}{P(B)} = \frac{\frac{2}{7}}{\frac{1}{2}} = \frac{4}{7}.$$

□

Partendo dalla definizione, è facile verificare le seguenti proprietà.

Proposizione 2.3: Proprietà elementari della probabilità condizionata

- *Se A e B sono mutuamente esclusivi, allora $P(A|B) = 0$.*
- *Se B implica A, allora $\mathrm{P}(A|B) = 1$.*
- *Se A implica B, allora $\mathrm{P}(A|B) = \mathrm{P}(A)/\mathrm{P}(B) \geq \mathrm{P}(A)$.*

Esercizio 2.2

Verificate le proprietà elencate nella Proposizione 2.3.

Esercizio 2.3

Mario pesca una carta da un mazzo di 52 carte francesi. Ada, Bernardo e Claudia scommettono rispettivamente su

A: qualsiasi figura (cioè J, Q o K) di seme rosso (cioè cuori o quadri),
B: qualsiasi figura di seme rosso, oppure una qualunque carta compresa tra 1 e 10 di seme nero,
C: qualunque asso.

Calcolare la probabilità di vittoria dei 3 giocatori.
Se poi Mario, senza svelare quale carta ha pescato, dice che è una carta di cuori, come cambia l'aspettativa di vittoria dei 3 giocatori?

Svolgimento. Gli eventi A, B e C contengono rispettivamente $3 \times 2 = 6$, $3 \times 2 + 10 \times 2 = 26$ e $1 \times 4 = 4$ esiti. Dunque le probabilità a priori dei 3 giocatori sono

$$\begin{aligned} P(A) &= \frac{6}{52} = \frac{3}{26} \approx 0.115 \ \text{ o } \ 11\%, \quad P(B) = \frac{6+20}{52} = \frac{1}{2} = 0.5 \ \text{ o } \ 50\%, \\ P(C) &= \frac{4}{52} = \frac{1}{13} \approx 0.077 \ \text{ o } \ 8\%, \end{aligned}$$

calcolate usando la nozione classica di probabilità.
Passiamo poi a calcolare la probabilità a posteriori, cioè dopo che abbiamo saputo che si è verificato l'evento rivelato da Mario, che indicheremo con la lettera M. M è costituito da 13 carte e dunque $P(M) = \frac{13}{52} = \frac{1}{4}$.
Ora $A \wedge M$ è rappresentato dalle 3 figure di cuori, quindi $P(A \wedge M) = \frac{3}{52}$ e da definizione

$$P(A|M) = \frac{\frac{3}{52}}{\frac{13}{52}} = \frac{3}{13} \approx 0.230 \ \text{ o } \ 23\%.$$

Anche $B \wedge M$ è rappresentato dalle 3 figure di cuori, quindi $P(B \wedge M) = \dfrac{3}{52}$ e

$$P(B|M) = \frac{\frac{3}{52}}{\frac{13}{52}} = \frac{3}{13} \approx 0.230 \;\text{ o }\; 23\%.$$

Infine $C \wedge M$ è costituito dal solo asso di cuori, quindi $P(C \wedge M) = \frac{1}{52}$ e

$$P(C|M) = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13} \approx 0.077 \;\text{ o }\; 8\%.$$

□

Notiamo come, in questo esercizio, la probabilità condizionata di A è maggiore della sua probabilità a priori, mentre quella di B è minore e quella di C è invariata. Questo fatto ci dà una misura dell'idea intuitiva che l'evento M è favorevole ad A, sfavorevole a B e indifferente a C. Diamo una definizione generale.

Definizione 2.4: Eventi indipendenti e dipendenti

In uno spazio di misura $(\Omega, \mathcal{E}, \mathrm{P})$, siano $A, B \in \mathcal{E}$ due eventi
Essi si dicono indipendenti se $\mathrm{P}(A|B) = \mathrm{P}(A)$. Altrimenti si dicono dipendenti; in particolare diremo che sono positivamente correlati se $\mathrm{P}(A|B) > \mathrm{P}(A)$, e viceversa negativamente correlati se $\mathrm{P}(A|B) < \mathrm{P}(A)$.

Nell'Esercizio 2.3 abbiamo visto che M è positivamente correlato con A, negativamente correlato con B e indipendente da C.

2.1. La legge della probabilità composta. Dalla definizione di probabilità condizionata discende il seguente teorema, che risulta utilissimo nel calcolo di esperimenti ripetuti o composti.

Teorema 2.5: Legge delle probabilità composte

In uno spazio di misura $(\Omega, \mathcal{E}, \mathrm{P})$, siano $A, B \in \mathcal{E}$ due eventi. Allora

$$P(A \wedge B) = P(A|B)\, P(B) = P(B|A)\, P(A) \tag{2.1}$$

e

$$\mathrm{P}(A \wedge B) = \mathrm{P}(A)\, \mathrm{P}(B) \tag{2.2}$$

se, e solo se, A e B sono indipendenti.
Inoltre se abbiamo qualsiasi numero di eventi $A_1, A_2, A_3 \ldots A_n$ si ha

$$P(A_1 \wedge A_2 \cdots A_n) = P(A_1)\, P(A_2|A_1)\, P(A_3|A_1 \wedge A_2) \cdots P(A_n|A_1 \wedge A_2 \ldots A_{n-1}). \tag{2.3}$$

Dimostrazione. Per dimostrare la prima uguaglianza in (2.1), cioè $P(A \wedge B) = P(A|B)\, P(B)$ basta prendere l'equazione definitiva di probabilità condizionata $P(A|B) \coloneqq \dfrac{P(A \wedge B)}{P(B)}$, moltiplicare ambo i membri per la quantità $P(B)$ e semplificare.
Analogamente, per dimostrare $P(A \wedge B) = P(B|A)\, P(A)$, partiamo dalla definizione $P(B|A) \coloneqq \dfrac{P(A \wedge B)}{P(A)}$, moltiplichiamo ambo i membri per la quantità $P(A)$ e semplifichiamo.

Inoltre la formula (2.2) vale se, e solo se, $\mathrm{P}(A|B) = \mathrm{P}(B)$, cioè appunto se A e B sono indipendenti. □

Esercizio 2.4

Abbiamo due urne. L'urna $\mathcal{A}$ contiene 3 palline rosse, 3 bianche e 5 nere; l'urna $\mathcal{B}$ contiene 6 palline bianche e 10 nere. Si estrae una pallina da $\mathcal{A}$ e una da $\mathcal{B}$. Calcolare le probalilità che

a) siano entrambe nere,
b) siano una rossa e una bianca,
c) non siano entrambe bianche.

Esercizio 2.5

Consideriamo il dado truccato dell'Esempio 1.10, in cui il numero sei ha probabilità $1/3$, il numero uno ha probabilità $1/12$ e tutti gli altri numeri hanno probabilità $7/48$. Luca, non sapendo che il dado è truccato, ha puntato sul numero cinque. Davide, consapevole che il dado è truccato, viene informato che il lancio ha avuto come esito un numero dispari. Che probabilità di vittoria attribuisce a Luca?

Esercizio 2.6

La domenica Susanna va al parco solo se non piove e la sua amica Cristina è libera. Per la prossima domenica sono previste piogge con probabilità del 20%. Cristina va a trovare i nonni una domenica su due, indipendentemente dal tempo. Con che probabilità Susanna andrà al parco domenica prossima?

Osservazione 2.6: Estrazioni ripetute

Quando facciamo estrazioni ripetute dalla stessa urna (o mazzo di carte), ogni evento che riguarda la successione di estrazioni può essere interpretato come l'intersezione di eventi che riguardano le singole estrazioni, dunque le formule (2.1) e (2.2) possono essere d'aiuto. Dobbiamo però chiarire chi sia la probabilità condizionata, in questo ambito. A tal fine è necessario precisare le regole del gioco; distinguiamo due diverse situazioni
Estrazioni ripetute con reintegro: *si effettua un'estrazione, si reintegra la situazione di partenza reinserendo nell'urna la biglia estratta, poi si procede con la seconda estrazione.*
In questo caso l'esito del primo turno non ha alcuna influenza sul secondo, sicché eventi riguardanti turni diversi sono indipendenti e vale la formula (2.2).
Estrazioni ripetute senza reintegro: *si effettua un'estrazione, si esclude la biglia estratta, poi si procede con la seconda estrazione.*
In questo caso l'esito della seconda estrazione dipende dall'esito della prima e dobbiamo utilizzare la formula generale (2.1). I due turni, infatti, avvengono in due spazi campione diversi: l'esito del primo turno determina lo spazio campione in cui avviene il secondo.

Esercizio 2.7: Estrazioni ripetute con e senza reintegro

Si effettuano di due estrazioni consecutive da un'urna contenente 7 biglie rosse e 3 nere. Calcolare le probabilità degli eventi

A: estrazione di due biglie rosse (con reintegro)
B: estrazione di due biglie rosse (senza reintegro)
C: estrazione di una biglia rossa e una nera (con reintegro)
D: estrazione di una biglia rossa e una nera (senza reintegro)

Svolgimento. Calcoliamo dapprima le probabilità degli eventi elementari

$$\begin{aligned} R &: \quad \text{estrazione di una biglia rossa (su un'unica estrazione)}, \quad \mathrm{P}(R) = 7/10, \\ N &: \quad \text{estrazione di una biglia nera (su un'unica estrazione)}, \quad \mathrm{P}(N) = 3/10. \end{aligned}$$

Indichiamo poi con R_1/R_2, rispettivamente, gli eventi "estrazione di una biglia rossa al primo/secondo turno". Ora, tanto l'evento A quanto l'evento B corrispondono all'evento intersezione $R_1 \wedge R_2$ e dunque possono essere calcolati usando le formule della probabilità composta del Teorema 2.5. Tuttavia, a seconda che si estragga con o senza reintegro, cambia la correlazione tra i due eventi, e quindi il risultato finale
Nel caso con reintegro gli eventi R_1 e R_2 sono indipendenti, pertanto usiamo la formula (2.2)

$$\mathrm{P}(A) = \underbrace{\mathrm{P}(R_1 \wedge R_2)}_{\text{eventi indipendenti}} = \mathrm{P}(R_1)\,\mathrm{P}(R_2) = \mathrm{P}(R)^2 = \frac{49}{100} = 0.49 \quad \text{o} \quad 49\%.$$

Nel caso senza reintegro, invece, dobbiamo usare la formula (2.1)

$$\mathrm{P}(B) = \underbrace{\mathrm{P}(R_1 \wedge R_2)}_{\text{eventi dipendenti}} = \mathrm{P}(R_1)\,\mathrm{P}(R_2|R_1).$$

Ci serve calcolare $\mathrm{P}(R_2|R_1)$, cioè la probabilità di estrarre una biglia rossa dall'urna, avendone già estratta una rossa al primo turno. Alla seconda estrazione abbiamo un'urna che contiene solo 9 biglie, di cui 6 sono rosse e 3 nere, dunque la formula della probabilità "classica" ci dà $\mathrm{P}(R_2|R_1) = 6/9 = 2/3$ e infine

$$P(B) = \underbrace{\mathrm{P}(R_1 \wedge R_2)}_{\text{eventi dipendenti}} = \frac{7}{10}\,\frac{2}{3} = \frac{7}{15} = 0.4\bar{6} \quad \text{o} \quad 47\%.$$

Passiamo ora a valutare la probabilità di estrarre due biglie di colore diverso. Usando le notazioni appena introdotte, sia gli eventi C che D possono essere decomposti come $(R_1 \wedge N_2) \vee (N_1 \wedge R_2)$. Poiché i due eventi $R_1 \wedge N_2$ e $N_1 \wedge R_2$ sono mutuamente esclusivi, si ha senz'altro che

$$\mathrm{P}\left((R_1 \wedge N_2) \vee (N_1 \wedge R_2)\right) = \mathrm{P}\left(R_1 \wedge N_2\right) + \mathrm{P}\left(N_1 \wedge R_2\right).$$

Poi, nel valutare le due probabilità rimaste, dovremo distinguere il caso con reintegro (eventi indipendenti) da quello senza reintegro (eventi dipendenti).
Se c'è reintegro abbiamo

$$\underbrace{\mathrm{P}(R_1 \wedge N_2)}_{\text{eventi indipendenti}} = \mathrm{P}(R)\,\mathrm{P}(N) = \frac{21}{100} = 0.21, \qquad \underbrace{\mathrm{P}(N_1 \wedge R_2)}_{\text{eventi indipendenti}} = \mathrm{P}(N)\,\mathrm{P}(R) = \frac{21}{100} = 0.21,$$

quindi $\mathrm{P}(C) = 2 \cdot \underbrace{\mathrm{P}(R_1 \wedge N_2)}_{\text{eventi indipendenti}} = \dfrac{42}{100} = 0.42 \quad \text{o} \quad 42\%.$

Invece se non c'è reintegro

$$\underbrace{\mathrm{P}(R_1 \wedge N_2)}_{\text{eventi dipendenti}} = \mathrm{P}(R)\,\mathrm{P}(N|R) = \frac{7}{10}\,\frac{3}{9} = \frac{7}{30} = 0.2\bar{3}.$$

anche se è meno evidente, anche ora $\mathrm{P}\left(N_1 \wedge R_2\right) = \mathrm{P}(R_1 \wedge N_2)$, infatti

$$\underbrace{\mathrm{P}(N_1 \wedge R_2)}_{\text{eventi dipendenti}} = \mathrm{P}(N)\,\mathrm{P}(R|N) = \frac{3}{10}\,\frac{7}{9} = \frac{7}{30} = 0.2\bar{3},$$

dunque concludiamo $\mathrm{P}(D) = 2 \cdot \underbrace{\mathrm{P}(R_1 \wedge N_2)}_{\text{eventi dipendenti}} = \dfrac{7}{15} = 0.4\bar{6} \quad \text{o} \quad 47\%.$ □

Osservazione 2.7

Con questo esercizio ci siamo accorti che scambiando l'ordine degli esiti di due (o più) estrazioni otteniamo eventi che hanno uguale probabilità. Questo è vero sia nel caso con reintegro, che in quello senza reintegro!
Ripercorrendo i conti svolti, è facile convincersi che è una proprietà generale, in probabilità classica.

Esercizio 2.8

Da un'urna contenente 4 palline bianche, 3 nere e 3 rosse vengono estratte consecutivamente due palline, senza reinserire la prima. Calcolare le probalilità che

a) siano entrambe nere,
b) siano una rossa e l'altra no,
c) almeno una sia rossa,
d) nessuna sia bianca.

2.2. Esperimenti indipendenti e formula di Bernoulli. Le estrazioni ripetute con reintegro sono un esempio di una successione di esperimenti aleatori che avvengono tutti esattamente nelle stesse condizioni. In statistica queste successioni di eventi hanno grossa rilevanza e vi si riferisce come *prove ripetute* o *esperimenti indipendenti.*

Immaginiamo di fare un esperimento che ha successo quando si verifica un certo evento che indicheremo con la lettera E. Indichiamo poi con la lettera p la probabilità che l'esperimento abbia successo, cioè $p \coloneqq P(E)$, e con $q \coloneqq P(\neg E) = 1 - p$ la probabilità di insuccesso. Se ripetiamo la prova due volte (sempre nelle stesse condizioni), la legge degli eventi indipendenti (2.2) afferma che

- la probabilità di successo in entrambe le prove è p^2
- la probabilità di insuccesso in entrambe le prove è q^2
- per calcolare la probabilità di avere un successo e un insuccesso, dobbiamo fare un piccolo ragionamento. Certamente la probabilità di avere successo nel primo esperimento e non nel secondo è pq. Ma questo non esaurisce le possibilità, può anche succedere di avere successo nel secondo esperimento e non nel primo, e questo accade sempre con probabilità pq. Ne deduciamo dunque che la probabilità che E si verifichi una volta sì e l'altra no è complessivamente $2pq$ (grazie all'assioma *(iii)*).

Vale la pena di osservare che sommando le probabilità di questi tre eventi otteniamo $p^2 + q^2 + 2pq = (p+q)^2 = 1^2 = 1$, come deve essere perchè essi esauriscono tutti i casi possibili e sono a due a due incompatibili.

Cosa succede se ripetiamo l'esperimento più volte? Ci risponde il cosiddetto "schema di Bernoulli", che enunciamo nel prossimo teorema.

Teorema 2.8: Formula di Bernoulli per le prove ripetute

Dato un esperimento aleatorio ripetibile n volte nelle stesse condizioni, siano p la probabilità di successo e $q = 1-p$ la probabilità di insuccesso. Allora la probabilità di ottenere esattamente k successi su n esperimenti è

$$\binom{n}{k} p^k q^{n-k}.$$

Dimostrazione. Si possono verificare tutti e soli gli eventi del tipo

$E_{k,n}$: l'esperimento ha successo esattamente k volte

con k che varia da 0 a n.
Come nell'esempio introduttivo ($n = 2$), è immediato calcolare che

$$P(E_{n,n}) = p^n, \qquad P(E_{0,n}) = q^n.$$

Cosa possiamo dire delle situazioni intermedie?
La probabilità che E si verifichi nei primi k esperimenti e non si verifichi in tutti gli $n-k$ esperimenti successivi è facilmente calcolabile con la formula (2.2) ed è

$$\underbrace{p\,p\ldots p}_{k\text{ volte}}\,\underbrace{q\,q\ldots q}_{n-k\text{ volte}} = p^k q^{n-k}.$$

In quanti altri modi può succedere che E si verifichi k volte su n? Ce lo dice il calcolo combinatorio: dobbiamo cambiare l'ordine delle lettere p (successo) e q (insuccesso) nella "parola" composta da

k lettere p e $n-k$ lettere q, quindi usiamo la formula per le permutazioni con ripetizioni della Proposizione 3.4 e otteniamo

$$P'_{k,n-k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

Riassumendo, per ottenere la probabilità di $E_{k,n}$ dobbiamo sommare la probabilità di questi $P'_{k,n-k}$ eventi e otteniamo

$$P(E_{k,n}) = \binom{n}{k} p^k q^{n-k}.$$

□

Osservazione 2.9

Se ripetiamo n volte un esperimento, certamente si deve verificare uno tra gli eventi $E_{k,n}$ con $k = 0, 1, \dots n$. Quindi l'evento somma logica $E_{0,n} \vee E_{1,n} \vee \cdots E_{n,n}$ è certo, cioè $P(E_{0,n} \vee E_{1,n} \vee \cdots E_{n,n}) = 1$. D'altra parte gli $E_{k,n}$ si escludono a vicenda, dunque l'assioma (iii) in Definizione 1.10) assicura che

$$P(E_{0,n} \vee E_{1,n} \vee \cdots E_{n,n}) = \sum_{k=0}^{n} P(E_{k,n}).$$

Se ora sostituiamo il valore di $P(E_{k,n})$ che abbiamo appena calcolato si ottiene

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = 1. \tag{2.4}$$

Ritroviamo in senso probabilistico la formula di Newton per la potenza n-esima di un binomio, infatti

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1,$$

perché per definizione $p + q = 1$.

In effetti il Teorema 2.8 fornisce una dimostrazione della formula di Newton.

Corollario 2.10: Dimostrazione probabilistica della formula di Newton per le potenze di un binomio

Per ogni numero naturale $n \in \mathbb{N}$ e $a, b \in \mathbb{R}$, con $a, b \geq 0$, si ha

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

Dimostrazione. Introduciamo la notazione

$$p = \frac{a}{a+b}, \quad q = \frac{b}{a+b}.$$

Poichè $p, q \geq 0$ e $p + q = 1$, vale lo schema di Bernoulli e dunque la formula (2.4). Sostituendo i valori che abbiamo scelto per p e q otteniamo

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{a}{a+b}\right)^k \left(\frac{b}{a+b}\right)^{n-k} = 1,$$

cioè $\displaystyle\sum_{k=0}^{n} \binom{n}{k} \frac{a^k b^{n-k}}{(a+b)^{k+n-k}} = 1$ e raccogliendo $\displaystyle\frac{1}{(a+b)^n} \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = 1.$

Moltiplicando entrambi i termini per $(a+b)^n$ otteniamo infine

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

Esercizio 2.9

Si fanno cinque estrazioni ripetute con reintegro da un'urna contenente 5 biglie rosse, 7 nere e 8 verdi. calcolare le probabilità degli eventi:

A: estrazione di esattamente 3 biglie nere,
B: estrazione di almeno 3 biglie nere.

Svolgimento. Poiché le estrazioni si ripetono dopo aver reinserito le biglie estratte nell'urna, tutte le estrazioni avvengono nelle stesse condizioni e possiamo usare la formula di Bernoulli. Calcoliamo prima le probabilità degli eventi elementari

successo: estrazione di una biglia nera (su un'unica estrazione), $p = \frac{7}{5+7+8} = \frac{7}{20}$,
insuccesso: estrazione di una biglia non nera (su un'unica estrazione), $q = 1 - \frac{7}{20} = \frac{13}{20}$.

L'evento A corrisponde ad avere 3 successi su 5 tentativi ($E_{3,5}$ se usiamo le notazioni che abbiamo introdotto nella dimostrazione del Teorema 2.8), quindi

$$\mathrm{P}(A) = P'_{3,2} \left(\frac{7}{20}\right)^3 \left(\frac{13}{20}\right)^2 = \frac{10 \cdot 7^3 \cdot 13^2}{20^5} \approx 0.18 \quad \text{o} \quad 18\%.$$

L'evento B corrisponde alla somma logica dell'evento $A = E_{3,5}$ con gli eventi $E_{4,5}$ ed $E_{5,5}$. Poiché sono mutuamente esclusivi, si ha $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(E_{4,5}) + \mathrm{P}(E_{5,5})$ e infine

$$\mathrm{P}(B) = 10\frac{7^3 \cdot 13^2}{20^5} + 5\frac{7^4 \cdot 13}{20^5} + \frac{7^5}{20^5} = \frac{7^3}{20^5}\left(10 \cdot 13^2 + 5 \cdot 7 \cdot 13 + 7^2\right) \approx 0.23 \quad \text{o} \quad 23\%.$$

□

Attenzione! Possiamo usare la formula di Bernoulli solo quando gli eventi sono indipendenti (in questo senso parliamo di esperimento ripetibile). Nel caso di estrazioni ripetute senza reintegro, mancano le condizioni per applicare il Teorema 2.8. Tuttavia il ragionamento che abbiamo usato per la sua dimostrazione ci fa da guida nello svolgimento di esercizi di questo tipo.

Esercizio 2.10

Facendo cinque estrazioni ripetute senza reintegro dalla stessa urna dell'Esercizio 2.9, qual è la probabilità dell'evento

A: estrazione di esattamente 3 biglie nere.

Svolgimento. Come prima, il "successo" nel nostro esperimento corrisponde all'estrazione di una biglia nera. Tuttavia a differenza dell'Esercizio 2.9 la probabilità di successo cambia da un'estrazione all'altra poiché le estrazioni non avvengono tutte nelle stesse condizioni, dato che le palline estratte in precedenza vengono tolte dall'urna. Per comodità schematizziamo con la lettera S il successo (cioè l'estrazione di una biglia nera) e con I l'insuccesso (cioè l'estrazione di una biglia di un altro colore). Calcoliamo, per cominciare, la probabilità dell'evento $SSSII$ (cioè successo nei primi tre turni di estrazione e insuccesso nel quarto e quinto). Ora dobbiamo usare la formula della probabilità composta (2.3)

$$P\{SSSII\} = P(S) \cdot P(S|S) \cdot P(S|SS) \cdot P(I|SSS) \cdot P(I|SSSI) = \frac{7}{20} \cdot \frac{6}{19} \cdot \frac{5}{18} \cdot \frac{13}{17} \cdot \frac{12}{16} \approx 0.0176.$$

Anche qui l'evento A si decompone nella somma logica di tutte quelle "catene di eventi" che corrispondono ad anagrammi della parola $SSSII$, cioè in cui abbiamo sempre 3 successi su 5, ma cambia l'ordine.

Già sappiamo che, complessivamente, sono $P'_{3,2} = \frac{5!}{3!2!} = 10$.

Ora però potrebbe sembrare che, cambiando l'ordine dei turni, la probabilità cambi a sua volta. Così non è: a titolo di esempio calcoliamo

$$P\{SISSI\} = P(S) \cdot P(I|S) \cdot P(S|SI) \cdot P(S|SIS) \cdot P(I|SISS) = \frac{7}{20} \cdot \frac{13}{19} \cdot \frac{6}{18} \cdot \frac{5}{17} \cdot \frac{12}{16} = \mathrm{P}\{SSSII\}.$$

Quindi possiamo colncludere che

$$P(A) = P'_{3,2} \cdot \mathrm{P}\{SSSII\} = 10\frac{7 \cdot 6 \cdot 5 \cdot 13 \cdot 12}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16} \approx 0.17 \quad \text{o} \quad 17\%.$$

□

Esercizio 2.11: Rifacciamo l'Esercizio 1.7

Calcolare la probabilità di fare ambo e quella di fare terno, giocando tre numeri al lotto sulla ruota di Napoli.

Svolgimento: Per fare un ambo, si deve verificare una stringa di eventi di tipo SSIII. Gli eventi di questo tipo sono $\mathcal{P}_{2,3} = \frac{5!}{3!\,2!} = 10$. Poiché le estrazioni del lotto avvengono senza reintegro, dovremo usare la formula (2.3), dunque

$$\mathrm{P}\{SSIII\} = \frac{3}{90}\frac{2}{89}\frac{87}{88}\frac{86}{87}\frac{85}{86} = \frac{3 \cdot 2 \cdot 85}{90 \cdot 89 \cdot 88 \cdot} \approx 0.0007.$$

Come nell'Esercizio 2.10 cambiando l'ordine tra successi e insuccesi ottengo sempre la stessa probabilità, quindi concludo che la probabilità di fare ambo è

$$\mathcal{P}_{2,3} \cdot \mathrm{P}\{SSIII\} = 10\frac{17}{23496} = 0.0072\dots \approx 7 \cdot 10^{-3} \quad \text{o} \quad 0.7\%.$$

Per fare terno, invece, si deve verificare una stringa di eventi di tipo SSSII. Gli eventi di questo tipo sono $\mathcal{P}_{3,2} = \frac{5!}{2!\,3!} = 10$ ed ognuno di essi ha probabilità

$$\mathrm{P}\{SSSII\} = \frac{3}{90}\frac{2}{89}\frac{1}{88}1\,1 \approx 0.000008.$$

Dunque la probabilità di fare terno è

$$\mathcal{P}_{3,2} \cdot \mathrm{P}\{SSSII\} \approx 0.000085 \approx 8 \cdot 10^{-5} \quad \text{o} \quad 0.008\%.$$

Esercizio 2.12

Tre persone pescano rispettivamente una carta da un mazzo di dieci carte numerate. Ognuno, dopo aver pescato la carta, la rimette nel mazzo.

(i) Qual è la probabilità che peschino tutti la stessa carta?
(ii) Qual è la probabilità che peschino tutti la carta numero 3?
(iii) Qual è la probabilità che due persone peschino la stessa carta e la persona rimanente ne peschi una differente?

2.3. Causa ed effetto: il Teorema di Bayes. Per cominciare illustriamo un'utile applicazione della legge della probabilità composta (2.1): la cosiddetta *legge di disintegrazione o delle probabilità totali.* Premettiamo una definizione

Definizione 2.11: Partizione

Dato uno spazio di probabilità $(\Omega, \mathcal{E}, P)$, una sua partizione è una famiglia di eventi $E_1, E_2, \dots E_n \in \mathcal{E}$ con le proprietà

- *gli eventi sono mutuamente esclusivi, cioè $E_i \wedge E_k = \emptyset$ se $i \neq k$,*
- *la loro somma logica esaurisce lo spazio campione, cioè $E_1 \vee E_2 \vee \cdots E_n = \Omega$.*

Ovviamente se $E_1, E_2, \dots E_n$ formano una partizione allora

$$P(E_1) + P(E_2) \cdots + P(E_n) = P(E_1 \vee E_2 \vee \cdots E_n) = 1.$$

Proposizione 2.12: Legge di disintegrazione o delle probabilità totali

Sia $E_1, E_2, \dots E_n$ una partizione in uno spazio di probabilità $(\Omega, \mathcal{E}, P)$. Allora per ogni evento $A \in \mathcal{E}$ si ha

$$P(A) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2) \cdots + P(A|E_n)P(E_n). \tag{2.5}$$

Dimostrazione. Poiché $E_1, E_2, \dots E_n$ esauriscono lo spazio campione, certamente

$$A = (A \wedge E_1) \vee (A \wedge E_2) \cdots \vee (A \wedge E_n).$$

D'altra parte, poiché $E_1, E_2, \dots E_n$ sono mutuamente esclusivi, lo sono anche $A \wedge E_1, A \wedge E_2, \dots A \wedge E_n$. Quindi l'assioma *(iii)* ci dice che

$$P(A) = P(A \wedge E_1) + (A \wedge E_2) \cdots + P(A \wedge E_n).$$

Ora ci viene in aiuto la legge della probabilità composta (2.1), che afferma che $P(A \wedge E_1) = P(A|E_1)P(E_1)$, e analogamente per tutti gli altri. Sostituendo nella formula precedente otteniamo infine (2.5). □

Osservazione 2.13

Ripercorrendo i passaggi della dimostrazione ci si accorge he non è necessario che $E_1, E_2, \dots E_n$ siano una partizone. In effetti basta

- *$E_1, E_2, \dots E_n$ sono mutuamente esclusivi,*
- *$A \subset E_1 \vee E_2 \cdots \vee E_n$.*

Esercizio 2.13

Marco deve fare l'esame di Matematica 2 e vuole sapere che probabilità ha di ricevere una domanda sugli argomenti di probabilità. I professori che interrogano sono due: prof. Tizio chiede probabilità a due studenti su tre, mentre prof. Caio la chiede a uno su due. Marco non sa con quale professore capiterà, però sa che prof. Tizio interroga tre studenti in un'ora mentre prof. Caio ne interroga solo due.

Svolgimento. Chiamiamo X l'evento di cui vogliamo conoscere la probabilità, cioè ricevere la domanda sugli argomenti di probabilità. In questo caso la partizione è data dai due professori Tizio e Caio, che indicheremo con le loro iniziali T e C. Poiché essi interrogano rispettivamente 3 e 2 studenti su 5, a priori sappiamo che $P(T) = \frac{3}{5}$ e $P(C) = \frac{2}{5}$. D'altra parte, la probabilità di ricevere la domanda di probabilità è $\frac{2}{3}$ se si viene interrogati da Tizio e $\frac{1}{2}$ se invece si fa l'esame con Caio. In simboli $P(X|T) = \frac{2}{3}$, $P(X|C) = \frac{1}{2}$. Finalmente possiamo usare la legge di disintegrazione (2.5) e calcolare

$$P(X) = P(X|T)P(T) + P(X|C)P(C) = \frac{2}{3}\frac{3}{5} + \frac{1}{2}\frac{2}{5} = \frac{3}{5} = 0.6 \text{ oppure } 60\%.$$

□

Esercizio 2.14

Se estraggo in successione senza reintegro due biglie da un'urna che contiene 3 biglie rosse, 3 nere e 4 bianche, qual è la probabilità che la seconda biglia estratta sia bianca?

Esercizio 2.15

Abbiamo due sacchetti, il primo contiene 7 biglie nere e 3 bianche, il secondo 4 biglie nere e 6 bianche. Scegliamo casualmente uno dei due sacchetti e peschiamo una biglia. Con quale probabilità è bianca?

Esercizio 2.16

In un sacchetto ci sono 7 gettoni numerati da 1 a 7. Si estrae un primo gettone e se reca un numero dispari viene lasciato fuori, mentre se è pari viene rimesso dentro. Qual è la probabilità che nella seconda estrazione esca un numero pari?

Introduciamo il Teorema di Bayes con un esempio. Abbiamo due urne: un'urna $\mathcal{U}_A$ che contiene 3 palline rosse e 2 verdi, e un'urna $\mathcal{U}_B$ che contiene 4 palline rosse e 5 verdi. Lasciamo al caso anche la scelta dell'urna da cui estrarre la pallina: lanciamo un dato e se esce un numero minore di 3 peschiamo una pallina dall'urna $\mathcal{U}_A$, mentre se esce un numero uguale o maggiore a 3 peschiamo una pallina dall'urna $\mathcal{U}_B$.
Ora mi chiedo: se è stata pescata una pallina rossa, qual è la probabilità che provenga dall'urna $\mathcal{U}_A$? In altri termini, conosciamo l'*effetto* (la pallina è rossa) e vogliamo risalire alla *causa* (l'urna da cui è stata pescata).

Cominciamo a fissare un po' di notazioni. Chiamiamo

A: estrazione dall'urna $\mathcal{U}_A$,
B: estrazione dall'urna $\mathcal{U}_B$,
R: estrazione di una pallina rossa,

Si noti che A e B costituiscono una partizione. Poiché la scelta dell'urna dipende dal lancio del dado, *a priori* (cioè prima che l'esperimento abbia inizio) abbiamo

$$P(A) = \frac{2}{6} = \frac{1}{3} \quad \text{e} \quad P(B) = \frac{4}{6} = \frac{2}{3}.$$

I dati del problema ci forniscono anche la probabilità di R condizionata all'urna, infatti sappiamo che abbiamo 3 possibilità su 5 di avere rosso se estraiamo da A, e 4 su 9 se invece estraiamo da B, in simboli

$$P(R|A) = \frac{3}{5}, \quad \text{e} \quad P(R|B) = \frac{4}{9}.$$

Da questo possiamo dedurre la probabilità a priori di R, grazie alla legge di disintegrazione (2.5)

$$P(R) = P(R|A)P(A) + P(R|B)P(B) = \frac{3}{5}\frac{1}{3} + \frac{4}{9}\frac{2}{3} = \frac{67}{135}.$$

Noi però ora ci stiamo interrogando su tutt'altro quesito: vogliamo conoscere $P(A|R)$ e $P(B|R)$. Facciamo un passo indietro e torniamo alla definizione di probabilità condizionata

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

Ricaviamo $P(A \wedge R)$ dalla legge della probabilità composta

$$P(A \wedge R) = P(A)P(R|A)$$

e sostituiamo

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

L'equazione che abbiamo appena ricavato è davvero notevole: ci permette di scambiare il ruolo di condizionante (causa) e condizionato (effetto), infatti nel termine di sinistra l'evento R ha il ruolo di condizionante, mentre nel termine di destra compare come evento condizionato. È, nella sostanza,

il Teorema di Bayes.
Concludiamo l'esercizio facendo i calcoli

$$P(A|R) = \frac{\frac{3}{5}\frac{1}{3}}{\frac{3}{5}\frac{1}{3}+\frac{4}{9}\frac{2}{3}} = \frac{\frac{1}{5}}{\frac{67}{135}} = \frac{27}{67} \approx 0.40 \quad \text{o } 40\%.$$

Passiamo ora ad enunciare in piena generalità il Teorema di Bayes.

Teorema 2.14: Teorema di Bayes

Sono dati uno spazio di misura $(\Omega, \mathcal{E}, \mathrm{P})$ *e degli eventi* $E_1, E_2, \ldots, E_n, A \in \mathcal{E}$. *Supponiamo che*

I. gli eventi $E_1, E_2, \ldots, E_n$ *sono mutuamente esclusivi (ovvero* $E_i \wedge E_k = \emptyset$ *se* $i \neq k$*),*
II. l'evento A *è possibile (ovvero* $\mathrm{P}(A) \neq 0$*),*
III. l'evento A *implica almeno uno tra* $E_1, E_2, \ldots, E_n$*, nel senso che* $A \subset \bigcup_{i=1}^{n} E_i$.

Allora per ogni $k = 1 \cdots n$ *si ha*

$$\mathrm{P}(E_k|A) = \frac{\mathrm{P}(A|E_k)\,\mathrm{P}(E_k)}{\sum_{i=1}^{n} \mathrm{P}(A|E_i)\,\mathrm{P}(E_i)}.$$

Dimostrazione. Cominciamo con l'osservare che, per definizione,

$$\mathrm{P}(E_k|A) = \frac{\mathrm{P}(A \wedge E_k)}{\mathrm{P}(A)}. \tag{2.6}$$

Per la legge delle probabilità composte, possiamo riscrivere il numeratore come

$$\mathrm{P}(A \wedge E_k) = \mathrm{P}(A|E_k) \cdot \mathrm{P}(E_k). \tag{2.7}$$

D'altra parte, grazie alle ipotesi *I.* e *III.* e all'Osservazione 2.13 vale la legge di disintegrazione

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A|E_i) \cdot \mathrm{P}(E_i). \tag{2.8}$$

Infine, sosituendo (2.7) al posto del numeratore e (2.8) al posto del denominatore in (2.6) otteniamo la tesi. □

L'importanza del Teorema di Bayes non sta tanto nell'elaborazione formale (peraltro, la sua dimostrazione è molto semplice), quanto nelle sue applicazioni alle leggi di induzione statistica. Proviamo ad interpretarlo in questa chiave: la famiglia $E_1, \ldots E_n$ rappresenta un insieme di cause che non si possono mai verificare insieme (sono mutuamente esclusive). Sappiamo che l'effetto A si è verificato: ciò è avvenuto come conseguenza di una (e solo una) delle possibili cause $E_1, \ldots E_n$, ma non sappiamo di quale. Per scoprirlo (in senso probabilistico), dobbiamo calcolare $\mathrm{P}(E_k|A)$, detta anche *verosimiglianza*. Il Teorema di Bayes ci dice che possiamo scoprire quale tra le cause è più verosimile se conosciamo

- le probabilità a priori delle varie cause (cioè tutte le $\mathrm{P}(E_i)$),
- le probabilità a posteriori dell'effetto A, condizionate a ciascuna delle cause (cioè $\mathrm{P}(A|E_i)$)).

Esempio 2.17: Diagnosi medica

Indichiamo con A l'insieme dei sintomi manifestati da un paziente e con $E_1, \dots E_n$ una famiglia di malattie. Il medico osserva i sintomi A e deve scoprire quale malattia li ha provocati con maggiore probabilità. Gli serve allora calcolare le probabilità condizionate $\mathrm{P}(E_k|A)$, per ogni k. Attraverso il teorema di Bayes, egli può farlo a partire da

- $\mathrm{P}(E_i)$, *le probabilità a priori delle varie malattie, che ricaverà dai dati statistici;*
- $\mathrm{P}(A|E_i)$, *le probabilità a posteriori dei sintomi A, che ricaverà a sua volta dai dati statistici cercando la frazione dei malati di E_i che hanno manifestato i sintomi A.*

Esempio 2.18: Il problema di Monty Hall

Il concorrente di un gioco a premi può scegliere tra tre porte: dietro una di esse c'è un'automobile, dietro le altre, due capre. Sceglie una porta, diciamo la numero 1, e il conduttore del gioco a premi, che sa cosa si nasconde dietro ciascuna porta, ne apre un'altra, diciamo la 3, rivelando una capra. Quindi domanda: "Vorresti scegliere la numero 2?" Conviene cambiare la tua scelta originale?

Si potrebbe pensare che, con due porte chiuse, si abbia una probabilità 1/2 per ognuna, e che quindi non ci sia motivo di cambiare porta. Non è questo il caso.

Chiamiamo l'evento che l'automobile si trovi dietro una certa porta rispettivamente E_1, E_2, e E_3. All'inizio, le nostre probabilità a priori sono

$$\mathrm{P}(E_1) = \mathrm{P}(E_2) = \mathrm{P}(E_3) = 1/3.$$

Come detto prima, la porta scelta è la numero 1. Chiamiamo poi A l'evento "il presentatore apre la porta 3". La sua probabilità priori sarà 1/2 (perché il presentatore può scegliere solo tra le porte 2 e 3). Ora noi sappiamo che si è verificato l'evento A, e vogliamo usare il Teorema di Bayes per valutare a posteriori le probabilità degli eventi E_1, E_2, E_3 (cioè la probabilità che l'automobile sia dietro alla porta 1, 2, 3).

- Se l'auto è dietro la porta 1, il presentatore è libero di scegliere la porta 2 o 3 casualmente. Pertanto, $\mathrm{P}(A|E_1) = 1/2$.
- Se l'auto è dietro la porta 2, il presentatore è obbligato ad aprire la porta 3. Pertanto $\mathrm{P}(A|E_2) = 1$.
- Se l'auto è dietro la porta 3, il presentatore sarà obbligato ad aprire la porta 2. Pertanto $\mathrm{P}(A|E_3) = 0$.

Verifichiamo, tra l' altro, che $\displaystyle\sum_{i=1}^{3} \mathrm{P}(A \wedge E_i) \cdot \mathrm{P}(E_i) = \frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2} = \mathrm{P}(A)$, cioè vale la legge di disintegrazione. Il teorema di Bayes afferma che

$$\mathrm{P}(E_1|A) = \frac{\mathrm{P}(A|E_1)\mathrm{P}(E_1)}{\sum\limits_{i=1}^{3} \mathrm{P}(A \wedge E_i) \cdot \mathrm{P}(E_i)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3},$$

$$\mathrm{P}(E_2|A) = \frac{\mathrm{P}(A|E_2)\mathrm{P}(E_2)}{\sum\limits_{i=1}^{3} \mathrm{P}(A \wedge E_i) \cdot \mathrm{P}(E_i)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3},$$

$$\mathrm{P}(E_3|A) = \frac{\mathrm{P}(A|E_3)\mathrm{P}(E_3)}{\sum\limits_{i=1}^{3} \mathrm{P}(A \wedge E_i) \cdot \mathrm{P}(E_i)} = \frac{0 \cdot \frac{1}{3}}{\frac{1}{2}} = 0.$$

Da ciò è evidente che, a posteriori, la probabilità che l'automobile sia dietro la porta 2 è maggiore di quella che sia dietro alla porta 1. Scegliere la porta 2 dopo che il presentatore ha aperto la porta

3 è equivalente a scegliere inizialmente entrambe le porte 2 e 3.
Se il conduttore non avesse saputo dov'è l'automobile, cosa sarebbe cambiato?

Esercizio 2.19

In un sacchetto ci sono 7 gettoni numerati da 1 a 7. Si estrae un primo gettone e se reca un numero dispari viene lasciato fuori, mentre se è pari viene rimesso dentro. Se nella seconda estrazione esce un numero pari, qual è la probabilità che il primo estratto fosse dispari?

Esercizio 2.20

In un gruppo di 30 persone, 10 sono donne. L'80% delle donne conosce la lingua inglese, mentre gli uomini che la conoscono sono 8. Calcolare la probabilità che una persona scelta a caso che conosce l'inglese sia uomo.

Esercizio 2.21

Un automobilista arriva ad un bivio e chiede indicazione ad una persona a caso. Ci sono due persone a cui può chiedere: A dice la verità 4 volte su 10, B invece 7 volte su 10.

a) Che probabilità ha di percorrere la strada giusta?
b) Se percorre la strada giusta, con che probabilità ha chiesto l'indicazione a B?

Esercizio 2.22

Abbiamo tre urne: la prima contiene 2 palline bianche e 3 rosse, la seconda 5 bianche e 3 rosse e la terza 4 bianche e 2 rosse. Scegliamo a caso un'urna ed estraiamo una pallina. Se la pallina è bianca, che probabilità c'è che provenga dalla seconda urna?

Esercizio 2.23

Un pezzo meccanico viene prodotto da due fabbriche. La fabbrica A produce il 40% del quantitativo totale, e il 98% della sua produzione è senza difetti. Il 7% dei pezzi prodotti dalla fabbrica B è difettoso.

a) Qual è la probabilità totale che un pezzo sia difettoso?
b) Se abbiamo ricevuto un pezzo difettoso, qual è la probabilità che provenga dalla fabbrica B?

CAPITOLO 3

Variabili aleatorie

1. Cos'è una variabile aleatoria

Una *variabile aleatoria (reale)* o *casuale* è una variabile che assume un ben determinato valore in funzione dell'esito di un esperimento casuale. Sono esempi di variabili aleatorie "la somma nel lancio di due dadi", "il numero di automobili che transitano per un dato incrocio nell'arco di una giornata", "il valore del dollaro".

Diamo una definizione rigorosa di variabile aleatoria: abbiamo detto che essa assume valore come conseguenza dell'esito, dunque essa sarà una variabile dipendente, che si esprime in funzione degli esiti. Precisamente

Definizione 1.1: Variabile aleatoria, funzioni distribuzione e probabilità

Dato uno spazio di probabilità $(\Omega, \mathcal{E}, \mathrm{P})$ *una variabile aleatoria o casuale è una funzione*

$$X : \Omega \to \mathbb{R}$$

misurabile, ovvero tale che per ogni intervallo $I \subset \mathbb{R}$ *l'evento*

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

appartiene alla σ*-algebra degli eventi misurabili (in simboli* $E_I \in \mathcal{E}$*).*
La funzione che ad ogni intervallo I *associa la probabilità* $P(E_I)$ *si dice legge della variabile aleatoria.*
Ci interessano principalmente due funzioni ausiliarie di tipo standard (cioè funzioni reali di variabile reale) che descrivono le probabilità inerenti alla variabile aleatoria:

- *La funzione reale* $F_X : \mathbb{R} \to \mathbb{R}$ *definita da*

$$F_X(t) \coloneqq \mathrm{P}(E_t) = \mathrm{P}(\{X \le t\}) \qquad \text{per ogni } t \in \mathbb{R}$$

  si dice funzione di distribuzione cumulativa (o anche di ripartizione) della variabile X*.*
- *La funzione reale* $p_X : \mathbb{R} \to \mathbb{R}$ *definita dalla legge*

$$p_X(t) \coloneqq \mathrm{P}(\{X = t\}) \qquad \text{per ogni } t \in \mathbb{R}$$

  si dice funzione di probabilità (o anche di densità discreta) della variabile X*.*

Qual é la differenza tra una variabile aleatoria e le funzioni a cui siamo abituati? In "teoria" nessuna (una variabile aleatoria *è* una funzione), però

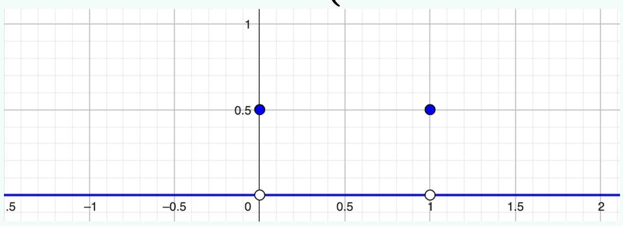
- in generale una variabile aleatoria assume valori reali, ma non è una funzione di variabile reale. Il suo dominio è uno spazio campione, come ad esempio le facce di una moneta o di un dado, o un complicato insieme di eventi macroeconomici se pensiamo al valore del dollaro.
- conoscere la legge di una variabile aleatoria non è soddisfacente. Nelle funzioni standard $f(x)$ la x ha il ruolo di *variabile indipendente* e rappresenta una grandezza che possiamo *scegliere* oppure *osservare*. Pensiamo ad esempio alla funzione che descrive la pressione. Se sono in un laboratorio e sto facendo un esperimento in cui ho tutti i parametri sotto controllo, faccio riferimento all'equazione di stato dei gas perfetti da cui ricavo la legge $P = (n\,R\,T)/V$ che descrive pressione P in funzione del numero di moli n, della temperatura T e del volume V (R è una costante che dipende dal gas). Se invece sono interessato alla pressione atmosferica che troverò tra una settimana in una località di villeggiatura, l'insieme di variabili che entrano in ballo non è controllabile né in numero né in rilevanza effettiva, quindi piuttosto che a una legge fisica (deterministica) mi affiderò

alle previsioni metereologiche che guardano la pressione come una variabile aleatoria e conoscono (dalla statistica) la probabilità con cui si distribuisce in un certo range di valori. Le funzioni che ci interessano sono quindi piuttosto la funzione di distribuzione e la funzione di probabilità, e queste sí sono funzioni reali di variabile reale in senso standard.

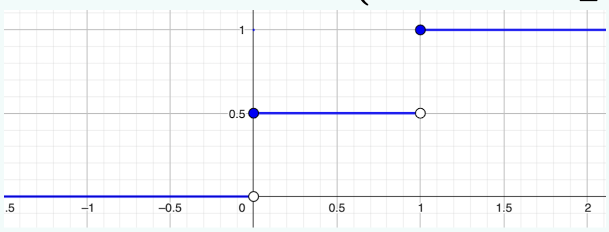
Esempio 1.1: Variabile aleatoria finita

Nello spazio di misura classico associato al lancio di una moneta, definiamo la variabile aleatoria X che vale 1 se esce testa, 0 se esce croce.

La funzione di probabilità è $p_X(t) = \begin{cases} 1/2 & \text{se } t = 0, \\ 1/2 & \text{se } t = 1, \\ 0 & \text{altrimenti} \end{cases}$

e ha grafico 

La funzione di distribuzione è $F_X(t) = \begin{cases} 0 & \text{se } t < 0, \\ 1/2 & \text{se } 0 \leq t < 1, \\ 1 & \text{se } t \geq 1 \end{cases}$

e ha grafico 

Notiamo che

- *la funzione di probabilità ha supporto finito, cioè è diversa da zero solo per un insieme finito di valori di t, precisamente i due valori (0 e 1) che può assumere la variabile aleatoria*
- *la somma dei valori non nulli di p_X dà 1*
- *la funzione di distribuzione non è continua: precisamente, ha un andamento a scalini crescenti, con discontinuità di salto in corrispondenza dei punti del supporto di p_X: l'altezza dei salti coincide con i corrispondenti valori di p_X.*
- *F_X coincide con 0 prima del primo punto del supporto di p_X e con 1 dopo l'ultimo punto.*

Esempio 1.2: Variabile aleatoria discreta

Si consideri la variabile aleatoria X che conta le immatricolazioni al Corso di Laurea in Informatica. Per sua stessa definizione, la variabile X assume valori nell'insieme $\mathbb{N}$ dei numeri naturali. Plausibili funzioni di probabilità e di distribuzione sono, rispettivamente,
In questo caso, il supporto di p_X non è finito, ma è comunque discreto perché sono ammessi solo i numeri naturali (e la serie dei valori corrispondenti deve dare 1). Di nuovo, la funzione di distribuzione ha andamento a scalini, con salti nei punti del supporto di p_X di altezza pari al corrispondente valore di p_X. Inoltre F_X vale 0 prima del primo valore ammissibile (che è 0), e ha $\lim_{t\to+\infty} F_X(t) = 1$.

Esempio 1.3: Variabile aleatoria continua

Si consideri la variabile aleatoria X che esprime il valore del dollaro (in euro). Questa variabile può assumere qualsiasi valore positivo. Una sua plausibile distribuzione è
Ora la funzione di distribuzione è continua*, è ancora* monotona crescente*, vale 0 prima del primo valore ammissibile (che è 0), e ha* $\lim\limits_{t\to+\infty} F_X(t) = 1$.
In questo caso (lo dimostreremo poi rigorosamente) la funzione di probabilità non ci dà alcuna informazione utile, perché è costantemente zero.

I tre esempi appena elencati illustrano i vari tipi di variabili aleatorie che incontreremo: i primi due sono inerenti a variabili aleatorie discrete (finita nel primo caso, infinita nel secondo), mentre il terzo esempio si riferisce ad una variabile aleatoria continua. In seguito esamineremo distintamente queste due situazioni. Per ora, mettiamo in evidenza i tratti comuni di tutte le variabili aleatorie.

Dalla definizione di funzione di distribuzione discende che:

Proposizione 1.2: Proprietà assiomatiche della distribuzione

Sia X una variabile aleatoria e $F_X(t) = \mathrm{P}(\{X \le t\})$ la sua funzione di distribuzione. Allora

i) La distribuzione assume valori compresi fra 0 e 1, cioè $0 \le F_X(t) \le 1$ *per ogni* $t \in \mathbb{R}$.
ii) La distribuzione è monotona crescente, cioè $F_X(t) \le F_X(s)$ *se* $t \le s$.
iii) La distribuzione ha limiti $\lim\limits_{t\to-\infty} F_X(t) = 0$ *e* $\lim\limits_{t\to+\infty} F_X(t) = 1$.
iv) La distribuzione è continua da destra, cioè $\lim\limits_{s\to t} F_X(s) = F_X(t)$ *per ogni* $t \in \mathbb{R}$.

Nota la funzione di distribuzione, è facile calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.3: Regole di calcolo con la distribuzione

Sia X una variabile aleatoria e $F_X(t) \coloneqq \mathrm{P}(\{X \le t\})$ la sua funzione di distribuzione. Allora

$$\mathrm{P}(\{X > t\}) = 1 - F_X(t) \qquad \text{per ogni } t \in \mathbb{R}, \tag{1.1}$$

$$\mathrm{P}(\{s < X \le t\}) = F_X(t) - F_X(s) \qquad \text{per ogni } s < t, \tag{1.2}$$

$$p_X(t) = \mathrm{P}(\{X = t\}) = F_X(t) - \lim_{s\to t^-} F_X(s) \qquad \text{per ogni } t \in \mathbb{R}. \tag{1.3}$$

Come conseguenza della proprietà (1.3) segue che

Corollario 1.4

Sia X una variabile aleatoria e $F_X(t) := \mathrm{P}(\{X \le t\})$ la sua funzione di distribuzione. Se F_X è una funzione continua, allora $p_X(t) = 0$ per ogni $t \in \mathbb{R}$.

Pertanto la funzione di probabilità si rivela utile solo nell'analisi delle variabili aleatorie discrete. Vediamo quali caratteristiche discendono dalla sua stessa definizione.

Proposizione 1.5: Proprietà assiomatiche della funzione di probabilità

Sia X una variabile aleatoria che assume valori in un insieme **discreto**, *cioè*

$$X \in \{x_1, x_2, \dots x_{\mathrm{N}}\} \qquad \textit{(caso finito)},$$

oppure

$$X \in \{x_1, x_2, \dots x_n, \dots\} \qquad \textit{(caso infinito)}.$$

Sia poi $p_X \coloneqq P(\{X = t\})$ la sua funzione di probabilità. Allora

i) La funzione di probabilità è non negativa, cioè $p_X(t) \geq 0$ per ogni $t \in \mathbb{R}$.
ii) La funzione di probabilità $p_X(t)$ è positiva solo se $t = x_n$, con $n = 1, \dots \mathrm{N}$ nel caso finito o $n \in \mathbb{N}$ nel caso infinito.
iii) La funzione di probabilità ha discontinuità eliminabili nei punti $t = x_n$, per $n = 1, 2, \dots \mathrm{N}$ nel caso finito o $n \in \mathbb{N}$ nel caso infinito.
iv) La somma dei valori assunti dalla funzione di probabilità è 1, cioè

$$\sum_{n=1}^{\mathrm{N}} p_X(x_n) = 1 \qquad \textit{(caso finito)},$$

$$\sum_{n=1}^{\infty} p_X(x_n) = 1 \qquad \textit{(caso infinito)}.$$

Nota la funzione di probabilità, è facile calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.6: Regole di calcolo con la funzione di probabilità

Sia X una variabile aleatoria discreta e $p_X(t) \coloneqq \mathrm{P}(\{X = t\})$ la sua funzione di probabilità. Allora

$$F(t) = \mathrm{P}(\{X \leq t\}) = \sum_{x_n \leq t} p_X(x_n) \qquad \textit{per ogni } t \in \mathbb{R}, \tag{1.4}$$

$$\mathrm{P}(\{X > t\}) = \sum_{x_n > t} p_X(x_n) \qquad \textit{per ogni } t \in \mathbb{R}, \tag{1.5}$$

$$\mathrm{P}(\{s < X \leq t\}) = \sum_{s < x_n \leq t} p_X(x_n) \qquad \textit{per ogni } s < t, \tag{1.6}$$

$$\mathrm{P}(\{s \leq X \leq t\}) = \sum_{s \leq x_n \leq t} p_X(x_n) \qquad \textit{per ogni } s < t. \tag{1.7}$$

Osservazione 1.7

La somme che compaiono in (1.5)–(1.7) *hanno sempre un numero finito di addendi nel caso di variabili discrete finite. Se, invece, X è una variabile discreta infinita, le somme in* (1.5) *e* (1.6) *possono indicare delle serie con infiniti termini. Esse saranno sempre convergenti in virtù della proprietà iv) della Proposizione 1.5.*

Nel caso di variabili aleatorie continue, è possibile definire un analogo della funzione di probabilità.

Definizione 1.8: Densità

Sia X una variabile aleatoria continua, e F_X la sua funzione di distribuzione. Se F_X è derivabile, definiamo funzione di densità di X la funzione reale $f_X : \mathbb{R} \to \mathbb{R}$ definita dalla legge

$$f_X(t) \coloneqq \frac{d}{dt} F_X(t).$$

Dalle proprietà assiomatiche della funzione di distribuzione segue

Proposizione 1.9: Proprietà assiomatiche della densità

Sia X una variabile aleatoria continua che ammette densità $f_X(t)$. Allora

i) La densità è non negativa, cioè $f_X(t) \geq 0$ per ogni $t \in \mathbb{R}$.

ii) La densità è integrabile in senso improprio su $\mathbb{R}$ nel senso che

$$\int_{-\infty}^{+\infty} f_X(t)\, dt \left(:= \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f_X(t)\, dt \right) = 1.$$

La funzione di densità ci permette di calcolare alcune probabilità inerenti alla variabile aleatoria:

Proposizione 1.10: Regole di calcolo con la densità

Sia X una variabile aleatoria di densità $f_X(t)$. Allora

$$F_X(t) = \mathrm{P}(\{X \leq t\}) = \int_{-\infty}^{t} f_X(r)\, dr \qquad \text{per ogni } t \in \mathbb{R}, \tag{1.8}$$

$$\mathrm{P}(\{X > t\}) = \int_{t}^{+\infty} f_X(r)\, dr \qquad \text{per ogni } t \in \mathbb{R}, \tag{1.9}$$

$$\mathrm{P}(\{s < X \leq t\}) = \mathrm{P}(\{s \leq X \leq t\}) = \int_{s}^{t} f_X(r)\, dr \qquad \text{per ogni } s < t. \tag{1.10}$$

2. Esempi notevoli di variabili aleatorie discrete

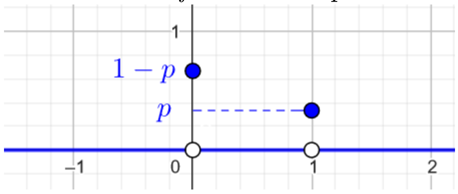
Passiamo ora in rivista alcune delle variabili aleatorie che più sono utilizzate nelle applicazioni. La più semplice di esse, la variabile Bernoulliana, altro non è che una generalizzazione della variabile introdotta nell'Esempio 1.1.

2.1. Variabile bernoulliana Bern(p)**.** Un esperimento aleatorio si dice *bernoulliano* se ha solo due possibili esiti: successo o insuccesso. Ad esempio, il lancio di una moneta costituisce un esperimento bernoulliano (in cui il successo può essere, per chiarire, l'uscita di testa). Anche il lancio di un dado a sei facce può essere visto come un esperimento bernoulliano se, ad esempio, interpretiamo come successo l'uscita del 6, e come insuccesso qualunque altro esito. Chiamiamo allora *parametro di un esperimento Bernoulliano* il numero p (compreso fra 0 e 1) che esprime la probabilità di successo. Il lancio di una moneta è dunque un esperimento bernoulliano di parametro 1/2, mentre il lancio di un dado a sei facce dell'esempio precedente costituisce un esperimento bernoulliano di parametro 1/6.

Dato un esperimento bernoulliano di parametro p, la variabile bernoulliana associata (che indicheremo nel seguito con Bern(p)) è la variabile che vale 1 se l'esito dell'esperimento è successo, 0 se è insuccesso.

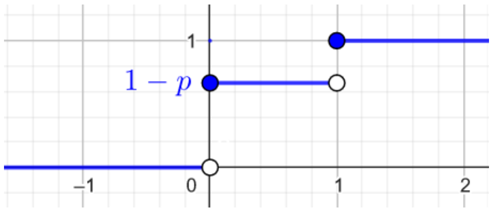
Si tratta chiaramente di una variabile discreta che può prendere un numero finito di valori (0 o 1) e che possiamo descrivere completamente mediante la sua *funzione di probabilità*:

$$p(t) = \begin{cases} p & \text{se } t = 1, \\ 1-p & \text{se } t = 0, \\ 0 & \text{altrimenti} \end{cases}$$



La *distribuzione di probabilità* è

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ 1-p & \text{se } 0 \leq t < 1, \\ 1 & \text{se } t \geq 1 \end{cases}$$



Esercizio 2.1

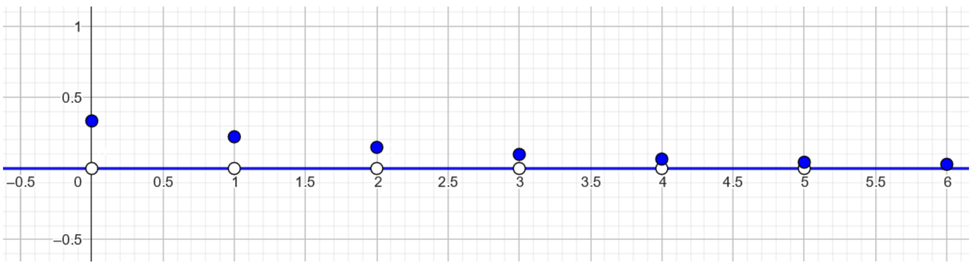
Ripetiamo 3 volte un esperimento bernoulliano di parametro $p = 2/5$, in modo che ogni nuovo esperimento sia indipendente dai risultati precedenti. Qual è la probabilità che si verifichino

a) esattamente due insuccessi,
b) due insuccessi nei primi due tentativi e poi un successo,
c) esattamente 2 successi,
d) almeno 2 successi,
e) non più di 2 successi,

2.2. Variabile geometrica Geom(p)**.** Ripetiamo (in modo indipendente) un esperimento bernoulliano di parametro p fino a che non otteniamo un successo. La variabile geometrica conta il numero di insuccessi prima del primo successo; essa può dunque assumere il valore 0 (successo al primo tentativo), 1 (insuccesso al primo tentativo e successo al secondo), 2, e così via: si tratta cioè di una variabile discreta che può assumere infiniti valori.

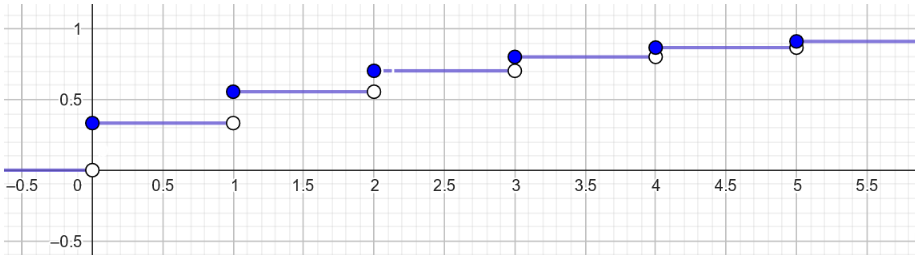
Generalizzando quanto visto nell'Esercizio 2.1, si calcola la sua *funzione di probabilità*

$$p(t) = \begin{cases} (1-p)^n p & \text{se } t = n \text{ è un intero} \\ 0 & \text{altrimenti,} \end{cases}$$



e di conseguenza la *distribuzione di probabilità*

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ p \sum_{k=0}^{n} (1-p)^k = 1 - (1-p)^{n+1} & \text{se } n \leq t < n+1 \end{cases}$$

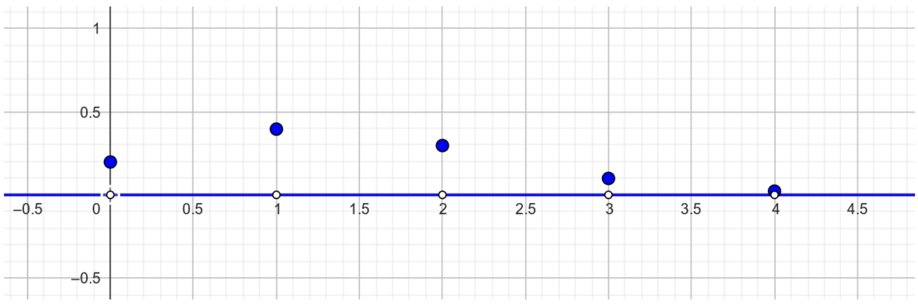


Esercizio 2.2

Verificare che la funzione $p(t)$ appena introdotta è effettivamente una funzione di probabilità, cioè che soddisfa le proprietà assiomatiche elencate nella Proposizione 1.5.
Preso poi un esperimento bernoulliano di parametro $p = 1/6$, calcolare la probabilità di avere successo entro il terzo tentativo (ovvero $\mathrm{P}(\{\mathrm{Geom}(1/6) \leq 3\})$) e la probabilità di avere successo fra il terzo e il quinto tentativo (ovvero $\mathrm{P}(\{3 \leq \mathrm{Geom}(1/6) \leq 5\})$).

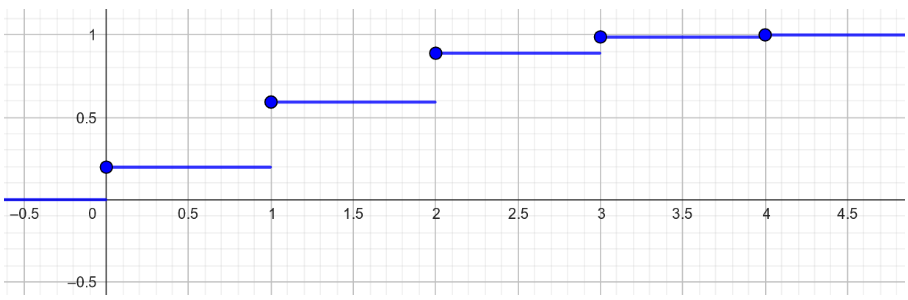
2.3. Variabile binomiale $\mathrm{Bin}(n, p)$**.** Ripetiamo n volte un esperimento bernoulliano in modo indipendente e contiamo quanti successi abbiamo ottenuto. La variabile binomiale $\mathrm{Bin}(n, p)$ fa proprio questo, a titolo di esempio vale 2 se abbiamo avuto esattamente 2 successi e $n-2$ insuccessi. Si tratta dunque di una variabile discreta che può assumere come valore ogni numero intero compreso fra 0 e n. Abbiamo già calcolato la sua funzione di probabilità: essa infatti è data dalla formula di Bernoulli (Teorema 2.8)

$$p(t) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{se } x = k \text{ con } k = 0, 1, \dots n, \\ 0 & \text{altrimenti} \end{cases}$$



Da questa ricaviamo la *distribuzione di probabilità*

$$F(t) = \begin{cases} 0 & \text{se } t < 0, \\ \displaystyle\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} & \text{se } k \leq t < k+1, \ \text{con } k = 0, 1, \dots n-1 \\ 1 & \text{se } t \geq n. \end{cases}$$



Esercizio 2.3

Verificare che la funzione $p(t)$ appena introdotta è effettivamente una funzione di probabilità.

Esercizio 2.4: Prove ripetute

Ripetiamo un esperimento bernoulliano di parametro p in modo che ogni nuovo esperimento sia indipendente dai precedenti. Calcoliamo la probabilità degli eventi

A: almeno un successo su n prove,
B: esattamente k successi su n prove,
C: primo successo all'n-ma prova.

Svolgimento. È possibile svolgere questo esercizio anche senza conoscere le variabili aleatorie, usando solo le proprietà degli eventi indipendenti e il calcolo combinatorio. Tuttavia le variabili aleatorie ci aiutano perché rispondono in modo "automatico" a quesiti di questo tipo senza dover ricavare ogni volta una nuova formula, ma scegliendo la variabile aleatoria notevole che fa al caso nostro.

A: Possiamo calcolare la probabilità dell'evento A in almeno due modi diversi.
1^o *modo:* Utilizziamo la variabile binomiale $\mathrm{Bin}(n,p)$, che conta quanti successi si ottengono in n prove indipendenti. L'evento A si verifica quando $\mathrm{Bin}(n,p)$ è strettamente maggiore di 0, dunque

$$P(A) = P\{\mathrm{Bin}(n,p) > 0\} \underset{(1.1)}{=} 1 - F_{\mathrm{Bin}(n,p)}(0) = 1 - \binom{n}{0} p^0 (1-p)^{n-0} = 1 - (1-p)^n.$$

2^o *modo:* Può convenire calcolare la probabilità della negazione di A, cioè di $\neg A$: "nessun successo in n prove indipendenti" = "n insuccessi in n prove indipendenti". Poiché la probabilità di insuccesso è $1-p$, il numero di insuccessi è ancora regolato da una variabile binomiale, che avrà però parametro $1-p$ anziché p. Ora $\neg A$ si verifica quando $\mathrm{Bin}(n, 1-p)$ è uguale a n, quindi dalla Definizione 1.1 sappiamo che

$$P(\neg A) = P\{\mathrm{Bin}(n, 1-p) = n\} = p_{\mathrm{Bin}(n,1-p)}(n) = \binom{n}{n}(1-p)^n p^{n-n} = (1-p)^n.$$

Ricaviamo infine $P(A) = 1 - P(\neg A) = 1 - (1-p)^n$.

B: L'evento B si verifica quando $\mathrm{Bin}(n,p)$ è uguale a k, dunque dalla Definizione 1.1 sappiamo che

$$P(B) = P\{\mathrm{Bin}(n,p) = k\} = p_{\mathrm{Bin}(n,p)}(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

C: In questo caso ci serve la variabile geometrica, che conta gli insuccessi consecutivi prima del primo successo. L'evento C si verifica se $\mathrm{Geom}(p) = n-1$, quindi

$$P(C) = P\{\mathrm{Geom}(p) = n-1\} = p_{\mathrm{Geom}(p)}(n-1) = (1-p)^{n-1}p.$$

□

Esercizio 2.5

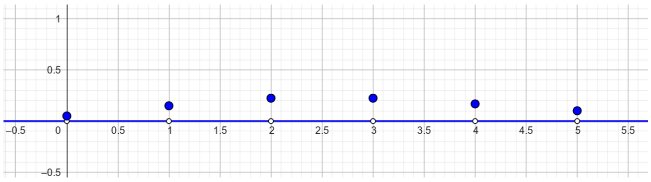
Sapendo che il 15% della popolazione è mancina, calcolare la probabilità che

A: Bisogna fermare almeno 20 persone per trovare un mancino
B: Bisogna fermare esattamente 20 persone per trovare un mancino
C: Bisogna fermare esattamente 20 persone per trovare 3 mancini

2.4. Variabile di Poisson $\mathcal{P}(\lambda)$. La variabile di Poisson è una variabile discreta infinita che assume valori nell'insieme $\mathbb{N}$ dei numeri naturali; essa viene utilizzata per "contare" gli eventi rari. Se $\lambda > 0$ rappresenta il numero di eventi che si verificano mediamente in un intervallo unitario di tempo, la variabile di Poisson si definisce a partire dalla sua funzione di probabilità

$$p(t) = \begin{cases} e^{-\lambda}\dfrac{\lambda^n}{n!} & \text{se } t = n \in \mathbb{N}, \\ 0 & \text{altrimenti.} \end{cases}$$

il cui grafico è

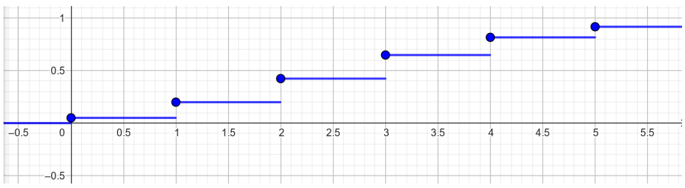


Si noti che p è effettivamente una funzione di probabilità, precisamente $e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=1$, poichè la serie di potenze $\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}$ è la serie di Mc Laurin della funzione $f(\lambda)=e^{\lambda}$.

Si ricava poi la funzione di distribuzione

$$F(t)=\begin{cases} 0 & \text{se } t<0,\\ e^{-\lambda}\sum\limits_{k=0}^{n}\frac{\lambda^k}{k!} & \text{se } n\le t<n+1 \text{ con } n\in\mathbb{N}\end{cases}$$

il cui grafico è



3. Esempi notevoli di variabili aleatorie continue

Si vedano, per ora, le dispense "Variabili aleatorie elementari"

4. Distribuzioni congiunte e variabili indipendenti

Talvolta è utile considerare congiuntamente due variabili aleatorie. Per farlo, ci servono alcune funzioni ausiliarie.

Definizione 4.1: distribuzione, probabilità e densità congiunte

Se X e Y sono due variabili aleatorie, chiamiamo funzione di **distribuzione congiunta** *di X e Y la funzione $F_{XY}:\mathbb{R}^2\to\mathbb{R}$ definita mediante*

$$F_{XY}(t,s)=\mathrm{P}(\{X\le t\}\wedge\{Y\le s\})\qquad \textit{per ogni } (t,s)\in\mathbb{R}^2.$$

Se X e Y sono variabili discrete, risulta utile introdurre la funzione di **probabilità congiunta**, *$p_{XY}:\mathbb{R}^2\to\mathbb{R}$ definita dalla legge*

$$p_{XY}(t,s)=\mathrm{P}(\{X=t\}\wedge\{Y=s\})\qquad \textit{per ogni } (t,s)\in\mathbb{R}^2.$$

Se X e Y sono variabili continue, definiamo invece la funzione di **densità congiunta**, *$f_{XY}:\mathbb{R}^2\to\mathbb{R}$ che soddisfa la proprietà*

$$\mathrm{P}(\{a\le X\le b\}\wedge\{c\le Y\le d\})=\int_a^b\int_c^d f_{XY}(t,s)dtds\qquad \textit{per tutti i numeri reali } a<b \textit{ e } c<d.$$

Se X e Y sono variabili discrete, diciamo per chiarire $X\in\{x_1,x_2,\dots x_{\mathrm{N}}\}$ e $Y\in\{y_1,y_2,\dots y_{\mathrm{M}}\}$, la probabilità congiunta p_{XY} è sempre nonnegativa, e positiva solo nei punti di (x_i,y_j) con $i=1,\dots\mathrm{N}$ e $j=1,\dots\mathrm{M}$, inoltre

$$\sum_{i=1}^{\mathrm{N}}\sum_{j=1}^{\mathrm{M}}p_{XY}(x_i,y_j)=1.$$

La funzione di probabilità della sola variabile X può essere ricostruita mediante la formula

probabilità **marginale**: $$p_X(x_i) = \sum_{j=1}^{\mathrm{M}} p_{XY}(x_i, y_j),$$

poiché $p_X(x_i) = \mathrm{P}(\{X = x_i\}) = \mathrm{P}(\{X = x_i\} \wedge \{-\infty < Y < +\infty\}) = \sum_{j=1}^{\mathrm{M}} p_{XY}(x_i, y_j)$. Formula analoga vale per la variabile Y.

Se X e Y sono variabili continue, la densità congiunta f_{XY} è sempre nonnegativa e

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(t,s)dtds = 1.$$

La funzione di densità della sola variabile X può essere ricostruita mediante la formula

densità **marginale**: $$f_X(t) = \int_{-\infty}^{+\infty} f_{XY}(t,s)ds.$$

Formula analoga vale per la variabile Y.

Esempio 4.1: lancio di un dado e di una moneta

Immaginiamo di lanciare un dado e una moneta. La variabile X legge la faccia del dado, la variabile Y prende il valore -1 se sulla moneta esce testa e +1 se esce croce. Possiamo considerare congiuntamente le variabili nello spazio campione $\Omega = \{(j,k) : j = 1,\dots 6$ e $k = \pm 1\}$, composto da 12 esiti equiprobabili. Dunque $p_{XY}(j,k) = 1/12$ per qualunque $j = 1,\dots 6$ e $k = \pm 1$. Possiamo ricorstruire la probabilità marginale di X e Y calcolando

$$\begin{aligned} p_X(j) =& p_{XY}(j,+1) + p_{XY}(j,-1) = \frac{1}{12} + \frac{1}{12} = \frac{1}{6} \\ p_Y(\pm 1) =& \sum_{j=1}^{6} p_{XY}(j,\pm 1) = \frac{6}{12} = \frac{1}{2}. \end{aligned}$$

Abbiamo in precedenza parlato di *eventi* indipendenti. Vogliamo ora estendere alle variabili aleatorie la nozione di indipendenza. Intuitivamente, due variabili si diranno indipendenti se gli eventi che riguardano solo una variabile sono indipendenti dagli eventi che riguardano solo l'altra. Non è necessario prendere in esame tutti gli eventi, bensì basta considerare solo gli eventi del tipo $\{X \le t\}$. Si ha dunque

Definizione 4.2: Variabili indipendenti

Due variabili aleatorie X e Y si dicono **indipendenti** *se, e solo se, per ogni $t, s \in \mathbb{R}$ gli eventi $\{X \le t\}$ e $\{Y \le s\}$ sono eventi indipendenti. Ricordando* (2.2), *ciò è equivalente ad affermare*

$$\mathrm{P}(\{X \le t\} \wedge \{Y \le s\}) = \mathrm{P}(\{X \le t\}) \cdot \mathrm{P}(\{Y \le s\}).$$

È possibile caratterizzare l'indipendenza mediante le funzioni ausiliarie.

Proposizione 4.3: Condizioni necessarie e sufficienti d'indipendenza

Siano X e Y due variabili aleatorie e indichiamo con F_X e F_Y le ripettive distribuzioni. Le variabili X e Y sono indipendenti se, e solo se,

$$F_{XY}(t,s) = F_X(t) \cdot F_Y(s) \qquad \text{per ogni } (t,s) \in \mathbb{R}^2.$$

Se poi X e Y sono due variabili discrete con funzione di probabilitià rispettive p_X e p_Y, esse sono indipendenti se, e solo se,

$$p_{XY}(t,s) = p_X(t) \cdot p_Y(s) \qquad \text{per ogni } (t,s) \in \mathbb{R}^2.$$

Se invece X e Y sono due variabili continue con densità rispettive f_X e f_Y, esse sono indipendenti se, e solo se,

$$f_{XY}(t,s) = f_X(t) \cdot f_Y(s) \qquad \text{per ogni } (t,s) \in \mathbb{R}^2.$$

Esempio 4.2: lancio di un dado e di una moneta - continua

È evidente che le variabili aleatorie dell'esempio Esempio 4.1 sono indipendenti. Dalla definizione vediamo che, comunque scelti $j = 1, \dots 6$ e $k \pm 1$, si ha

$$p_{XY}(j,k) = \frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2} = p_X(j)p_Y(k).$$

Esempio 4.3: incidenza di una malattia a seconda del genere

Su un campione di 20000 persone, 200 sono state riscontrate affette da una certa malattia. Dunque l'incidenza della malattia è $200/20000 = 1\%$. Per studiare l'effetto del genere sulla malattia, il campione è stato composto con 10000 uomini e 10000 donne. Dei 200 malati, 136 sono uomini e 64 donne. Notiamo dunque che c'è una correlazione tra la probabilità di sviluppare la malattia e il genere. Possiamo inquadrare la questione in termini di variabili aleatorie. Definiamo

- *X vale 1 se si è malati e 0 se si è sani. Interpretando l'incidenza della malattia nel nostro campione come una probabilità, abbiamo $p_X(1) = 1\%$, $p_X(0) = 99\%$.*
- *Y vale 1 se si è uomini e 0 se si è donne. Interpretando di nuovo la frequenza nel nostro campione come probabilità abbiamo $p_Y(1) = p_Y(0) = 50\%$.*

Guardando congiuntamente le variabili, dai nostri dati ricaviamo che

$$p_{XY}(1,1) = \frac{\#\ \text{uomini malati}}{\#\ \text{totale}} = \frac{136}{20000} = 0.68\%, \quad p_{XY}(1,0) = \frac{\#\ \text{donne malate}}{\#\ \text{totale}} = \frac{64}{20000} = 0.32\%,$$
$$p_{XY}(0,1) = \frac{\#\ \text{uomini sani}}{\#\ \text{totale}} = \frac{9864}{20000} = 49.32\%, \quad p_{XY}(0,0) = \frac{\#\ \text{donne sane}}{\#\ \text{totale}} = \frac{9936}{20000} = 49.68\%.$$

Queste variabili **non** *sono indipendenti, infatti $p_X(1) \cdot p_Y(0) = 1\% \cdot 50\% = \frac{50}{10000} = 0.5\%$ è diverso da $p_{XY}(1,0) = 0.32\%$.*

5. Valore atteso, varianza e scarto quadratico medio

Spendiamo due parole sul concetto di **media** in *statistica*. Se vogliamo conoscere il numero medio di figli in una famiglia italiana, il buon senso ci suggerisce di fare il rapporto tra il numero di figli (n) e il numero di famiglie (N). In *statistica*, si prende un campione di N famiglie, e le si suddivide in base al numero di figli. Se chiamiamo N_0, N_1, N_2,... il numero di famiglie con 0, 1, 2,... figli, allora il numero totale di famiglie è $\mathrm{N} = \mathrm{N}_0 + \mathrm{N}_1 + \mathrm{N}_2 + \cdots$, mentre il numero totale di figli è $n = 0 \cdot \mathrm{N}_0 + 1 \cdot \mathrm{N}_1 + 2 \cdot \mathrm{N}_2 + \cdots$. Il numero medio di figli è dato allora da

$$\bar{n} = \frac{n}{N} = \frac{0 \cdot n_0 + 1 \cdot n_1 + 2 \cdot n_2 + \cdots}{\mathrm{N}} = 0 \cdot \frac{n_0}{\mathrm{N}} + 1 \cdot \frac{n_1}{\mathrm{N}} + 2 \cdot \frac{n_2}{\mathrm{N}} + \cdots.$$

Nell'ultima espressione, le quantità $\frac{\mathrm{N}_0}{\mathrm{N}}$, $\frac{\mathrm{N}_1}{\mathrm{N}}$,... rappresentano la *frequenza* di famiglie con 0,1,... figli. Se poi immaginiamo di avere un campione infinito (cioè estrapoliamo il limite per $\mathrm{N} \to \infty$), tali frequenze approssimano la probabilità dei rispettivi eventi, secondo la concezione frequentista.

Possiamo allora riformulare il discorso in termini *probabilistici*. Consideriamo la variabile aleatoria X che esprime il numero di figli in una famiglia italiana scelta a caso; si tratta di una variabile discreta che prende valori in $\mathbb{N}$. La sua funzione di probabilità p_X è plausibilmente espressa dal limite delle frequenze

$$p_X(i) = \lim_{\mathrm{N}\to+\infty} \frac{n_i}{\mathrm{N}} \qquad \text{per } i = 0, 1, \dots$$

Il corrispondente limite della media statistica viene interpretato come **valor medio** della variabile X; scriviamo in formule

$$E(X) = \lim_{\mathrm{N}\to+\infty} \bar{n} = \lim_{\mathrm{N}\to+\infty} \left(0 \cdot \frac{n_0}{\mathrm{N}} + 1 \cdot \frac{n_1}{\mathrm{N}} + 2 \cdot \frac{n_2}{\mathrm{N}} + \cdots\right) = 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + \cdots .$$

Introduciamo ora una definizione rigorosa, alla luce di questo esempio.

Definizione 5.1: Valore atteso o medio

Se X è una variabile aleatoria <u>discreta finita</u>, $X \in \{x_1, x_2, \dots x_{\mathrm{N}}\}$, e p_X la sua funzione di probabilità, si definisce valor medio o valore atteso o speranza matematica di X il <u>numero reale</u>

$$E(X) \coloneqq x_1 p_X(x_1) + x_2 p_X(x_2) \cdots + x_{\mathrm{N}} p_X(x_{\mathrm{N}}).$$

Se X una variabile aleatoria <u>discreta infinita</u>, $X \in \{x_1, x_2, \dots x_n, \dots\}$, e p_X la sua funzione di probabilità, la somma precedente diventa una serie. Precisamente, quando la serie di termine generico $x_n p_X(x_n)$ è assolutamente convergente, il valore di tale serie è detto valor medio o valore atteso o speranza matematica di X, in formule

$$E(X) \coloneqq \sum_{n=1}^{\infty} x_n p_X(x_n).$$

Infine, se X è una variabile aleatoria <u>continua</u> e f_X la sua funzione di densità, la somma diventa un integrale. Precisamente, quando la funzione $t\, f_X(t)$ è assolutamente integrabile in senso improprio su $(-\infty, +\infty)$, il valore di tale integrale è detto valor medio o valore atteso o speranza matematica di X, in formule

$$E(X) \coloneqq \int_{-\infty}^{+\infty} t\, f_X(t)\, dt.$$

Esempio 5.1

Sia X la variabile aleatoria che esprime l'esito del lancio di un dado perfetto a 6 facce, che prende i valori $\{1, 2, \dots 6\}$ e ha funzione di probabilità

$$p_X(t) = \begin{cases} 1/6 & \text{se } t = 1, 2, \dots 6 \\ 0 & \text{altrimenti.} \end{cases}$$

Il suo valor medio è

$$E(X) = 1\, p_X(1) + 2\, p_X(2) + \cdots 6\, p_X(6) = \frac{1 + 2 + \cdots 6}{6} = \frac{21}{6} = \frac{7}{2}.$$

Si noti che il valor medio non rientra tra i valori effettivamente assunti dalla variabile aleatoria.

Esempio 5.2

Calcoliamo il valore atteso della variabile bernoulliana introdotta nel paragrafo 2.1. Si ha

$$E(\mathrm{Bern}(p)) = (1-p)\,0 + p\,1 = p.$$

Esempio 5.3

Calcoliamo il valore atteso della variabile geometrica $\mathrm{Geom}(p)$ *definita nel paragrafo 2.2. Per definizione*

$$E(\mathrm{Geom}(p)) = \sum_{n=0}^{\infty} n\,(1-p)^n p = p \sum_{n=0}^{\infty} n\,(1-p)^n = p \sum_{n=1}^{\infty} n\,(1-p)^n.$$

Ma per ogni n *fissato si ha* $n = (n-1)+1$ *e dunque*

$$\sum_{n=1}^{\infty} n\,(1-p)^n = \sum_{n=1}^{\infty} ((n-1)+1)\,(1-p)^n = \sum_{n=1}^{\infty} (n-1)\,(1-p)^n + \sum_{n=1}^{\infty} (1-p)^n.$$

Calcoliamo separatamente queste due serie. Ponendo $m = n-1$ *si ottiene*

$$\sum_{n=1}^{\infty} (n-1)\,(1-p)^n = \sum_{m=0}^{\infty} m\,(1-p)^{m+1} = (1-p) \sum_{m=0}^{\infty} m\,(1-p)^m = \frac{1-p}{p} E(\mathrm{Geom}(p)).$$

D'altra parte

$$\sum_{n=1}^{\infty} (1-p)^n = \sum_{n=0}^{\infty} (1-p)^n - 1 = \frac{1}{1-(1-p)} - 1 = \frac{1}{p} - 1 = \frac{1-p}{p},$$

ricordando la serie geometrica di ragione $1-p$. *Ricapitolando*

$$E(\mathrm{Geom}(p)) = p \sum_{n=1}^{\infty} n\,(1-p)^n = p\left(\frac{1-p}{p} E(\mathrm{Geom}(p)) + \frac{1-p}{p}\right) = (1-p)\,E(\mathrm{Geom}(p)) + 1 - p.$$

Esaminiamo l'uguaglianza fra primo e ultimo membro e ricaviamo

$$(1-1+p)\,E(\mathrm{Geom}(p)) = 1-p,$$

da cui $E(\mathrm{Geom}(p)) = (1-p)/p$.

Il seguente risultato afferma che il valore atteso è una grandezza lineare.

Proposizione 5.2: Linearità del valore atteso

Siano X *e* Y *due variabili aleatorie e* $a, b, c \in \mathbb{R}$. *Allora*

$$E(a\,X + b\,Y + c) = a\,E(X) + b\,E(Y) + c.$$

Osserviamo esplicitamente che, come è facile verificare dalla definizione, il valore atteso di una costante è pari alla costante.

Esempio 5.4

Calcoliamo il valor medio della variabile binomiale introdotta nel paragrafo 2.3. Tale variabile altro non è che la somma di n *variabili bernoulliane di parametro* p *(una per ogni ripetizione della prova), dunque*

$$E(\mathrm{Bin}(n,p)) = n\,E(\mathrm{Bern}(p)) = n\,p.$$

Esempio 5.5

I valori attesi delle principali variabili aleatorie continue sono calcolati nelle dispense "Variabili aleatorie elementari".

Il valore atteso è lineare rispetto alla *somma* di variabili aleatorie, ma non rispetto al *prodotto*: non è vero che, in generale, $E(XY) = E(X)\,E(Y)$. L'uguaglianza vale solo per le variabili indipendenti, come afferma il teorema seguente.

Teorema 5.3: Valore atteso del prodotto di variabili indipendenti

Se X e Y sono due variabili indipendenti, allora

$$E(XY) = E(X)\,E(Y).$$

Dimostrazione. Faremo la dimostrazione nel caso in cui X e Y sono variabili discrete, ovvero $X \in \{x_1, \dots x_n\}$ e $Y \in \{y_1, \dots y_m\}$, e indicheremo con p_X e p_Y le rispettive funzioni di probabilità. La variabile XY è ancora una variabile discreta che può assumere i valori $\{x_i y_j\ :\ i = 1, \dots n \text{ e } j = 1, \dots m\}$ e per definizione

$$E(XY) = \sum_{t\in\mathbb{R}} t\,q(t), \tag{5.1}$$

dove q designa la sua funzione di probabilità $q(t) = \mathrm{P}\{XY = t\}$[1]. Se $t = 0$, $q(0)$ non dà alcun contributo al calcolo del valore atteso secondo la formula (5.1). Se, invece, $t \neq 0$, possiamo esprimere $q(t)$ in termini delle due funzioni p_X e p_Y. Cominciamo con l'osservare che l'evento $\{XY = t\}$ si può decomporre nella somma logica di eventi mutuamente esclusivi

$$\{XY = t\} = \bigcup_{s\in\mathbb{R}\setminus\{0\}} \{X = s \text{ e } Y = t/s\},$$

pertanto

$$q(t) = \mathrm{P}\{XY = t\} = \sum_{s\in\mathbb{R}\setminus\{0\}} \mathrm{P}\{X = s \text{ e } Y = t/s\}. \tag{5.2}$$

D'altronde $\{X = s \text{ e } Y = t/s\} = \{X = s\} \wedge \{Y = t/s\}$ e, poiché le variabili X e Y sono indipendenti, si ha

$$\mathrm{P}\left(\{X = s\} \wedge \{Y = t/s\}\right) = \mathrm{P}\{X = s\}\,\mathrm{P}\{Y = t/s\} = p_X(s)\,p_Y(t/s).$$

Sostituendo questa uguaglianza nella formula (5.2) si ottiene

$$q(t) = \sum_{s\in\mathbb{R}\setminus\{0\}} p_X(s)\,p_Y\left(\frac{t}{s}\right),$$

e quindi, tornando alla formula (5.1),

$$E(XY) = \sum_{t\in\mathbb{R}\setminus\{0\}} \sum_{s\in\mathbb{R}\setminus\{0\}} t\,p_X(s)\,p_Y\left(\frac{t}{s}\right).$$

[1] si osservi che la somma scritta è finita, poiché $q(t) \neq 0$ solo per un numero finito di valori di t

Ora, moltiplicando e dividendo per s si ottiene

$$\begin{aligned} E(XY) &= \sum_{t\in\mathbb{R}\setminus\{0\}} \sum_{s\in\mathbb{R}\setminus\{0\}} s\, p_X(s)\, \frac{t}{s}\, p_Y\left(\frac{t}{s}\right) \\ &= \sum_{s\in\mathbb{R}\setminus\{0\}} s\, p_X(s) \sum_{t\in\mathbb{R}\setminus\{0\}} \frac{t}{s}\, p_Y\left(\frac{t}{s}\right) \\ &= \sum_{s\in\mathbb{R}\setminus\{0\}} s\, p_X(s) \sum_{r\in\mathbb{R}\setminus\{0\}} r\, p_Y(r) = E(X)\, E(Y). \end{aligned}$$

□

Come abbiamo già sottolineato, Il valore atteso ci dà un'idea di qual è il valore *medio* intorno a cui gli effettivi valori assunti dalla variabile aleatoria si distribuiscono. Non é in grado di dirci, però, quanto la variabile aleatoria può discostarsi da questo valore. Consideriamo due variabili aleatorie X e Y definite come segue:

X vale sempre zero. Quindi $p_X(0) = 1$, $p_X(t) = 0$ se $t \neq 0$ e $E(X) = 0 \cdot 1 = 0$

Y può prendere come valore qualunque numero intero $k \in \mathbb{Z}$, $k \neq 0$, con $p_Y(k) = c/k^4$, dove $c = \left(2 \sum_{k=1}^{\infty} \frac{1}{k^4}\right)^{-1}$. Quindi

$$E(Y) = \sum_{k\in\mathbb{Z}\setminus\{0\}} k \cdot \frac{c}{k^4} = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{c}{k^3} = \sum_{k=1}^{\infty} \frac{c}{k^3} + \sum_{k=1}^{\infty} \frac{c}{(-k)^3} = \sum_{k=1}^{\infty} \frac{c}{k^3} - \sum_{k=1}^{\infty} \frac{c}{k^3} = 0.$$

Anche se Y ha valore atteso pari a zero, è impossibile che prenda il valore zero, e c'è invece probabilità non nulla che assuma valori arbitrariamente grandi.

Per meglio descrivere una variabile aleatoria, quindi, serve anche sapere "quanto" può discostarsi dal suo valore atteso. La grandezza che misura questo scostamento è la *varianza* viene definito come segue.

Definizione 5.4: Varianza

Sia X una variabile aleatoria con valor medio finito e poniamo $\mu = E(X) \in \mathbb{R}$. Si definisce varianza di X il numero reale nonnegativo

$$\sigma^2(X) = E((X-\mu)^2).$$

La radice quadrata della varianza è spesso indicata come deviazione standard o scarto quadratico medio.

Con un po' di lavoro si può ricavare un'utile formula di calcolo

Proposizione 5.5: Formula per il calcolo della varianza

Sia X una variabile aleatoria di valore atteso finito; indichiamo con $\mu = E(X)$ il suo valore atteso. Allora

$$\sigma^2(X) = \begin{cases} \sum_{n=1}^{N} (x_n - \mu)^2 p_X(x_n) & \text{nel caso discreto finito,} \\ \sum_{n=1}^{\infty} (x_n - \mu)^2 p_X(x_n) & \text{nel caso discreto infinito,} \\ \int_{-\infty}^{+\infty} (t-\mu)^2 f_X(t)\, dt & \text{nel caso continuo.} \end{cases}$$

Se ora torniamo alle variabili X e Y di prima, vediamo che

$$\sigma^2(X) = (0-0)\cdot 1 = 0,$$
$$\sigma^2(Y) = \sum_{k\in\mathbb{Z}\setminus\{0\}} (k-0)^2 \cdot \frac{c}{k^4} = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{c}{k^2} = 2c\sum_{k=1}^{\infty}\frac{1}{k^2} > 0$$

Esempio 5.6

Calcoliamo la varianza della variabile bernoulliana Bern(p) *definita nel paragrafo 2.1. Sappiamo che* $E(\mathrm{Bern}(p)) = p$, *poi*

$$\sigma^2(\mathrm{Bern}(p)) = (0-p)^2(1-p) + (1-p)^2 p = p\,(1-p).$$

Per chiarirci il significato della varianza vediamo una nota disuguaglianza.

Proposizione 5.6: Disuguaglianza di Chebychev

Se X è una variabile aleatoria di valore atteso μ e varianza σ^2, allora

$$\mathrm{P}(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

per ogni $\epsilon > 0$.

Dimostrazione. Consideriamo solo il caso in cui X è una variabile continua che ammette densità $f(t)$. Poiché $\{|X-\mu| \geq \epsilon\} = \{X \leq \mu-\epsilon\} \vee \{X \geq \mu+\epsilon\}$ si ha

$$\mathrm{P}(|X-\mu| \geq \epsilon) = \int_{-\infty}^{\mu-\epsilon} f(t)dt + \int_{\mu+\epsilon}^{+\infty} f(t)dt. \tag{5.3}$$

D'altra parte

$$\begin{aligned}\frac{\sigma^2}{\epsilon^2} &= \int_{-\infty}^{+\infty} \frac{(t-\mu)^2}{\epsilon^2} f(t)dt \\ &= \int_{-\infty}^{\mu-\epsilon} \frac{(t-\mu)^2}{\epsilon^2} f(t)dt + \int_{\mu-\epsilon}^{\mu+\epsilon} \frac{(t-\mu)^2}{\epsilon^2} f(t)dt + \int_{\mu+\epsilon}^{+\infty} \frac{(t-\mu)^2}{\epsilon^2} f(t)dt.\end{aligned}$$

Ora, il secondo integrale è maggiore o uguale di zero perché tale è l'integrando. Inoltre, nel primo e nel secondo intervallo di integrazione si ha $|t-\mu| \geq \epsilon$ e dunque $(t-\mu)^2/\epsilon^2 \geq 1$. Ne segue che

$$\frac{\sigma^2}{\epsilon^2} \geq \int_{-\infty}^{\mu-\epsilon} f(t)dt + \int_{\mu+\epsilon}^{+\infty} f(t)dt = \mathrm{P}(|X-\mu| \geq \epsilon)$$

per la (5.3). □

Elenchiamo alcune utili proprietà della varianza

Proposizione 5.7

Se X è una variabile aleatoria di valor medio finito e $a, b \in \mathbb{R}$, allora

$$\sigma^2(X) = E(X)^2 - (E(X))^2, \tag{5.4}$$

$$\sigma^2(X) = 0 \quad \text{se, e solo se, } X \text{ è costante (con probabilità 1),} \tag{5.5}$$

$$\sigma^2(aX+b) = a^2\,\sigma^2(X). \tag{5.6}$$

Dimostrazione. Per verificare la proprietà (5.4), poniamo $\mu = E(X)$ e calcoliamo

$$\sigma^2(X) = E\left((X-\mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right)$$

Applicando la Proposizione 5.2 si ha

$$= E(X^2) - 2\mu\, E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2$$

□

Definizione 5.8: Variabile standardizzata

Se X è una variabile casuale di valor medio μ e deviazione σ, chiamiamo variabile standardizzata associata a X la variabile casuale

$$X^* = \frac{X-\mu}{\sigma}.$$

Utilizzando le proprietà elencate nelle Proposizione 5.2 e Proposizione 5.7 si verifica facilmente che

$$E(X^*) = 0, \qquad \sigma^2(X^*) = 1.$$

A titolo di esempio, vediamo come è fatta la bernoulliana standardizzata. Poiché sappiamo che $E(\mathrm{Bern}(p)) = p$ e $\sigma^2(\mathrm{Bern}(p)) = p(1-p)$, la standardizzata sarà $\frac{1}{\sqrt{p(1-p)}}(\mathrm{Bern}(p) - p)$, cioè la variabile che può prendere il valore $\sqrt{\frac{1-p}{p}}$ con probabilità p o il valore $-\sqrt{\frac{p}{1-p}}$ con probabilità $1-p$.

In generale, la varianza non è lineare. Tuttavia, la varianza della somma di due variabili indipendenti è la somma delle varianze.

Teorema 5.9: Varianza della somma di variabili indipendenti

Se X e Y sono due variabili indipendenti, allora

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y).$$

Dimostrazione. Per la formula si ha

$$\begin{aligned}
\sigma^2(X+Y) =& E((X+Y)^2) - (E(X) + E(Y))^2 \\
=& E(X^2 + Y^2 + 2XY) - (E(X))^2 - (E(Y))^2 - 2E(X)\,E(Y) \\
=& E(X^2) + E(Y^2) + 2E(XY) - (E(X))^2 - (E(Y))^2 - 2E(X)\,E(Y),
\end{aligned}$$

dopo aver utilizzato la linearità del valore atteso Proposizione 5.2. Raccogliendo abbiamo

$$\begin{aligned}
\sigma^2(X+Y) =& \left(E(X^2) - (E(X))^2\right) + \left(E(Y^2) - (E(Y))^2\right) + 2\left(E(XY) - E(X)\,E(Y)\right) \\
=& \sigma^2(X) + \sigma^2(Y).
\end{aligned}$$

Nell'ultima uguaglianza abbiamo usato di nuovo la formula (5) per il calcolo di $\sigma^2(X)$ e $\sigma^2(Y)$; inoltre abbiamo utilizzato che $E(XY) = E(X)\,E(Y)$ per il Teorema 5.3. □

Esempio 5.7

Calcoliamo la varianza della variabile binomiale $\mathrm{Bin}(n,p)$ *introdotta nell'Esempio 2.3. Poiché essa è la somma di n variabili bernoulliane di parametro p mutuamente indipendenti, segue dal Teorema 5.9 che*

$$\sigma^2(\mathrm{Bin}(n,p)) = n\,\sigma^2(\mathrm{Bern}(p)) = n\,p\,(1-p).$$

6. Convergenza di variabili aleatorie: la legge dei grandi numeri e il teorema del limite centrale

Secondo la nozione frequentista di probabilità, ripetendo un gran numero di volte un esperimento, la frequenza con cui un dato evento A si verifica è "all'incirca" uguale alla sua probabilità. Questa affermazione è nota come "legge empirica dei grandi numeri". Andiamo ora ad enunciare un teorema che formalizza questa legge empirica.

Dato un qualunque evento A, possiamo definire una variabile che vale 1 se l'evento si verifica, 0 altrimenti. È questa una variabile di tipo bernoulliano di parametro $p = \mathrm{P}(A)$. Immaginiamo ora di ripetere il nostro esperimento n volte in modo indipendente e di contare quante volte l'evento si verifica. Otteniamo così una variabile binomiale che indicheremo in seguito con

$$X_n = \mathrm{Bin}(n, p).$$

La frequenza (statistica) con cui l'evento A si verifica è ora data da

$$\mathrm{fq}_n(A) = \frac{X_n}{n}.$$

La legge empirica dei grandi numeri afferma dunque che

$$\lim_{n\to\infty} \frac{X_n}{n} = \mathrm{P}(A).$$

Ma cosa significa questa scrittura? Come va inteso il limite di una successione di variabili aleatorie? Intuitivamente, possiamo dire che una successione di variabili X_n tende a X se, per grandi valori di n, siamo pressoché certi che X_n e X abbiano valori vicini (o, viceversa, è pressoché impossibile che abbiano valori lontani). Formalmente, si ha la seguente definizione

Definizione 6.1: Convergenza in probabilità

Una successione di variabili casuali $(X_n)_{n\in\mathbb{N}}$ **converge in probabilità** *alla variabile* X *se per ogni* $\epsilon > 0$ *si ha*

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| < \epsilon) = 1,$$

o equivalentemente

$$\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0.$$

In simboli scriveremo $X_n \xrightarrow{\mathrm{P}} X$.

Siamo ora pronti ad enunciare

Teorema 6.2: Legge dei grandi numeri

Sia A *un evento. Se* X_n *è il numero di volte che l'evento* A *si verifica in* n *prove indipendenti, allora*

$$\frac{X_n}{n} \xrightarrow{\mathrm{P}} \mathrm{P}(A).$$

Dimostrazione. Detta $p = \mathrm{P}(A)$, dobbiamo dimostrare che per ogni $\epsilon > 0$ si ha

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) = 0.$$

Poiché, come abbiamo osservato in precedenza, X_n è una variabile di tipo binomiale di parametri n e p, si ha che

$$\begin{aligned}\mu =&E\left(\frac{X_n}{n}\right) = \frac{E(X_n)}{n} = \frac{np}{n} = p,\\ \sigma^2 =&\sigma^2\left(\frac{X_n}{n}\right) = \frac{\sigma^2(X_n)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.\end{aligned}$$

Applichiamo ora la disuguaglianza di Chebichev alla variabile X_n/n, si ottiene

$$\mathrm{P}\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

D'altra parte per costruzione $\mathrm{P}(|X_n/n - p| \geq \epsilon) \geq 0$. Segue allora dal teorema dei due carabinieri che

$$0 \leq \lim_{n\to\infty} \mathrm{P}\left(\left|\frac{X_n}{n} - p\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{p(1-p)}{n\epsilon^2} = 0,$$

cioè appunto la nostra tesi. □

Passiamo ora a trattare un altro teorema fondamentale, noto come "teorema del limite centrale". Si tratta anche in questo caso di un teorema di approssimazione, ma si fa ricorso ad un altro concetto di limite.

Definizione 6.3: Convergenza in legge

Una successione di variabili casuali $(X_n)_{n\in\mathbb{N}}$ **converge in legge** *alla variabile X se convergono puntualmente le rispettive funzioni di distribuzione, precisamente se*

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$$

per ogni t in cui F_X è continua. In simboli scriveremo $X_n \xrightarrow{L} X$.

Teorema 6.4: Teorema del limite centrale

Sia X_n una successione di variabili casuali mutuamente indipendenti, tutte con la stessa funzione di distribuzione F. In particolare, esse hanno uguale valor medio $E(X_n) = \mu$ e ugual varianza $\sigma^2(X_n) = \sigma^2$ per ogni $n \in \mathbb{N}$. Definiamo poi la variabile

$$Z_n = \frac{\sum\limits_{k=1}^{n} X_n - n\mu}{\sigma\sqrt{n}}.$$

Allora Z_n converge in legge alla variabile normale standard $N(0,1)$, ovvero

$$Z_n \xrightarrow{L} N(0,1).$$

Consideriamo la somma parziale delle X_n, cioè $\sum\limits_{k=1}^{n} X_n$; si osservi che, poichè le X_n sono indipendenti, si ha

$$E\left(\sum_{k=1}^{n} X_n\right) = \sum_{k=1}^{n} E(X_n) = n\mu, \qquad \sigma^2\left(\sum_{k=1}^{n} X_n\right) = \sum_{k=1}^{n} \sigma^2(X_n) = n\sigma^2.$$

Pertanto Z_n è la variabile standardizzata di tale somma parziale. Se le X_n sono tutte variabili normali distribuite secondo $N(\mu, \sigma^2)$, allora Z_n è la variabile normale standard per ogni $n \in \mathbb{N}$. Il teorema del limite centrale afferma che, in generale, anche se le X_n <u>non</u> sono gaussiane, la loro somma standardizzata approssima la variabile normale standard.

Dal punto di vista applicativo, se una variabile casuale può essere vista come somma di molte variabili casuali equidistribuite (cioè con la stessa funzione di distribuzione), allora la sua distribuzione è "quasi" normale. Ecco perché l'uso della distribuzione normale è molto frequente nei problemi di tipo statistico.

Esempio 6.1: Approssimazione normale della distribuzione binomiale

Abbiamo osservato in precedenza che la variabile binomiale $\mathrm{Bin}(n, p)$ *è la somma di* n *variabili bernoulliane* $\mathrm{Bern}(p)$ *indipendenti fra loro: indichiamole con* $X_1, X_2, \dots X_n$ *(dove l'indice rappresenta la prova a cui si riferiscono). Sappiamo che* $\mu = E(X_k) = p$, $\sigma = \sigma(X_k) = \sqrt{p(1-p)}$, $E(\mathrm{Bin}(n,p)) = np$, $\sigma(\mathrm{Bin}(n,p)) = \sqrt{np(1-p)}$. *Dunque la variabile standardizzata della binomiale è*

$$Z_n = \frac{\mathrm{Bin}(n,p) - np}{\sqrt{np(1-p)}} = \frac{\sum\limits_{k=1}^{n} X_n - n\mu}{\sigma\sqrt{n}}$$

e il Teorema del Limite Centrale afferma che, per n *grande, essa può essere approssimata dalla variabile gaussiana standard.*
In modo equivalente, per n *grande possiamo approssimare la funzione della binomiale* $\mathrm{Bin}(n, p)$ *con la distribuzione normale* $N(np, \sqrt{np(1-p)})$.