SAPIENZA Università di Roma

INTRODUZIONE
AL
CALCOLO DELLE PROBABILITÀ

Fabio Spizzichino e Giovanna Nappo

Indice

Introduzione

Il presente testo prende spunto dall'esperienza maturata nell'ambito dell'insegnamento in corsi introduttivi alla Probabilità, a partire dall'inizio degli anni 2000 e in Anni Accademici diversi. Più precisamente, si tratta del risultato di successivi aggiornamenti ed elaborazioni di appunti di lezioni da noi rispettivamente svolte per *Probabilità 1* al primo anno del Corso di Laurea Triennale in Matematica, presso l'Università Sapienza di Roma.

Le elaborazioni man mano apportate hanno però avuto anche, come ulteriore punto di riferimento, le esigenze didattiche di un pubblico più ampio, ed in particolare degli studenti di altri corsi universitari, introduttivi alla Probabilità, quale il corso di Calcolo delle Probabilità per informatici. In ogni caso si tiene conto che le conoscenze matematiche degli studenti, ai quali ci si rivolge, sono ad un livello più elementare rispetto agli argomenti che gli studenti di Matematica affronteranno negli anni successivi.

Ci siamo dunque posti un duplice obiettivo. Da una parte - come in molti altri testi - si mira a fornire un'introduzione al linguaggio, alla problematica e alle nozioni fondamentali della probabilità. D'altra parte, pensando specialmente agli studenti che proseguiranno lo studio della matematica, della probabilità e dei processi stocastici in particolare, è opportuno presentare le nozioni fondamentali in una forma che, seppure completamente elementare, già mostra le linee fondamentali della trattazione con il linguaggio della teoria della misura.

Per tale ragioni, in definitiva, gli argomenti presentati vengono suddivisi in due diverse parti.

Nella prima parte ci si limita a considerare spazi di probabilità in cui lo *spazio campione* Ω è un insieme finito ed in cui l'algebra degli eventi coincide con l'insieme $\mathcal{P}(\Omega)$ delle parti di Ω.

Nella seconda parte vengono considerate situazioni più generali, in cui Ω non è finito oppure, pur essendo esso finito, l'algebra degli eventi considerati non coincide con l'insieme $\mathcal{P}(\Omega)$.

Nella prima parte si mira a consolidare, nella mente dello studente, l'idea fondamentale che un qualunque *evento* può essere visto come un sottoinsieme di un opportuno spazio Ω e che una qualunque variabile aleatoria può essere vista come una funzione definita su Ω.

Ovviamente, tale impostazione non affatto è l'unica possibile, ma riteniamo che essa debba risultare fondamentalmente chiara al principiante, in quanto essa è alla base degli attuali sviluppi della teoria della probabilità e costituisce uno strumento essenziale per *comunicare* le relative idee in ambito matematico.

In particolare, ciò permette di definire direttamente il valore atteso (di una variabile aleatoria) sullo spazio Ω e di dimostrare quindi la proprietà di linearità in modo diretto, anche prima di studiare le distribuzioni di probabilità e -specialmente- le distribuzioni congiunte.

Nella seconda parte, dedicata ai casi più generali, viene introdotta la proprietà di additività numerabile e la conseguente proprietà di continuità.

Alcune indicazioni ad uso degli studenti:
Per il momento segnaliamo che tra gli esempio, quelli in cui compare l'asterisco sono sotto forma di esercizio con soluzione: si consiglia di provare a pensare alla loro soluzione prima di guardare la soluzione.
Il testo contiene delle note di approfondimento messe in riquadri: a volte si tratta di richiami, a volte di suggerimenti, a volte si tratta di chiarimenti e spiegazioni importanti.

Per segnalazioni di eventuali sviste, errori di stampa, incoerenza delle notazioni, o anche per suggerimenti si prega di inviare un messaggio a

Giovanna Nappo: e-mail `nappo@mat.uniroma1.it`

Roma, giugno 2019

Avvertenze per gli studenti

Oltre a questi appunti ci sono/saranno altri appunti di Giovanna Nappo riassuntivi e/o sostitutivi su vari argomenti, in particolare su

NUMERI DI OCCUPAZIONE
SCHEMA di VARIABILI ALEATORIE
VALORE ATTESO CONDIZIONALE
APPROSSIMAZIONE NORMALE
CATENE DI MARKOV

Trovate/troverete tali appunti sui siti e-learning del corso, e precisamente

2022-23 PROBABILITA' 1 (L-Z)
al link:
https://elearning.uniroma1.it/course/view.php?id=16183

2023-24 PROBABILITA' 1 (L-Z)
al link:
https://elearning.uniroma1.it/course/view.php?id=17792

CALCOLO delle PROBABILITA' (canale 1) a.a. 2024-25
al link:
https://elearning.uniroma1.it/course/view.php?id=18505

Infine che gli studenti devono tenere presente che c'è anche un testo

Berger-Caravenna-Dai Pra: Probabilità
che era reperibile gratuitamente sul web (per gli studenti della Sapienza) al link:
https://link.springer.com/book/10.1007/978-88-470-4006-9
ma ora non lo è più.

Roma, settembre 2024

1 Fenomeni aleatori; spazio dei risultati elementari di un esperimento

1.1 Operazioni logiche su eventi e interpretazione insiemistica

Iniziamo con una discussione euristica mirante a giustificare la definizione della nozione di *spazio finito di probabilità*, che verrà data nella prossima lezione.

Come punto di partenza, pensiamo ad un esperimento che possa dar luogo a diversi risultati possibili. Possiamo essere interessati a diverse situazioni prodotte dall'esperimento stesso.

Esempio 1.1. *Consideriamo l'esperimento consistente nell'osservazione dei punti ottenuti dal lancio di una coppia di dadi a sei facce. Esempi di situazioni diverse sono*
$A = \{$*Il punteggio ottenuto dal primo dado è minore o uguale al punteggio del secondo dado*$\}$
$B = \{$*La somma dei punteggi ottenuti è pari*$\}$
$C = \{$*Il punteggio ottenuto dal primo dado è (strettamente) maggiore di* $3\}$
$D = \{$*Il punteggio ottenuto dal primo dado è 2 e il punteggio ottenuto dal secondo dado è 5*$\}$
Più sinteticamente, indicando i due punteggi con i simboli $X_1,\ X_2$*, possiamo scrivere:*
$A = \{X_1 \leq X_2\}$, $B = \{X_1 + X_2$ *pari*$\}$, $C = \{X_1 > 3\}$, $D = \{X_1 = 2, X_2 = 5\}$.

Tali situazioni verranno chiamate ***eventi***. Possiamo vedere un evento come una proposizione relativa al modo di risultare di tale esperimento.

Indichiamo, per il momento, con il simbolo $\mathscr{E}$ *(= E maiuscolo corsivo)* la famiglia dei possibili eventi distinti in un esperimento. Come è facile rendersi conto (e verificheremo presto), la famiglia $\mathscr{E}$ costituita da tutti gli eventi nel precedente Esempio 1.1 è una famiglia finita. In quanto immediatamente segue, ci limiteremo ancora a considerare esperimenti per cui $\mathscr{E}$ è una ***famiglia finita***; successivamente, tale limitazione verrà rimossa.

È anche facile rendersi conto che, all'interno della famiglia $\mathscr{E}$, è naturale introdurre le operazioni di ***somma logica*** (oppure ***OR***), di ***prodotto logico*** (oppure ***AND***) e di ***negazione*** (oppure ***NOT***), che verranno rispettivamente indicate (per il momento) con i simboli $\bigvee, \bigwedge, \neg$; siano $E_1,\ E_2,\ E$ eventi appartenenti ad $\mathscr{E}$, allora

la somma logica $E_1 \bigvee E_2$ coincide con l'evento:

$$\{\textit{si è verificato almeno uno dei due eventi } E_1 \textit{ ed} E_2\}$$

il prodotto logico $E_1 \bigwedge E_2$ coincide con l'evento:

$$\{\textit{si sono verificati entrambi i due eventi } E_1 \textit{ ed } E_2\}$$

la negazione $\neg E$ coincide con l'evento:

$$\{\textit{non si è verificato l'evento } E\}.$$

Definizione 1.1. *Un evento* $E \in \mathscr{E}$ *si dice* **composto** *se esistono almeno due eventi* $E_1, E_2 \in \mathscr{E}$*, tali che*

$$E = E_1 \bigvee E_2, \quad E \neq E_1, E \neq E_2.$$

Un evento che non sia composto si dice **semplice** *o* **elementare***.*

Esempio 1.2. *Nell'esperimento del lancio di un dado a sei facce, l'evento $E = \{X_1 > 3\}$ è un evento composto. In tale esperimento gli eventi semplici sono dati da*

$$\{X_1 = 1\},\ \{X_1 = 2\}, ..., \ \{X_1 = 6\},$$

e l'evento E si riscrive come

$$\{X_1 > 3\} = \{X_1 = 4\} \bigvee \{X_1 = 5\} \bigvee \{X_1 = 6\}.$$

Nell'esperimento del lancio di una coppia di dadi gli eventi semplici sono invece quelli del tipo

$$\{X_1 = h,\ X_2 = k\} \qquad h = 1, 2, ..., 6;\ \ k = 1, 2, ..., 6$$

ed un evento del tipo $\{X_1 = h\}$ risulta essere un evento composto, in quanto possiamo scrivere

$$\{X_1 = h\} = \bigvee_{k=1}^{6} \{X_1 = h, X_2 = k\}.$$

Abbiamo supposto che la famiglia $\mathscr{E}$ sia finita, e quindi anche la famiglia degli eventi elementari è finita. Sia N il numero degli eventi elementari. Indichiamo ora con i simboli $\omega_1, ..., \omega_N$ gli eventi elementari in un esperimento.

Definizione 1.2. *L'insieme $\Omega = \{\omega_1, ..., \omega_N\}$ che ha come punti gli eventi elementari di un esperimento viene detto* **spazio campione**, *per quell'esperimento.*

***Osservazione* 1.1.** È facile verificare che ogni evento composto $E \in \mathscr{E}$ si decompone in uno ed in un sol modo (a meno dell'ordine) come somma logica di un numero finito di eventi elementari.

Indichiamo con il simbolo $\mathscr{P}(\Omega)$ la ***famiglia delle parti*** di Ω (ossia la famiglia dei sottoinsiemi di Ω) e, per $E \in \mathscr{P}(\Omega)$, indichiamo con $|E|$ la ***cardinalità*** di E (ossia il numero degli elementi di E)

Ad esempio nel lancio di un dado a 4 facce, numerate da 1 a 4, possiamo porre
ω_1 l'evento elementare $\{$*il punteggio ottenuto dal lancio del dado è 1*$\}$
ω_2 l'evento elementare $\{$*il punteggio ottenuto dal lancio del dado è 2*$\}$
ω_3 l'evento elementare $\{$*il punteggio ottenuto dal lancio del dado è 3*$\}$
ω_4 l'evento elementare $\{$*il punteggio ottenuto dal lancio del dado è 4*$\}$
in modo che $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, e la famiglia $\mathscr{P}(\Omega)$ delle parti di Ω è l'insieme i cui elementi sono tutti i sottoinsiemi di Ω:

$$\begin{aligned}\mathscr{P}(\Omega) = \Big\{ &\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\},\\ &\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_3, \omega_4\},\\ &\{\omega_1, \omega_2, \omega_3\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \Big\}\end{aligned}$$

$\emptyset$ è l'insieme vuoto, che non contiene elementi e quindi ha cardinalità 0,
$\{\omega_1\}$, $\{\omega_2\}$, $\{\omega_3\}$, $\{\omega_4\}$ sono i singoletti, che contengono un unico elemento e quindi hanno cardinalità 1
$\{\omega_1, \omega_2\}$, $\{\omega_1, \omega_3\}$, $\{\omega_1, \omega_4\}$, $\{\omega_2, \omega_3\}$, $\{\omega_2, \omega_4\}$, $\{\omega_3, \omega_4\}$ sono i sottoinsiemi con due elementi e quindi hanno cardinalità 2,
$\{\omega_1, \omega_2, \omega_3\}$, $\{\omega_1, \omega_2, \omega_4\}$, $\{\omega_1, \omega_3, \omega_4\}$ sono i sottoinsiemi con tre elementi e quindi hanno cardinalità 3,
$\{\omega_1, \omega_2, \omega_3, \omega_4\}$ coincide con Ω, ma è pensato come sottoinsieme di se stesso, contiene 4 elementi e quindi ha cardinalità 4.

.

Esempio 1.3. *Un'urna inizialmente contiene quattro oggetti numerati da* 1 *a* 4*. Vuotiamo l'urna facendo quattro successive estrazioni senza reinserimento, osservando di volta in volta il numero indicato sull'oggetto estratto.*

Si ha $\Omega = \big\{\text{permutazioni}^{1} \text{ di } \{1,2,3,4\}\big\}$, $|\Omega| = 24$.

Nell'Esempio 1.3 possiamo indicare sinteticamente ogni evento elementare con una quaterna ordinata dei numeri 1,2,3,4, in modo che siano tutti diversi tra loro. Ad esempio l'evento elementare

$\{$*il primo numero estratto è il 3, il secondo è l'1, il terzo è il 2, il quarto è il 4*$\}$

viene indicato come la permutazione $(3,1,2,4)$*. Tutte le permutazioni di* $\{1,2,3,4\}$ *sono elencate qui sotto*

$$\begin{array}{l}
(1,2,3,4),(1,2,4,3),(1,3,2,4),(1,3,4,2),(1,4,2,3),(1,4,3,2),\\
(2,1,3,4),(2,1,4,3),(2,3,1,4),(2,3,4,1),(2,4,1,3),(2,4,3,1),\\
(3,1,2,4),(3,1,4,2),(3,2,1,4),(3,2,4,1),(3,4,1,2),(3,4,2,1),\\
(4,1,2,3),(4,1,3,2),(4,2,1,3),(4,2,3,1),(4,3,1,2),(4,3,2,1)
\end{array}$$

vale la pena osservare che l'evento

E=$\{$il primo numero estratto è l'1 e il terzo è il 3$\}$

è un evento composto, in quanto è la somma logica dei due eventi

E_1=$\{$il primo estratto è l'1, il secondo è il 2, il terzo è il 3, il quarto è il 4$\}$
E_2=$\{$il primo estratto è l'1, il secondo è il 4, il terzo è il 3, il quarto è il 2$\}$

Nello studio della probabilità si è interessati a quei casi in cui vi sia una *situazione di incertezza* (cioè di mancanza di completa informazione) circa il modo di risultare dell'esperimento stesso. Ciò significa che non sappiamo a priori quale effettivamente si realizzerà fra i diversi risultati elementari possibili. In tali casi parleremo di fenomeni aleatori o di *esperimenti aleatori*. Parleremo dunque di *esperimento aleatorio* quando non sappiamo quali eventi saranno verificati e quali risulteranno falsi.

In tale ambito è importante sottolineare la seguente terminologia:
diremo che

un evento composto E *si è verificato*
se e solo se
si è verificato un evento elementare ω_i *della decomposizione di* E,
anche se non è noto quale esattamente si sia verificato.

Esempio 1.4. *Si lancia un dado a sei facce; si ha* $\Omega = \{\omega_1, ..., \omega_6\}$*, dove* ω_i *rappresenta l'evento elementare* "il dado mostra la faccia i"*. Supponiamo si verifichi* ω_4*, allora, ad esempio,*
si sono verificati anche gli eventi composti:

$$\{X \leq 5\},\ \{X \textit{pari}\},\ \{X > 2\},$$

mentre non sono verificati gli eventi composti

$$\{X > 5\},\ \{X \textit{dispari}\},\ \{X \leq 3\},\ \{X \textit{ numero primo}\},$$

[1] Per la definizione di permutazione vedere più avanti la Lezione 3 ed in particolare il Sezione 3.2.

Con tale terminologia si ha che, dati due eventi $E_1, E_2 \in \mathscr{E}$,

(a) $E_1 \bigvee E_2$ si verifica se e solo se è verificato un evento elementare ω_i che si presenti nella decomposizione di E_1 oppure in quella di E_2

(b) $E_1 \bigwedge E_2$ si verifica se e solo se è verificato un evento elementare ω_i che si presenti sia nella decomposizione di E_1 sia in quella di E_2

(c) $\neg E_1$ si verifica se e solo se è verificato un evento elementare ω_i che non sia presente nella decomposizione di E_1

***Osservazione* 1.2 (Eventi come sottoinsiemi di Ω).** Per definizione, i punti dello spazio Ω sono gli "eventi semplici" o "risultati elementari" dell'esperimento considerato.

Notiamo ora che sussiste una corrispondenza biunivoca fra sottoinsiemi di Ω, costituiti da più di un elemento, e gli eventi composti: basta infatti associare, ad un evento composto, l'insieme costituito dagli eventi semplici che lo compongono; viceversa ad un sottoinsieme di Ω possiamo associare l'evento composto che si ottiene come somma logica degli elementi (eventi semplici) in esso contenuti.

Ad un evento semplice $\omega_i \in \Omega$, facciamo corrispondere l'insieme $\{\omega_i\} \in \mathscr{P}(\Omega)$ (ricordiamo che un insieme che ha un solo elemento è detto *singoletto*, o anche con il termine inglese *singleton*)

Dato un evento $E \in \mathscr{E}$, indichiamo per comodità con $\mathcal{H}(E)$ il sottoinsieme di Ω individuato secondo quanto appena detto.

Nell'Esempio 1.3 dell'estrazione senza reinserimento di 4 palline numerate da 1 a 4, possiamo identificare Ω con l'insieme di tutte le permutazioni di $\{1,2,3,4\}$, *(per la definizione precisa di permutazione vedere la successiva Lezione 3 sul Calcolo Combinatorio)* ossia

$$\begin{aligned}\Omega = \{&(1,2,3,4),(1,2,4,3),(1,3,2,4),(1,3,4,2),(1,4,2,3),(1,4,3,2),\\&(2,1,3,4),(2,1,4,3),(2,3,1,4),(2,3,4,1),(2,4,1,3),(2,4,3,1),\\&(3,1,2,4),(3,1,4,2),(3,2,1,4),(3,2,4,1),(3,4,1,2),(3,4,2,1),\\&(4,1,2,3),(4,1,3,2),(4,2,1,3),(4,2,3,1),(4,3,1,2),(4,3,2,1)\,\}\end{aligned}$$

e quindi, per esempio, all'evento $E = \{$*il primo numero estratto è l'1 e il terzo è il 3*$\}$ corrisponde il sottoinsieme

$$\mathcal{H}(E) = \{(1,2,3,4),(1,4,3,2)\},$$

mentre all'evento $F = \{$*il secondo numero estratto è il 3*$\}$ corrisponde il sottoinsieme

$$\mathcal{H}(F) = \{(1,3,2,4),(1,3,4,2),(2,3,1,4),(2,3,4,1),(4,3,1,2),(4,3,2,1)\}.$$

***Osservazione* 1.3 (Operazioni su eventi e operazioni su sottoinsiemi).** Consideriamo di nuovo la corrispondenza biunivoca $\mathcal{H}$ fra eventi e sottoinsiemi di Ω, stabilita nella precedente ***Osservazione* 1.2**:

$$\mathscr{E} \overset{\mathcal{H}}{\Leftrightarrow} \mathscr{P}(\Omega)\,.$$

Ci si rende facilmente conto, da quanto detto sopra, che, in tale corrispondenza biunivoca fra eventi e sottoinsiemi, le operazioni $\bigvee, \bigwedge, \neg$ (definite su $\mathscr{E}$) vengono rispettivamente trasformate nelle operazioni booleane di unione $\cup$, di intersezione $\cap$, e di passaggio al complementare $\complement$ (definite su $\mathscr{P}(\Omega)$, la famiglia delle parti di Ω); infatti, traducendo "in formule" i precedenti punti (a), (b) e (c) potremo scrivere, per degli arbitrari $E_1,\ E_2,\ E \in \mathscr{E}$,

$$\mathcal{H}\left(E_1 \bigvee E_2\right) = \mathcal{H}(E_1) \cup \mathcal{H}(E_2),$$

$$\mathcal{H}\left(E_1 \bigwedge E_2\right) = \mathcal{H}(E_1) \cap \mathcal{H}(E_2),$$

$$\mathcal{H}(\neg E) = \complement(\mathcal{H}(E)).$$

1.2 Interpretazione "logica" di nozioni di tipo insiemistico

Da questo momento in poi, quindi, potremo **identificare** "eventi" e sottoinsiemi di Ω e dunque lasceremo cadere l'uso dei simboli $\mathscr{E}, \bigvee, \bigwedge, \neg, \mathcal{H}(\cdot)$; continueremo la trattazione utilizzando solo le nozioni di sottoinsieme di Ω e di operazioni booleane fra sottoinsiemi.

Sempre nell'Esempio 1.3 dell'estrazione di 4 palline numerate da 1 a 4, scriveremo quindi più semplicemente

$$E = \{\textit{il primo numero estratto è l'1 e il terzo è il 3}\} = \{(1,2,3,4),(1,4,3,2)\}$$

$$F = \{\textit{il secondo numero estratto è il 3}\}$$
$$= \{(1,3,2,4),(1,3,4,2),(2,3,1,4),(2,3,4,1),(4,3,1,2),(4,3,2,1)\}.$$

Ad esempio, per indicare l'evento $E \bigvee F$ al quale corrisponderebbe l'insieme $\mathcal{H}(E) \cup \mathcal{H}(F)$, scriveremo semplicemente $E \cup F$.

Dovremo però continuare ad aver presente il significato di tipo "logico" che stiamo dando a tali nozioni, nel contesto dell'analisi di fenomeni aleatori. In tale ambito, risulta naturale attribuire un'interpretazione di tipo "logico" a varie semplici nozioni di tipo insiemistico; a tale proposito vediamo intanto lo specchietto presentato qui di seguito.

(i) $A \subseteq B$ significa che ogni evento elementare che rende verificato A rende verificato anche B
ovvero il verificarsi di A implica il verificarsi di B e dunque
interpretiamo la relazione $A \subseteq B$ come "***A implica B***",

(ii) Ω è un evento vero qualunque evento elementare si verifichi,
in quanto esso contiene tutti gli eventi elementari e dunque
interpretiamo Ω come l'***evento certo***

(iii) l'insieme vuoto $\emptyset$ è un evento che non è mai verificato
in quanto non contiene alcuno degli eventi elementari possibili e dunque
interpretiamo $\emptyset$ come l'***evento impossibile***

(iv) $A \cup B = \Omega$ significa che
l'evento costituito dal verificarsi di almeno uno dei due eventi A o B coincide con l'evento certo Ω
e dunque
interpretiamo $A \cup B = \Omega$ come ***A e B sono esaustivi***
ovvero è certo che se ne verifichi ***almeno*** uno dei due

(v) $A \cap B = \emptyset$ significa che
l'evento costituito dal verificarsi di entrambi gli eventi A e B coincide con l'evento impossibile $\emptyset$
e dunque
interpretiamo la condizione $A \cap B = \emptyset$ come ***A e B sono incompatibili***
ossia è certo che se ne verifichi ***al più*** uno dei due eventi
ovvero è certo che se ne verifichi ***al massimo*** uno dei due eventi.

Passiamo ora ad analizzare il significato "logico" della nozione di ***partizione dell'evento certo***.

Definizione 1.3 (Partizione dell'evento certto). *Consideriamo una collezione di sottoinsiemi* $\mathcal{H} = \{H_1, ..., H_m\}$ *dello spazio campione* Ω $(H_\ell \in \mathscr{P}(\Omega), \ell = 1, ..., m)$*. Tale collezione costituisce una* **partizione di Ω** *se e solo se gli insiemi* $H_1, ..., H_m$ *sono disgiunti a due a due e la loro unione coincide con* Ω*, ossia*

$$H_{\ell_1} \cap H_{\ell_2} = \emptyset, \text{ per } \ell_1 \neq \ell_2; \quad e \quad \bigcup_{\ell=1}^{m} H_\ell = \Omega.$$

Gli insiemi H_ℓ *vengono anche detti* **elementi della partizione** $\mathcal{H}$*.*
Interpretando $H_1, ..., H_m$ *come eventi, abbiamo che essi sono a* **due a due incompatibili** *(cioè è* impossibile che se ne possono verificare due contemporaneamente*) e, d'altra parte, essi sono* **esaustivi** *(*è certo che se ne verifichi almeno uno*).*
In altre parole è certo che si verifichi uno ed uno soltanto degli eventi $H_1, ..., H_m$*.*

Nell'Esempio 1.3 dell'estrazione senza reinserimento di 4 palline numerate da 1 a 4, in cui abbiamo identificato Ω con l'insieme di tutte le permutazioni di $\{1, 2, 3, 4\}$, la famiglia $\mathcal{H} = \{H_1, H_2, H_3, H_4\}$, dove

$$H_1 = \{\textit{Il primo numero estratto è l'1}\} \quad H_2 = \{\textit{Il primo numero estratto è il 2}\}$$

$$H_3 = \{\textit{Il primo numero estratto è il 3}\} \quad H_4 = \{\textit{Il primo numero estratto è il 4}\}$$

è una partizione dell'evento certo Ω. In termini di insiemi si ha

$$\begin{aligned}
H_1 &= \{(1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (1, 3, 4, 2), (1, 4, 2, 3), (1, 4, 3, 2)\},\\
H_2 &= \{(2, 1, 3, 4), (2, 1, 4, 3), (2, 3, 1, 4), (2, 3, 4, 1), (2, 4, 1, 3), (2, 4, 3, 1)\},\\
H_3 &= \{(3, 1, 2, 4), (3, 1, 4, 2), (3, 2, 1, 4), (3, 2, 4, 1), (3, 4, 1, 2), (3, 4, 2, 1)\},\\
H_4 &= \{(4, 1, 2, 3), (4, 1, 3, 2), (4, 2, 1, 3), (4, 2, 3, 1), (4, 3, 1, 2), (4, 3, 2, 1)\}
\end{aligned}$$

e chiaramente gli elementi della partizione sono **esaustivi**, ossia l'unione degli elementi della partizione è tutto Ω, ossia

$$\Omega = H_1 \cup H_2 \cup H_3 \cup H_4,$$

e sono **incompatibili a due a due**, ossia gli insiemi sono disgiunti a due a due, ossia

$$H_1 \cap H_2 = \emptyset, \ H_1 \cap H_3 = \emptyset, \ H_1 \cap H_4 = \emptyset, \ H_2 \cap H_3 = \emptyset, \ H_2 \cap H_4 = \emptyset, \ H_3 \cap H_4 = \emptyset.$$

Ricordiamo infine le seguenti proprietà basilari delle operazioni booleane su insiemi:

Proprietà distributiva dell'unione rispetto all'intersezione:

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

Proprietà distributiva dell'intersezione rispetto all'unione:

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

Legge di De Morgan:

$$\complement(A \cap B) = \complement(A) \cup \complement(B) \quad \text{o equivalentemente} \quad A \cap B = \complement\big(\complement(A) \cup \complement(B)\big).$$

Per brevità d'ora in poi scriveremo $\overline{A}$ invece di $\complement(A)$, per indicare il complementare (o la "negazione") di un insieme $A \in \mathscr{P}(\Omega)$ per cui la Legge di De Morgan si riscrive

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \quad \text{o equivalentemente} \quad A \cap B = \overline{(\overline{A} \cup \overline{B})}. \tag{1}$$

Anche tali proprietà ammettono una semplice interpretazione logica, ad esempio, per quanto riguarda la legge di De Morgan, l'interpretazione è la seguente:
negare il verificarsi sia di A che di B equivale a richiedere il verificarsi di almeno una tra la negazione di A e la negazione di B.
Le proprietà precedenti si estendono per induzione a famiglie finite di insiemi $\{A_i, i \in I\}$ (ma valgono anche per famiglie infinite):

$$\Big(\bigcup_{i\in I} A_i\Big) \cap C = \bigcup_{i\in I}(A_i \cap C),$$

$$\Big(\bigcap_{i\in I} A_i\Big) \cup C = \bigcap_{i\in I}(A_i \cup C),$$

$$\complement\Big(\bigcap_{i\in I} A_i\Big) = \bigcup_{i\in I}\complement(A_i) \quad \text{o con l'altra notazione,} \quad \overline{\Big(\bigcap_{i\in I} A_i\Big)} = \bigcup_{i\in I}\overline{A_i}.$$

1.3 Esercizi di verifica

Esercizio 1.1. Consideriamo l'esperimento consistente nel lancio di una pallina nel gioco della roulette. In tale esperimento è naturale porre

$$\Omega = \{0, 1, ..., 36\}$$

e vedere i risultati *manque, passe, noir, rouge, pair, unpair,* come altrettanti eventi composti Supponiamo che nell'esperimento si verifichi l'evento elementare $\{16\}$. Quale degli eventi composti sopra elencati è verificato e quale no?

Come funziona la roulette

Si ricorda che la Roulette è una ruota con trentasette settori numerati da zero a trentasei. Una pallina viene fatta girare e alla fine si ferma su uno di questi numeri. Inoltre puntare su *manque* significa puntare su un numero tra 1 e 18, puntare su *passe* significa puntare su un numero tra 19 e 36, puntare su *noir* significa puntare su un numero nero, puntare su *rouge* significa puntare su un numero rosso, ed analogamente per *pair*, ovvero pari e *unpair*, ovvero dispari. Ai fini della soluzione dell'esercizio, è importante sapere che il 16 è rosso.

Esercizio 1.2. Dati due eventi A e B, scrivere l'espressione dell'evento:

$$\{\textit{si verifica esattamente un solo evento fra } A \text{ e } B\}$$

in termini delle operazioni booleane di unione $\cup$, di intersezione $\cap$ e di passaggio al complementare.

Esercizio 1.3. Siano A, B e C eventi. Scrivere le espressioni degli eventi:
(a) Almeno due tra questi si verificano;
(b) Esattamente due tra questi si verificano;
(c) Al più due tra questi si verificano;
(d) Esattamente uno tra questi si verifica.

Esercizio 1.4. Un'urna contiene oggetti di tipo A ed oggetti di tipo B; si eseguono due successive estrazioni dall'urna e si definiscono, per $i = 1, 2$, gli eventi:

$$A_i = \{\textit{oggetto di tipo } \mathrm{A} \textit{ alla } i\textit{-esima estrazione}\}.$$

In termini di operazioni booleane su A_1 e A_2, scrivere l'espressione per l'evento

$$\{\textit{gli oggetti risultanti dalle due successive estrazioni sono dello stesso tipo}\}.$$

Esercizio 1.5. Un'urna contiene esattamente quattro elementi di tipo A e tre elementi di tipo B; da tale urna si effettuano tre successive estrazioni senza reinserimento, registrando il tipo dell'elemento via via estratto.
(a) Elencare gli eventi elementari in questo esperimento e contate quanti sono.
(b) Quali e quanti sono, fra tali eventi elementari, quelli che realizzano l'evento

$$\{\textit{almeno due elementi di tipo } \mathrm{B} \textit{ fra i tre elementi estratti}\}?$$

(c) Quali e quanti sono, fra tali eventi elementari, quelli che realizzano l'evento

$$\{\textit{almeno due elementi di tipo } \mathrm{B}\} \cup \{\textit{l'elemento estratto alla seconda estrazione è di tipo } \mathrm{B}\}$$

Esercizio 1.6. Consideriamo di nuovo l'urna del precedente Esercizio 1.5. Si effettuano sette successive estrazioni senza reinserimento (cioè l'urna viene progressivamente svuotata), registrando anche in questo caso soltanto il tipo dell'elemento via via estratto (tutti gli elementi di tipo A sono indistinguibili fra di loro, e tutti gli elementi di tipo B sono indistinguibili fra di loro).
Elencare gli eventi elementari in questo esperimento e contare quanti sono.

Esercizio 1.7. Consideriamo di nuovo l'urna del precedente Esercizio 1.5. Questa volta però le tre estrazioni sono effettuate con reinserimento.
(a) Elencare anche in questo caso gli eventi elementari.
(b) Dove risiede la differenza fra le due situazioni di estrazioni con e senza reinserimento?

Per rispondere alla domanda (b) *del precendente Esercizio 1.7 e alle domande dei successivi esercizi servono degli elementi di calcolo combinatorio. Tali argomenti saranno affrontati nella Lezione 3, inoltre i problemi relativi alle estrazioni da urne saranno esaminati in generale nella Lezione 6*

Esercizio 1.8. Consideriamo ora il caso in cui vengono effettuate sette estrazioni con reinserimento sempre dall'urna del precedente Esercizio 1.5. Quanti sono gli elementi elementari?

Esercizio 1.9. Una moneta viene lanciata due volte, registrando ogni volta se il risultato sia stato testa o croce.
(a) Elencare gli eventi elementari possibili, in questo esperimento, e contare quanto vale $|\Omega|$, la cardinalità di Ω.
(b) Qual è la cardinalità di $\mathscr{P}(\Omega)$, l'insieme delle parti di Ω? Cioè, quanti sono in tutto gli eventi, contando sia quelli semplici, quelli composti e quelli "banali" $\emptyset$ e Ω?

Esercizio 1.10. *Sia* $\Omega = \{1, 2, 3, 4, 5, 6\}$ *e siano* $H_1 = \{1, 5\}$, $H_2 = \{2, 3, 4\}$ *e* $H_3 = \{3, 6\}$ *tre sottoinsiemi che identifichiamo con tre eventi.*
(a) Gli eventi H_1, H_2 *ed* H_3 *sono esaustivi?*
(b) Gli eventi H_1, H_2 *ed* H_3 *sono incompatibili a due a due?*
(c) Gli eventi H_1, H_2 *ed* H_3 *formano una partizione (dell'evento certo)?*

Proprietà delle operazioni booleane

Proprietà distributiva dell'unione rispetto all'intersezione

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

Iniziamo dimostrando che $(A \cup B) \cap C \subset (A \cap C) \cup (B \cap C)$.
Sia ω un elemento di $(A \cup B) \cap C$, ossia ω è sicuramente un elemento di C e inoltre ω è un elemento di A oppure ω è un elemento di B, in formule $\omega \in C$ e [$\omega \in A$ oppure $\omega \in B$]: se $\omega \in A$ allora $\omega \in A \cap C$ e quindi $\omega \in (A \cap C) \cup (B \cap C)$; analogamente se $\omega \in B$ allora $\omega \in B \cap C$ e quindi $\omega \in (A \cap C) \cup (B \cap C)$.
Passiamo ora a $(A \cap C) \cup (B \cap C) \subset (A \cup B) \cap C$.
Sia ω un elemento di $(A \cap C) \cup (B \cap C)$, ossia ω è un elemento di $A \cap C$ oppure è un elemento di $B \cap C$. Nel primo caso, è un elemento di A, oppure, nel secondo caso, è un elemento di B, e quindi appartiene ad almeno uno tra A e B, ossia appartiene alla loro unione $A \cup B$; inoltre, in entrambi i casi ω è un elemento di C, e quindi appartiene ad $(A \cup B) \cap C$.
La proprietà distributiva dell'intersezione rispetto all'unione, ossia $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ si dimostra in modo analogo.
Legge di De Morgan:

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \quad \text{o equivalentemente} \quad A \cap B = \overline{(\overline{A} \cup \overline{B})}.$$

Affermare che ω non è un elemento di $A \cap B$, significa affermare che non può essere contemporaneamente un elemento di A e un elemento di B, ossia ω non appartiene ad A oppure non appartiene a B, ossia [$\omega \in \overline{A}$ oppure $\omega \in \overline{B}$], cioè $\omega \in \overline{A} \cup \overline{B}$, e abbiamo dimostrato che $\overline{A \cap B} \subset \overline{A} \cup \overline{B}$.

Viceversa se $\omega \in \overline{A} \cup \overline{B}$, allora [$\omega \in \overline{A}$ oppure $\omega \in \overline{B}$], e ciò implica che ω non può appartenere sia ad A che a B, ossia che ω non è un elemento di $A \cap B$, ovvero è un elemento del complementare di $A \cap B$. Abbiamo così dimostrato che $\overline{A} \cup \overline{B} \subset \overline{A \cap B}$.

L'equivalenza tra le due formulazioni dipende dal fatto che il complementare del complementare di un insieme E è l'insieme E stesso: $\overline{\overline{E}} = E$.

Quest'ultima proprietà discende dal principio del *tertium non datur* (ovvero *non è ammessa una terza possibilità*) per la quale [o $\omega \in E$ o $\omega \notin E$]. Di conseguenza se non vale che $\omega \in \overline{E}$ allora necessariamente deve valere $\omega \in E$ (ovvero $\overline{\overline{E}} \subset E$) mentre se $\omega \in E$ allora non può valere che $\omega \in \overline{E}$ (ovvero $E \subset \overline{\overline{E}}$).

Differenza e differenza simmetrica fra due insiemi

Dati due insiemi A e B si definisce la *differenza tra* A *e* B come l'insieme degli elementi di A che non appartengono a B

$$A \setminus B = A \setminus (A \cap B) := \{\omega \in A : \ \omega \notin B\}$$

e che si definisce la *differenza tra* A *e* B come l'insieme degli elementi che appartengono ad almeno uno tra A e B, ma non appartengono ad entrambi, ossia gli elementi di $A \cup B$ che non appartengono a $A \cap B$:

$$A \Delta B = (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A) = \{\omega \in A \cup B : \ \omega \notin A \cap B\}$$

Infine ricordiamo che scrivere $[o \ A \ o \ B]$ è diverso da scrivere $[A \ o \ B]$:
$[o \ A \ o \ B] = [aut \ A \ aut \ B] = A \Delta B = (A \cup B) \setminus A \cap B$, mentre $[A \ o \ B] = A \cup B$.

1.4 Sintesi della Lezione 1

Gli eventi vengono espressi di solito con proposizioni che possono essere vere o no, si possono verificare o no. Ad esempio "esce 1 nel lancio di un dado". Le relazioni tra questi sono definite dalle operazioni logiche AND, OR e NOT (anche denotate rispettivamente come $\wedge$, $\vee$, $\neg$). Nel caso in cui siamo interessati a eventi che riguardano una particolare situazione (come il lancio di un dado) ci sono degli eventi (detti eventi elementari) che non si possono scrivere come al somma logica di altri due eventi si usa descrivere gli eventi di interesse tramite un insieme[2], solitamente denotato con la lettera greca Ω (Omega maiuscola, i cui elementi sono solitamente denotati[3] con la lettera greca ω (omega minuscolo).
Ad ogni evento si fa corrispondere un sottoinsieme di Ω e le operazioni logiche diventano operazioni booleane su insiemi secondo la seguente corrispondenza $AND \leftrightarrow \cap$, $OR \leftrightarrow \cup$, $\neg \leftrightarrow$ "passaggio al complementare":

$$A \vee B \leftrightarrow A \cap B, \quad A \wedge B \leftrightarrow A \cup B, \quad \neg A \leftrightarrow \overline{A} = \Omega \setminus A = A^c.$$

Ad esempio $\Omega = \{1, 2, 3, 4, 5, 6\}$, 4 è un elemento di Ω e l'evento "esce un numero pari" è rappresentato dal sottoinsieme $\{2, 4, 6\}$, mentre "esce un numero minore di 5" è rappresentato dal sottoinsieme $\{1, 2, 3, 4\}$. L'evento ("esce un numero pari" OR "esce un numero minore di 5") è rappresentato dal sottoinsieme $\{2, 4, 6\} \cup \{1, 2, 3, 4\} = \{1, 2, 3, 4, 6\}$, etc.
Anche se concettualmente un evento è diverso da un insieme nel seguito useremo la stessa lettera[4] per denotare sia l'evento che il sottoinsieme che lo rappresenta. Sappiamo che si può verificare uno e uno solo degli eventi rappresentati dai singoletti $\{\omega\}$. Inoltre, si dice che un evento E si è verificato se e solo se si è verificato uno degli elementi appartenenti ad E.
Ad esempio si dice *si è verificato l'evento "esce un numero pari"* significa che si è verificato uno degli eventi elementari $\{2\}$, oppure $\{4\}$, oppure $\{6\}$, ovvero si è verificato un $\omega \in \{2, 4, 6\}$, ma non sappiamo quale.

Oltre alla corrispondenza tra le operazioni logiche tra proposizioni e le operazioni booleane, c'è anche una corrispondenza biunivoca tra "A implica B" e $A \subseteq B$.

Altro concetto fondamentale è il concetto di **partizione** (di Ω): una partizione è una famiglia di eventi esaustivi e incompatibili a due a due. Gli eventi di una famiglia sono esaustivi se e solo se sicuramente se ne verifica almeno uno, mentre sono incompatibili a due a due se e solo se comunque presi due eventi della famiglia, questi due eventi non si possono verificare contemporaneamente. In termini degli insiemi che li rappresentano, la proprietà di essere esaustivi significa che la loro unione coincide con Ω, mentre la proprietà di essere incompatibili a due a due, significa che comunque se prendono due, questi due insiemi hanno intersezione vuota.

Infine sono fondamentali le Leggi di De Morgan: La negazione di si verificano tuttli gli eventi A_i, con $i \in I$, è almeno uno degli eventi A_i non si verifica, ossia si verifica almeno uno tra gli eventi "Negazione di A_i" (ossia in termini di insiemi, l'insieme complementare di A_i denotato da $\overline{A}_i = A_i^c$. In formule con la notazione degli insiemi e delle operazioni booleane

$$\overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A}_i \quad \text{o equivalentemente} \quad \bigcap_{i \in I} \overline{A_i} \bigcap_{i \in I} \overline{A}_i = \bigcup_{i \in I} A_i$$

La seconda formulazione si legge (in italiano) "*non si verifica nessuno degli* A_i (o anche *nessun insuccesso*) coincide con *non è vero che si verifica almeno uno degli* "A_i (o anche *non è vero che si ha almeno un successo*).

[2] Un modo per ottenere l'insieme Ω è quello di considerare l'insieme i cui elementi sono gli eventi elementari. Ricordiamo che un evento E è elementare se non è possibile trovare altri due eventi E_1 ed E_2 diversi tra loro e da E per i quali $E = E_1 \vee E_2$. Tuttavia ci possono essere anche altre rappresentazioni.

[3] Attenzione a non confondere la lettera greca ω (omega minuscolo) e w (doppia v).

[4] Anche se non è una regola scritta, in genere per gli eventi si usano le lettere in STAMPATELLO MAIUSCOLO. Inoltre in genere si usano le prime lettere dell'alfabeto latino, quindi A, B, C, ..., G, H, mentre si evita di usare le ultime lettere dell'alfabeto, che come vedremo sono usate in genere per denotare le variabili aleatorie.

2 Spazi finiti di probabilità

2.1 Prime definizioni e proprietà

Introduciamo ora il concetto di **probabilità**. Come vedremo, tale concetto permette di formalizzare il problema di esprimere uno stato di incertezza circa il modo di risultare di un esperimento aleatorio

Sia dato uno spazio campione Ω e sia $\mathcal{P}(\Omega)$ la famiglia delle sue parti. In linea con la precedente Lezione 1, assumiamo che Ω sia un insieme finito,

Definizione 2.1. (provvisoria) *Una* **misura di probabilità** *o, più semplicemente, una* **probabilità** *su* $(\Omega, \mathcal{P}(\Omega))$ *è una funzione che soddisfa i seguenti assiomi*[5]

i) $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$

ii) $\mathbb{P}(\Omega) = 1$ *(***condizione di normalizzazione***)*

iii) se $E_1 \cap E_2 = \emptyset$ *allora* $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$, *(***proprietà di additività***).*

Definizione 2.2. *Uno* **spazio finito di probabilità** *è una terna* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *dove* Ω *è un insieme finito,* $\mathcal{P}(\Omega)$ *è la famiglia delle parti di* Ω *e* $\mathbb{P}$ *è una misura di probabilità su* $(\Omega, \mathcal{P}(\Omega))$ *.*

***Osservazione* 2.1.** Prima di proseguire è opportuno citare il fatto che esistono diverse possibili *interpretazioni* del termine probabilità: ad esempio probabilità *classiche*, *frequentistiche*, *soggettivistiche, etc...*

Non rientra nei nostri scopi soffermarci sul significato e la portata di tali interpretazioni; per quanto ci riguarda ci basta accennare al fatto che, all'interno di ciascuna di dette interpretazioni, è giustificato imporre che la probabilità soddisfi le condizioni *i), ii), iii)* della Definizione 2.1.

Su tale base possiamo imporre tali condizioni come ***assiomi*** e procedere in modo appunto ***assiomatico***; e di tali assiomi vedremo fra poco alcune conseguenze immediate.

Vediamo prima alcuni esempi di probabilità tramite due esercizi: nel primo vediamo un caso particolare di probabilità classica, mentre nel secondo vedremo un esempio di una funzione che è effettivamente una probabilità e un esempio di una funzione che invece non lo è .

Esercizio proposto 2.1. *Pensiamo all'esperimento del lancio di un dado a sei facce con*

$$\Omega = \{\omega_1, ..., \omega_6\},$$

dove ω_i *rappresenta l'evento elementare il dado mostra la faccia* i*, ovvero indicando con* X *il valore della faccia* $\{\omega_i\} = \{X = i\}$*.*

Poniamo su Ω *la probabilità classica (numero dei casi favorevoli diviso il numero dei casi possibili), ossia*

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{6}.$$

Verificare inoltre che $\mathbb{P}(\cdot)$ *soddisfa gli assiomi* $i)$, $ii)$, $iii)$ *e calcolare*

$$\mathbb{P}(\{X \leq 2\} \cup \{X \geq 5\}), \quad \mathbb{P}(\{X \geq 3\} \cap \{X \leq 4\}).$$

Chiaramente basta rappresentare gli eventi precedenti come sottoinsiemi di Ω *e calcolarne la cardinalità; ad esempio* $\{X \leq 2\} \cup \{X \geq 5\}) = \{\omega_1, \omega_2\} \cup \{\omega_5, \omega_6\}$

[5] Su alcuni testi la proprietà *i)* è sostituita dalla **proprietà di non negatività**, ossia:
i') $\mathbb{P} : \mathcal{P}(\Omega) \to \mathbb{R}$, con la proprietà che $\mathbb{P}(E) \geq 0$ per ogni $E \in \mathcal{P}(\Omega)$.
È facile vedere che le proprietà *i'), ii)* e *iii) implicano che* $\mathbb{P}(E) \in [0, 1]$.

Esempio* 2.1. Mostrare che tra le due seguenti funzioni $\mathbb{P}_1$ e $\mathbb{P}_2$, solo $\mathbb{P}_1$ è una (misura di) probabilità su $\Omega, \mathscr{P}(\Omega)$, con $\Omega = \{a, b, c, d\}$:

$$\begin{aligned}
\mathbb{P}_1 :\ \mathscr{P}(\Omega) &\longrightarrow [0,1]\\
\emptyset &\mapsto \mathbb{P}_1(\emptyset) = 0\\
\{a\} &\mapsto \mathbb{P}_1(\{a\}) = 1/10\\
\{b\} &\mapsto \mathbb{P}_1(\{b\}) = 2/10\\
\{c\} &\mapsto \mathbb{P}_1(\{c\}) = 3/10\\
\{d\} &\mapsto \mathbb{P}_1(\{d\}) = 4/10\\
\{a,b\} &\mapsto \mathbb{P}_1(\{a,b\}) = 3/10\\
\{a,c\} &\mapsto \mathbb{P}_1(\{a,c\}) = 4/10\\
\{a,d\} &\mapsto \mathbb{P}_1(\{a,d\}) = 5/10\\
\{b,c\} &\mapsto \mathbb{P}_1(\{b,c\}) = 5/10\\
\{b,d\} &\mapsto \mathbb{P}_1(\{b,d\}) = 6/10\\
\{c,d\} &\mapsto \mathbb{P}_1(\{c,d\}) = 7/10\\
\{a,b,c\} &\mapsto \mathbb{P}_1(\{a,b,c\}) = 6/10\\
\{a,b,d\} &\mapsto \mathbb{P}_1(\{a,b,d\}) = 7/10\\
\{a,c,d\} &\mapsto \mathbb{P}_1(\{a,c,d\}) = 8/10\\
\{b,c,d\} &\mapsto \mathbb{P}_1(\{b,c,d\}) = 9/10\\
\{a,b,c,d\} &\mapsto \mathbb{P}_1(\{a,b,c,d\}) = 1
\end{aligned}
\qquad
\begin{aligned}
\mathbb{P}_2 :\ \mathscr{P}(\Omega) &\longrightarrow [0,1]\\
\emptyset &\mapsto \mathbb{P}_2(\emptyset) = 0\\
\{a\} &\mapsto \mathbb{P}_2(\{a\}) = 1/6\\
\{b\} &\mapsto \mathbb{P}_2(\{b\}) = 1/6\\
\{c\} &\mapsto \mathbb{P}_2(\{c\}) = 1/6\\
\{d\} &\mapsto \mathbb{P}_2(\{d\}) = 1/3\\
\{a,b\} &\mapsto \mathbb{P}_2(\{a,b\}) = 1/3\\
\{a,c\} &\mapsto \mathbb{P}_2(\{a,c\}) = 1/3\\
\{a,d\} &\mapsto \mathbb{P}_2(\{a,d\}) = 1/2\\
\{b,c\} &\mapsto \mathbb{P}_2(\{b,c\}) = 1/3\\
\{b,d\} &\mapsto \mathbb{P}_2(\{b,d\}) = 1/2\\
\{c,d\} &\mapsto \mathbb{P}_2(\{c,d\}) = 1/2\\
\{a,b,c\} &\mapsto \mathbb{P}_2(\{a,b,c\}) = 1/2\\
\{a,b,d\} &\mapsto \mathbb{P}_2(\{a,b,d\}) = 2/3\\
\{a,c,d\} &\mapsto \mathbb{P}_2(\{a,c,d\}) = 2/3\\
\{b,c,d\} &\mapsto \mathbb{P}_2(\{b,c,d\}) = 2/3\\
\{a,b,c,d\} &\mapsto \mathbb{P}_2(\{a,b,c,d\}) = 1
\end{aligned}$$

*È immediato verificare che $\mathbb{P}_2$ **non è una probabilità**: infatti, anche se le condizioni $i)$ e $ii)$ sono verificate, in quanto $\mathbb{P}_2(E) \in [0,1]$ per ogni $E \subset \Omega$ e $\mathbb{P}_2(\Omega) = \mathbb{P}_2(\{a,b,c,d\}) = 1$, tuttavia non vale la proprietà di additività $iii)$: a questo scopo basta trovare un evento composto $E = E_1 \cup E_2$, con $E_1 \cap E_2 = \emptyset$ e tale che $\mathbb{P}_2(E) = \mathbb{P}(E_1 \cup E_2) \neq \mathbb{P}(E_1) + \mathbb{P}(E_2)$.*
Osserviamo infatti che, ad esempio, $\{a,b,c,d\} = \{a,b,c\} \cup \{d\}$, gli eventi $\{a,b,c\}$ e $\{d\}$ sono eventi incompatibili: quindi, se $\mathbb{P}_2$ fosse una probabilità allora si dovrebbe avere $\mathbb{P}_2(\{a,b,c,d\}) = \mathbb{P}_2(\{a,b,c\}) + \mathbb{P}(\{d\})$, ma invece

$$1 = \mathbb{P}_2(\{a,b,c,d\}) \neq \mathbb{P}_2(\{a,b,c\}) + \mathbb{P}(\{d\}) = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}.$$

Invece per stabilire che $\mathbb{P}_1$ è effettivamente una (misura di) probabilità, dobbiamo necessariamente controllare che vale la proprietà di additività per i seguenti insiemi

$$\begin{aligned}
&\{a,b\} = \{a\} \cup \{b\},\ \{a,c\} = \{a\} \cup \{c\},\ \{a,d\} = \{a\} \cup \{d\},\\
&\{b,c\} = \{b\} \cup \{c\},\ \{b,d\} = \{b\} \cup \{d\},\ \{c,d\} = \{c\} \cup \{d\},\\
&\{a,b,c\} = \{a,b\} \cup \{c\} = \{a,c\} \cup \{b\} = \{b,c\} \cup \{a\},\\
&\{a,b,d\} = \{a,b\} \cup \{d\} = \{a,d\} \cup \{b\} = \{b,d\} \cup \{a\}\\
&\{a,c,d\} = \{a,c\} \cup \{d\} = \{a,d\} \cup \{c\} = \{c,d\} \cup \{a\}\\
&\{b,c,d\} = \{b,c\} \cup \{d\} = \{b,d\} \cup \{c\} = \{c,d\} \cup \{b\}\\
&\{a,b,c,d\} = \{a,b,c\} \cup \{d\} = \{a,b,d\} \cup \{c\} = \{a,c,d\} \cup \{b\} = \{b,c,d\} \cup \{a\}\\
&\qquad\qquad \{a,b\} \cup \{c,d\} = \{a,c\} \cup \{b,d\} = \{a,d\} \cup \{b,c\}
\end{aligned}$$

In altre parole dobbiamo verificare che

$$\begin{aligned}
&\mathbb{P}_1(\{a,b\}) = \mathbb{P}_1(\{a\} \cup \{b\}) = \mathbb{P}_1(\{a\}) + \mathbb{P}_1(\{b\}),\\
&\mathbb{P}_1(\{a,c\}) = \mathbb{P}_1(\{a\} \cup \{c\}) = \mathbb{P}_1(\{a\}) + \mathbb{P}(\{c\}),\\
&\mathbb{P}_1(\{a,d\}) = \mathbb{P}_1(\{a\} \cup \{d\}) = \mathbb{P}_1(\{a\}) + \mathbb{P}(\{d\}),
\end{aligned}$$

e così via. Lasciamo al lettore tutte le altre verifiche.

Dal precedente Esempio 2.1 il lettore potrebbe pensare che sia veramente lungo e complicato verificare se una funzione $\mathbb{P}$ soddisfa la proprietà di additività, come vedremo c'è un modo più semplice per ottenere che è effettivamente una probabilità e anche più in generale di costruire una probabilità nel caso di spazi campione Ω finiti (si veda la successiva Proposizione 2.1).

Come già ricordato, in quanto segue consideriamo ancora il caso di spazio campione Ω finito e, se N è la cardinalità di Ω, poniamo

$$\Omega = \{\omega_1, ..., \omega_N\}.$$

Elenchiamo ora alcune proprietà della probabilità, che risultano conseguenze immediate degli assiomi $i)$, $ii)$, $iii)$ della Definizione 2.1.

Prima di tutto notiamo che la proprietà *iii)* di additività si generalizza al caso di n eventi disgiunti a due a due

iii') **(proprietà di additività finita)** *Siano $E_1, ..., E_n \in \mathcal{P}(\Omega)$ disgiunti (o incompatibili) a due a due, ovvero tali che*

$$E_i \cap E_j = \emptyset, \quad \text{per } i,j \in \{1,2,\cdots,n\}, \text{ con } i \neq j; \tag{2}$$

allora si ha

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{i=1}^{n} \mathbb{P}(E_i). \tag{3}$$

(La dimostrazione si ottiene facilmente per induzione su n.)

Proprietà di additività per $n = 3$ eventi

Se E_1, E_2, E_3 sono eventi disgiunti a due a due, ossia $E_1 \cap E_2 = \emptyset$, $E_1 \cap E_3 = \emptyset$, $E_2 \cap E_3 = \emptyset$, allora

$$E_1 \cup E_2 \cup E_3 = (E_1 \cup E_2) \cup E_3, \quad \text{e} \quad (E_1 \cup E_2) \cap E_3 = \underbrace{(E_1 \cap E_3)}_{=\emptyset} \cup \underbrace{(E_2 \cap E_3)}_{=\emptyset} = \emptyset$$

e quindi per la proprietà di additività, applicata ai due eventi disgiunti $E_1 \cup E_2$ e E_3 si ottiene

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}\big((E_1 \cup E_2) \cup E_3\big) = \mathbb{P}(E_1 \cup E_2) + \mathbb{P}(E_3)$$

e applicando di nuovo la proprietà di additività ai due eventi disgiunti E_1 ed E_2 si ottiene che

$$\mathbb{P}(E_1 \cup E_2 \cup E_3) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \mathbb{P}(E_3)$$

Le relazioni elencate qui di seguito costituiscono delle immediate conseguenze degli assiomi della probabilità: si consiglia vivamente il lettore di provare a dimostrarle autonomamente, prima di vedere le note alle pagine 15 e 16.

(a) Ponendo

$$p(\omega_i) = \mathbb{P}(\{\omega_i\}), \qquad i = 1, ..., N,$$

per ogni $E \in \mathcal{P}(\Omega)$ risulta[6]

$$\mathbb{P}(E) = \sum_{\substack{i=1 \\ i:\,\omega_i \in E}}^{N} p(\omega_i) = \sum_{\omega \in E} p(\omega). \tag{4}$$

[6] La somma $\sum_{\omega_i \in E} p(\omega_i)$ va intesa come la somma sugli indici $i \in \{1, \dots, N\}$ tali che $\omega_i \in E$.

(b) Per ogni $E \in \mathcal{P}(\Omega)$ risulta

$$\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E), \tag{5}$$

dove ricordiamo che $\overline{E}$ indica il complementare (o la "negazione") di un evento $E \in \mathcal{P}(\Omega)$.

(c) L'evento impossibile ha probabilità nulla, ovvero

$$\mathbb{P}(\emptyset) = 0 \tag{6}$$

(d) (proprietà di monotonia) Siano $A, B \in \mathcal{P}(\Omega)$ tali che $A \subseteq B$; allora risulta

$$\mathbb{P}(A) \leq \mathbb{P}(B). \tag{7}$$

Inoltre, ricordando che $B \setminus A = B \cap \overline{A}$, e che $A \subseteq B$ risulta

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)$$

(e) (formula di inclusione ed esclusione per due eventi) Per due arbitrari $A, B \in \mathcal{P}(\Omega)$ risulta

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{8}$$

(f) Siano $H_1, ..., H_n \in \mathcal{P}(\Omega)$ tali che

$$\bigcup_{i=1}^{n} H_i = \Omega, \quad \text{e} \quad H_i \cap H_j = \emptyset \quad \text{per } i \neq j, \tag{9}$$

ossia la collezione $H_1, ..., H_n$ è una partizione dell'evento certo; allora si ha la condizione

$$\sum_{i=1}^{n} \mathbb{P}(H_i) = 1. \tag{10}$$

In particolare, ricordando che $p(\omega_i) = \mathbb{P}(\{\omega_i\})$, $i = 1, ..., N$, e prendendo $H_i = \{\omega_i\}$, deve risultare

$$p(\omega_i) \geq 0 \qquad \text{e} \qquad \sum_{i=1}^{N} p(\omega_i) = 1. \tag{11}$$

Ulteriori conseguenze degli assiomi della probabilità verranno viste in seguito, dopo aver introdotto il concetto di probabilità condizionata (si veda la Lezione 4).

Le precedenti proprietà **(a)-(f)** si possono dimostrare a partire dagli assiomi della Definizione 2.1 di probabilità in uno spazio finito.

Una volta dimostrata la (2.3), ossia la proprietà **(a)**, le altre proprietà **(b)-(f)** possono essere verificate a partire da tale proprietà (lasciamo questa verifica al lettore come esercizio).

Dimostrazione della proprietà (a) con la Definizione 2.1

Chiaramente $E = \cup_{\omega \in E}\{\omega\}$, ovvero, se indichiamo con $n = |E|$, e con $\{j_1, j_2, \ldots, j_n\}$ il sottoinsieme degli indici $\{1, 2, \ldots, N\}$, tali che

$$E = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_n}\},$$

si ha

$$E = \bigcup_{k=1}^{n} \{\omega_{j_k}\} = \bigcup_{k=1}^{n} E_k.$$

dove gli insiemi $E_k := \{\omega_{j_k}\}$ sono chiaramente incompatibili a due a due e quindi dalla proprietà di additività finita *iii')*, ed in particolare dalla (3), si ottiene la (2.3), ossia che

$$\mathbb{P}(E) = \sum_{k=1}^{n} \mathbb{P}(E_k) = \sum_{k=1}^{n} \mathbb{P}\left(\{\omega_{j_k}\}\right).$$

Tuttavia è importante osservare che le proprietà **(b)-(f)** si possono ottenere anche dalla seguente relazione

$$\text{per ogni } A \text{ e } B \text{ si ha} \qquad \boxed{\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}),} \tag{12}$$

la cui dimostrazione dipende solo dall'osservazione che

$$A = (A \cap B) \cup (A \cap \overline{B}),$$

che gli eventi $A \cap B$ e $A \cap \overline{B}$ sono incompatibili (ossia la loro intersezione è l'insieme vuoto) e infine dalla proprietà *iii)* di additività per una probabilità.
Ovviamente vale anche una formula analoga in cui si scambia il ruolo di A e di B:

$$\boxed{\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap \overline{A}) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B).}$$

.

Diamo qui la dimostrazione di $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$: infatti $A = (A \cap B) \cup (A \cap \overline{B})$

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

dove nell'ultima uguaglianza abbiamo usato la proprietà distributiva dell'unione rispetto all'intersezione. Inoltre è immediato verificare che $A \cap B$ e $A \cap \overline{B}$ sono disgiunti

$$(A \cap B) \cap (A \cap \overline{B}) = (A \cap B) \cap (A \cap \overline{B}) = A \cap B \cap \overline{B} = A \cap \emptyset = \emptyset$$

Quindi per la proprietà di additività

$$\mathbb{P}(A) = \mathbb{P}\left((A \cap B) \cap (A \cap \overline{B})\right) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$$

Dimostrazioni dei punti (b)–(f) attraverso la (12)

La proprietà **(b)** che $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$ si ottiene prendendo $A = \Omega$ e $B = E$ per cui $A \cap B = \Omega \cap E = E$ e $A \cap \overline{B} = \Omega \cap \overline{E} = \overline{E}$, da cui, per la proprietà (12) cioè $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$, si ha

$$\underbrace{\mathbb{P}(\Omega)}_{\overset{ii)}{=}1} \overset{(12)}{=} \mathbb{P}(\Omega \cap E) + \mathbb{P}(\Omega \cap \overline{E}) = \mathbb{P}(E) + \mathbb{P}(\overline{E}) \overset{ii)}{\Longleftrightarrow} 1 = \mathbb{P}(E) + \mathbb{P}(\overline{E})$$

che è equivalente alla proprietà **(b)**.

La proprietà **(c)** che $\mathbb{P}(\emptyset) = 0$, a sua volta si può derivare dalla **(b)** osservando che l'evento impossibile è il complementare dell'evento certo, ossia $\emptyset = \overline{\Omega}$, e l'assioma $ii)$ delle probabilità, ossia la proprietà di normalizzazione $\mathbb{P}(\Omega) = 1$: $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.
oppure
si può dimostrare osservando che, $\emptyset = \emptyset \cup \emptyset$ e che, ovviamente, $\emptyset \cap \emptyset = \emptyset$, per cui dall'assioma $iii)$ delle probabilità, ossia la proprietà di additività (Definizione 2.1),

$$\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = \mathbb{P}(\emptyset) + \mathbb{P}(\emptyset),$$

da cui si ottiene immediatamente che $\mathbb{P}(\emptyset) = 0$.

La proprietà **(d)** che se $A \subset B$ allora $\mathbb{P}(A) \leq \mathbb{P}(B)$ (monotonia della probabilità) invece deriva dall'osservare che, se $A \subset B$, allora $A \cap B = A$, e quindi $\mathbb{P}(A) = \mathbb{P}(A \cap B)$ e che, scambiando il ruolo di A e B nella (12), si ha

$$\mathbb{P}(B) = \mathbb{P}(\underbrace{B \cap A}_{=A}) + \mathbb{P}(B \cap \overline{A}) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \cap \overline{A})}_{\geq 0} \geq \mathbb{P}(A).$$

Inoltre, dal fatto che, se $A \subset B$ allora $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A})$ e che $B \setminus A = B \cap \overline{A}$, si ottiene che,

$$\text{se } A \subset B \text{ allora} \quad \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

La proprietà **(e)** formula di inclusione-esclusione per due eventi) si può dimostrare a partire sempre dalla (12) (utilizzata anche scambiando il ruolo di A e B) e osservando che $A \cup B$ si scrive come l'unione di tre eventi incompatibili a due a due

$$A \cup B = (A \cap B) \cup (A \cap \overline{B}) \cup (B \cap \overline{A}),$$

e quindi

$$\mathbb{P}(A \cup B) = \underbrace{\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})}_{\mathbb{P}(A)} + \underbrace{\mathbb{P}(B \cap \overline{A})}_{\mathbb{P}(B) - \mathbb{P}(A \cap B)} = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

(Per la generalizzazione vi veda la nota di approfondimento a pagina 21)

Infine la proprietà **(f)** che, se H_1, H_2,..., H_m formano una partizione dell'evento certo, allora $\sum_{\ell=1}^{m} \mathbb{P}(H_\ell) = 1$ è immediata conseguenza degli assiomi, e precisamente dell'assioma $i)$ e dell'immediata conseguenza della proprietà $iii')$ di additività finita: infatti da una parte, essendo H_1, H_2,..., H_m incompatibili, ossia disgiunti a due a due, si ha che

$$\mathbb{P}\,\Big(\bigcup_{\ell=1}^{m} H_\ell\Big) = \sum_{\ell=1}^{m} \mathbb{P}(H_\ell)$$

dall'altra, essendo H_1, H_2,..., H_m esaustivi, ossia $\cup_{\ell=1}^{m} H_\ell = \Omega$, grazie all'assioma $ii)$, si ha che $\mathbb{P}\left(\cup_{\ell=1}^{m} H_\ell\right) = \mathbb{P}(\Omega) = 1$.

***Osservazione* 2.2.** Nel caso in cui Ω è un insieme finito, possiamo guardare alla probabilità nei due modi, apparentemente diversi ma sostanzialmente equivalenti, che verranno illustrati qui di seguito (teniamo presente il fatto che ciascun punto di Ω può essere visto come un particolare sottoinsieme, cioè come un sottoinsieme composto da un solo elemento) :

MODO 1) Prima definiamo $\mathbb{P}$ come una funzione di insieme, cioè

$$\mathbb{P}: \ \mathscr{P}(\Omega) \to [0,1]; \quad E \mapsto \mathbb{P}(E)$$

che soddisfi gli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1, e poi definiamo la funzione di punto

$$p: \ \Omega \to [0,1]; \quad \omega_i \mapsto p(\omega_i) := \mathbb{P}(\{\omega_i\}).$$

Questa funzione dovrà soddisfare le condizioni

$$p(\omega_i) \geq 0, \qquad \sum_{\omega\in\Omega} p(\omega) = 1. \quad \text{ovvero} \quad \sum_{i=1}^{N} p(\omega_i) = 1. \tag{13}$$

MODO 2) Prima definiamo una funzione di punto

$$p: \ \Omega \to [0,1]; \ \omega_i \to p(\omega_i),$$

che soddisfi le condizioni (13) e poi definiamo una funzione di insieme

$$E \mapsto \mathbb{P}(E) := \sum_{\omega\in E} p(\omega), \tag{14}$$

ossia attraverso la precedente formula (2.3).

Proposizione* 2.1.** *La funzione di insieme definita come nel* ***MODO 2) *qui sopra è una probabilità sull'insieme finito Ω, cioè soddisfa gli assiomi i), ii), iii) della Definizione 2.1.*

Esercizio proposto 2.2. *Dimostrare la precedente* ***Proposizione 2.1****.*

Esempio 2.2 (Un esempio concreto di spazio di probabilità finito)**.** *Sia $\Omega = \{a,b,c,d\}$ siano $p(a) = 1/8$, $p(b) = 1/4$, $p(c) = 1/2$ $p(d) = 1/8$, ossia*

$$\begin{aligned} p: \ \{a,b,c,d\} &\to [0,\infty) \\ a \mapsto p(a) &= \frac{1}{8} \\ b \mapsto p(b) &= \frac{1}{4} \\ c \mapsto p(c) &= \frac{1}{2} \\ d \mapsto p(d) &= \frac{1}{8} \end{aligned}$$

Chiaramente $p(a)$, $p(b)$, $p(c)$, $p(d) \geq 0$ e $p(a)+p(b)+p(c)+p(d) = 1$. Allora la probabilità definita sull'insieme delle parti di $\Omega = \{a,b,c,d\}$ dalla precedente formula 14, ossia $\mathbb{P}(E) := \sum_{\omega\in E} p(\omega)$, è data dalla funzione che è specificata nella seguente tabella.

Tabella della Probabilità definita nell'Esempio 2.2
(un esempio concreto di spazio di probabilità finito)

Sia $\Omega = \{a, b, c, d\}$ siano $p(a) = 1/8$, $p(b) = 1/4$, $p(c) = 1/2$ $p(d) = 1/8$, allora la seguente funzione è una probabilità:

$$\mathbb{P} : \mathscr{P}(\{a, b, c, d\}) \longrightarrow [0, 1], \quad E \mapsto \mathbb{P}(E) := \sum_{\omega \in E} p(\omega);$$

ovvero, in dettaglio, la seguente funzione $\mathbb{P}$ è una probablità

$$\begin{aligned}
\emptyset &\mapsto \mathbb{P}(\emptyset) = 0 \\
\{a\} &\mapsto \mathbb{P}(\{a\}) = p(a) = 1/8 \\
\{b\} &\mapsto \mathbb{P}(\{b\}) = p(b) = 1/4 \\
\{c\} &\mapsto \mathbb{P}(\{c\}) = p(c) = 1/2 \\
\{d\} &\mapsto \mathbb{P}(\{d\}) = p(d) = 1/8 \\
\{a, b\} &\mapsto \mathbb{P}(\{a, b\}) = p(a) + p(b) = 1/8 + 1/4 = 3/8 \\
\{a, c\} &\mapsto \mathbb{P}(\{a, c\}) = p(a) + p(c) = 1/8 + 1/2 = 5/8 \\
\{a, d\} &\mapsto \mathbb{P}(\{a, d\}) = p(a) + p(d) = 1/8 + 1/8 = 1/4 \\
\{a, b\} &\mapsto \mathbb{P}(\{b, c\}) = p(b) + p(c) = 1/4 + 1/2 = 3/4 \\
\{b, d\} &\mapsto \mathbb{P}(\{b, d\}) = p(b) + p(d) = 1/4 + 1/8 = 3/8 \\
\{c, d\} &\mapsto \mathbb{P}(\{c, d\}) = p(c) + p(d) = 1/2 + 1/8 = 5/8 \\
\{a, b, c\} &\mapsto \mathbb{P}(\{a, b, c\}) = p(a) + p(b) + p(c) = 1/8 + 1/4 + 1/2 = 7/8 \\
\{a, b, d\} &\mapsto \mathbb{P}(\{a, b, d\}) = p(a) + p(b) + p(d) = 1/8 + 1/4 + 1/8 = 1/2 \\
\{a, c, d\} &\mapsto \mathbb{P}(\{a, c, d\}) = p(a) + p(c) + p(d) = 1/8 + 1/2 + 1/8 = 3/4 \\
\{b, c, d\} &\mapsto \mathbb{P}(\{b, c, d\}) = p(b) + p(c) + p(d) = 1/4 + 1/2 + 1/8 = 7/8 \\
\{a, b, c, d\} &\mapsto \mathbb{P}(\{a, b, c, d\}) = p(a) + p(b) + p(c) + p(d) = 1/8 + 1/4 + 1/2 + 1/8 = 1
\end{aligned}$$

***Osservazione* 2.3 (Probabilità definite a meno di un fattore di proporzionalità).** Una misura di probabilità sullo spazio $\Omega = \{\omega_1, ..., \omega_N\}$ è individuata quando vengano assegnati i numeri $p(\omega_i) = \mathbb{P}(\{\omega_i\})$, per $i = 1, 2, ..., N$, soddisfacenti le condizioni (13). Supponiamo ora che $p(\omega_i)$, $i = 1, 2, ..., N$, siano assegnati a meno di una costante di proporzionalità; supponiamo cioè che esiste una costante positiva K e che siano assegnati dei numeri g_i ($i = 1, 2, ..., N$), tali che

$$\text{per ogni } i = 1, 2, ..., N, \text{ vale} \quad p(\omega_i) = K \cdot g_i \tag{15}$$

Sommando su $i = 1, 2, ..., N$, si ottiene $\sum_{i=1}^N p(\omega_i) = \sum_{i=1}^N K \cdot g_i = K \sum_{i=1}^N \cdot g_i$ e quindi, dalla condizione di normalizzazione (18), si ricava il valore della costante K, detta anche fattore di proporzionalità:

$$K = \frac{1}{\sum_{j=1}^N g_j}; \quad \Rightarrow \quad p(\omega_i) = \frac{g_i}{\sum_{j=1}^N g_j}.$$

Notiamo che si usa esprimere brevemente la condizione (15) usando il seguente simbolismo:

$$\boxed{p(\omega_i) \propto g_i} \quad \Big(\Leftrightarrow \quad \exists\, K \text{ tale che } \forall\, i = 1, 2, ..., N, \quad p(\omega_i) = K \cdot g_i \Big)$$

che si legge $p(\omega_i)$ è proporzionale a g_i.

Esempio* 2.3 (dado non equilibrato). *Un dado ha sei facce numerate da 1 a 6; esso è pesato in modo tale che ciascuna faccia abbia una probabilità di presentarsi (in un singolo lancio) proporzionale al suo valore. Sia*

$$A = \{\textit{si presenta un numero pari}\}.$$

Trovare $\mathbb{P}(A)$.

Soluzione: Si ha $\Omega = \{\omega_1, ..., \omega_6\}$ e vogliamo imporre

$$p(\omega_i) = K \cdot i, \qquad i = 1, ..., 6,$$

essendo K una costante positiva da determinare imponendo la condizione di normalizzazione (18). Si ottiene dunque

$$1 = p(\omega_1) + p(\omega_2) + p(\omega_3) + p(\omega_4) + p(\omega_5) + p(\omega_6) = K \cdot 1 + K \cdot 2 + K \cdot 3 + K \cdot 4 + K \cdot 5 + K \cdot 6 = K \cdot 21,$$

quindi $K = 1/21$,

$$p(\omega_i) = \frac{i}{21}, \quad i = 1, ..., 6,$$

e

$$\mathbb{P}(A) = p(\omega_2) + p(\omega_4) + p(\omega_6) = \frac{12}{21} = \frac{4}{7}.$$

□

Ricordiamo qui il fatto che in generale

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

La dimostrazione si ottiene facilmente per induzione, ma conviene pensare al seguente ragionamento:

$$\sum_{k=1}^{n} k = 1 + 2 + \cdots + (n-1) + n = n + (n-1) + \cdots + 2 + 1$$

e quindi, da una parte

$$\big[1 + 2 + \cdots + (n-1) + n\big] + \big[n + (n-1) + \cdots + 2 + 1\big] = 2 \sum_{k=1}^{n} k$$

dall'altra

$$\begin{array}{ccccccccccccc}
1 & + & 2 & + & \cdots & + & (n-1) & + & n & = & \sum_{k=1}^n k \\
+ & & + & & + & & + & & + & & + \\
n & + & (n-1) & + & \cdots & + & 2 & + & 1 & = & \sum_{k=1}^n k \\
\hline
(n+1) & + & (n+1) & + & \cdots & + & (n+1) & + & (n+1) & = & n(n+1)
\end{array}$$

da cui

$$2 \sum_{k=1}^{n} k = n(n+1).$$

2.2 Esercizi di verifica

Esercizio 2.1. Un dado è pesato in modo tale che la probabilità di avere un punto pari è il doppio della probabilità di avere un punto dispari. Qual è la probabilità di avere punto pari?

Esercizio 2.2. Siano A e B due eventi tali che

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap \overline{B}) = \mathbb{P}(\overline{A} \cap B) = \mathbb{P}(\overline{A} \cap \overline{B}).$$

(a) Quanto vale $\mathbb{P}(A \cap B)$?
(b) Qual è la probabilità che, fra A e B, se ne verifichi almeno uno?
(c) Qual è la probabilità che se ne verifichi esattamente uno?

Esercizio 2.3. Siano E_1 ed E_2 due eventi. Mostrare che la condizione $\mathbb{P}(E_1 \cap \overline{E}_2) = \mathbb{P}(\overline{E}_1 \cap E_2)$ implica $\mathbb{P}(E_1) = \mathbb{P}(E_2)$.
Per fissare le idee potete pensare che una moneta venga lanciata due volte e che

$$E_i = \{\textit{testa all'i-esimo lancio}\}, \quad i = 1, 2.$$

Supponete ora che da un'urna contenente 3 palline azzurre e 2 bianche vengono effettuate due estrazioni senza reinserimento e che

$$A_i = \{\textit{ all'i-sima estrazione esce una pallina azzurra}\}, \quad i = 1, 2.$$

Mostrare che $\mathbb{P}(A_1) = \mathbb{P}(A_2)$

Esercizio 2.4. Siano $A, B,$ e C tre eventi. Dopo aver dimostrato la formula (8) di inclusione-esclusione per due eventi, dimostrare che vale la formula di inclusione-esclusione per tre eventi

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

(questa formula costituisce un caso particolare della formula di inclusione - esclusione per n eventi: si veda la nota nella pagina seguente)

Esercizio 2.5. Siano $A, B,$ e C tre eventi tali che

$$\mathbb{P}\left(A \cap B \cap C\right) = \mathbb{P}\left(A \cap \overline{B} \cap C\right) = 0.1,$$

$$\mathbb{P}\left(\overline{A} \cap B \cap C\right) = \mathbb{P}\left(A \cap B \cap \overline{C}\right) = 0.15,$$

$$\mathbb{P}\left(A \cap \overline{B} \cap \overline{C}\right) = \mathbb{P}\left(\overline{A} \cap B \cap \overline{C}\right) = \mathbb{P}\left(\overline{A} \cap \overline{B} \cap C\right) = 0.05.$$

Calcolare
(a) $\mathbb{P}\left(A\right), \mathbb{P}\left(B\right), \mathbb{P}\left(C\right)$
(b) $\mathbb{P}\left(A \cup B\right), \mathbb{P}\left(A \cup C\right)$
(c) $\mathbb{P}\left(A \cup B \cup C\right).$

FORMULA DI INCLUSIONE-ESCLUSIONE

Dati n eventi A_1, A_2, A_n, la probabilità che se ne verifichi almeno uno tra tali eventi è data dalla seguente formula:

$$\begin{aligned}
&\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) \\
&- [\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \cdots + \mathbb{P}(A_i \cap A_j) + \cdots + \mathbb{P}(A_{n-1} \cap A_n)] \\
&+ \cdots \\
&+ (-1)^{k-1} \sum_{\{i_1, i_2, \cdots, i_k\}} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\
&+ \cdots \\
&+ (-1)^{n-1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1, i_2, \cdots, i_k\}} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}),
\end{aligned}$$

dove, per k generico, le somme sono estese a tutti i sottoinsiemi $\{i_1, i_2, \cdots, i_k\}$ di $\{1, 2, \cdots, n\}$, composti da k elementi (nella prossima Lezione 3, calcoleremo il numero complessivo di tali sottoinsiemi, al variare di n e di k).

Per il caso $n = 2$ basta osservare che, da una parte

$$A_1 \cup A_2 = (A_1 \cap \overline{A}_2) \cup (A_1 \cap A_2) \cup (\overline{A}_1 \cap A_2),$$

per cui $$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \cap \overline{A}_2) + \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(\overline{A}_1 \cap A_2),$$

e che, dall'altra parte,

$$A_1 = (A_1 \cap \overline{A}_2) \cup (A_1 \cap A_2) \quad \text{e} \quad A_2 = (A_1 \cap A_2) \cup (\overline{A}_1 \cap A_2),$$

per cui
$$\begin{aligned}
&\mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \\
&= \mathbb{P}(A_1 \cap \overline{A}_2) + \mathbb{P}(A_1 \cap A_2) + \cancel{\mathbb{P}(A_1 \cap A_2)} + \mathbb{P}(\overline{A}_1 \cap A_2) - \cancel{\mathbb{P}(A_1 \cap A_2)}
\end{aligned}$$

Per ottenere il caso $n = 3$, iniziamo osservando che

$$A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3 \text{ e che } (A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3).$$

Quindi, usando la formula di inclusione/esclusione per due eventi si ha

$$\begin{aligned}
&\mathbb{P}\left((A_1 \cup A_2) \cup A_3\right) = \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}\left((A_1 \cup A_2) \cap A_3\right) \\
&= \mathbb{P}(A_1 \cup A_2) + \mathbb{P}(A_3) - \mathbb{P}\left((A_1 \cap A_3) \cup (A_2 \cap A_3)\right) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_3) \\
&\quad - \left[\mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3) - \mathbb{P}\left(\underbrace{(A_1 \cap A_3) \cap (A_2 \cap A_3)}_{=A_1 \cap A_2 \cap A_3}\right)\right] \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) \\
&\quad + \mathbb{P}(A_1 \cap A_2 \cap A_3)
\end{aligned}$$

La dimostrazione di tale formula per ogni n si ottiene per induzione osservando che $A_1 \cup A_2 \cup \cdots \cup A_{n-1} \cup A_n = (A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cup A_n$, e che quindi

$$\begin{aligned}
&\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \\
&= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + \mathbb{P}(A_n) - \mathbb{P}\left((A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cap A_n\right) \\
&= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) + \mathbb{P}(A_n) \\
&\quad - \mathbb{P}\left((A_1 \cap A_n) \cup (A_2 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)\right)
\end{aligned}$$

ed infine utilizzando il passo induttivo per $n - 1$, anche per calcolare la probabilità di $(A_1 \cap A_n) \cup (A_2 \cap A_n) \cup \cdots \cup (A_{n-1} \cap A_n)$.

Dimostrazione della Proposizione 2.1

Si tratta di dimostrare che la funzione sull'insieme delle parti $\mathscr{P}(\Omega)$ definita come nella formula (2.3), soddisfa gli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1, e quindi è una (misura di) probabilità su $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$.
$i)$ si tratta di verificare che per ogni $E = \{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}\} \subset \Omega$ allora $\mathbb{P}(E) \in [0,1]$, e chiaramente

$$\mathbb{P}(E) \stackrel{(def)}{=} \sum_{h=1}^{m} p(\omega_{j_h}) = \sum_{\omega \in E} p(\omega) \geq 0$$

in quanto $p(\omega) \geq 0$ per ogni $\omega \in \Omega$ e anche $\mathbb{P}(E) \leq 1$, in quanto

$$\mathbb{P}(E) \stackrel{(def)}{=} \sum_{h=1}^{m} p(\omega_{j_h}) = \sum_{\omega \in E} p(\omega) \leq \sum_{\omega \in \Omega} p(\omega) = \sum_{i=1}^{N} p(\omega_i) = 1.$$

$ii)$ Si tratta di verificare la condizione di normalizzazione, ossia che $\mathbb{P}(\Omega) = 1$, e chiaramente

$$\mathbb{P}(\Omega) \stackrel{(def)}{=} \sum_{\omega \in \Omega} p(\omega) = \sum_{i=1}^{N} p(\omega_i) = 1.$$

$iii)$ Si tratta di verificare la proprietà di additività, ossia che per ogni $E = \{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}\} \subset \Omega$ ed $F = \{\omega_{k_1}, \omega_{k_2}, ..., \omega_{k_r}\} \subset \Omega$ tali che $E \cap F = \emptyset$ allora $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$, e infatti
se $E \cap F = \emptyset$ allora

$$E \cup F = \overbrace{\{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}\}}^{=E} \cup \overbrace{\{\omega_{k_1}, \omega_{k_2}, ..., \omega_{k_r}\}}^{=F} = \{\omega_{j_1}, \omega_{j_2}, ..., \omega_{j_m}, \omega_{k_1}, \omega_{k_2}, ..., \omega_{k_r}\}$$

e quindi

$$\mathbb{P}(E \cup F) = \overbrace{p(\omega_{j_1}) + p(\omega_{j_2}) + ... + p(\omega_{j_m})}^{=\mathbb{P}(E)} + \overbrace{p(\omega_{k_1}) + p(\omega_{k_2}) + ... + p(\omega_{k_r})}^{=\mathbb{P}(F)} =$$
$$= \sum_{h=1}^{m} p(\omega_{j_h}) + \sum_{\ell=1}^{r} p(\omega_{k_\ell}) = \mathbb{P}(E) + \mathbb{P}(F)$$

2.3 Sintesi della Lezione 2

In questa lezione si considera solamente il caso in cui le situazioni di interesse sono in numero finito e quindi Ω è un insieme finito e gli eventi di interesse sono rappresentati da un qualunque sottoinsieme di Ω, ossia qualunque elemento della famiglia delle parti di Ω, famiglia denotata da $\mathcal{P}(\Omega)$.
Da una parte ci sono tre proprietà fondamentali, gli assiomi, dalle quali si possono ottenere le proprietà principali delle probabilità. Gli assiomi (nel caso in cui Ω è finito, sono
1) la probabilità di un evento è un numero compreso tra 0 e 1
(ovvero: per ogni $A \subset \Omega$ si ha $\mathbb{P}(A) \in [0,1]$)
2) **(proprietà di normalizzazione)** la probabilità dell'evento certo è 1
(ovvero $\mathbb{P}(\Omega) = 1$)
3) **(proprietà di additività finita, per due eventi)** dati due eventi incompatibilie la probabilità che se ne verifichi almeno uno dei due è la somma delle rispettive probabilità
(ovvero se $A_1, A_2 \subset \Omega$, con $A_1 \cap A_2 = \emptyset$ allora $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$)

Da questi assiomi si deducono tutte le seguenti proprietà
(proprietà di additività finita) Siano dati n eventi $E_1, ..., E_n$ incompatibili a due a due, la probabilità che se ne verifichi almeno uno è la somma delle rispettive probabilità
(ovvero se gli insiemi che li rappresentano $E_1, ..., E_n \in \mathcal{P}(\Omega)$ sono disgiunti a due a due, ossia $E_i \cap E_j = \emptyset$, per $i, j \in \{1, 2, \cdots, n\}$, con $i \neq j$ allora si ha $\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}(E_i)$.
(propietà di base) Per ogni A e $B \in \mathcal{P}(\Omega)$ si ha
$\boxed{\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})}$ e $\boxed{\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)}$
e quindi, ricordando che $B \setminus A = B \cap \overline{A}$, risulta $\boxed{\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)}$
(a) Ponendo $p(\omega_i) := \mathbb{P}(\{\omega_i\})$, $i = 1, ..., N$, per ogni $E \in \mathcal{P}(\Omega)$ risulta $\mathbb{P}(E) = \sum_{\substack{i=1\\ i:\,\omega_i \in E}}^N p(\omega_i) = \sum_{\omega \in E} p(\omega)$.
(b) Per ogni $E \in \mathcal{P}(\Omega)$ risulta $\mathbb{P}\left(\overline{E}\right) = 1 - \mathbb{P}(E)$.
(c) L'evento impossibile ha probabilità nulla, ovvero $\mathbb{P}(\emptyset) = 0$.
(d) **(proprietà di monotonia)** Se il verificarsi di A implica il verificarsi di B allora $\mathbb{P}(A) \leq \mathbb{P}(B)$
(in quanto se A, $B \in \mathcal{P}(\Omega)$ sono gli insiemi che li rappresentano e se si ha che $A \subseteq B$, allora dalla proprietà di base si ottiene che $0 \leq \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A)$, in quanto $A \subset B$ equivale a $A \cap B = A$)
(e) **(formula di inclusione ed esclusione per due eventi)** Dati due eventi, la probabilità che se ne verifichi almeno uno è la somma delle rispettive probabilità meno la probabilità che si verifichino entrambi *ossia, per due arbitrari $A, B \in \mathcal{P}(\Omega)$ risulta $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*
(f) Siano $H_1, ..., H_n \in \mathcal{P}(\Omega)$ tali che

$$\bigcup_{i=1}^n H_i = \Omega, \quad \text{e} \quad H_i \cap H_j = \emptyset \quad \text{per } i \neq j, \tag{16}$$

ossia la collezione $H_1, ..., H_n$ è una partizione dell'evento certo; allora si ha la condizione

$$\sum_{i=1}^n \mathbb{P}(H_i) = 1. \tag{17}$$

In particolare, ricordando che $p(\omega_i) = \mathbb{P}(\{\omega_i\})$, $i = 1, ..., N$, e prendendo $H_i = \{\omega_i\}$, deve risultare

$$p(\omega_i) \geq 0 \qquad \text{e} \qquad \sum_{i=1}^N p(\omega_i) = 1. \tag{18}$$

Ulteriori conseguenze degli assiomi della probabilità verranno viste in seguito, dopo aver introdotto il concetto di probabilità condizionata (si veda la Lezione 4).

3 Probabilità "classiche (o uniformi)" e calcolo combinatorio

3.1 Probabilità "classiche (o uniformi) "

Qui ci soffermiamo a trattare alcuni casi particolari, ma molto rilevanti, di spazi di probabilità finiti. Sia dunque

$$\Omega = \{\omega_1, ..., \omega_N\},$$

e supponiamo che si voglia porre

$$p(\omega_i) = K, \qquad \forall \omega_i \in \Omega \tag{19}$$

per un'opportuna costante positiva K. Si vuole cioè imporre che tutti i risultati elementari siano, fra di loro, equiprobabili.

Ci riferiremo a tale caso dicendo che si ha una ***distribuzione di probabilità uniforme sugli eventi elementari***.

Confrontando la posizione (19) con la condizione di normalizzazione (18), otteniamo immediatamente

$$p(\omega_i) = \frac{1}{|\Omega|} = \frac{1}{N}, \qquad i = 1, ..., N$$

e da ciò segue, ricordando la formula (2.3) della precedente Lezione 2,

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{N}, \qquad \forall E \in \mathscr{P}(\Omega). \tag{20}$$

Esempio* 3.1. *L'addetto ad un guardaroba restituisce a caso n ombrelli che gli sono stati consegnati; qual è la probabilità che il secondo cliente abbia indietro il suo proprio ombrello?*

Soluzione: Si ha che Ω è costituito dalle permutazioni[7] di n elementi; dunque $|\Omega| = n!$ L'evento

$$E = \{\textit{Il secondo cliente riceve indietro il suo ombrello}\}$$

è un evento composto, costituito da tutte le permutazioni che tengono fisso il secondo elemento; tali permutazioni sono in numero di $(n-1)!$, corrispondente al numero delle possibili permutazioni dei restanti $(n-1)$ elementi. L'espressione "a caso" vuole significare che tutte le permutazioni sono da considerare equiprobabili fra di loro. Dunque

$$\mathbb{P}(E) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

□

Osservazione* 3.1.** La formula (20) esprime il fatto che, nel caso in cui tutti gli eventi elementari di uno spazio finito sono equiprobabili, la probabilità di un generico evento composto si calcola quale ***rapporto fra casi favorevoli e casi possibili.

Si faccia attenzione al fatto che la (20) *non costituisce una definizione del concetto di probabilità, ma soltanto una formula per il suo calcolo in un caso particolare*. A questo proposito ricordiamo che, nella precedente Lezione 2, abbiamo analizzato il concetto di probabilità in modo assiomatico e, in ***Osservazione* 2.3**, abbiamo ricavato un modo generale di ottenere una probabilità: la formula (20) è il caso particolare in cui gli eventi sono equiprobabili, cioè, nella formula (15), il valore g_i è costante.

Osservazione* 3.2.** Nel caso in cui si imponga la condizione (19), il calcolo della probabilità di un evento composto E si riduce al problema, ***combinatorio, di individuare $N = |\Omega|$ e $|E|$.

[7] Per la definizione di permutazione vedere più avanti la Sezione 3.2.

3.2 Calcolo combinatorio: primi elementi

Facendo seguito alle precedenti ***Osservazione 3.1*** e ***Osservazione 3.2***, ci rivolgiamo ora a richiamare succintamente alcune nozioni basilari di calcolo combinatorio, che risultano indispensabili per affrontare i primi problemi di calcolo delle probabilità.

Le formule che verranno presentate si ricavano facilmente tramite applicazione del ***principio di induzione finita***.

Iniziamo innanzitutto ricordando due fatti fondamentali:

a) Due insiemi finiti hanno la stessa cardinalità se e solo se fra essi è possibile stabilire una ***corrispondenza biunivoca***.

b) Dati due arbitrari insiemi A e B, si definisce ***prodotto cartesiano*** di A per B l'insieme costituito dalle coppie **ordinate** (a, b) dove $a \in A$ e $b \in B$; indichiamo tale insieme con il simbolo $A \times B$. Nel caso in cui A e B sono insiemi finiti, risulta

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

Supponiamo di aver fissato un arbitrario insieme A costituito da n elementi:

$$A = \{a_1, ..., a_n\}.$$

Disposizioni con ripetizione di classe k di n elementi.

Una disposizione con ripetizione di classe k degli n elementi di A non è altro che una k-upla ordinata degli elementi stessi.

Tali disposizioni costituiscono dunque l'insieme $A^k = \overbrace{A \times A \times \cdots \times A}^{k\, volte}$, e si ha $|A^k| = n^k$.

Disposizioni senza ripetizione di classe k di n elementi e permutazioni di n elementi

Le disposizioni senza ripetizione di classe k degli n elementi sono le k-uple **ordinate** costituite da elementi di A, ***tutti diversi fra loro*** (quindi necessariamente $k \leq n$).

Tali disposizioni costituiscono un sottoinsieme, dell'insieme A^k, di cardinalità[8]

$$\overbrace{n(n-1)...(n-(k-1))}^{k\, fattori} = n(n-1)...(n-k+1) = \frac{n!}{(n-k)!},$$

dove si è usata la notazione *n fattoriale, ovvero* $n! = n(n-1) \cdots 3 \cdot 2 \cdot 1$.

Nel caso in cui si ponga $k = n$, si ottengono le ***permutazioni*** degli elementi di A. Di conseguenza il numero delle permutazioni di n elementi è $n!$.

Combinazioni di classe k di n elementi

Ci sono diversi modi di darne la definizione.
Le combinazioni di classe $\boldsymbol{k}$ di $\boldsymbol{n}$ elementi sono le k-uple **non ordinate** costituite da elementi di A, ***tutti diversi fra loro*** (quindi necessariamente $k \leq n$).
In altre parole si tratta di classi di equivalenza di disposizioni senza ripetizione di classe k di n elementi, modulo la relazione di equivalenza costituita dal considerare equivalenti due disposizioni che contengono gli stessi elementi, eventualmente in ordine diverso.
Alternativamente una combinazione di classe k di n elementi si può definire come ***un sottoinsieme di cardinalità k di un insieme di cardinalità n***.

Ad esempio, se consideriamo l'insieme $A = \{a, b, c, d, e, f, g\}$, che ha $n = 7$ elementi, preso $k = 3$, la terna ordinata (a, d, f) è una disposizione di 7 elementi di classe 3 e differisce dalla terna ordinata (d, f, a), appunto in quanto l'ordine è cambiato. Invece la combinazione $\{a, d, f\}$ coincide con la combinazione

[8] Si veda la nota di approfondimento a pagina 41.

$\{d, f, a\}$: si noti che per le disposizioni si usano le parentesi tonde, come per i vettori, in cui appunto l'ordine ha un'importanza fondamentale, mentre per le combinazioni si usano le parentesi graffe, come per gli insiemi e i sottoinsiemi, in cui invece l'ordine non ha affatto importanza.

Se C_k^n indica il numero delle combinazioni di classe k di n elementi, D_k^n indica il numero delle disposizioni senza ripetizione di classe k di n elementi, e P_k indica il numero delle permutazioni di k elementi, è immediato che $C_0^n = C_n^n = 1$, inoltre è facile convincersi che C_k^n=C_{n-k}^n e che (vedere la nota di pagina 42)

$$D_k^n = C_k^n \cdot P_k,$$

Dalla formula precedente, tenendo conto che $P_k = k!$ e $D_k^n = \frac{n!}{(n-k)!}$, si ricava immediatamente $\frac{n!}{(n-k)!} = C_k^n \cdot k!$.
Il numero complessivo di tali combinazioni è dunque dato da

$$C_k^n = \frac{1}{k!} \cdot \frac{n!}{(n-k)!}.$$

Come usuale si pone

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

Il numero $\binom{n}{k}$ prende il nome di ***coefficiente binomiale n sopra k*** (o anche n su k, o anche scegli k tra n).
In questo modo il numero C_k^n delle combinazioni di n elementi di classe k coincide con il coefficiente binomiale $\binom{n}{k}$, ossia

$$C_k^n = \binom{n}{k}.$$

A proposito di coefficienti binomiali si usa la seguente convenzione: per ogni numero naturale n, si pone

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

come è ovvio, sia tenendo presente la convenzione $0! = 1$, sia tenendo presente il fatto che l'unico sottoinsieme di cardinalità zero è l'insieme vuoto.

Esempio 3.2. *Consideriamo un circolo costituito da n persone e supponiamo di dover eleggere un* presidente*, un* segretario *e un* tesoriere.

Se pensiamo di scegliere tre persone diverse, ognuna con la sua specifica carica, ciascuna scelta coincide con una disposizione senza ripetizione di classe 3 degli n elementi; abbiamo $n(n-1)(n-2)$ possibili scelte.

Se pensiamo che ogni carica è assegnata con una votazione indipendente dalle altre, si possono avere anche delle ripetizioni (cioè è ammesso un cumulo delle cariche); in tal caso ciascuna possibile scelta coincide con una disposizione di classe 3, con ripetizione, degli n elementi; abbiamo n^3 possibili scelte.

Se pensiamo di eleggere complessivamente una terna di persone diverse, senza attribuire una specifica carica a ciascuna di loro, ma incaricandoli complessivamente dei compiti di presidente, di segretario e di tesoriere, ciascuna possibile scelta coincide con una combinazione di classe 3 degli n elementi; abbiamo, in tal caso $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$ possibili scelte.

Esercizio proposto 3.1. *Dimostrare che $\sum_{k=0}^{n} C_k^n = 2^n$, senza usare la formula della potenza del binomio di Newton.*
suggerimento: *vedere la nota di approfondimento a pagina 38.*

Ai fini di calcolare esplicitamente $\binom{n}{k}$, per ogni n e k interi con $0 \leq k \leq n$, può essere comodo osservare che

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-(k-2)) \cdot (n-(k-1))}{k(k-1)\ldots 2 \cdot 1} = \prod_{i=0}^{k-1} \frac{n-i}{k-i}.$$

Ad esempio, utilizzando anche la proprietà

$$C_k^n = C_{n-k}^n \quad \Leftrightarrow \quad \binom{n}{k} = \binom{n}{n-k},$$

possiamo calcolare facilmente $\binom{10}{7}$ con i seguenti passaggi:

$$\binom{10}{7} = \binom{10}{3} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = \frac{\overset{5}{\cancel{10}} \cdot \overset{3}{\cancel{9}} \cdot 8}{\cancel{3} \cdot \cancel{2} \cdot 1} = 120.$$

IMPORTANTE: nel seguito useremo i coefficienti binomiali **solamente per valori interi n e con $0 \leq k \leq n$.**

Tuttavia, segnaliamo come curiosità, che è possibile estendere il simbolo di coefficiente binomiale anche a tutti i numeri reali α, fermo restando $k \geq 0$ intero, nel seguente modo:

$$\binom{\alpha}{0} = 1, \qquad \binom{\alpha}{k} = \frac{\alpha \cdot (\alpha-1) \cdot \cdots \cdot (\alpha-(k-1))}{k(k-1)\ldots 2 \cdot 1} = \prod_{i=0}^{k-1} \frac{\alpha-i}{k-i}, \qquad k \geq 1.$$

Ciò, tra l'altro, permette **(ma noi non lo faremo)** di usare il simbolo $\binom{n}{k}$ anche nel caso di n intero ma con $k > n$: in tale caso si ottiene che $\binom{n}{k} = \prod_{i=0}^{k-1} \frac{n-i}{k-i} = 0$ in quanto il fattore $\frac{n-i}{k-i}$ per $i = n$ si annulla.

3.3 Alcuni classici esempi

Consideriamo ora qualche semplice e classico esempio di probabilità combinatorie.

Esempio* 3.3 (Problema del compleanno)**.** *Qual è la probabilità che, fra M persone scelte a caso, ve ne siano almeno due che festeggiano il compleanno nello stesso giorno? (Si supponga l'anno costituito da 365 giorni e che vi sia una situazione di simmetria rispetto alle nascite).*

Soluzione: Calcoliamo la probabilità dell'evento complementare

$$\overline{E} = \{\textit{Le M persone festeggiano il compleanno in tutti giorni diversi}\}$$

Lo spazio Ω è costituito dalle disposizioni con ripetizione di classe M di 365 elementi (i giorni dell'anno solare); mentre $\overline{E}$ è un evento composto, costituito da tutte le disposizioni senza ripetizione di classe M di 365 elementi. Quindi

$$\mathbb{P}(\overline{E}) = \frac{365 \cdot 364 \cdot \ldots \cdot (365 - M + 1)}{(365)^M}$$

e la probabilità cercata è fornita da $\mathbb{P}(E) = 1 - \mathbb{P}(\overline{E})$. Indichiamo ora tale probabilità con $\mathbb{P}_M(E)$ per mettere in evidenza la sua dipendenza dal valore di M. Ovviamente $\mathbb{P}_M(E)$ è una funzione crescente di M ed è interessante notare che si ha $\mathbb{P}_M(E) > \frac{1}{2}$ per $M > 22$, in particolare si ha $\mathbb{P}_{22}(E) \simeq 0.4756$ mentre

$\mathbb{P}_{23}(E) \simeq 0.5072$.
Osserviamo infine che abbiamo implicitamente assunto $M \le 365$: se fosse $M > 365$ ci sarebbero sicuramente almeno due persone con la stessa data di nascita, e si avrebbe $\mathbb{P}_M(E) = 1$.

□

Esempio* 3.4 ("Paradosso del Cavalier De Méré"). *È più probabile*[9] *ottenere almeno un asso in 4 lanci consecutivi di un dado o un doppio asso in 24 lanci consecutivi di una coppia di dadi?*

Soluzione: Anche qui conviene calcolare le probabilità dei due eventi complementari:

$$\mathbb{P}(\{\textit{almeno un asso in 4 lanci}\}) = 1 - \mathbb{P}(\{\textit{nessun asso in 4 lanci}\})$$
$$\mathbb{P}(\{\textit{almeno un doppio asso in 24 lanci}\}) = 1 - \mathbb{P}(\{\textit{nessun doppio asso in 24 lanci}\}).$$

I risultati possibili nei 4 lanci del dado sono rappresentati dalle disposizioni con ripetizione di classe 4 di 6 elementi; in altre parole, possiamo rappresentare Ω come lo spazio delle quaterne ordinate (x_1, x_2, x_3, x_4) con $x_i \in \{1, ..., 6\}$. Dunque $|\Omega| = 6^4$. Gli eventi elementari che costituiscono l'evento composto $\{$*nessun asso in 4 lanci*$\}$ corrispondono, invece, alle disposizioni con ripetizione di classe 4 dei 5 elementi $\{2, ..., 6\}$. Si ha quindi

$$\mathbb{P}(\{\textit{almeno un asso in 4 lanci}\}) = 1 - \frac{5^4}{6^4} \simeq 0.52.$$

Analogamente si ottiene

$$\mathbb{P}(\{\textit{almeno un doppio asso in 24 lanci}\}) = 1 - \frac{35^{24}}{36^{24}} \simeq 0.49.$$

□

Esempio* 3.5. *Un gruppo di $4n$ persone comprende $2n$ ragazzi e $2n$ ragazze. Vengono formate a caso due squadre di $2n$ persone ciascuna.*
(a) Qual è la probabilità che tutte le ragazze si trovino nella stessa squadra e tutti i ragazzi nella squadra avversaria?
(b) Qual è la probabilità che ciascuna squadra sia, all'opposto, composta esattamente da n ragazzi ed n ragazze?

Soluzione: Qui il generico evento elementare è specificato da un modo di scegliere $2n$ oggetti (i componenti della prima squadra) da un insieme di $4n$ oggetti; dunque la cardinalità dello spazio degli eventi elementari Ω è data da $\binom{4n}{2n}$. □

Suggerimento: in effetti si può pensare che le ragazze siano numerate come $f_1, f_2, \ldots, f_{2n}$ ed i ragazzi come $m_1, m_2, \ldots, m_{2n}$. Per specificare la prima squadra basta prendere un sottoinsieme di

$$R = \{f_1, f_2, \ldots, f_{2n}, m_1, m_2, \ldots, m_{2n}\}$$

di cardinalità $2n$, ovvero

$$\Omega = \{\text{combinazioni dei } 4n \text{ elementi di } R \text{ di classe } 2n\}, \qquad \text{con } |\Omega| = \binom{4n}{2n}$$

Nel caso *(a)* due soli eventi elementari sono favorevoli e dunque la probabilità cercata è $\frac{2}{\binom{4n}{2n}}$.

Suggerimento: i due casi favorevoli sono le due combinazioni $\{f_1, f_2, \ldots, f_{2n}\}$ e$\{m_1, m_2, \ldots, m_{2n}\}$.

[9] Il quesito fu posto a Pascal dal Cavalier De MMéré nel 1654. Questo esempio verrà ripreso successivamente (si veda pagina 121), quando si parlerà di valore atteso e si chiarirà il motivo per cui questo esempio è noto come un paradosso.

Nel caso *(b)* gli eventi elementari favorevoli[10] sono in numero di $\binom{2n}{n}\binom{2n}{n}$, corrispondente al numero dei modi in cui si possono scegliere n ragazze dal gruppo di tutte le $2n$ e n ragazzi dal gruppo di tutti i $2n$. La probabilità cercata è dunque data da

$$\frac{\binom{2n}{n}\binom{2n}{n}}{\binom{4n}{2n}}.$$

□

Esempio* 3.6. *Supponiamo che una moneta perfetta venga lanciata n volte. Per $h \le n$, qual è la probabilità di nessuna testa sui primi h lanci?*

Soluzione: Possiamo schematizzare gli eventi elementari in questo esperimento come gli elementi dell'insieme $\{0,1\}^n$ cioè come n-uple con elementi uguali a 0 (croce) o uguali a 1 (testa), (ad esempio l'evento elementare $\omega = (0,0,1,1,0,...,0)$ coincide con il fatto che i primi due lanci danno croce, poi si hanno consecutivamente due risultati testa, e poi in tutti i successivi lanci si ottiene ancora croce); dunque si ha $|\Omega| = 2^n$.

L'evento $\{$*nessuna testa sui primi h lanci*$\}$ è allora l'evento composto

$$E = \{\omega \in \Omega: \ \omega = (0,0,...,0,x_{h+1},...,x_n)\,, \text{ con } (x_{h+1},...,x_n) \in \{0,1\}^{n-h}\}.$$

Traduciamo la condizione che la moneta sia perfetta con la posizione $p(\omega) = \frac{1}{2^n}$, per ogni $\omega \in \Omega$.
Si ha $|E| = 2^{n-h}$ e dunque $\mathbb{P}(E) = \frac{2^{n-h}}{2^n} = \left(\frac{1}{2}\right)^h$. □

Esempio* 3.7. *Qual è la probabilità di k risultati testa negli n lanci di una moneta?*

Soluzione: Si ha lo stesso spazio di probabilità dell'esercizio precedente; questa volta $|E| = \binom{n}{k}$ e dunque $\mathbb{P}(E) = \frac{\binom{n}{k}}{2^n}$; come vedremo in seguito si tratta di un caso particolare di probabilità ***binomiali***. □

Esempio* 3.8. *Trovare la probabilità di k voti per lo schieramento A in un sondaggio elettorale di ampiezza n in un gruppo di M elettori di cui è noto che m_1 votano per A e $m_2 = M - m_1$ votano per B.*

Soluzione: Si è sottinteso che gli n elettori siano stati selezionati senza reinserimento. L'esperimento consiste dunque nel selezionare un sottoinsieme di cardinalità n (*il campione*) dall'insieme degli M elettori (*la popolazione*) e quindi $|\Omega| = \binom{M}{n}$. Si sottointende che il sondaggio sia condotto in modo casuale, cioè che ogni "campione" abbia uguale probabilità $\frac{1}{\binom{M}{n}}$ di essere estratto. Fra tali "campioni", ve ne sono $\binom{m_1}{k}\cdot\binom{m_2}{n-k}$ che contengono k elettori per A e $(n-k)$ per B. Infatti, ci sono $\binom{m_1}{k}$ modi di selezionare k elettori fra i votanti per A, ci sono $\binom{m_2}{n-k}$ modi di selezionare $(n-k)$ elettori fra i votanti per B e, inoltre, una qualunque scelta di k elettori fra i votanti per A e di $(n-k)$ elettori fra i votanti per B dà luogo ad una n-upla di elettori (un campione) che contiene k elettori per A.

Dunque la probabilità cercata è data da

$$\frac{\binom{m_1}{k}\cdot\binom{m_2}{n-k}}{\binom{M}{n}} = \frac{\binom{m_1}{k}\cdot\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}}.$$

Come vedremo in seguito, si tratta di un caso particolare di probabilità ***ipergeometriche***. Osserviamo che i valori possibili per k devono rispettare la condizione

$$0 \le k \le m_1, \quad 0 \le n-k \le m_2, \quad \textit{con} \quad n \le M = m_1 + m_2,$$

che, dopo semplici passaggi, diviene

$$0 \vee (n-m_2) = \max(0, n-m_2) \le k \le \min(n, m_1) = n \wedge m_1.$$

□

[10] Per convincersene si consiglia il lettore di considerare il caso $n = 1$ ed $n = 2$, elencando esplicitamente sia tutti i casi possibili che tutti i casi favorevoli.

Esercizio proposto 3.2. *Siano M, m_1, n e k numeri assegnati e tali che*

$$m_1 < M, \quad n < M, \quad \max(0, m_1 + n - M) \leq k \leq \min(n, m_1).$$

Verificate l'identità

$$\frac{\binom{m_1}{k}\binom{M-m_1}{n-k}}{\binom{M}{n}} = \frac{\binom{n}{k}\binom{M-n}{m_1-k}}{\binom{M}{m_1}}.$$

Riuscite a darne un'interpretazione probabilistica?

Interpretazione dell'identità

$$\frac{\binom{m_1}{k}\binom{M-m_1}{n-k}}{\binom{M}{n}} = \frac{\binom{n}{k}\binom{M-n}{m_1-k}}{\binom{M}{m_1}},$$

del precedente Esercizio-proposto 3.2

Pensiamo ad esempio di avere un'urna e di estrarre tutte le palline in ordine: si suppone come al solito che l'urna contenga m_1 palline di tipo A ed $m_2 = M - m_1$ di tipo B. Siamo interessati alla probabilità che tra le prime n (con $1 \le n \le M$) estratte ce ne siano (esattamente) k di tipo A (e quindi $n - k$ di tipo B). Un esperimento si può descrivere annotando solo l'ordine con cui vengono estratti i tipi. Ad esempio se $m_1 = 3$ ed $m_2 = 2$, i risultati possibili sono

(A, A, A, B, B)	(A, A, B, A, B)	(A, A, B, B, A)	(A, B, A, A, B)	(A, B, A, B, A)
(A, B, B, A, A)	(B, A, A, A, B)	(B, A, A, B, A)	(B, A, B, A, A)	(B, B, A, A, A)

Non è difficile convincersi che tutti i casi possibili sono equiprobabili e che sono in tutto tanti quante sono le combinazioni di 5 elementi di classe 3, infatti basta specificare le posizioni in cui sono uscite le palline di tipo A. Con questa corrispondenza i casi precedenti sono in corrispondenza con

$\{1,2,3\}$	$\{1,2,4\}$	$\{1,2,5\}$	$\{1,3,4\}$	$\{1,3,5\}$
$\{1,4,5\}$	$\{2,3,4\}$	$\{2,3,5\}$	$\{2,4,5\}$	$\{3,4,5\}$

Più in generale si ottiene che i casi possibili sono appunto

$$C^M_{m_1} = \binom{M}{m_1},$$

in quanto per elencare tutti i casi possibili basta specificare solo le m_1 posizioni occupate dalle A su tutte le M posizioni.
Per quanto riguarda i casi favorevoli, iniziamo di nuovo con il caso particolare dell'urna precedente e con $n = 3$ e $k = 2$. I casi favorevoli sono

(A, A, B, A, B)	(A, A, B, B, A)
(A, B, A, A, B)	(A, B, A, B, A)
(B, A, A, A, B)	(B, A, A, B, A)

e si ottengono nel seguente modo: basta specificare quali sono le due A tra le prime 3 posizioni, e dove si trova la terza A tra le ultime $2 = 5 - 3$ posizioni. Il numero totale risulta quindi $\binom{3}{2} \cdot \binom{2}{1} = 3 \cdot 2 = 6$. Si osservi inoltre che AAB, ABA e BAA sono tutti gli anagrammi composti da due A e un B e che AB e BA sono gli anagrammi composti da un A e un B.

Più in generale per elencare tutti casi favorevoli all'evento che tra le prime n estratte ce ne siano k di tipo A, basta specificare le k posizioni tra le prime n in cui si trovano le A (e questo si può fare in $C^n_k = \binom{n}{k}$ modi) e inoltre specificare le $m_1 - k$ posizioni occupate dalle A, nelle ultime $M - n$ posizioni (e questo si può fare in $C^{M-n}_{m_1-k} = \binom{M-n}{m_1-k}$ modi). In totale quindi si hanno

$$\binom{n}{k} \cdot \binom{M-n}{m_1-k}$$

casi favorevoli.

3.4 Alcune proprietà dei coefficienti binomiali.

Nello studio del Calcolo delle Probabilità è opportuno tenere presente alcune identità fondamentali riguardanti i coefficienti binomiali. Ne presentiamo intanto alcune qui di seguito.

Si noti innanzitutto che per $k = 0$ e per $k = n$ ovviamente si ha $\binom{n}{0} = \binom{n}{n} = 1$ e che $\binom{n}{k} = \binom{n}{n-k}$, per $k = 0, 1, \ldots, n$. Un'altra semplice, ma non immediata, identità è la seguente (nota come formula di Stiefel)

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \tag{21}$$

purché $k - 1 \geq 0$ e $k \leq n - 1$, ovvero per $1 \leq k \leq n - 1$. Per la verifica di tali formule basta sviluppare i coefficienti binomiali (provare come esercizio).

Qui vogliamo comunque anche darne una semplice dimostrazione probabilistica[11], ricordando quanto visto nel precedente Esempio 3.7, relativo ad n lanci di una moneta.

Poniamo E ={*si ottengono k risultati testa in n lanci di una moneta perfetta*}. Sappiamo che la probabilità di ottenere tale risultato è uguale a $\frac{\binom{n}{k}}{2^n}$. D'altra parte, ponendo

$$\begin{aligned} E_1 &= \{\textit{(k-1) teste sui primi (n-1) lanci}\} \cap \{\textit{testa all'n-esimo lancio}\}\,, \\ E_2 &= \{\textit{k teste sui primi (n-1) lanci}\} \cap \{\textit{croce all'n-esimo lancio}\}\,, \end{aligned}$$

possiamo anche scrivere $E = E_1 \cup E_2$, e, essendo chiaramente $E_1 \cap E_2 = \emptyset$,

$$\mathbb{P}(E) = \mathbb{P}(E_1) + \mathbb{P}(E_2). \tag{22}$$

Ora possiamo notare che gli eventi composti E_1 ed E_2 hanno rispettivamente cardinalità uguale a $\binom{n-1}{k-1}$ e $\binom{n-1}{k}$; e dunque la (22) diventa

$$\frac{\binom{n}{k}}{2^n} = \frac{\binom{n-1}{k-1}}{2^n} + \frac{\binom{n-1}{k}}{2^n}.$$

L'identità (21) è in particolare alla base della costruzione del ben noto ***Triangolo di Tartaglia***.

$$\begin{array}{c}
1 \\
\swarrow\ \searrow \\
1 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 2 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 3 \quad 3 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1 \\
\swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow\ \swarrow\ \searrow \\
1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1
\end{array}$$

Triangolo di Tartaglia (o di Pascal)

[11]Si veda anche la nota di approfondimento a pagina 39

Utilizzando (21) è anche facile, per $0 \le k \le n$, ottenere la seguente uguaglianza

$$\sum_{r=k}^{n} \binom{r}{k} = \binom{n+1}{k+1}$$

È ben noto (e comunque si verifica immediatamente per induzione, utilizzando la (21)) che i coefficienti binomiali intervengono come segue nello sviluppo della potenza di un binomio: siano a, b due arbitrari numeri reali non nulli e sia n un numero naturale; allora risulta

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \cdot b^{n-k}. \tag{23}$$

Ponendo nella (23) $a = x, b = 1$, otteniamo l'identità

$$(x+1)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{24}$$

In particolare ponendo $a = b = 1$, otteniamo

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Tenendo presente che $\binom{n}{k}$ coincide con il numero di sottoinsiemi di cardinalità k contenuti in un insieme composto da n elementi, otteniamo che 2^n è uguale alla cardinalità della famiglia delle parti di un insieme di n elementi. Dunque se in un esperimento vi sono n eventi elementari, vi sono allora in tutto 2^n eventi fra elementari, composti e contando anche l'evento certo e quello impossibile (si veda anche la nota di approfondimento a pagina 38) .

Un'altra utile identità è l'***identità di Vandermonde***, ricordata nella seguente proposizione.

***Proposizione* 3.1** (Identità di Vandermonde)**.** *Per qualunque terna di numeri naturali r, s, n con $n \le r + s$ si ha:*

$$\sum_{k=0\vee(n-s)}^{n\wedge r} \binom{r}{k}\binom{s}{n-k} = \binom{r+s}{n}, \tag{25}$$

dove la somma è estesa a tutti gli indici k per i quali $0 \le k \le r$ e $0 \le n-k \le s$.

Dimostrazione. L'identità di Vandermonde può essere verificata, sia dandole un'interpretazione tramite le estrazioni da un'urna, sia con una dimostrazione analitica.

Dimostrazione con le estrazioni da un'urna. Supponiamo di avere un'urna che contiene $r + s$ palline numerate da 1 a $r + s$, di cui r rosse ed s color senape. Le palline inoltre sono numerate in modo che le palline rosse hanno i numeri da 1 ad r e le palline color senape sono numerate da $r + 1$ fino ad $r + s$. Vogliamo effettuare l'estrazione di n palline in blocco. Ovviamente dobbiamo supporre $n \le r + s$ e contare il numero m dei modi di estrarre in blocco n palline.

Un modo è caratterizzato dai numeri delle palline estratte e quindi è facile convincersi che i modi di estrarre n palline dall'urna sono tanti quanti i sottoinsiemi di $\{1, 2, ..., r + s\}$ di cardinalità n, ossia $m = C_n^{r+s} = \binom{r+s}{n}$.

D'altra parte possiamo anche suddividere tali modi raggruppandoli per il numero di palline rosse che contengono: quindi, posto ρ_k il numero delle estrazioni che contengono esattamente k palline rosse e, di conseguenza, $n-k$ color senape, appare chiaro che $C_n^{r+s} = \binom{r+s}{n}$ si può ottenere come la somma di tali numeri ρ_k, ossia $m = \sum_k \rho_k$.

Per ottenere ρ_k, ossia per contare in quanti modi possiamo estrarre esattamente k palline rosse e, di conseguenza, $n-k$ color senape, basta osservare che ci sono $C_k^r = \binom{r}{k}$ modi di estrarre (esattamente) k palline rosse e $C_{n-k}^s = \binom{s}{n-k}$ modi di estrarre (esattamente) $n-k$ palline color senape, e quindi $\rho_k = \binom{r}{k}\binom{s}{n-k}$. Osserviamo inoltre che necessariamente $0 \le k \le r$ e $0 \le n-k \le s$, in quanto non è possibile estrarre più di r palline rosse e neanche più di s palline color senape.

A questo punto l'identità di Vendermonde è immediata in quanto equivale alla relazione $m = \sum_k \rho_k$.

Dimostrazione analitica Iniziamo osservando intanto quanto segue: per ogni x reale

$$(1+x)^{r+s} = (1+x)^r \cdot (1+x)^s.$$

Da una parte, utilizzando la (24) con $n = r$, e $n = s$, e la proprietà che il prodotto di sommatorie è la sommatoria del prodotto (si veda il richiamo nel riquadro a pagina 35)

$$\begin{aligned}(1+x)^r \cdot (1+x)^s &= \left(\sum_{k=0}^{r}\binom{r}{k}x^k\right)\cdot\left(\sum_{h=0}^{s}\binom{s}{h}x^h\right)\\ &= \sum_{k=0}^{r}\sum_{h=0}^{s}\binom{r}{k}\binom{s}{h}x^{k+h} = \sum_{n=0}^{r+s}\sum_{\substack{0\le k\le r,\,0\le h\le s\\ k+h=n}}\binom{r}{k}\binom{s}{h}x^{k+h}\\ &= \sum_{n=0}^{r+s}\sum_{0\le k\le r,\,0\le n-k\le s}\binom{r}{k}\binom{s}{n-k}x^n = \sum_{n=0}^{r+s}\left(\sum_{k=0\vee(n-s)}^{r\wedge n}\binom{r}{k}\binom{s}{n-k}\right)x^n.\end{aligned}$$

D'altra parte, per la formula della potenza del binomio, ossia per la (24) con $n = r+s$,

$$(1+x)^{r+s} = \sum_{n=0}^{r+s}\binom{r+s}{n}x^n,$$

e la (25) si ottiene confrontando termine a termine tali due sviluppi[12]. $\square$

Ponendo, ad esempio, in particolare nell'identità di Vandermonde (25), $r \le s$ e $n = s$, e utilizzando il fatto che $\binom{s}{s-k} = \binom{s}{k}$ si ottiene

$$\sum_{k=0}^{r}\binom{r}{k}\binom{s}{k} = \binom{r+s}{s}$$

e potremo dunque anche scrivere

$$\sum_{k=0}^{r\wedge s}\binom{r}{k}\binom{s}{k} = \binom{r+s}{s} = \binom{r+s}{r}, \tag{26}$$

che per $r = s = n$ diviene

$$\sum_{k=0}^{n}\binom{n}{k}^2 = \binom{2n}{n}. \tag{27}$$

[12] Il nome di Vandermonde è legato anche al determinante di una matrice e che potrebbe essere usato per dimostrare che due polinomi con lo stesso grado d coincidono se e solo se hanno lo stesso valore in d punti $x_1, x_2, \ldots, x_d$, tutti distinti fra loro.

Esercizio proposto 3.3. Una moneta perfetta viene lanciata r volte da Renato ed s volte da Stefano. Si ponga
$X =$ numero dei lanci in cui Renato ottiene il risultato testa
$Y =$ numero dei lanci in cui Stefano ottiene il risultato testa.
Calcolare la probabilità dei seguenti eventi ($r > 3, s > 3$).
(a) $\{X = 3\} \cap \{Y = 3\}$ (b) $\{X = 3\} \cup \{Y = 3\}$ (c) $\{\max(X, Y) = 3\}$ (d) $\{X = Y\}$.

Suggerimento per la soluzione dell'Esercizio proposto 3.3

Per calcolare la probabilità dell'evento $\{X = Y\}$ si consiglia di considerare il fatto che la moneta è perfetta e quindi, ai fini del calcolo delle probabilità è equivalente scambiare successo (esce testa) con insuccesso (esce croce) nel caso di uno dei due giocatori, ad esempio Stefano.
In altre parole si sta considerando che la probabilità di $\{X = Y\}$ è la stessa di $\{X = Y'\}$, dove $Y' = s - Y$ è il numero di croci ottenute da Stefano.
Ma allora $\{X = Y'\} = \{X = s - Y\} = \{X + Y = s\}$ e ciò permette di ottenere immediatamente, e con un ragionamento probabilistico, la relazione (26) (nota anche come formula di Vandermonde).

Richiamo: PROPRIETÀ DISTRIBUTIVA DELLA SOMMA RISPETTO AL PRODOTTO

Dati R ed S interi ($R,\ S \geq 1$), e $a_1, \ldots, a_R$ e $b_1, \ldots, b_S$, numeri reali, la sommatoria di tutti i prodotti del tipo $a_r\, b_s$ coincide con il prodotto delle somme di a_r per le somme di b_r:

$$\sum_{r=1}^{R} \sum_{s=1}^{S} a_r b_s = \left(\sum_{r=1}^{R} a_r \right) \left(\sum_{s=1}^{S} b_s \right)$$

Infatti basta osservare che

$$\begin{aligned}
\textstyle\sum_{r=1}^{R} \sum_{s=1}^{S} a_r b_s &= a_1 b_1 + a_1 b_2 + \cdots + a_1 b_S \\
&\quad + a_2 b_1 + a_2 b_2 + \cdots + a_2 b_S \\
&\quad + \cdots \\
&\quad + a_R b_1 + a_R b_2 + \cdots + a_R b_S \\
\\
&= a_1 \big(b_1 + b_2 + \cdots + b_S \big) + a_2 \big(b_1 + b_2 + \cdots + b_S \big) \\
&\quad + \cdots + a_R \big(b_1 + b_2 + \cdots + b_S \big) \\
\\
&= \textstyle\sum_{r=1}^{R} a_r \big(\sum_{s=1}^{S} b_s \big) = \Big(\sum_{s=1}^{S} b_s \Big) \Big(\sum_{r=1}^{R} a_r \Big).
\end{aligned}$$

Abbiamo usato questa proprietà nella dimostrazione analitica della formula di Vandermonde, ma sottolineiamo che tale proprietà è fondamentale in molte altre occasioni.

3.5 Il problema delle concordanze

Supponiamo di avere la seguente situazione: una segretaria ha n lettere indirizzate a n persone distinte ed n buste con già scritti gli n indirizzi: le cadono tutte le n lettere e quindi mette le n lettere a caso nelle n buste.

(i) **Quanto vale la probabilità $p(n)$ che nessuna lettera sia nella busta corrispondente?**

(ii) **Quanto vale la probabilità $q(n)$ che ci sia almeno una lettera nella busta corrispondente?**

In altre parole, se diciamo che c'è una *concordanza* quando una lettera viene messa nella busta corrispondente $p(n)$ è la probabilità che non ci siano concordanze, e $q(n)$ è la probabilità che ci sia almeno una concordanza. La probabilità in ***(i)*** è quindi il complemento a 1 della probabilità in ***(ii)***, ossia

$$p(n) = 1 - q(n)$$

Si tratta di un problema classico che ha anche altre diverse interpretazioni, ad esempio si potrebbe considerare una situazione analoga a quella dell'Esempio 3.1.

Tutte però, in modo astratto, possono essere ricondotte alla seguente situazione: data un'urna che contiene n palline numerate da 1 ad n, si estraggono le palline una alla volta contando il numero d'ordine di estrazione. Se, per un $h \in \{1, 2, ..., n\}$, si estrae la pallina con il numero h proprio all'h-sima estrazione si dice che c'è una ***concordanza***.

In altre parole stiamo assumendo che lo spazio Ω sia l'insieme di tutte le permutazioni $(j_1, j_2, ..., j_n)$ e si ha almeno una concordanza se $j_h = h$ per almeno un $h \in \{1, 2, ..., n\}$. Infine stiamo assumendo che le $n!$ permutazioni abbiano tutte la stessa probabilità.

Per calcolare $q(n)$ poniamo

$$A_h = \{\text{la lettera } h\text{-sima è nella busta } h\text{-sima}\},$$

di modo che per la formula di inclusione ed esclusione (vedere la nota di approfondimento a pagina 21 nella Lezione 2)

$$\begin{aligned} q(n) &= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1, i_2, \cdots, i_k\}} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \end{aligned}$$

ora, qualunque sia la combinazione $\{i_1, i_2, \cdots, i_k\}$ di $\{1, 2, \cdots, n\}$, si ha[13]

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \frac{(n-k)!}{n!}$$

[13] È importante osservare che stiamo chiedendo che a ciascuno dei posti i_1, i_2, $\cdots$, i_k compaia il numero corrispondente, ma non stiamo escludendo che nei posti rimanenti ci possa essere qualche altro *punto fisso* .
Quindi, una volta fissati i k valori i_1, i_2, ..., i_k dei posti fissi, nei rimanenti $n-k$ posti possiamo mettere i rimanenti valori in tutti gli $(n-k)!$ ordinamenti possibili.
Si consiglia di pensare ad un esempio particolare, ad esempio $n = 7$, $k = 3$ e $i_1 = 2$, $i_2 = 3, i_3 = 5$: stiamo contando le permutazioni di $\{1, 2, 3, 4, 5, 6, 7\}$ del tipo $(j_1, 2, 3, j_4, 5, j_6, j_7)$ dove j_1, j_4, j_6, j_7 possono variare solo tra i numeri $\{1, 4, 6, 7\}$ e di fatto (j_1, j_4, j_6, j_7) è una permutazione degli elementi $\{1, 4, 6, 7\}$, che formano appunto un insieme di $7-3 = 4$ elementi e sono quindi in totale $(7-3)! = 4!$ permutazioni.

e quindi, come vedremo, la somma

$$\sum_{\{i_1,i_2,\cdots,i_k\}} \mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}) = \frac{1}{k!}.$$

Infatti si tratta della somma fatta su tutte le $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ combinazioni $\{i_1, i_2, \cdots, i_k\}$ degli n elementi $\{1, 2, ..., n\}$ di classe k di numeri tutti uguali a $\frac{(n-k)!}{n!}$, e quindi si ha

$$\sum_{\{i_1,i_2,\cdots,i_k\}} \mathbb{P}(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k}) = \frac{n!}{k!(n-k)!}\cdot\frac{(n-k)!}{n!} = \frac{1}{k!}.$$

Di conseguenza la probabilità di almeno una concordanza vale

$$q(n) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{k-1}\frac{1}{k!} + \cdots + (-1)^{n-1}\frac{1}{n!}$$

e quindi la probabilità di nessuna concordanza vale

$$\begin{aligned} p(n) = 1 - q(n) &= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \cdots + (-1)^k\frac{1}{k!} + \cdots + (-1)^n\frac{1}{n!} \\ &= \sum_{k=0}^{n}(-1)^k\frac{1}{k!}. \end{aligned}$$

È interessante notare che $p(n)$ converge (per n che tende ad infinito) ad e^{-1}: basta ricordare lo sviluppo in serie di Taylor della funzione e^x in $x_0 = 0$, ovvero lo sviluppo di Mac Laurin:

$$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}, \qquad \text{da cui} \qquad e^{-x} = \sum_{k=0}^{\infty}(-1)^k\frac{x^k}{k!}.$$

Per controllo possiamo calcolare $p(n)$ per $n = 3,\ 4,\ 5,\ 6,\ 7$:

$$\begin{aligned} p(3) &= 1 - 1/1! + 1/2! - 1/3! = 1/2 - 1/6 = 1/3 \\ &\approx 0,3333 \\ p(4) &= 1 - 1/1! + 1/2! - 1/3! + 1/4! = p(3) + 1/4! = 1/3 + 1/24 = 9/24 \\ &= 0,375 \\ p(5) &= 1 - 1/1! + 1/2! - 1/3! + 1/4! - 1/5! = p(4) - 1/5! = 9/24 - 1/120 = 44/120 \\ &\approx 0,3666 \\ p(6) &= p(5) + 1/6! = 44/120 + 1/720 = 265/720 \\ &\approx 0,3680 \\ p(7) &= p(6) - 1/7! = 265/720 - 1/5040 = 1856/5040 \\ &\approx 0,3682 \end{aligned}$$

Vale la pena di osservare che il valore di $e^{-1} = 1/e \approx 0,367879441$.

È interessante osservare che $n!p(n)$ è il numero delle permutazioni $(i_1, i_2, ...i_n)$ di n elementi che non presentano punti fissi, ossia per le quali, qualunque sia h, il valore i_h è diverso da h, ovvero il numero delle permutazioni di n elementi senza concordanze. Infatti, indicando con $\kappa_n(0)$ il numero delle permutazioni senza concordanze, in teoria, avremmo potuto calcolare la probabilità classica di nessuna concordanza come $p(n) = \frac{\kappa_n(0)}{n!}$. Invece, al contrario, poiché abbiamo calcolato $p(n)$, possiamo ottenere il numero $\kappa_n(0)$ delle permutazioni senza concordanze moltiplicando la probabilità $p(n)$ per $n!$: $\kappa_n(0) = p(n)\cdot n! = n!\sum_{k=0}^{n}(-1)^k\frac{1}{k!}$.

3.6 Approfondimenti sui primi elementi di calcolo combinatorio

Questa sezione contiene delle dimostrazioni alternative di alcune delle formule di calcolo combinatorio precedentemente verificate con i metodi tradizionali dell'Analisi Matematica. L'idea è di ottenere alcune relazioni come, ad esempio, la formula di Stiefel, partendo dalle definizioni di permutazioni, di disposizioni, e di combinazioni, senza far uso del valore esplicito dei valori di P_k, D_k^n e C_k^n, ossia, rispettivamente, del numero P_k delle permutazioni di k elementi, del numero D_k^n delle disposizioni di n elementi di classe k e del numero C_k^n delle combinazioni di n elementi di classe k.
Iniziamo con una semplice osservazione:

L'insieme delle parti di $\{a_1, \cdots, a_n\}$ ha cardinalità 2^n

Denotiamo con $\mathscr{P}(\{a_1, \cdots, a_n\})$ l'insieme delle parti di $\{a_1, \cdots, a_n\}$. Per mostrare che $|\mathscr{P}(\{a_1, \cdots, a_n\})| = 2^n$ possiamo procedere come segue: $\mathscr{P}(\{a_1, \cdots, a_n\})$ è in corrispondenza biunivoca con l'insieme delle funzioni $g : \{a_1, \cdots, a_n\} \mapsto \{0, 1\}$. La corrispondenza è data da $G \longleftrightarrow g = \mathbf{1}_G$, dove

$$\mathbf{1}_G(a_i) = 1, \qquad se \quad a_i \in G$$

$$\mathbf{1}_G(a_i) = 0, \qquad se \quad a_i \notin G.$$

Queste ultime sono tante quanti sono gli elementi del prodotto cartesiano di $\{0, 1\} \times \{0, 1\} \times \cdots \{0, 1\}$ (n volte) e sono quindi 2^n. Per capire meglio la corrispondenza poniamo $n = 4$ e consideriamo il sottoinsieme $G = \{a_1, a_3\}$:

$$\{a_1, a_3\} \qquad \Longleftrightarrow \qquad \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 0 & 1 & 0 \end{pmatrix} \qquad \Longleftrightarrow \qquad (1, 0, 1, 0) \in \{0, 1\}^4.$$

Ancora, ad esempio $(0, 0, 0, 0)$ corrisponde all'insieme vuoto, e $(1, 1, 1, 1)$ a tutto $A = \{a_1, a_2, a_3, a_4\}$. Nel caso generale di $A = \{a_1, a_2, \ldots, a_n\}$, la corrispondenza tra i sottoinsiemi di A e gli elementi di $\{0, 1\}^n$ dovrebbe essere quindi chiara, e di conseguenza l'uguaglianza $|\mathscr{P}(\{a_1, \cdots, a_n\})| = 2^n$.

Una dimostrazione elementare di
$\sum_{k=0}^n C_k^n = \sum_{k=0}^n \binom{n}{k} = 2^n$

La relazione $\sum_{k=0}^n \binom{n}{k} = 2^n$ si può dedurre anche nel seguente modo (anche senza conoscere la formula della potenza del binomio di Newton): osserviamo che l'insieme delle parti $\mathscr{P}(\{a_1, \cdots, a_n\}$ di $\{a_1, \cdots, a_n\}$ è l'unione delle famiglie dei sottoinsiemi di cardinalità k al variare di k da 0 ad n. Ricordando che con C_k^n indichiamo il numero delle combinazioni di n elementi di classe k, o equivalentemente il numero dei sottoinsiemi di cardinalità k di un insieme di cardinalità n, possiamo affermare che la cardinalità dell'insieme delle parti, $|\mathscr{P}(\{a_1, \cdots, a_n\})|$, soddisfa la relazione seguente

$$\sum_{k=0}^n C_k^n = |\mathscr{P}(\{a_1, \cdots, a_n\})|$$

e quindi $\sum_{k=0}^n C_k^n = 2^n$, dato che $|\mathscr{P}(\{a_1, \cdots, a_n\})| = 2^n$. Per ottenere che $\sum_{k=0}^n \binom{n}{k} = 2^n$ basta osservare che $C_k^n = \binom{n}{k}$. Per la dimostrazione dettagliata di questa relazione vedere le prossime note di approfondimento alle pagine 40- 42.

$$C_k^n = C_{n-k}^n$$

La proprietà $C_k^n = C_{n-k}^n$ deriva immediatamente dalla definizione di combinazione come un sottoinsieme di cardinalità k di un insieme di cardinalità n.
Per verificare tale proprietà basta trovare una corrispondenza biunivoca tra la famiglia dei sottoinsiemi di cardinalità k (che contiene C_k^n sottoinsiemi) con la famiglia dei sottoinsiemi di cardinalità $n-k$ (che contiene C_{n-k}^n sottoinsiemi). A tale scopo basta associare a ciascun sottoinsieme di cardinalità il suo complementare che ha cardinalità $n-k$. Ovviamente vale anche il viceversa: ad ogni sottoinsieme di cardinalità $n-k$ possiamo associare il suo complementare che ha cardinalità $n-(n-k)=k$.

Dimostrazione della formula di Stiefel, senza l'uso dei fattoriali

Prendendo come interpretazione di $\binom{n}{k}$ il numero C_k^n dei sottoinsiemi di cardinalità k di un insieme $\boldsymbol{A = \{a_1,\, a_2, \cdots a_{n-1},\, a_n\}}$ di cardinalità n, la formula di Stiefel diviene

$$C_k^n = C_{k-1}^{n-1} + C_k^{n-1}.$$

Per dimostrare la precedente uguaglianza si può ragionare anche nel seguente modo:
(attenzione, si noti che ***non è necessario conoscere il valore esplicito di*** $\boldsymbol{C_k^n}$, cioè nella dimostrazione ***non si usa*** il fatto che $C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$)

I sottoinsiemi di cardinalità k si possono dividere in due classi:
(1) i sottoinsiemi C di A che contengono a_n
(2) i sottoinsiemi D di A che non contengono a_n.
Quindi il numero C_k^n dei sottoinsiemi di cardinalità k si può esprimere come la somma del numero dei sottoinsiemi del primo tipo e del numero dei sottoinsiemi del secondo tipo.
D'altra parte
(i) gli insiemi C del primo tipo si possono esprimere come $C = C' \cup \{a_n\}$, con C' sottoinsieme di $\{a_1, a_2, \cdots a_{n-1}\}$ che ha cardinalità $k-1$, e quindi sono tanti quanti i sottoinsiemi di cardinalità $k-1$ di un insieme di cardinalità $n-1$,
e quindi sono in tutto esattamente C_{k-1}^{n-1},
(ii) gli insiemi D del secondo tipo, sono sottoinsiemi di $\{a_1, a_2, \cdots a_{n-1}\}$ di cardinalità k di un insieme di cardinalità $n-1$,
e quindi sono in tutto esattamente C_k^{n-1}.

Come si contano le disposizioni-I

In questa nota vogliamo vedere come si può arrivare a contare sia le disposizioni che le combinazioni. Iniziamo con un esempio.
Consideriamo $n = 5$ e $k = 1, 2, 3$ e proviamo a scrivere tutte le disposizioni di 5 elementi di classe k. Indichiamo il nostro insieme di 5 elementi con $\{1, 2, 3, 4, 5\}$.
Ovviamente tutte le disposizioni di classe 1 sono $D_1^5 = 5$:

(1)	(2)	(3)	(4)	(5)

Per ottenere le disposizioni di classe 2 si può procedere mettendo insieme le disposizioni che inziano per 1, quelle che iniziano per 2, etc. come segue

(1,2)	(2,1)	(3,1)	(4,1)	(5,1)
(1,3)	(2,3)	(3,2)	(4,2)	(5,2)
(1,4)	(2,4)	(3,4)	(4,3)	(5,3)
(1,5)	(2,5)	(3,5)	(4,5)	(5,4)

Vengono quindi 20 disposizioni, in quanto si tratta di 5 colonne ciascuna di lunghezza 4: infatti una volta scelto il primo elemento, rimangono $5-1=4$ scelte per il secondo elemento.
Per ottenere le disposizioni di classe 3 possiamo ancora dividere le disposizioni mettendo insieme le disposizioni che iniziano per una disposizione (i, j) di classe 2. Le disposizioni di classe 3 che iniziano per tali elementi sono del tipo (i, j, k) con k che varia nell'insieme $\{1, 2, 3, 4, 5\} \backslash \{i, j\}$ di cardinalità $5-2=3$. Otteremo quindi in totale $60 = 20 \cdot 3$ disposizioni.

(1,2,3) (1,2,4) (1,2,5)	(2,1,3) (2,1,4) (2,1,5)	(3,1,2) (3,1,4) (3,1,5)	(4,1,2) (4,1,3) (4,1,5)	(5,1,2) (5,1,3) (5,1,4)
(1,3,2) (1,3,4) (1,3,5)	(2,3,1) (2,3,4) (2,3,5)	(3,2,1) (3,2,4) (3,2,5)	(4,2,1) (4,2,3) (4,2,5)	(5,2,1) (5,2,3) (5,2,4)
(1,4,2) (1,4,3) (1,4,5)	(2,4,1) (2,4,3) (2,4,5)	(3,4,1) (3,4,2) (3,4,5)	(4,3,1) (4,3,2) (4,3,5)	(5,3,1) (5,3,2) (5,3,4)
(1,5,2) (1,5,3) (1,5,4)	(2,5,1) (2,5,3) (2,5,4)	(3,5,1) (3,5,2) (3,5,4)	(4,5,1) (4,5,2) (4,5,3)	(5,4,1) (5,4,2) (5,4,3)

Come si contano le disposizioni-II

Quanto fatto nella nota di approfondimento precedente si generalizza al caso n e k (con $1 \le k \le n$) ottenendo

$$D_1^n = n, \qquad \text{e la formula ricorsiva} \qquad D_k^n = D_{k-1}^n\big(n-(k-1)\big)$$

da cui si ricava immediatamente che

$$D_1^n = n, \quad D_2^n = n(n-1), \quad D_3^n = n(n-1)(n-2), \quad D_k^n = n(n-1)\cdots\big(n-(k-1)\big),$$

ossia $\quad D_k^n = \frac{n!}{(n-k)!} \qquad$ e per le permutazioni $\qquad P_n = D_n^n = n!$

Relazioni tra disposizioni e combinazioni-I

Abbiamo visto che per trovare il numero C_k^n delle combinazioni di n elementi di classe k, basta osservare che

$$D_k^n = C_k^n \cdot P_k,$$

da cui immediatamente

$$C_k^n = \frac{D_k^n}{P_k} = \frac{n!}{k!\,(n-k)!}.$$

Iniziamo con il caso particolare $n = 5$ e $k = 3$. Riscriviamo tutte le 60 disposizioni mettendo in ogni riga tutte quelle che contengono gli stessi tre elementi, e poi permutiamo i tre elementi in tutti i modo possibili.

{1, 2, 3}	(1, 2, 3)	(1, 3, 2)	(2, 1, 3)	(2, 3, 1)	(3, 1, 2)	(3, 2, 1)
{1, 2, 4}	(1, 2, 4)	(1, 4, 2)	(2, 1, 4)	(2, 4, 1)	(4, 1, 2)	(4, 2, 1)
{1, 2, 5}	(1, 2, 5)	(1, 5, 2)	(2, 1, 5)	(2, 5, 1)	(5, 1, 2)	(5, 2, 1)
{1, 3, 4}	(1, 3, 4)	(1, 4, 3)	(3, 1, 4)	(3, 4, 1)	(4, 1, 3)	(4, 3, 1)
{1, 3, 5}	(1, 3, 5)	(1, 5, 3)	(3, 1, 5)	(3, 5, 1)	(5, 1, 3)	(5, 3, 1)
{1, 4, 5}	(1, 4, 5)	(1, 5, 4)	(4, 1, 5)	(4, 5, 1)	(5, 1, 4)	(5, 4, 1)
{2, 3, 4}	(2, 3, 4)	(2, 4, 3)	(3, 2, 4)	(3, 4, 2)	(4, 2, 3)	(4, 3, 2)
{2, 3, 5}	(2, 3, 5)	(2, 5, 3)	(3, 2, 5)	(3, 5, 2)	(5, 2, 3)	(5, 3, 2)
{2, 4, 5}	(2, 4, 5)	(2, 5, 4)	(4, 2, 5)	(4, 5, 2)	(5, 2, 4)	(5, 4, 2)
{3, 4, 5}	(3, 4, 5)	(3, 5, 4)	(4, 3, 5)	(4, 5, 3)	(5, 3, 4)	(5, 4, 3)

Abbiamo ottenuto quindi una tabella con 10 righe, ossia il numero dei sottoinsiemi di cardinalità 3 dell'insieme $\{1, 2, 3, 4, 5\}$ e 6 colonne, in quanto ogni riga ha 6 elementi, in quanto 6 sono le permutazioni di tre elementi. Questa tabella contiene effettivamente tutte le 60 disposizioni.

Relazioni tra disposizioni e combinazioni-II

In generale, per dimostrare che il numero delle disposizioni D_k^n è il prodotto del numero delle combinazioni C_k^n per il numero delle permutazioni P_k si può procedere osservando che le disposizioni di n elementi di classe k si possono raggruppare in modo che ciascun gruppo è formato da disposizioni che contengono gli stessi elementi, ma differiscono solo per l'ordine: ognuno di tali gruppi individua quindi il sottoinsieme degli elementi comuni, ovvero una combinazione di classe k di n elementi.
Chiaramente ciascuno dei C_k^n gruppi è composto dallo stesso numero (P_k) di disposizioni: da ciascuna combinazione di classe k di n elementi si ottengono P_k disposizioni diverse, permutando tra loro i k elementi distinti che compongono la combinazione. Quest'ultima osservazione significa appunto che vale la relazione $D_k^n = C_k^n \cdot P_k$.

Principio fondamentale del calcolo combinatorio

Come nel caso del calcolo della cardinalità dell'insieme delle Disposizioni, ossia del numero D_k^n delle disposizioni senza ripetizione, può accadere che gli elementi di un insieme G possano essere determinati con in h passi uno dopo l'altro, e in modo che per ciascun passo il numero delle possibili scelte sia un numero prefissato: al primo passo ci siano m_1 scelte possibili, al secondo passo ce ne siano $m_2, \ldots$, all'h-simo passo ce ne siano m_h. Se inoltre accade che sequenze distinte delle h scelte determinano elementi distinti dell'insieme G, in altre parole due sequenze diverse di scelte danno luogo a due elementi diversi di G, allora

$$|G| = m_1 \cdot m_2 \cdots m_h.$$

Non diamo una dimostrazione formale di questa affermazione, detta appunto **Principio fondamentale del calcolo combinatorio** è spesso usato ma bisogna stare attenti a controllare bene che effettivamente le h scelte diverse diano effettivamente luogo ad elementi diversi, altrimenti si può incorrere in errori.

3.7 Esercizi di verifica

Esercizio 3.1. Le lettere AAMMM vengono ordinate a caso. Qual è la probabilità di ottenere la parola MAMMA?

Esercizio 3.2. Si fanno n lanci di una moneta perfetta. Per $1 \leq h \leq n$, qual è la probabilità di ottenere il risultato testa per la prima volta all'h-esimo lancio?

Esercizio 3.3. Da un'urna, che contiene 6 oggetti numerati da 1 a 6, si estraggono a caso tre oggetti contemporaneamente. Qual è la probabilità che il minimo numero estratto sia superiore a 2?

Esercizio 3.4. In una mano del gioco della roulette si punta su $\{$*pair*$\}$, $\{$*passe*$\}$, $\{$*16*$\}$. Qual è la probabilità di vincere almeno una di queste puntate?

Esercizio 3.5. Qual è la probabilità che il numero 16 esca almeno una volta su cinque mani del gioco della roulette?

Esercizio 3.6. Qual è la probabilità che esca il numero 16 in una delle cinque estrazioni su una ruota del lotto? (Si estrae senza reinserimento da un'urna contenente i numeri $\{$1,2, ..., 90$\}$).

Esercizio 3.7. Qual è la probabilità che esca la coppia di numeri 16 e 48 nelle cinque estrazioni su una ruota del lotto?

Esercizio 3.8. Qual è la probabilità che esca la terna di numeri 16, 48, 90 nelle cinque estrazioni su una ruota del lotto?

Esercizio 3.9. Vengono lanciati contemporaneamente 5 dadi perfetti.
Calcolate la probabilità degli eventi elencati qui di seguito:
(a) $\{$*tutti i dadi danno punteggi diversi fra loro*$\}$
(b) $\{$*due dadi danno punteggi uguali fra loro e gli altri tre danno punteggi tutti diversi*$\}$ ("coppia")
(c) $\{$*tre dadi danno punteggi uguali fra loro e gli altri due danno due punteggi diversi*$\}$ ("tris")
(d) $\{$*quattro dadi danno punteggi uguali fra loro e uno da un punteggio diverso*$\}$ ("poker")
(e) $\{$*tutti i dadi danno lo stesso punteggio*$\}$ ("jazzi")
(f) $\{$*due diverse coppie di punteggi fra loro uguali e un punteggio diverso dagli altri due*$\}$ ("doppia coppia")
(g) $\{$*tre punteggi uguali fra loro e gli altri due uguali fra loro e diversi dal precedente*$\}$ ("full").

Esercizio 3.10. Riformulare da soli ove possibile, con gli opportuni cambiamenti, e poi risolvere l'analogo dell'esercizio precedente per il caso del lancio di soli tre dadi. (Questo esercizio si può saltare se non si sono trovate eccessive difficoltà a risolvere completamente l'esercizio precedente).

Esercizio 3.11. Un servizio da tè consiste di quattro tazzine e quattro piattini con due tazzine e due piattini di un colore e i rimanenti di un altro colore. Le tazzine sono poste a caso sopra i piattini. Calcolare le probabilità degli eventi:

$\{$*Nessuna tazzina è su un piattino dello stesso colore*$\}$

$\{$*Una sola tazzina è su un piattino dello stesso colore*$\}$

$\{$*Due sole tazzine sono su un piattino dello stesso colore*$\}$

Calcolare la probabilità dell'evento

$\{$*Nessuna tazzina su un piattino dello stesso colore*$\}$

se il servizio è composto di quattro tazzine e quattro piattini di quattro colori diversi.

4 Probabilità condizionate

In questa lezione verrà introdotta la definizione di probabilità condizionata e ne verranno illustrate alcune conseguenze immediate: la ***Formula delle probabilità composte***, la ***Formula delle probabilità totali*** e la ***Formula di Bayes***.

Abbiamo già visto che, in uno spazio di probabilità finito, la teoria della probabilità è in effetti già tutta contenuta nella formula (2.3), che mostra come si ottenga la probabilità di un evento composto, una volta assegnate le probabilità a ciascun evento elementare. Si vedrà comunque che le formule che verranno ottenute nel seguito (e la nozione di indipendenza stocastica che verrà illustrata a partire dalla prossima lezione) costituiscono spesso una guida al ragionamento probabilistico, che può rivelarsi complementare all'uso della formula (2.3)). Tali nozioni infatti permettono, alcune volte, di assegnare probabilità ad eventi composti (oppure di calcolarle sulla base di probabilità assegnate ad altri eventi) in modo più diretto, senza necessariamente far intervenire tutta la collezione degli eventi semplici. Vedremo nelle successive lezioni, in particolare, come si possano risolvere, in modo alternativo, alcuni degli esercizi affrontati nella lezione precedente.

Prima di iniziare tale studio è opportuno ricordare il significato "logico" della nozione di ***partizione di un insieme***.

Prima di tutto ricordiamo che una collezione di sottoinsiemi $H_1, ..., H_m$ dello spazio Ω ($H_\ell \in \mathcal{P}(\Omega)$, $\ell = 1, ..., m$), costituisce una partizione di Ω, se e solo se

$$\bigcup_{\ell=1}^{m} H_\ell = \Omega; \qquad H_{\ell_1} \cap H_{\ell_2} = \emptyset, \text{ per } \ell_1 \neq \ell_2.$$

Interpretando $H_1, ..., H_m$ come eventi, la condizione che $\bigcup_{\ell=1}^{m} H_\ell = \Omega$ si traduce dicendo che sono eventi ***esaustivi*** (è certo che se ne verifichi almeno uno*) e, d'altra parte, la condizione che come insiemi sono disgiunti a due a due si traduce dicenfo che sono eventi a* **due a due incompatibili** *(cioè è* certo che non se ne possono verificare due contemporaneamente*);* dunque: è certo che si verifichi uno ed uno soltanto degli eventi $H_1, ..., H_m$ (la nostra situazione di incertezza risiede nel fatto che non sappiamo **quale** di essi sia verificato).

Per i nostri scopi è importante osservare che, per qualunque evento $E \in \mathcal{P}(\Omega)$, possiamo scrivere

$$E = E \cap \Omega = E \cap \Big(\cup_{\ell=1}^{m} H\ell \Big),$$

e quindi, grazie alla proprietà distributiva dell'unione rispetto all'intersezione,

$$E = (E \cap H_1) \cup ... \cup (E \cap H_m)$$

e dunque, tenendo conto che, per ogni $i \neq j$ si ha $(E \cap H_i) \cap (E \cap H_j) \subset H_i \cap H_j = \emptyset$, grazie agli assiomi della probabilità, e in particolare per la proprietà di additività, si ha che

$$\mathbb{P}(E) = \mathbb{P}(E \cap H_1) + ... + \mathbb{P}(E \cap H_m) . \tag{28}$$

Nel caso in cui $E = \Omega$ ritroviamo, come caso particolare, la formula se $H_1, ..., H_m$ costituisce una partizione di Ω, allora deve risultare

$$\sum_{\ell=1}^{m} \mathbb{P}(H_\ell) = 1,$$

risultato che avevamo ottenuto come immediata conseguenza degli assiomi della probabilità.

4.1 Definizione di probabilità condizionata

Cominciamo con un esempio

Esempio* 4.1. *In un lancio di un dado a sei facce, quale probabilità dobbiamo assegnare all'evento $A = \{X \ dispari\}$, sapendo che si è verificato l'evento $B = \{X \geq 2\}$?*

Soluzione: Tutti gli eventi elementari

$$\{X = 1\}, \quad \{X = 2\}, ..., \quad \{X = 6\}$$

sono inizialmente ritenuti equiprobabili.

Sapere che si è verificato l'evento B equivale a sapere che si è verificato uno dei seguenti eventi elementari:

$$\{X = 2\}, ..., \{X = 6\} \quad \text{(non sappiamo però “quale”)}.$$

È naturale, a questo punto, **assumere** quanto segue: ***l'informazione che si è verificato l'evento B non modifica la situazione di equiprobabilità fra gli eventi $\{X = 2\}$,...,$\{X = 6\}$.***

A seguito di tale informazione, quindi, la probabilità di osservare l'evento A deve essere dunque valutata come la probabilità del verificarsi di uno fra 2 eventi elementari favorevoli su un totale di 5 eventi elementari possibili, equiprobabili fra loro; valuteremo quindi tale probabilità “condizionata” uguale a $\frac{2}{5}$.

La soluzione del precedente esempio mostra che, nel caso di un numero finito di eventi elementari equiprobabili, è naturale imporre che la probabilità da attribuire ad un evento A, quando si sappia per certo che si è verificato un evento B, sia data da

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

Ciò suggerisce la seguente definizione:

Definizione 4.1 (Probabilità condizionata)**.** *Siano E ed H due eventi, con $\mathbb{P}(H) > 0$. Viene detta* **probabilità condizionata di E dato H**, *ed indicata con il simbolo $\mathbb{P}(E|H)$, la quantità*

$$\mathbb{P}(E|H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}. \tag{29}$$

***Osservazione* 4.1 (di carattere euristico).** All'interno di ciascuna delle interpretazioni della probabilità (classica, frequentista, soggettivista, ...) cui si è accennato in precedenza, il numero $\mathbb{P}(E|H)$ definito nella (29) coincide effettivamente con la probabilità che, coerentemente con tale interpretazione, dovremmo attribuire al verificarsi di E, se sapessimo che si è verificato H.

Ciò costituisce la motivazione per definire “assiomaticamente” la nozione di probabilità condizionata attraverso la (29).

Esercizio proposto 4.1 (continuazione dell'Esempio 2.3, dado non equilibrato)**.** *Un dado ha sei facce numerate da 1 a 6; esso è pesato in modo tale che ciascuna faccia abbia una probabilità di presentarsi (in un singolo lancio) proporzionale al suo valore. Siano*

$$A = \{\text{Si presenta un numero pari}\}, \qquad B = \{\text{Si presenta un numero primo}\}.$$

Calcolare $\mathbb{P}(B|A)$ e $\mathbb{P}(A|B)$.

***Osservazione* 4.2.** Nel precedente Esercizio proposto 4.1 la domanda *Calcolare* $\mathbb{P}(B|A)$, si potrebbe tradurre a parole come segue:
Calcolare la probabilità condizionata che si presenti un numero primo ***dato che*** *si è presentato un numero pari*
oppure
Calcolare la probabilità condizionata che si presenti un numero primo ***sapendo che*** *si è presentato un numero pari.*
Tuttavia va sottolineato che quando la probabilità è seguita dalle parole **dato che** oppure **sapendo che** è chiaro che si tratta di probabilità condizionate e quindi di solito la parola *condizionata* viene omessa, e quindi si scrive invece, ad esempio
Calcolare la probabilità che si presenti un numero primo ***sapendo che*** *si è presentato un numero pari.*
Infine per due eventi qualsiasi E ed H spesso si dice semplicemente, **omettendo la parola condizionata** *Calcolare la probabilità di* E *dato* H oppure *Calcolare la probabilità che avrebbe l'evento* E *se si sapesse che si è verificato l'evento* H.

Interpretazione frequentista delle probabilità condizionate

Supponiamo di aver lanciato un dado per 100 volte e di aver ottenuto
16 volte il numero **1**, per 21 volte il numero **2**, per 19 volte il numero **3**, per 11 volte il numero **4**, per 16 volte il numero **5**, e infine per 17=100-(16+21+19+11+16) volte il numero **6**. Sappiamo quindi la frequenza relativa di ciascuno dei numeri da **1** a **6**:

$$f(\mathbf{1}) = \frac{16}{100}, \quad f(\mathbf{2}) = \frac{21}{100}, \quad f(\mathbf{3}) = \frac{19}{100}, \quad f(\mathbf{4}) = \frac{11}{100}, \quad f(\mathbf{5}) = \frac{16}{100}, \quad f(\mathbf{6}) = \frac{17}{100}.$$

Inoltre la frequenza relativa di B=*esce un numero primo* è data da

$$f(B) = f(\mathbf{2}) + f(\mathbf{3}) + f(\mathbf{5}) = \frac{21+19+16}{100} = \frac{56}{100},$$

e la frequenza relativa di A=*esce un numero pari* è data da

$$f(A) = f(\mathbf{2}) + f(\mathbf{4}) + f(\mathbf{6}) = \frac{21+11+17}{100} = \frac{49}{100},$$

e infine, dato che **2** l'unico numero pari e primo

$$f(A \cap B) = f(\mathbf{2}) = \frac{21}{100}.$$

Se però siamo interessati a calcolare la frequenza relativa di A dato B, ossia la frequenza con cui esce un numero pari, dato che è uscito un numero primo (ma non sappiamo quale), dovremmo restringere le nostre osservazioni ai soli casi in cui sono usciti i numeri **2**,**3** o **5**, ossia i 21+19+16=56 casi relativi e quindi, calcoleremmo tale frequenza relativa come

$$\frac{21}{56} = \frac{\frac{21}{100}}{\frac{56}{100}} = \frac{\frac{21}{100}}{\frac{21}{100} + \frac{19}{100} + \frac{16}{100}} = \frac{f(\mathbf{2})}{f(\mathbf{2}) + f(\mathbf{3}) + f(\mathbf{5})}.$$

Nella prossima sezione vedremo le semplici, ma importanti, conseguenze della definizione di probabilità condizionata, già menzionate in precedenza.

4.2 Conseguenze immediate della definizione di probabilità condizionata

4.2.1 Formula delle probabilità composte

Dalla definizione di probabilità condizionata si ottiene immediatamente che, se $\mathbb{P}(E_1) > 0$, allora

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\, \mathbb{P}(E_1). \tag{30}$$

Per ottenere la formula (30) basta moltiplicare per $\mathbb{P}(E_1)$ ambo i membri della formula $\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)}$.

Questa formula si può generalizzare.

***Proposizione* 4.1 (Formula delle probabilità composte).** *Consideriamo n eventi E_1, E_2,..., E_n, tali che*

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0. \tag{31}$$

Si ha

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2) \cdots \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}), \tag{32}$$

detta formula delle probabilità composte.

Dimostrazione. Iniziamo con l'osservare che, essendo $E_1 \supseteq E_1 \cap E_2 \supseteq \cdots \supseteq E_1 \cap E_2 \cap \cdots \cap E_{n-1}$, per la proprietà di monotonia della probabilità si ha $\mathbb{P}(E_1) \geq \mathbb{P}(E_1 \cap E_2) \geq \cdots \geq \mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$. Quindi la condizione (31) implica che $\mathbb{P}(E_1) > 0$, $\mathbb{P}(E_1 \cap E_2) > 0$, ..., $\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$, per cui il prodotto a destra della (32) ha senso.

L'uguaglianza (32) segue immediatamente dalla definizione di probabilità condizionata: possiamo scrivere

$$\mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_n) = \mathbb{P}(E_n|E_1 \cap E_2 \cap \ldots \cap E_{n-1})\, \mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-1})$$

A sua volta $\mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-1})$ può essere scritto come

$$\mathbb{P}(E_{n-1}|E_1 \cap E_2 \cap \ldots \cap E_{n-2})\, \mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_{n-2}).$$

La dimostrazione quindi si ottiene facilmente proseguendo così di seguito, fino a scrivere

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\, \mathbb{P}(E_1).$$

La precedente dimostrazione si può formalizzare utilizzando il principio di induzione.
Il caso $n = 2$ corrisponde alla formula (30). Supposta vera la formula delle probabilità composte per $n-1$ eventi, ovvero

$$\begin{aligned}
&\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) = \\
&= \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2) \cdot \cdots \cdot \mathbb{P}(E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2}).
\end{aligned}$$

mostriamo ora che vale per n eventi, e infatti

$$\begin{aligned}
&\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n) \\
&= \mathbb{P}\left((E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \cap E_n\right) \\
&= \mathbb{P}\left(E_1 \cap E_2 \cap \cdots \cap E_{n-1}\right) \cdot \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}) \\
&= \left[\mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \mathbb{P}(E_3|E_1 \cap E_2) \cdot \cdots \cdot \mathbb{P}(E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2})\right] \\
&\quad \cdot \mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}).
\end{aligned}$$

□

La formula delle probabilità composte di solito viene usata per trovare la probabilità dell'intersezione di un numero finito di eventi, specialmente quando è più facile valutare le probabilità condizionate rispetto alle probabilità dell'intersezione. Il prototipo di situazioni simili sono le estrazioni di palline una dopo l'altra, in successione, da un'urna che contiene palline di colore diverso. Ad esempio se l'urna contiene 4 palline arancioni, 2 palline bianche e 3 celesti, e si fanno 3 estrazioni senza reinserimento, e se indichiamo con

$$\begin{aligned} A_i &= \{\text{esce una pallina arancione all'}i\text{-sima estrazione}\}, \\ B_i &= \{\text{esce una pallina bianca all'}i\text{-sima estrazione}\}, \\ C_i &= \{\text{esce una pallina celeste all'}i\text{-sima estrazione}\} \end{aligned}$$

allora, per calcolare la probabilità che *la prima pallina estratta sia arancione, la seconda bianca e la terza arancione*$= A_1 \cap B_2 \cap A_3$, possiamo usare la formula delle probabilità composte

$$\mathbb{P}(A_1 \cap B_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(B_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap B_2) = \frac{4}{9} \cdot \frac{2}{8} \cdot \frac{3}{7}$$

in quanto sappiamo calcolare facilmente $\mathbb{P}(A_1) = \frac{4}{9}$, $\mathbb{P}(B_2|A_1) = \frac{2}{8}$, infatti nell'urna sono rimaste in totale 8 palline di cui 2 bianche, ed infine $\mathbb{P}(A_3|A_1 \cap B_2) = \frac{3}{7}$, in quanto dopo le prime due estrazioni di una arancione e una bianca, sono rimaste in totale 7 palline, di cui $3 = 4 - 1$ arancioni.

Questa idea viene illustrata anche nel seguente esempio (confrontare anche le osservazioni al successivo Esempio 5.2 della Lezione 5, e ancora l'Osservazione 1 della Lezione 6).

Esempio* 4.2. *Un uomo ha un mazzo di n chiavi, una sola delle quali apre la porta di casa. Egli prova le chiavi a caso ad una ad una, escludendo dal mazzo quelle già provate, finché non trova la chiave giusta. Fissato k, con $k \in \{1, 2, ..., n\}$, vogliamo trovare la probabilità dell'evento*

$$F_k = \{\textit{chiave giusta al } k\textit{-esimo tentativo}\}$$

Soluzione: Scriviamo F_k come intersezione di diversi eventi come segue:

$$F_k = \{\textit{chiave errata al } 1^o \textit{ tentativo}\} \cap \{\textit{ chiave errata al } 2^o \textit{ tentativo}\} \cap ... \cap \{\textit{chiave giusta al } k^{-simo} \textit{ tentativo}\}.$$

In simboli, ponendo $C_h = \{\textit{chiave giusta all'}h^{simo} \textit{ tentativo}\}$

$$F_k = \overline{C}_1 \cap \overline{C}_2 \cap \cdots \cap \overline{C}_{k-1} \cap C_k$$

e quindi, utilizzando la formula delle probabilità composte possiamo scrivere dunque

$$\begin{aligned} \mathbb{P}(F_k) &= \mathbb{P}(\overline{C}_1)\,\mathbb{P}(\overline{C}_2|\overline{C}_1) \cdots \mathbb{P}(\overline{C}_{k-1}|\overline{C}_1 \cap \overline{C}_2 \cap \cdots \cap \overline{C}_{k-2})\,\mathbb{P}(C_k|\overline{C}_1 \cap \overline{C}_2 \cap \cdots \cap \overline{C}_{k-1}) \\ &= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot ... \cdot \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}. \end{aligned}$$

Va sottolineato il fatto che
$\mathbb{P}(C_1)$ si potrebbe leggere come: Probabilità che *al 1^o tentativo la chiave sia giusta;*
$\mathbb{P}(C_2|\overline{C}_1)$ si potrebbe leggere come: Probabilità che *al 2^o tentativo la chiave sia giusta* sapendo che *la chiave scelta al 1^o tentativo è errata;*
$\mathbb{P}(C_3|\overline{C}_1 \cap \overline{C}_2)$ si potrebbe leggere come: Probabilità che *al 3^o tentativo la chiave sia giusta* sapendo che *sia la chiave scelta al 1^o tentativo è errata, sia la chiave scelta al 2^o tentativo è errata*, e così via.

4.2.2 Formula delle probabilità totali

Come abbiamo visto all'inizio di questa lezione, data una partizione $\{H_1, ..., H_m\}$ è possibile scrivere la probabilità di un evento E come la somma delle probabilità $\mathbb{P}(E \cap H_\ell)$ (formula (28)). Come nel caso della formula delle probabilità composte, anche la formula delle probabilità totali si mostra particolarmente utile quando è più semplice valutare le probabilità condizionate $\mathbb{P}(E|H_\ell)$ che quelle dell'intersezione $\mathbb{P}(E \cap H_\ell)$.

***Proposizione* 4.2 (Formula delle probabilità totali).** *Sia $\{H_1, ..., H_m\}$ una partizione di Ω, con $\mathbb{P}(H_\ell) > 0$, $\ell = 1, ..., m$. Per un qualunque evento $E \in \mathcal{P}(\Omega)$, risulta*

$$\mathbb{P}(E) = \mathbb{P}\left(E|H_1\right)\mathbb{P}\left(H_1\right) + ... + \mathbb{P}\left(E|H_m\right)\mathbb{P}\left(H_m\right),$$

o più brevemente

$$\mathbb{P}(E) = \sum_{k=1}^{m} \mathbb{P}\left(E|H_k\right)\mathbb{P}\left(H_k\right).$$

Dimostrazione. Per la dimostrazione basta ricordare la precedente formula (28), e applicare la formula delle probabilità composte per ciascuno degli eventi $\mathbb{P}(E \cap H_k)$, $k = 1, 2, ..., m$. Tuttavia, per comodità del lettore e per sottolineare l'importanza del fatto che gli eventi $H_1, H_2, ..., H_m$ formano una partizione, ne diamo qui la dimostrazione completa (anche se la prima parte è una ripetizione).
Gli eventi $H_1, ..., H_m$ sono esaustivi, e ciò significa che, se pensati come insiemi, la loro unione coincide con Ω: $\cup_{\ell=1}^{m} H\ell = \Omega$, e quindi, per qualunque evento $E \in \mathcal{P}(\Omega)$, possiamo scrivere

$$E = E \cap \Omega = E \cap \left(\cup_{\ell=1}^{m} H\ell \right)$$

e quindi, per la proprietà distributiva dell'unione rispetto all'intersezione

$$E = \left(E \cap H_1\right) \cup ... \cup \left(E \cap H_m\right),$$

Inoltre, gli eventi H_ℓ sono incompatibili a due a due, ossia, come eventi sono disgiunti a due a due, ossia $H_i \cap H_j = \emptyset$ per ogni $i \neq j$, possiamo affermare che anche gli eventi $E \cap H_\ell$ sono incompatibili a due a due: infatti per ogni $i \neq j$ si ha $(E \cap H_i) \cap (E \cap H_j) \subset H_i \cap H_j = \emptyset$. Di conseguenza, grazie alla proprietà di additività si ha la (28), ossia si ha che

$$\mathbb{P}(E) = \mathbb{P}\left(E \cap H_1\right) + ... + \mathbb{P}\left(E \cap H_m\right) = \sum_{k=1}^{m} \mathbb{P}(E \cap H_k).$$

e infine

$$\mathbb{P}(E) = \sum_{k=1}^{m} \mathbb{P}(E \cap H_k) = \sum_{k=1}^{m} \mathbb{P}(E|H_k)\,\mathbb{P}(H_k).$$

in quanto, grazie all'ipotesi $\mathbb{P}(H_\ell) > 0$, per ogni $\ell = 1, ..., m$, si ha che $\mathbb{P}\left(E \cap H_\ell\right) = \mathbb{P}\left(E|H_\ell\right)\mathbb{P}\left(H_\ell\right)$. □

Esempio* 4.3 (Estrazione di un numero al lotto)**.** *Qual è la probabilità che sia uguale a 16 il secondo estratto su una ruota del lotto?*

Soluzione: Indichiamo con X_1, X_2 i valori rispettivamente ottenuti nella prima e nella seconda estrazione e poniamo

$$A_1 = \{X_1 = 16\}, \quad A_2 = \{X_2 = 16\}.$$

Sia

$$E = A_2 = \{X_2 = 16\},$$

e consideriamo la partizione $\{H, \overline{H}\}$, dove

$$H = A_1 = \{X_1 = 16\}, \quad \text{(e quindi } \overline{H} = \overline{A}_1 = \{X_1 \neq 16\})$$

e applichiamo la formula delle probabilità totali; si ha così

$$\mathbb{P}(E) = \mathbb{P}(E|H)\,\mathbb{P}(H) + \mathbb{P}(E|\overline{H})\,\mathbb{P}(\overline{H}) \quad \Leftrightarrow \quad \mathbb{P}(A_2) = \mathbb{P}(A_2|A_1)\,\mathbb{P}(A_1) + \mathbb{P}(A_2|\overline{A}_1)\,\mathbb{P}(\overline{A}_1),$$

da cui otteniamo facilmente

$$\mathbb{P}(E) = \mathbb{P}(A_2) = 0 \times \frac{1}{90} + \frac{1}{89} \times \frac{89}{90} = \frac{1}{90}.$$

Osserviamo che come conseguenza $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \frac{1}{90}$, ossia che la probabilità che il numero 16 esca alla prima estrazione è uguale alla probabilità che il numero 16 esca alla seconda estrazione. Si osservi che quando si vuole calcolare $\mathbb{P}(A_2)$ si sottintende che si vuole calcolare la probabilità che il 16 esca alla seconda estrazione prima di iniziare le estrazioni (o comunque senza essere a conoscenza di quali numeri sono usciti alla prima estrazione). Quindi è ovviamente diverso da $\mathbb{P}(A_2|\overline{A_1}) = \frac{1}{89}$, che invece rappresenta la probabilità che esca alla seconda estrazione, sapendo che non è uscito il numero 16.

Esempio* 4.4. *Una moneta perfetta viene lanciata n volte. Qual è la probabilità di ottenere un numero pari di risultati testa?*

Soluzione: Poniamo, per $k = 1, 2, ..., n$

$$E_k = \{\text{numero pari di risultati testa sui primi } k \text{ lanci}\}.$$

Si capisce subito che, se la moneta è perfetta, si ha, per motivi di simmetria,

$$\mathbb{P}(E_n) = \mathbb{P}(\overline{E}_n) = \frac{1}{2}.$$

È però utile anche ragionare come segue, adoperando la formula delle probabilità totali:

$$\mathbb{P}(E_n) = \mathbb{P}(E_{n-1}) \cdot \mathbb{P}(E_n|E_{n-1}) + \mathbb{P}(\overline{E}_{n-1}) \cdot \mathbb{P}(E_n|\overline{E}_{n-1}), \tag{33}$$

e risulta

$$\mathbb{P}(E_n|E_{n-1}) = \mathbb{P}(\{\textit{risultato croce all'}n\textit{-esimo lancio}\}|E_{n-1})$$
$$\mathbb{P}(E_n|\overline{E}_{n-1}) = \mathbb{P}(\{\textit{risultato testa all'}n\textit{-esimo lancio}\}|\overline{E}_{n-1})$$

Il fatto che la moneta sia perfetta ci porta ora a valutare:

$$\mathbb{P}(E_n|E_{n-1}) = \mathbb{P}(E_n|\overline{E}_{n-1}) = \frac{1}{2}.$$

Otteniamo dunque dalla (33)

$$\mathbb{P}(E_n) = \frac{1}{2}\big(\mathbb{P}(E_{n-1}) + \mathbb{P}(\overline{E}_{n-1})\big) = \frac{1}{2}.$$

4.2.3 Formula di Bayes

Applicando la definizione di probabilità condizionata e poi la formula delle probabilità composte si ottiene, se $\mathbb{P}(E) > 0$ e $\mathbb{P}(H) > 0$,

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

che rappresenta la ***forma elementare della formula di Bayes***. Quando $H = H_\ell$, dove $\{H_1, ..., H_m\}$ è una partizione dell'evento certo questa formula si generalizza nel seguente modo.

***Proposizione* 4.3 (Formula di Bayes).** *Sia ancora $\{H_1, ..., H_m\}$ una partizione di Ω, con $\mathbb{P}(H_\ell) > 0$, $\ell = 1, ..., m$. Per un qualunque evento $E \in \mathscr{P}(\Omega)$, risulta*

$$\mathbb{P}(H_\ell|E) = \frac{\mathbb{P}(E|H_\ell)\,\mathbb{P}(H_\ell)}{\sum_{r=1}^{m}\mathbb{P}(E|H_r)\,\mathbb{P}(H_r)}, \quad \ell = 1, ..., m. \tag{34}$$

Dimostrazione. Per la definizione di probabilità condizionata, si ha

$$\mathbb{P}(H_\ell|E) = \frac{\mathbb{P}(H_\ell \cap E)}{\mathbb{P}(E)}.$$

Applicando la formula delle probabilità composte al numeratore del membro a destra otteniamo

$$\mathbb{P}(H_\ell \cap E) = \mathbb{P}(E|H_\ell)\,\mathbb{P}(H_\ell);$$

applicando la formula delle probabilità totali al denominatore otteniamo

$$\mathbb{P}(E) = \sum_{r=1}^{m}\mathbb{P}(E|H_r)\,\mathbb{P}(H_r).$$

□

Osservazione* 4.3 (Interpretazione euristica della formula di Bayes).** La formula di Bayes trova naturale applicazione nei problemi in cui si debba analizzare come "un'osservazione" o, più precisamente, come l'osservazione del verificarsi di un evento*** porti a modificare lo stato di informazione sugli eventi di una partizione; spesso problemi di tale tipo sono originati da questioni di "inferenza statistica".

Fissiamo l'attenzione su una partizione di Ω, $\mathcal{H} = \{H_1, ..., H_m\}$: sappiamo che è verificato uno ed uno soltanto degli eventi $H_1, ..., H_m$, ma non sappiamo quale.
Attribuiamo, rispettivamente, probabilità $\mathbb{P}(H_1), ..., \mathbb{P}(H_m)$ a ciascuno di tali eventi (possiamo pensare che tali probabilità esprimano il nostro stato di informazione "iniziale" sulla partizione $\mathcal{H}$).
Supponiamo successivamente di avere l'informazione che è verificato l'evento E e ci chiediamo come, in conseguenza di ciò, si debbano modificare le probabilità da attribuire agli eventi $H_1, ..., H_m$ (cioè come ciò modifichi il nostro stato di informazione "iniziale" su $\mathcal{H}$).
Tali "nuove" probabilità coincideranno con le probabilità condizionate $\mathbb{P}(H_\ell|E)$ ($\ell = 1, ..., m$), che vanno calcolate attraverso la formula (34).

In conclusione la formula di Bayes può essere vista come la regola secondo cui lo stato di informazione su $\mathcal{H}$ si modifica sulla base dell'osservazione dell'evento E.

A volte la formula di Bayes viene detta **formula delle probabilità inverse** o anche **formula delle probabilità delle cause dato l'effetto**.
L'idea è che gli eventi H_k, $k = 1, 2, ..., m$, sono le diverse *cause* che possono dare luogo al verificarsi di E, che viene pensato come un *effetto*. Si suppone di conoscere con quale probabilità si verifica ciascuna delle *cause* H_k.
Per questo motivo le $\mathbb{P}(H_k)$ sono anche dette **probabilità a priori** di H_k.
Si suppone anche di conoscere con quale probabilità si verifica E a *causa* di H_k, ossia si suppone di conoscere $\mathbb{P}(E|H_k)$. In altre parole si sa quanto è **verosimile** il verificarsi di E a *causa* di H_k.
Per questo motivo $\mathbb{P}(E|H_k)$ viene anche detta **verosimiglianza** di E dato H_k.
Se si osserva l'effetto E, per ogni $k = 1, 2, .., m$, le probabilità condizionate $\mathbb{P}(H_k|E)$ ci danno quindi la probabilità della causa H_k dato l'effetto E.
Per questo motivo $\mathbb{P}(H_k|E)$ viene anche detta **probabilità a posteriori** di H_k, ossia la probabilità dopo aver osservato l'effetto E.

Il seguente esempio costituisce un paradigma di tale uso della formula di Bayes; seguendo infatti la logica illustrata in tale esempio, la formula di Bayes può essere applicata in molti altri problemi, sostanzialmente analoghi, suggeriti in diversi campi di applicazione della teoria delle probabilità.

Esempio* 4.5. *In un lotto di pezzi (che risultano, all'apparenza, simili) vi sono elementi di tipo A, B e C, rispettivamente nelle proporzioni del 50%, 30%, 20%.*

Quelli di tipo A hanno una probabilità del 10% di guastarsi durante il loro utilizzo. Le analoghe probabilità per quelli di tipo B e C sono rispettivamente del 15% e del 18%, rispettivamente. Viene scelto un pezzo a caso dal lotto. Quale probabilità si deve attribuire al fatto che esso sia di tipo C, se si osserva un suo guasto durante l'utilizzo?

Soluzione: Poniamo

$$E := \{\textit{si osserva un guasto del pezzo scelto durante l'utilizzo}\}$$

$$H_1 := \{\textit{il pezzo scelto è di tipo } A\}, \quad H_2 := \{\textit{il pezzo scelto è di tipo } B\}, \quad H_3 := \{\textit{il pezzo scelto è di tipo } C\}.$$

La condizione che il pezzo sia scelto "a caso" si traduce nella assegnazione di probabilità

$$\mathbb{P}(H_1) = 0.5, \quad \mathbb{P}(H_2) = 0.3, \quad \mathbb{P}(H_3) = 0.2.$$

Gli altri dati del problema forniscono:

$$\mathbb{P}(E|H_1) = 0.10, \quad \mathbb{P}(E|H_2) = 0.15, \quad \mathbb{P}(E|H_3) = 0.18.$$

Dalla formula delle probabilità totali otteniamo

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(H_1)\,\mathbb{P}(E|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(E|H_2) + \mathbb{P}(H_3)\,\mathbb{P}(E|H_3)\\ &= 0.5 \times 0.10 + 0.3 \times 0.15 + 0.2 \times 0.18 = 0.131\end{aligned}$$

e la formula di Bayes fornisce:

$$\mathbb{P}(H_1|E) = \frac{50}{131}, \quad \mathbb{P}(H_2|E) = \frac{45}{131}, \quad \mathbb{P}(H_3|E) = \frac{36}{131}.$$

□

In vista del fatto che le probabilità condizionate di un evento A dato E rappresentano le probabilità che dovremmo usare, nel momento in cui ci venisse data l'informazione che l'evento E si è verificato, È del tutto naturale aspettarsi che, fissato E, al variare di A, $\mathbb{P}(A|E)$ sia anch'essa una probabilità:

***Proposizione* 4.4.** *Sia* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ *uno spazio di probabilità finito e sia* $E \in \mathcal{P}(\Omega)$ *tale che* $\mathbb{P}(E) > 0$. *Allora* $\mathbb{P}_E$ *la funzione definita da*

$$\mathbb{P}_E : \mathcal{P}(\Omega) \mapsto [0,1];\ A \mapsto \mathbb{P}_E(A) := \mathbb{P}(A|E) = \frac{\mathbb{P}(A \cap E)}{\mathbb{P}(E)}.$$

risulta essere effettivamente una probabilità.

La verifica che $\mathbb{P}_E$ così definita è una misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$ è molto semplice e la lasciamo al lettore:

Esercizio proposto 4.2 ($\mathbb{P}_E$ è una probabilità)**.** *Controllare che* $\mathbb{P}_E$ *verifica gli assiomi i), ii) e iii) della Definizione 2.1, ovvero che*
i) $\mathbb{P}_E(A) \in [0,1]$
ii) $\mathbb{P}_E(\Omega) = 1$,
iii) se A_1 *e* A_2 *sono eventi incompatibili, cioè* $A_1 \cap A_2 = \emptyset$, *allora* $\mathbb{P}_E(A_1 \cup A_2) = \mathbb{P}_E(A_1) + \mathbb{P}_E(A_2)$.

Suggerimenti per Esercizio proposto 4.2

Per il punto *i)* basta osservare che $A \cap E \subset E$ e utilizzare la proprietà di monotonia della probabilità $\mathbb{P}$.
Per il punto *ii)* basta osservare che $\Omega \cap E = E$
Per il punto *iii)* basta osservare che $(A_1 \cup A_2) \cap E = (A_1 \cap E) \cup (A_2 \cap E)$, che se $A_1 \cap A_2 = \emptyset$ allora anche $(A_1 \cap E) \cap (A_2 \cap E)$, e infine usare la proprietà di additività delle probabilità.

***Osservazione* 4.4.** La formula di Bayes (34) può essere più brevemente scritta nella forma

$$\mathbb{P}(H_\ell|E) \propto \mathbb{P}(H_\ell) \cdot \mathbb{P}\left(E|H_\ell\right), \qquad \ell = 1, ..., m,$$

ossia, a parole, le probabilità $\mathbb{P}(H_\ell|E)$ sono proporzionali a $\mathbb{P}(H_\ell) \cdot \mathbb{P}\left(E|H_\ell\right)$, ovvero esiste un valore K tale che

$$\mathbb{P}(H_\ell|E) = K\ \mathbb{P}(H_\ell) \cdot \mathbb{P}\left(E|H_\ell\right), \qquad \ell = 1, ..., m.$$

Notiamo che, essendo $\mathcal{H} = \{H_1, H_2, ..., H_m\}$ una partizione di Ω, sappiamo già a priori che deve risultare

$$\sum_{\ell=1}^{m} \mathbb{P}\left(H_\ell|E\right) = 1.$$

Infatti, considerando quanto detto nell'***Osservazione 4.3***, e nell'Esercizio Proposto 4.2, vale

$$\sum_{\ell=1}^{m} \mathbb{P}\left(H_\ell|E\right) = \sum_{\ell=1}^{m} \mathbb{P}_E\left(H_\ell\right) = 1.$$

La quantità $K = \dfrac{1}{\mathbb{P}(E)} = \dfrac{1}{\sum_{r=1}^{m} \mathbb{P}\left(E|H_r\right) \mathbb{P}\left(H_r\right)}$ ha dunque il ruolo di ***costante di normalizzazione***.

Esempio* 4.6. ***(Un esempio medico)*** *Siamo nel 1990, agli inizi del problema dell'HIV, e siamo nella seguente situazione: c'è un test (poco costoso) che permette di verificare la sieropositività per l'HIV, ma che non è infallibile, nel senso che può accadere di risultare sieropositivi, anche senza essere stati colpiti dal virus. È noto che un individuo colpito dal virus HIV risulta sieropositivo in 999 casi su 1000, mentre un individuo "sano" risulta sieropositivo in un caso su 100. Infine è noto che, nella popolazione di New York, la percentuale dei colpiti dal virus è dello $0,6\%$. John, che è stato scelto a caso tra gli abitanti di New York, si sottopone al test. Calcolare la probabilità che John risulti sieropositivo, e, nel caso in cui sia risultato sieropositivo, calcolare la probabilità che sia stato colpito veramente dal virus HIV. Ripetere i calcoli considerando la sieronegatività.*

Soluzione: Iniziamo esplicitando le probabilità sulla base dei dati del problema. Indicando con S^+ l'evento $\{John\, risulta\ sieropositivo\}$, con S^- l'evento complementare $\{John\ risulta\, sieronegativo\}$ e con H l'evento $\{John\, ha\, il\, virus\, HIV\}$, abbiamo

$$\mathbb{P}(H) = \frac{6}{1000}, \qquad \mathbb{P}(\overline{H}) = \frac{994}{1000},$$

$$\mathbb{P}(S^+|H) = \frac{999}{1000}, \quad \big(\text{quindi } \mathbb{P}(S^-|H) = \frac{1}{1000}\big), \quad \mathbb{P}(S^+|\overline{H}) = \frac{1}{100}, \quad \big(\text{quindi } \mathbb{P}(S^-|\overline{H}) = \frac{99}{100}\big).$$

Di conseguenza

$$\mathbb{P}(S^+) = \mathbb{P}(H)\, \mathbb{P}(S^+|H) + \mathbb{P}(\overline{H})\, \mathbb{P}(S^+|\overline{H}) = \frac{6}{1000}\,\frac{999}{1000} + \frac{994}{1000}\,\frac{1}{100} = 0,015934,$$

e analogamente

$$\mathbb{P}(S^-) = \mathbb{P}(H)\,\mathbb{P}(S^-|H) + \mathbb{P}(\overline{H})\,\mathbb{P}(S^-|\overline{H}) = \frac{6}{1000}\,\frac{1}{1000} + \frac{994}{1000}\,\frac{99}{100} = 0,984066\big(= 1 - \mathbb{P}(S^+)\big).$$

Dalla formula di Bayes si ha

$$\begin{aligned}\mathbb{P}(H|S^+) &= \frac{\mathbb{P}(H)\,\mathbb{P}(S^+|H)}{\mathbb{P}(H)\,\mathbb{P}(S^+|H) + \mathbb{P}(\overline{H})\,\mathbb{P}(S^+|\overline{H})} = \frac{\frac{6}{1000}\,\frac{999}{1000}}{\frac{6}{1000}\,\frac{999}{1000} + \frac{994}{1000}\,\frac{1}{100}} = \frac{\frac{6}{1000}\,\frac{999}{1000}}{\frac{6}{1000}\,\frac{999}{1000} + \frac{994}{1000}\,\frac{10}{1000}} \\ &= \frac{6 \cdot 999}{6 \cdot 999 + 994 \cdot 10} \simeq 0,37617672\end{aligned}$$

e analogamente

$$\begin{aligned}\mathbb{P}(H|S^-) &= \frac{\mathbb{P}(H)\,\mathbb{P}(S^-|H)}{\mathbb{P}(H)\,\mathbb{P}(S^-|H) + \mathbb{P}(\overline{H})\,\mathbb{P}(S^-|\overline{H})} = \frac{\frac{6}{1000}\,\frac{1}{1000}}{\frac{6}{1000}\,\frac{1}{1000} + \frac{994}{1000}\,\frac{99}{100}} = \frac{\frac{6}{1000}\,\frac{1}{1000}}{\frac{6}{1000}\,\frac{1}{1000} + \frac{994}{1000}\,\frac{990}{1000}} \\ &= \frac{6 \cdot 1}{6 \cdot 1 + 994 \cdot 990} \simeq 0,00000609.\end{aligned}$$

La probabilità $\mathbb{P}(H|S^+)$ ottenuta può sembrare sorprendentemente bassa: alcune osservazioni a questo proposito sono contenute nella seguente nota.

Alcune osservazioni sull'esempio medico 4.6

Come abbiamo già detto, la probabilità $\mathbb{P}(H|S^+)$ può sembrare sorprendentemente bassa, ma il risultato appare meno strano se si pone l'attenzione su due fatti:
(1) John non è stato sottoposto al test perché si era esposto al rischio di infezione, ma è stato scelto a caso.
(2) Il test non è simmetrico, nel senso che dà errore con frequenza diversa a seconda del caso in cui uno sia affetto dal virus o no: infatti la probabilità di errore nel caso di individuo sano, ossia la probabilità di risultare positivo al test, nonostante non si sia malati $\mathbb{P}(S^+|\overline{H}) = \frac{1}{100}$, è dieci volte il valore $\mathbb{P}(S^-|H) = \frac{1}{1000}$, ossia la probabilità di errore nel caso di individuo infetto.

Per quanto riguarda il punto 1), supponiamo invece che John sia stato a contatto con una o più persone malate di HIV. La probabilità (a priori) di essere stato colpito dal virus non sarebbe stata uguale a $6/1000$, ma molto più alta. A titolo di esempio supponiamo che sia $\mathbb{P}(H) = \mathbb{P}(\overline{H}) = 1/2$, allora l'essere risultato positivo al test avrebbe comportato il seguente calcolo:

$$\begin{aligned}\mathbb{P}(H|S^+) &= \frac{\mathbb{P}(H)\,\mathbb{P}(S^+|H)}{\mathbb{P}(H)\,\mathbb{P}(S^+|H) + \mathbb{P}(\overline{H})\,\mathbb{P}(S^+|\overline{H})} \\ &= \frac{\frac{1}{2}\,\frac{999}{1000}}{\frac{1}{2}\,\frac{999}{1000} + \frac{1}{2}\,\frac{10}{1000}} = \frac{999}{999+10} \simeq 0,99.\end{aligned}$$

Sembra quindi che questo test sia abbastanza buono nel caso in cui si abbiano seri motivi per sospettare un contagio (ovviamente sarà bene in ogni caso fare altre indagini).
Tuttavia, il risultato ottenuto nel caso di un individuo scelto a caso suggerisce il dubbio se un tale test sia valido per fare uno screening per una popolazione, nel senso che sembra dare troppi falsi allarmi. Questo accade perché il test non è simmetrico, nel senso del punto 2). Se la asimmetria non fosse stata così grande il test sarebbe stato migliore per uno screening. A titolo di esempio, sempre nel caso in cui John fosse scelto a caso, cioè con $\mathbb{P}(H) = \frac{6}{1000}$, $\mathbb{P}(\overline{H}) = \frac{994}{1000}$, la precisione del test nel caso di infezione fosse sempre la stessa, cioè $\mathbb{P}(S^-|H) = \frac{1}{1000}$, ma la precisione del test fosse migliore nel caso di non infezione, ad esempio se fosse $\mathbb{P}(S^+|\overline{H}) = \frac{1{,}5}{1000}$, allora si avrebbe

$$\begin{aligned}\mathbb{P}(H|S^+) &= \frac{\mathbb{P}(H)\,\mathbb{P}(S^+|H)}{\mathbb{P}(H)\,\mathbb{P}(S^+|H) + \mathbb{P}(\overline{H})\,\mathbb{P}(S^+|\overline{H})} \\ &= \frac{\frac{6}{1000}\,\frac{999}{1000}}{\frac{6}{1000}\,\frac{999}{1000} + \frac{994}{1000}\,\frac{1{,}5}{1000}} = \frac{6\cdot 999}{6\cdot 999 + 994\cdot 1,5} \simeq 0,8008,\end{aligned}$$

che è una probabilità abbastanza alta (ovviamente anche in questo caso sarebbero opportune ulteriori indagini).

4.3 Esercizi di verifica

Esercizio 4.1. A e B sono due eventi tali che

$$\mathbb{P}(A \cap B) = 0.3, \quad \mathbb{P}(A \cap \overline{B}) = 0.2, \quad \mathbb{P}(\overline{A} \cap B) = 0.1.$$

Calcolare $\mathbb{P}(A \cup B)$, $\mathbb{P}(A|B)$.

Esercizio 4.2. Vengono estratti a caso, senza reinserimento, due elementi dall'insieme $\{1, 2, ..., 9\}$. Poniamo

$$A_i := \{X_i \textit{ pari}\}, \; i = 1, 2.$$

Utilizzando la formula delle probabilità composte, calcolare le probabilità degli eventi $A_1 \cap A_2$, $A_1 \cap \overline{A}_2$, $\overline{A}_1 \cap A_2$, $\overline{A}_1 \cap \overline{A}_2$.

Esercizio 4.3. Nel lancio di due dadi, qual è la probabilità condizionata che nessuno dei due punteggi sia superiore a 4 sapendo che la somma dei due punteggi è uguale a 7?

Esercizio 4.4. Abbiamo due urne: l'urna U contiene una sola pallina gialla ed r palline rosse; l'urna V contiene una sola pallina rossa ed r palline gialle. Viene scelta a caso una fra queste due urne e ne estraiamo (ancora a caso) una pallina.
(a) Calcolare la probabilità dell'evento

$$E := \{\textit{la pallina estratta è gialla}\}$$

(b) Condizionatamente all'osservazione dell'evento E, qual è la probabilità di aver eseguito le estrazioni dall'urna U?

Esercizio 4.5. Nel gioco della roulette, qual è la probabilità condizionata del risultato *pair*, dato che si è ottenuto il risultato *passe*? (Ricordiamo che il risultato $\{0\}$, non è *passe*, ne' *manque*, ne' *pair*, ne' *unpair*).

Esercizio 4.6. Un'urna contiene 3 palline rosse e 7 bianche; si esegue un'estrazione casuale e se ne reinserisce una pallina di colore opposto a quella estratta; si procede quindi ad una successiva estrazione casuale.
(a) Qual è la probabilità di una pallina rossa alla seconda estrazione?
(b) Sapendo che le palline estratte nelle due successive estrazioni sono dello stesso colore, qual è la probabilità che siano entrambe bianche?

Esercizio 4.7. Sto organizzando un appuntamento per una cena fra amici per questa sera. Non riesco a raggiungere Emilio per telefono e chiedo a Aldo e a Bruno di provare ad avvertirlo. Aldo e a Bruno proveranno separatamente ad avvertirlo, Aldo inviandogli un messaggio di posta elettronica e Bruno inviando un messaggio sul telefono cellulare. Dò le seguenti valutazioni di probabilità

$$\mathbb{P}(\{\textit{Emilio leggerà la sua posta elettronica}\}) = 0.7$$

$$\mathbb{P}(\{\textit{Emilio riceverà il messaggio sul suo cellulare}\}) = 0.8$$

$$\mathbb{P}(\{\textit{Emilio leggerà la posta elettronica e riceverà il messaggio sul cellulare}\}) = 0.56.$$

(a) Come devo valutare la probabilità che Emilio venga all'appuntamento?
(b) Dato che Emilio effettivamente si presenta all'appuntamento, come devo valutare la probabilità che egli abbia letto la sua posta elettronica?

Esercizio 4.8. Relativamente a due eventi A, B, suppongo di aver assegnato le probabilità

$$\mathbb{P}(A), \quad \mathbb{P}(B), \quad \mathbb{P}(A|B).$$

Come devo valutare la probabilità che si sia verificato A, condizionatamente all'informazione che si è verificato almeno uno fra i due eventi A e B?

5 Correlazione e indipendenza fra eventi

5.1 Il caso di due eventi: correlazione positiva, negativa e indipendenza

Riprendiamo l'Esempio 4.5 della precedente Lezione 4, sull'estrazione da un lotto di pezzi apparentemente identici. Come ci si poteva già aspettare intuitivamente prima di svolgere i calcoli, risulta

$$\mathbb{P}(H_1|E) < \mathbb{P}(H_1), \quad \mathbb{P}(H_3|E) > \mathbb{P}(H_3).$$

Da tale osservazione prendiamo spunto per formulare le seguenti definizioni.

Definizione 5.1. *Siano A e B due eventi, con $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$. A e B si dicono* **correlati positivamente** *se risulta*

$$\mathbb{P}(A|B) > \mathbb{P}(A).$$

Notiamo che tale condizione è equivalente a

$$\mathbb{P}(A \cap B) > \mathbb{P}(A) \cdot \mathbb{P}(B)$$

e che, dunque, tale relazione è simmetrica.

Definizione 5.2. *Due eventi A e B, con $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, si dicono* **correlati negativamente** *se risulta*

$$\mathbb{P}(A|B) < \mathbb{P}(A)$$

oppure

$$\mathbb{P}(A \cap B) < \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Definizione 5.3. *Due eventi A e B si dicono* **stocasticamente indipendenti** *se risulta*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Notiamo che non abbiamo richiesto necessariamente la condizione $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$. Se tale condizione è verificata e A e B sono indipendenti allora risulta

$$\mathbb{P}(A|B) = \mathbb{P}(A), \quad \mathbb{P}(B|A) = \mathbb{P}(B).$$

Esempio 5.1. *Consideriamo l'esperimento relativo al lancio di due dadi, prendendo $\Omega = \big\{(i,j), \quad i,j \in \{1,2,3,4,5,6\}\big\}$. Imponiamo la condizione che ciascuno dei trentasei eventi elementari possibili abbia probabilità $\mathbb{P}(\{(i,j)\}) = \frac{1}{36}$. Denotiamo con X_1 il risultato del primo dado, e con X_2 quello del secondo dado. Consideriamo gli eventi composti $E_1 = \{X_1 \text{ pari}\}$, $E_2 = \{X_1 + X_2 \text{ pari}\}$, $E_3 = \{X_1 + X_2 \leq 4\}$, $E_4 = \{X_1 \leq 2\}$, $E_5 = \{\max(X_1, X_2) > 3\}$.*

È facile verificare che risulta:
E_1 ed E_2 sono stocasticamente indipendenti,
E_3 ed E_4 sono correlati positivamente,
E_3 ed E_5 sono correlati negativamente.

È importante a questo punto tener presente quanto segue.

***Osservazione* 5.1.** Consideriamo ancora, a mò di esempio, l'esperimento del lancio di due dadi. Si verifica immediatamente che assegnare uguali probabilità $\frac{1}{36}$ a tutti gli eventi elementari implica che gli eventi (composti) del tipo $\{X_1 = i\}$, $\{X_2 = j\}$ sono tutti equiprobabili con probabilità uguale ad $\frac{1}{6}$ e $\{X_1 = i\}$, $\{X_2 = j\}$, per $1 \leq i \leq 6$, $1 \leq j \leq 6$, costituiscono coppie di eventi indipendenti. Anche le coppie di eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$, dove I e J sono sottoinsiemi di $\{1, 2, 3, 4, 5, 6\}$, sono indipendenti (vedere la nota alla pagina seguente)

È d'altra parte immediato verificare anche il viceversa, cioè che la condizione

$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = j) = \frac{1}{6}$$

insieme all'indipendenza stocastica per tutte le coppie $\{X_1 = i\}$, $\{X_2 = j\}$ implica che tutti gli eventi elementari $\{X_1 = i\} \cap \{X_2 = j\}$ hanno probabilità uguale ad $\frac{1}{36}$.

Questa equivalenza prefigura un fatto piuttosto generale: *spesso l'assegnazione di probabilità non avviene imponendo le probabilità degli eventi elementari ma, piuttosto, imponendo che certi eventi composti abbiano probabilità assegnate ed imponendo l'indipendenza stocastica fra opportune coppie di eventi.*

Ciò può permettere di individuare quali debbano essere le corrispondenti probabilità per tutti gli eventi semplici (e quindi, attraverso la (2.3), per tutti gli eventi composti) oppure può permettere di individuare quali siano le probabilità almeno per certi eventi cui siamo effettivamente interessati.

Esempio* 5.2. *Il modello del lancio di due dadi dell'Esempio 5.1 può essere costruito alternativamente nel seguente modo. Assumiamo che*

$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = j) = \frac{1}{6}, \qquad 1 \leq i \leq 6, \quad 1 \leq j \leq 6$$

e che

$$\mathbb{P}(\{X_1 \in I\} \cap \{X_2 \in J\}) = \mathbb{P}(X_1 \in I) \cdot \mathbb{P}(X_2 \in J),$$

essendo I, J una arbitraria coppia di sottoinsiemi di $\{1, 2, ..., 6\}$.

Calcolare la probabilità dell'evento $E = \{\text{almeno un punteggio} \geq 5 \text{ nei due lanci}\}$.

Soluzione: Per calcolare $\mathbb{P}(E)$ basta osservare che $E = \{X_1 \geq 5\} \cup \{X_2 \geq 5\}$ ed utilizzare la formula di inclusione ed esclusione. Va inoltre osservato che $\mathbb{P}(X_1 \geq 5) = \mathbb{P}(X_2 \geq 5) = \frac{2}{6} = \frac{1}{3}$ e che, per ipotesi, $\mathbb{P}(\{X_1 \geq 5\} \cap \{X_2 \geq 5)\}) = \mathbb{P}(\{X_1 \geq 5\})\, \mathbb{P}(\{X_2 \geq 5)\})$:

$$\begin{aligned}
\mathbb{P}(E) &= \mathbb{P}(\{(X_1 \geq 5\} \cup \{X_2 \geq 5)\}) \\
&= \mathbb{P}(X_1 \geq 5) + \mathbb{P}(X_2 \geq 5) - \mathbb{P}(\{X_1 \geq 5\} \cap \{X_2 \geq 5)\}) \\
&= 2\,\mathbb{P}(X_1 \geq 5) - [\mathbb{P}(X_1 \geq 5)]^2 = \frac{2 \cdot 1}{3} - \frac{1}{9} = \frac{5}{9}.
\end{aligned}$$

Osserviamo che abbiamo calcolato tale probabilità senza usare la formula (2.3) che richiede di specificare quali sono gli eventi elementari che costituiscono l'evento di interesse (e quindi senza ricorrere ad un discorso di tipo combinatorio). Abbiamo invece calcolato tale probabilità imponendo i valori delle probabilità per alcuni eventi e di indipendenza stocastica fra certe coppie di eventi; tale procedimento è più sintetico, e ciò può costituire una caratteristica importante nei casi in cui $|\Omega|$ è un numero molto grande.

Osserviamo che

$$\{X_1 = i\} = \cup_{j=1}^{6}\{X_1 = i, X_2 = j\} = \{(i,1),(i,2),(i,3),(i,4),(i,5),(i,6)\}$$

e quindi, indicando con $\mathbb{P}_c$ la probabilità classica si ha

$$\mathbb{P}_c(\{X_1 = i\}) = \sum_{j=1}^{6} \mathbb{P}_c(\{X_1 = i, X_2 = j\}) = \frac{6}{36} = \frac{1}{6}.$$

Similmente $\mathbb{P}_c(\{X_2 = j\}) = 6/36 = 1/6$.
D'altra parte $\{X_1 = i\} \cap \{X_2 = j\} = \{(i,j)\}$ e quindi

$$\mathbb{P}_c(\{X_1 = i\} \cap \{X_2 = j\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}_c(\{X_1 = i\})\,\mathbb{P}_c(\{X_2 = j\})$$

ossia rispetto alla probabilità classica $\mathbb{P}_c$ gli eventi $\{X_1 = i\}$, $\{X_2 = j\}$ sono tutti equiprobabili ed indipendenti (stocasticamente).

Osserviamo inoltre che, per ogni I e J sottoinsiemi di $\{1,2,3,4,5,6\}$, si ha

$$\{X_1 \in I\} \cap \{X_2 \in J\} = \{(i,j) : \ i \in I, j \in J) = I \times J$$

e

$$\{X_1 \in I\} = \cup_{i \in I}\{X_1 = i\} \quad \text{e} \quad \{X_2 \in J\} = \cup_{j \in J}\{X_2 = j\},$$

da cui

$$\mathbb{P}_c(\{X_1 \in I\} \cap \{X_2 \in J\}) = \mathbb{P}(I \times J) = \frac{|I \times J|}{36} = \frac{|I| \cdot |J|}{36}$$

e

$$\mathbb{P}_c(\{X_1 \in I\}) = \sum_{i \in I} \mathbb{P}_c(\{X_1 = i\}) = \frac{|I|}{6} \text{ e } \mathbb{P}_c(\{X_2 \in J\}) = \sum_{j \in J} \mathbb{P}_c(\{X_2 = j\}) = \frac{|J|}{6}.$$

Di conseguenza possiamo affermare che per la probabilità classica tutte le coppie di eventi $\{X_1 \in I\}$ e $\{X_2 \in J\}$, al variare di I e J, sono indipendenti:

$$\mathbb{P}_c(\{X_1 \in I\} \cap \{X_2 \in J\}) = \frac{|I| \cdot |J|}{36} = \frac{|I|}{6} \cdot \frac{|J|}{6} = \mathbb{P}_c(\{X_1 \in I\})\ \mathbb{P}_c(\{X_2 \in J\}).$$

Viceversa se $\mathbb{P}$ denota una probabilità tale che

$$\mathbb{P}(\{X_1 = i\}) = \mathbb{P}(\{X_2 = j\}) = \frac{1}{6}$$

e che

$$\mathbb{P}(\{X_1 \in I\} \cap \{X_2 \in J\}) = \mathbb{P}(\{X_1 \in I\})\ \mathbb{P}(\{X_2 \in J\}), \quad \text{per ogni } I \text{ e } J,$$

allora

$$\mathbb{P} = \mathbb{P}_c\,.$$

Infatti, prendendo $I = \{i\}$ e $J = \{j\}$, le precedenti condizioni implicano che

$$\mathbb{P}(\{X_1 = i\} \cap \{X_2 = j\}) = \mathbb{P}(X_1 = i)\ \mathbb{P}(X_2 = j) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36};$$

quindi

$$\mathbb{P}(\{(i,j)\}) = \frac{1}{36} = \mathbb{P}_c(\{(i,j)\}), \quad \text{per ogni elemento } (i,j) \text{ di } \Omega.$$

Considerando che due probabilità coincidono se e solo se coincidono sui singoletti, abbiamo quindi mostrato che $\mathbb{P}$ coincide con la probabilità classica $\mathbb{P}_c$.

Esempio* 5.3. *Tizio ha comprato due biglietti per ciascuna di due diverse lotterie. Sono stati emessi 700 biglietti per ciascuna lotteria e, in ogni lotteria, vengono estratti tre biglietti vincenti. Quale probabilità ha di vincere almeno un premio?*

Soluzione: Si sottointende che, per ciascuna lotteria, le estrazioni sono casuali e senza reinserimento; si sottointende inoltre che vi sia indipendenza stocastica fra quello che succede nelle due lotterie diverse: dunque, ponendo $E = \{\textit{Tizio vince almeno un premio}\}$ ed $E_i = \{\textit{ Tizio vince almeno un premio nella lotteria } i\}$, con $i = 1, 2$, si ha

$$\mathbb{P}(E) = \mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2) \overset{E_1 \perp E_2}{=} \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1)\,\mathbb{P}(E_2)$$
$$\overset{\mathbb{P}(E_1)=\mathbb{P}(E_2)}{=} 2\,\mathbb{P}(E_1) - [\mathbb{P}(E_1)]^2\,.$$

Per quanto riguarda il calcolo di $\mathbb{P}(E_1)$, osserviamo che, applicando la formula delle probabilità composte, si ottiene

$$\mathbb{P}(E_1) = 1 - \mathbb{P}\left(\overline{E}_1\right) = 1 - \frac{697}{700} \cdot \frac{696}{699}.$$

Suggerimento per il calcolo della probabilità $\mathbb{P}(\overline{E}_1)$

Per calcolare la probabilità di $\overline{E}_1$ possiamo ragionare in due modi:

(I MODO) Posto A_1^i l'evento all'estrazione i-sima non viene estratto nessuno dei due biglietti si ha

$$\mathbb{P}\left(\overline{E}_1\right) = \mathbb{P}(A_1^1 \cap A_1^2 \cap A_1^3) = \mathbb{P}(A_1^1)\,\mathbb{P}(A_1^2|A_1^1)\,\mathbb{P}(A_1^3|A_1^1 \cap A_1^2) = \frac{698}{700}\,\frac{697}{699}\,\frac{696}{698} = \frac{697}{700} \cdot \frac{696}{699},$$

in quanto ci si riconduce a tre estrazioni senza reinserimento da un'urna che contiene 2 palline bianche (i due numeri dei due biglietti posseduti da Tizio, e 698 rosse, tutti gli altri: l'evento E_1 corrisponde allora all'evento nelle tre estrazioni escono solo palline rosse.

(II MODO) È interessante notare anche che a questo risultato si può arrivare anche pensando che invece si tratti di estrazioni in blocco, ossia di scegliere 2 biglietti tra i 700 di cui si sa che esattamente 3 sono vincenti, come sarebbe logico in un gratta e vinci, o una lotteria in cui si comprano biglietti su cui può essere scritta la frase *Non hai vinto ritenta* oppure *Hai vinto!*. Allora la situazione si riconduce a due estrazioni senza reinserimento da un'urna che contiene 3 palline verdi e 697 arancioni, e l'evento $\overline{E}_1$ diviene l'evento $\{$*nelle due estrazioni si estraggono solo palline arancioni*$\}$: la probabilità di non vincere diviene allora immediatamente

$$\frac{\binom{3}{0}\binom{697}{2}}{\binom{700}{2}} = \frac{697}{700} \cdot \frac{696}{699}.$$

Per finire notiamo che l'indipendenza di due eventi non è una proprietà intrisenca degli eventi ma è determinata dalla probabilità assegnata, come mostra il seguente esempio:

Esempio 5.4. *Come nell'Esempio 5.1, consideriamo l'esperimento relativo al lancio di due dadi, prendendo* $\Omega = \big\{(i, j),\ \ i, j \in \{1, 2, 3, 4, 5, 6\}\big\}$. *Imponiamo invece la condizione che i trentasei eventi elementari possibili abbiano probabilità* $\mathbb{Q}(\{(i, j)\}) = C\, i \cdot j$. *Osserviamo che, per mettere in evidenza la differenza con la situazione dell'Esempio 5.1, abbiamo usato il simbolo* $\mathbb{Q}$ *per denotare la misura di probabilità, invece di* $\mathbb{P}$. *Per calcolare il valore* *di C basta imporre che $\sum_{i=1}^{6}\sum_{j=1}^{6} Ci \cdot j = 1$, e dei semplici calcoli mostrano che $C = \frac{1}{21^2}$. Denotiamo con X_1 il risultato del primo dado, e con X_2 quello del secondo dado e, come nell'Esempio 5.1, consideriamo gli eventi composti $E_1 = \{X_1 \text{ pari}\}$ ed $E_2 = \{X_1 + X_2 \text{ pari}\}$. È facile calcolare che $\mathbb{Q}(X_1 = i) = \frac{i}{21}$, per $i \in \{1, 2, 3, 4, 5, 6\}$, e che*

$$\mathbb{Q}(E_1) = \frac{2+4+6}{21} = \frac{4}{7},$$

$$\mathbb{Q}(E_2) = \frac{1(1+3+5)+2(2+4+6)+3(1+3+5)+4(2+4+6)+5(1+3+5)+6(2+4+6)}{21^2} = \frac{25}{49},$$

$$\mathbb{Q}(E_1 \cap E_2) = \frac{2(2+4+6)+4(2+4+6)+6(2+4+6)}{21^2} = \frac{16}{49} \neq \mathbb{Q}(E_1) \cdot \mathbb{Q}(E_2),$$

e quindi, gli eventi E_1 ed E_2 non sono stocasticamente indipendenti rispetto alla misura di probabilità $\mathbb{Q}$, mentre lo sono rispetto alla misura di probabilità $\mathbb{P}$ dell'Esempio 5.1.

In quanto segue approfondiremo alcuni aspetti critici della nozione di indipendenza stocastica fra eventi. D'ora in poi verrà utilizzato il simbolo $A \perp\!\!\!\perp B$ per indicare l'indipendenza stocastica fra due eventi A e B.

5.2 Indipendenza fra due partizioni e fra due algebre di eventi

In uno spazio di probabilità, consideriamo due eventi A e B e le loro rispettive negazioni $\overline{A}$ e $\overline{B}$.

È facile verificare che le seguenti relazioni sono fra di loro equivalenti:

$$A \perp\!\!\!\perp B, \quad \overline{A} \perp\!\!\!\perp B, \quad A \perp\!\!\!\perp \overline{B}, \quad \overline{A} \perp\!\!\!\perp \overline{B},$$

ad esempio, assumiamo $A \perp\!\!\!\perp B$ e mostriamo che $\overline{A} \perp\!\!\!\perp B$; infatti possiamo scrivere

$$\begin{aligned}\mathbb{P}\left(\overline{A} \cap B\right) &= \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A) \cdot \mathbb{P}(B) \\ &= \mathbb{P}(B) \cdot [1 - \mathbb{P}(A)] = \mathbb{P}(B) \cdot \mathbb{P}\left(\overline{A}\right),\end{aligned}$$

e dunque $\overline{A} \perp\!\!\!\perp B$.

Possiamo riassumere quanto sopra affermando che prendendo un qualunque evento della partizione $\{A, \overline{A}\}$ ed un qualunque evento della partizione $\{B, \overline{B}\}$ otteniamo una coppia di eventi indipendenti.

Ciò suggerisce la seguente definizione.

Definizione 5.4. *Siano $\mathcal{A} = \{A_1, A_2, ..., A_n\}$ e $\mathcal{B} = \{B_1, B_2, ..., B_m\}$ due diverse partizioni (finite) di uno stesso spazio campione Ω. $\mathcal{A}$ e $\mathcal{B}$ sono due* **partizioni indipendenti** *se risulta*

$$A_i \perp\!\!\!\perp B_j, \quad \forall\, i = 1, ..., n;\ j = 1, 2, ..., m.$$

Esempio 5.5. *Consideriamo di nuovo l'esperimento del lancio di due dadi perfetti (o ben equilibrati) e gli eventi*

$$A_i := \{X_1 = i\}, \quad i = 1, 2, ..., 6; \qquad B_j := \{X_2 = j\}, \quad j = 1, 2, ..., 6.$$

La condizione che tutti gli eventi elementari siano equiprobabili implica, come è immediato verificare (confrontare anche l' ***Osservazione 1****), che le partizioni $\mathcal{A} := \{A_1, A_2, ..., A_6\}$ e $\mathcal{B} := \{B_1, B_2, ..., B_6\}$ sono indipendenti.*

Esercizio proposto 5.1. *Nelle condizioni del precedente Esempio 5.5, verificare che gli eventi $E = \{X_1 \leq 4\}$ ed $F = \{X_2 \text{ pari}\}$ sono indipedenti.*

Esercizio proposto 5.2. *Lanciamo due dadi e denotiamo con X_1 il risultato del primo dado, e con X_2 quello del secondo dado. Poniamo $A_i = \{X_1 = i\}$ e $B_j = \{X_2 = j\}$, per $i,\ j \in \{1,2,3,4,5,6\}$.*
Supponiamo che i dadi siano entrambi truccati in modo che $\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = Ci$, $i \in \{1,2,3,4,5,6\}$ e che le le partizioni $\mathcal{A} := \{A_1, A_2, ..., A_n\}$ e $\mathcal{B} := \{B_1, B_2, ..., B_m\}$ siano partizioni indipendenti.
(a) *Verificare che gli eventi $E = \{X_1 \leq 4\}$ ed $F = \{X_2 \text{ pari}\}$ sono indipendenti.*
(b) *Verificare che comunque presi $I,\ J \subseteq \{1,2,3,4,5,6\}$ gli eventi $E = \{X_1 \in I\}$ ed $F = \{X_2 \in J\}$ sono indipedenti.*

I precedenti esercizi sono un caso particolare del seguente risultato più generale:

***Proposizione* 5.1.** *Siano $\mathcal{A} = \{A_1, A_2, ..., A_n\}$ e $\mathcal{B} = \{B_1, B_2, ..., B_m\}$ due partizioni (finite) indipendenti di Ω. Allora, comunque scelti due eventi*

$$E = \bigcup_{i\in I} A_i \quad e \quad F = \bigcup_{j\in J} B_j,$$

dove $I = \{i_1, i_2, \cdots, i_k\} \subseteq \{1, 2, \cdots, n\}$ e $J = \{j_1, j_2, \cdots, j_h\} \subseteq \{1, 2, \cdots, m\}$, risulta

$$E \perp\!\!\!\perp F.$$

Dimostrazione. Siano

$$I = \{i_1, i_2, \cdots, i_k\} \subseteq \{1, 2, \cdots, n\} \quad \text{e} \quad J = \{j_1, j_2, \cdots, j_h\} \subseteq \{1, 2, \cdots, m\},$$

posto

$$E = \bigcup_{i\in I} A_i = \bigcup_{r=1}^{k} A_{i_r}, \quad \text{ed} \quad F = \bigcup_{j\in J} B_j = \bigcup_{s=1}^{h} B_{j_s},$$

allora

$$E \cap F = \bigcup_{i\in I} \bigcup_{j\in J} A_i \cap B_j = \bigcup_{r=1}^{k} \bigcup_{s=1}^{h} A_{i_r} \cap B_{j_s}.$$

Si osservi che finora non abbiamo utilizzato ne' il fatto che $\mathcal{A}$ e $\mathcal{B}$ sono partizioni, ne' il fatto che sono partizioni indipendenti.
Utilizziamo ora il fatto che $\mathcal{A} = \{A_1, A_2, ..., A_n\}$ e $\mathcal{B} = \{B_1, B_2, ..., B_m\}$ sono due partizioni (finite) dell'evento certo. Allora

$$\mathbb{P}(E \cap F) = \sum_{r=1}^{k} \sum_{s=1}^{h} \mathbb{P}(A_{i_r} \cap B_{j_s}),$$

come si deduce immediatamente dalla proprietà di additività delle probabilità, e dal fatto, che essendo $\mathcal{A}$ e $\mathcal{B}$ partizioni, se $(i', j') \neq (i'', j'')$ allora gli eventi del tipo $A_{i'} \cap B_{j'}$ e $A_{i''} \cap B_{j''}$ sono disgiunti.
Si osservi che per il momento abbiamo utilizzato solo il fatto che $\mathcal{A}$ e $\mathcal{B}$ sono partizioni, ma non che sono partizioni indipendenti.
A questo punto si arriva immediatamente al risultato osservando che, se le due partizioni sono indipendenti, allora

$$\mathbb{P}(E \cap F) = \sum_{r=1}^{k} \sum_{s=1}^{h} \mathbb{P}(A_{i_r} \cap B_{j_s}) = \sum_{r=1}^{k} \sum_{s=1}^{h} \mathbb{P}(A_{i_r})\, \mathbb{P}(B_{j_s}),$$

e che, dall'altra parte,

$$\mathbb{P}(E)\ \mathbb{P}(F) = \left(\sum_{r=1}^{k} \mathbb{P}(A_{i_r})\right)\left(\sum_{s=1}^{h} \mathbb{P}(B_{j_s})\right).$$

Per ottenere l'indipendenza, basta poi osservare che, dalla proprietà distributiva della somma rispetto al prodotto (vedere la nota a pagina 35) si ottiene quindi

$$\mathbb{P}(E \cap F) = \mathbb{P}(E)\ \mathbb{P}(F),$$

□

***Osservazione* 5.2.** La nozione di indipendenza fra partizioni ha, nella teoria della probabilità, un importante significato concettuale, su cui ritorneremo in seguito. Per il momento ci limitiamo ad accennare che, in un certo senso, la nozione di indipendenza stocastica esprime una relazione che si addice ad una coppia di partizioni piuttosto che ad una coppia di eventi. In ogni caso vedremo presto che la nozione di indipendenza fra due partizioni ci servirà per definire in modo semplice e concettualmente efficiente la nozione di indipendenza fra due *variabili aleatorie* (Lezioni 7 e 8).

In tale prospettiva è utile presentare qui le seguenti nozioni.

Definizione 5.5 (Algebra). *Una famiglia $\mathcal{G}$ di eventi di Ω (dunque $\mathcal{G} \subseteq \mathcal{P}(\Omega)$) è un'***algebra***, se sono verificate le seguenti proprietà*
i) $\Omega \in \mathcal{G}$
ii) $E \in \mathcal{G} \Rightarrow \overline{E} \in \mathcal{G}$
iii) $E_1, E_2 \in \mathcal{G} \Rightarrow E_1 \cup E_2 \in \mathcal{G}$

È ovvio che se $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ è un'algebra allora si ha anche che $\emptyset \in \mathcal{G}$ (usare le proprietà *i)* e *ii) con* $E = \Omega$) *e inoltre* che $E_1, E_2 \in \mathcal{G} \Rightarrow E_1 \cap E_2 \in \mathcal{G}$ (usare la Legge di De Morgan (1) e le proprietà *ii)*, *iii)*).

Esercizio proposto 5.3. *Si dimostri che l'insieme delle parti $\mathcal{P}(\Omega)$ è un'algebra.*

Esercizio proposto 5.4. *Data una partizione dell'evento certo $\mathcal{A} = \{A_1,\ A_2,\ ...,\ A_n\}$, si dimostri che la famiglia degli insiemi $E \in \mathcal{P}(\Omega)$ che sono le unioni di eventi della partizione è un'algebra: in formule si tratta della famiglia degli eventi del tipo*

$$E = \bigcup_{i \in I} A_i = A_{i_1} \cup A_{i_2} \cup \cdots \cup A_{i_k}, \quad \textit{con } I = \{i_1, \cdots, i_k\} \subseteq \{1, 2, \cdots, n\},$$

con la convenzione che se $I = \emptyset$ allora $E = \bigcup_{i \in I} A_i = \emptyset$.

Definizione 5.6 (Algebra generata da una famiglia di eventi). *Sia $\mathcal{A} = \{A_1,\ A_2, \cdots,\ A_n\}$ una famiglia di eventi in uno spazio campione Ω. Si definisce* **algebra generata** *da $\mathcal{A}$, la famiglia $\mathcal{G}(\mathcal{A})$ di eventi di Ω caratterizzata dalle seguenti proprietà:*
* $\mathcal{G}(\mathcal{A})$ *è un'algebra*
** $\mathcal{A} \subseteq \mathcal{G}(\mathcal{A})$
*** *se $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ è un'algebra e $\mathcal{A} \subseteq \mathcal{C}$ allora $\mathcal{G}(\mathcal{A}) \subseteq \mathcal{C}$.*

Possiamo dire cioè che $\mathcal{G}(\mathcal{A})$ è la più piccola famiglia di sottoinsiemi di Ω che abbia contemporaneamente le due proprietà di essere un'algebra e di contenere al suo interno tutti i sottoinsiemi della famiglia $\mathcal{A}$.

Esercizio proposto 5.5 (Algebra generata da una partizione)**.** *Si dimostri che se $\mathcal{A} = \{A_1,\ A_2, ..., \ A_n\}$ è una partizione (finita) dell'evento certo, l'algebra degli insiemi $E \in \mathscr{P}(\Omega)$ che sono le unioni di eventi della partizione è l'algebra generata da $\mathcal{A}$, ovvero che*

$$\mathcal{G}(\mathcal{A}) = \big\{E = \bigcup_{i \in I} A_i,\ I \subseteq \{1, 2, ..., n\}\big\}$$

Vale in proposito la seguente riformulazione della precedente ***Proposizione*** **5.1**

Proposizione **5.2.** *Siano $\mathcal{A}$ e $\mathcal{B}$ due partizioni (finite) indipendenti di Ω. Allora, comunque scelti due eventi $E \in \mathcal{G}(\mathcal{A}), F \in \mathcal{G}(\mathcal{B})$ risulta $E \perp\!\!\!\perp F$.*

Come applicazione della precedente ***Proposizione*** **5.2** si suggerisce di risolvere il seguente esercizio.

Esercizio proposto 5.6. *Siano $\mathcal{A} = \{A_1, A_2, ..., A_6\}$ e $\mathcal{B} = \{B_1, B_2, ..., B_6\}$ le partizioni dell'Esempio 5.5, relativo al lancio di due dadi. Si verifichi che*
(a) $\mathcal{G}(\mathcal{A}) = \big\{\{X_1 \in I\},\ per\ I \subseteq \{1, 2, 3, 4, 5, 6\}\big\}$;
(b) $\mathcal{G}(\mathcal{B}) = \big\{\{X_2 \in J\},\ per\ J \subseteq \{1, 2, 3, 4, 5, 6\}\big\}$;
Come nell'Esempio 5.5 si assuma la sola condizione che tutti gli eventi elementari sono equiprobabili, in modo che le due partizioni $\mathcal{A}$ e $\mathcal{B}$ sono indipendenti. Utilizzando la ***Proposizione*** *5.2 e i precedenti punti (a) e (b) si verifichi che*
(c) qualunque siano $I,\ J \subseteq \{1, 2, 3, 4, 5, 6\}$, gli eventi $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti.

5.3 Indipendenza completa e prove bernoulliane

Dovrebbe essere abbastanza chiaro il seguente significato intuitivo della condizione di indipendenza stocastica fra due eventi A e B: A e B sono indipendenti se il sapere con certezza che si è verificato B, o anche il sapere con certezza che non si è verificato B, non modifica le aspettative circa il verificarsi, o meno, dell'evento A.

Ovviamente si tratta di un concetto limite, di una condizione ideale, che viene assunta quale ipotesi di lavoro per ottenere delle rilevanti semplificazioni nell'analisi di un problema reale (è utile, per fare un'analogia, pensare ad esempio al concetto di *punto* in Geometria o di *punto materiale* in Meccanica: si tratta di una condizione limite, mai realizzata, ma che viene assunta ogni qualvolta sia accettabile entro una discreta approssimazione).

Come si è già detto, tale nozione è fondamentale nella costruzione di modelli probabilistici. Infatti, in pratica, nell'assegnare una misura di probabilità su uno spazio campione, si parte molto spesso dall'individuazione di famiglie di eventi a ciascuno dei quali si impone uguale probabilità e a coppie di eventi tra i quali si impone l'indipendenza stocastica, ad esempio, si assume in genere che le estrazioni del lotto su ruote diverse in una stessa settimana siano fenomeni indipendenti fra di loro, i successivi lanci di una moneta perfetta siano indipendenti fra di loro, etc.... In base a tali posizioni si può dedurre quanto vale la probabilità di alcuni eventi composti, interessanti nel problema stesso.

Vedremo comunque presto che vi sono delle situazioni naturali in cui la condizione di indipendenza è palesemente contraddetta; per ora accenniamo soltanto che ciò accade nei casi in cui una situazione di mancanza di informazione fa sì che ciascun evento osservato contenga un forte valore informativo, che si riflette sulle aspettative relative ad altri eventi connessi. Tale punto verrà sviluppato nella successiva Lezione 11 (in particolare per eventi condizionatamente indipendenti).

Veniamo ora ad aspetti tecnici della nozione di indipendenza. Dobbiamo rilevare a questo proposito che la definizione precedentemente formulata, si rivela non adeguata ad esprimere compiutamente una condizione di indipendenza reciproca fra molti eventi diversi.

Ciò è efficacemente illustrato dal seguente semplice esempio.

Esempio 5.6. *Riprendiamo ancora una volta il caso del lancio di due dadi e consideriamo gli eventi:* $A := \{X_1 \text{ pari}\}$, $B := \{X_2 \text{ pari}\}$, $C := \{X_1 + X_2 \text{ dispari}\}$. *Imponendo la condizione di equiprobabilità fra gli eventi elementari abbiamo le relazioni:* $A \perp\!\!\!\perp B$, $A \perp\!\!\!\perp C$, $B \perp\!\!\!\perp C$. *Notiamo però che ovviamente risulta*

$$\mathbb{P}(C|A \cap B) = 0.$$

Tale conclusione contrasta, naturalmente, con il significato di indipendenza e mostra l'esigenza di una definizione appropriata per il caso di più di due eventi.

Si dà allora le seguente definizione. Sia $\{E_1, E_2, ..., E_n\}$ una famiglia di eventi in uno stesso spazio di probabilità e consideriamo le partizioni $\mathcal{P}_1 := \{E_1, \overline{E}_1\}, ..., \mathcal{P}_n := \{E_n, \overline{E}_n\}$.

Definizione 5.7. *Gli eventi* $E_1, E_2, ..., E_n$ *sono una famiglia di eventi* **completamente (o globalmente) indipendenti** *se, per ogni* m *tale che* $2 \le m \le n$*, comunque presi degli indici* $\{j_1, ..., j_m\}$ *dall'insieme* $\{1, 2, ..., n\}$ *e comunque scelti degli eventi* $\widetilde{E}_{j_i} \in \mathcal{P}_{j_i}$ *(dove* $\widetilde{E}_{j_i} = E_{j_i}$ *oppure* $\widetilde{E}_{j_i} = \overline{E}_{j_i}$*) risulta*[14]

$$\mathbb{P}(\widetilde{E}_{j_1} \cap \widetilde{E}_{j_2} \cap ... \cap \widetilde{E}_{j_m}) = \mathbb{P}(\widetilde{E}_{j_1}) \cdot \mathbb{P}(\widetilde{E}_{j_2}) \cdot ... \cdot \mathbb{P}(\widetilde{E}_{j_m}). \tag{35}$$

Chiaramente la precedente definizione implica le seguenti due condizioni (36) e (37):

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap ... \cap E_{j_m}) = \mathbb{P}(E_{j_1}) \cdot \mathbb{P}(E_{j_2}) \cdot ... \cdot \mathbb{P}(E_{j_m}) \tag{36}$$
$$\text{per ogni } \{j_1, ..., j_m\} \subseteq \{1, 2, ..., n\}, \text{ con } 2 \le m \le n,$$

che si ottiene dalla (35) considerando solo i casi in cui $\widetilde{E}_{j_i} = E_{j_i}$ (e nessun caso in cui $\widetilde{E}_{j_i} = \overline{E}_{j_i}$). Si tratta di $2^n - n - 1$ condizioni, in quanto sono tante quanti i sottoinsiemi di $\{1, 2, ..., n\}$ esclusi l'insieme vuoto e i singoletti.

$$\mathbb{P}(\widetilde{E}_1 \cap \widetilde{E}_2 \cap ... \cap \widetilde{E}_n) = \mathbb{P}(\widetilde{E}_1) \cdot \mathbb{P}(\widetilde{E}_2) \cdot ... \cdot \mathbb{P}(\widetilde{E}_n), \qquad \text{con } \widetilde{E}_i = E_i \text{ oppure } \widetilde{E}_i = \overline{E}_i, \tag{37}$$

che si ottiene dalla (35) considerando solo il caso $m = n$, ma lasciando che $\widetilde{E}_j$ possa essere sia E_j che $\overline{E}_j$. Si tratta quindi di 2^n condizioni.

Per capire bene queste definizioni conviene esaminare il caso di $n = 3$ eventi.

L'INDIPENDENZA COMPLETA (o GLOBALE) NEL CASO DI $n = 3$ EVENTI

Dati tre eventi A, B, C, la condizione (36), equivale alle seguenti 4 condizioni:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C),$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C),\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C).$$

Invece la condizione (37) equivale alle seguenti 8 condizioni:

$$\begin{aligned}
&\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C), && \mathbb{P}(A \cap B \cap \overline{C}) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(\overline{C}),\\
&\mathbb{P}(A \cap \overline{B} \cap C) = \mathbb{P}(A)\,\mathbb{P}(\overline{B})\,\mathbb{P}(C), && \mathbb{P}(A \cap \overline{B} \cap \overline{C}) = \mathbb{P}(A)\,\mathbb{P}(\overline{B})\,\mathbb{P}(\overline{C}),\\
&\mathbb{P}(\overline{A} \cap B \cap C) = \mathbb{P}(\overline{A})\,\mathbb{P}(B)\,\mathbb{P}(C), && \mathbb{P}(\overline{A} \cap B \cap \overline{C}) = \mathbb{P}(\overline{A})\,\mathbb{P}(B)\,\mathbb{P}(\overline{C}),\\
&\mathbb{P}(\overline{A} \cap \overline{B} \cap C) = \mathbb{P}(\overline{A})\,\mathbb{P}(\overline{B})\,\mathbb{P}(C), && \mathbb{P}(\overline{A} \cap \overline{B} \cap \overline{C}) = \mathbb{P}(\overline{A})\,\mathbb{P}(\overline{B})\,\mathbb{P}(\overline{C});
\end{aligned}$$

infine la condizione (35) equivale a chiedere, oltre alle precedenti 8 condizioni, anche le seguenti 12 condizioni

$$\begin{aligned}
&\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B), && \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C), && \mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C),\\
&\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)\,\mathbb{P}(\overline{B}), && \mathbb{P}(A \cap \overline{C}) = \mathbb{P}(A)\,\mathbb{P}(\overline{C}), && \mathbb{P}(B \cap \overline{C}) = \mathbb{P}(B)\,\mathbb{P}(\overline{C}),\\
&\mathbb{P}(\overline{A} \cap B) = \mathbb{P}(\overline{A})\,\mathbb{P}(B), && \mathbb{P}(\overline{A} \cap C) = \mathbb{P}(\overline{A})\,\mathbb{P}(C), && \mathbb{P}(\overline{B} \cap C) = \mathbb{P}(\overline{B})\,\mathbb{P}(C),\\
&\mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(\overline{A})\,\mathbb{P}(\overline{B}), && \mathbb{P}(\overline{A} \cap \overline{C}) = \mathbb{P}(\overline{A})\,\mathbb{P}(\overline{C}), && \mathbb{P}(\overline{B} \cap \overline{C}) = \mathbb{P}(\overline{B})\,\mathbb{P}(\overline{C}),
\end{aligned}$$

[14] Per ogni m e $\{j_1, ..., j_m\}$ si tratta di 2^m condizioni, e quindi in totale le condizioni sono $\sum_{m=2}^{n} \binom{n}{m} \cdot 2^m = \sum_{m=1}^{n} \binom{n}{m} \cdot 2^m \cdot 1^{n-m} - \binom{n}{1} \cdot 2 - \binom{n}{0} \cdot 2^0 = (2+1)^n - 2 \cdot n - 1 = 3^n - 2n - 1.$, e per $n = 3$ si hanno $3^3 - 2 \cdot 3 - 1 = 27 - 6 - 1 = 20$ condizioni.

Va inoltre detto che solo tre delle quattro condizioni non implicano la quarta condizione: ossia esistono esempi di tre eventi per i quali si ha

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C),\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\,\mathbb{P}(C),$$

ma $\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$. Esistono anche esempi in cui

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B),\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\,\mathbb{P}(C),\ \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C),$$

ma $\mathbb{P}(B \cap C) \neq \mathbb{P}(B)\,\mathbb{P}(C)$.

ESEMPIO 1 Lanciamo due dadi ben equilibrati, posto X_1 il valore del primo dado e X_2 il valore del secondo dado,

$$A = \{X_1 \textit{ pari}\}, \quad B = \{X_2 \textit{ pari}\}, \quad C = \{X_1 + X_2 \textit{ dispari}\},$$

le tre condizioni sono soddisfatte in quanto

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

Tuttavia $A \cap B \cap C = \emptyset$ e quindi $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = 1/8$.

ESEMPIO 2 Un'urna ha 8 palline numerate da 1 a 8, e facciamo un'estrazione casuale. Posto X il numero estratto possiamo considerare $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$

$$A = \{X \leq 4\} = \{1, 2, 3, 4\}, \quad B = \{X \textit{ dispari}\} = \{1, 3, 5, 7\},$$
$$C = \{X \textit{ primo}\} = \{2, 3, 5, 7\}$$

Allora, chiaramente $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$,

$$A \cap B \cap C = \{3\}, \quad A \cap B = \{1, 3\}, \quad A \cap C = \{2, 3\}, \quad B \cap C = \{3, 5, 7\},$$

e quindi

$$\mathbb{P}(A \cap B \cap C) = 1/8 = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C),$$

$$\mathbb{P}(A \cap B) = 1/4 = \mathbb{P}(A)\,\mathbb{P}(B), \quad \mathbb{P}(A \cap C) = 1/4 = \mathbb{P}(A)\,\mathbb{P}(C),$$

ma A, B e C **non sono tre eventi indipendenti** in quanto

$$\mathbb{P}(B \cap C) = 3/8 \neq \mathbb{P}(B)\,\mathbb{P}(C) = 1/4.$$

***Osservazione* 5.3.** È importante notare che in alcuni testi la definizione di famiglia di eventi completamente indipendenti è data attraverso la relazione (36), mentre in altri è data attraverso la (37). Ciò è dovuto al fatto che le relazioni (36) e (37) sono equivalenti, ed entrambe implicano la (35) (risultando quindi equivalenti alla (35)). Ciò è ovvio per $n = 2$, come detto all'inizio della Sezione 5.2.

Non diamo qui la dimostrazione di questa proprietà, cioè l'equivalenza tra le tre proprietà (35), (36) e (37), ma proponiamo al lettore il seguente esercizio.

Esercizio proposto 5.7. *Nel caso di $n = 3$ eventi A, B e C, dimostrare l'equivalenza tra* (35), (36) *e* (37).
Suggerimento: *la soluzione è basata sull'osservazione che per tre eventi A, B, C si ha*

$$A \cap B = (A \cap B \cap C) \cup (A \cap B \cap \overline{C}),$$

e che quindi

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap B \cap C) + \mathbb{P}(A \cap B \cap \overline{C}),$$

da cui anche

$$\mathbb{P}(A \cap B \cap \overline{C}) = \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B \cap C).$$

Un caso assai particolare ma di notevole interesse è quello individuato dalla seguente definizione di ***schema di Bernoulli*** o delle ***prove di Bernoulli*** (detto anche delle ***prove ripetute***).

Definizione 5.8 (Schema di Bernoulli, o prove bernoulliane)**.** *Gli eventi $E_1, E_2, ..., E_n$ costituiscono delle* **prove bernoulliane** *se sono completamente indipendenti ed hanno tutti una stessa probabilità θ, con $0 < \theta < 1$.*

Se $E_1, E_2, ..., E_n$ costituiscono delle ***prove bernoulliane*** si ha dunque, in particolare, per $m \leq n$ e per qualunque sottoinsieme di indici $J = \{j_1, ..., j_m\} \subseteq \{1, 2, ..., n\}$,

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap ... \cap E_{j_m}) = \theta^m.$$

Inoltre, se $\{i_1, i_2, ..., i_k\} \subseteq \overline{J} = \{1, 2, ..., n\} \setminus \{j_1, ..., j_m\}$ allora

$$\mathbb{P}(E_{j_1} \cap E_{j_2} \cap ... \cap E_{j_m} \cap \overline{E}_{i_1} \cap \overline{E}_{i_2} ... \cap \overline{E}_{i_k}) = \theta^m \, (1 - \theta)^k.$$

5.4 Indipendenza completa di partizioni

Abbiamo già osservato che la nozione di indipendenza completa tra n eventi E_i, $i = 1, \dots, n$, si può esprimere attraverso una proprietà che coinvolge gli elementi delle partizioni $\mathcal{P}_i$, $i = 1, \dots, n$, generate da tali eventi, ossia $\mathcal{P}_i = \{E_i, \overline{E}_i\}$.

Date n partizioni $\mathcal{P}_i = \{H^i_1, H^i_2, \dots, H^i_{n_i}\}$, $i = 1, \dots, \ell$, si può generalizzare immediatamente la definizione di indipendenza completa:

Definizione 5.9. *Le partizioni $\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_n$ sono una famiglia di partizioni* **completamente (o globalmente)** indipendenti *se comunque scelto m (con $2 \leq m \leq n$), comunque estratti degli indici $\{j_1, ..., j_m\}$ dall'insieme $\{1, 2, ..., n\}$ e comunque scelti degli eventi $F_{j_i} \in \mathcal{P}_{j_i}$, per $i = 1, \dots m$, risulta*

$$\mathbb{P}(F_{j_1} \cap F_{j_2} \cap ... \cap F_{j_m}) = \mathbb{P}(F_{j_1}) \cdot \mathbb{P}(F_{j_2}) \cdot ... \cdot \mathbb{P}(F_{j_m}). \tag{38}$$

Chiaramente la precedente definizione implica la seguente condizione, che si ottiene considerando solo il caso $m = n$:
comunque scelti $F_i \in \mathcal{P}_i$, per $i = 1, \dots n$,

$$\mathbb{P}(F_1 \cap F_2 \cap ... \cap F_n) = \mathbb{P}(F_1) \cdot \mathbb{P}(F_2) \cdot ... \cdot \mathbb{P}(F_n). \tag{39}$$

Si può dimostrare che la condizione (39), per $F_i \in \mathcal{P}_i$, implica la condizione (38), ossia che le due precedenti condizioni sono equivalenti. Si può anche dimostrare che tali condizioni implicano la condizione (39), ma con, al posto degli elementi F_i delle partizioni $\mathcal{P}_i$, gli eventi G_i che appartengono a $\mathcal{G}(\mathcal{P}_i)$, l'algebra generata dalla partizione $\mathcal{P}_i$. Più precisamente vale la seguente generalizzazione della ***Proposizione* 5.2**.

***Proposizione* 5.3.** *Siano $\mathcal{P}_1, \mathcal{P}_2, ..., \mathcal{P}_n$ una famiglia di partizioni completamente (o globalmente) indipendenti. Allora, comunque scelti n eventi $G_i \in \mathcal{G}(\mathcal{P}_i)$, per $i = 1, \dots, n$, risulta*

$$\mathbb{P}(G_1 \cap G_2 \cap ... \cap G_n) = \mathbb{P}(G_1) \cdot \mathbb{P}(G_2) \cdot ... \cdot \mathbb{P}(G_n). \tag{40}$$

Non ne diamo la dimostrazione, tuttavia facciamo notare che dalla (40) si può ottenere la seguente relazione: comunque scelto m (con $2 \leq m \leq n$), comunque estratti degli indici $\{j_1, ..., j_m\}$ dall'insieme $\{1, 2, ..., n\}$ e comunque scelti degli eventi $G_{j_i} \in \mathcal{G}\big(\mathcal{P}_{j_i}\big)$, per $i = 1, \dots m$, risulta

$$\mathbb{P}(G_{j_1} \cap G_{j_2} \cap ... \cap G_{j_m}) = \mathbb{P}(G_{j_1}) \cdot \mathbb{P}(G_{j_2}) \cdot ... \cdot \mathbb{P}(G_{j_m}).$$

Basta infatti considerare $G_\ell = \Omega$ per $\ell \notin \{j_1, \dots, j_m\}$ e applicare la (40).

Esempio 5.7. *Si lanciano tre dadi e sia X_i il risultato dell'i-esimo lancio. Si considerino le partizioni*

$$\mathcal{P}_i = \big\{\{X_i = 1\}, \{X_i = 2\}, \{X_i = 3\}, \{X_i = 4\}, \{X_i = 5\}, \{X_i = 6\}\big\}, \qquad i = 1, 2, 3.$$

Chiaramente, comunque scelti j_1, j_2, e j_3 si ha

$$\mathbb{P}(\{X_1 = j_1,\, X_2 = j_2,\, X_3 = j_3\}) = \frac{1}{216} = \mathbb{P}(\{X_1 = j_1\})\, \mathbb{P}(\{X_2 = j_2\})\, \mathbb{P}(\{X_3 = j_3\})$$

*e quindi, osservando che ciascuno degli eventi $\{X_1 = j_1, X_2 = j_2, X_3 = j_3\}$ coincide con l'evento $\{X_1 = j_i\} \cap \{X_2 = j_2\} \cap \{X_3 = j_3\}$, e tenuto conto del fatto che la relazione (39) è verificata, e che è equivalente alla (36), le tre partizioni sono indipendenti. Inoltre, grazie alla **Proposizione** 5.3, risulta anche che, comunque scelti tre sottoinsiemi I_1, I_2 ed I_3 di $\{1, 2, 3, 4, 5, 6\}$, si ha*

$$\mathbb{P}\big(\{X_1 \in I_1\} \cap \{X_2 \in I_2\} \cap \{X_3 \in I_3\}\big) = \mathbb{P}\big(\{X_1 \in I_1\}\big)\, \mathbb{P}\big(\{X_2 \in I_2\}\big)\, \mathbb{P}\big(\{X_3 \in I_3\}\big).$$

5.5 Esercizi di verifica

Esercizio 5.1. A e B sono due eventi tali che

$$\mathbb{P}(A \cap B) = 0.3, \quad \mathbb{P}(A \cap \overline{B}) = 0.2, \quad \mathbb{P}(\overline{A} \cap B) = 0.1.$$

Verificare se A e B sono stocasticamente indipendenti.

Esercizio 5.2. Mostrare che se A e B sono due eventi indipendenti e $A \subseteq B$, allora si ha $\mathbb{P}(A) = 0$, oppure $\mathbb{P}(B) = 1$.

Esercizio 5.3. Siano A e B due eventi fra loro incompatibili. Mostrare che A e B risultano stocasticamente indipendenti se e solo se almeno uno di essi ha probabilità nulla.

Esercizio 5.4. Indichiamo con X ed Y i punteggi ottenuti nel lancio di due dadi a sei facce. Poniamo

$$A := \{\max(X, Y) < 5\}, \quad B := \{\min(X, Y) > 3\}$$

(a) Calcolare $\mathbb{P}(A \cap B)$, $\mathbb{P}(A \cup B)$, $\mathbb{P}(A|B)$, $\mathbb{P}(B|A)$.
(b) Gli eventi A e B sono indipendenti? A e B sono incompatibili?

Esercizio 5.5. Consideriamo i due risultati *pair* e *passe* nel gioco della roulette. Sono stocasticamente indipendenti, correlati positivamente o correlati negativamente?

Esercizio 5.6. Indichiamo con X un numero selezionato a caso nell'insieme dei primi 120 numeri naturali e consideriamo gli eventi

$$E := \{X \textit{ pari}\}, \quad F := \{X \textit{ divisibile per 3}\}.$$

Fra E ed F sussiste correlazione positiva, negativa o indipendenza stocastica?

Esercizio 5.7. (a) Qual è la probabilità che il numero 16 venga estratto su una data ruota del lotto in una fissata giornata?
(b) Qual è la probabilità che il numero 16 non venga mai estratto su una data ruota del lotto per n giornate consecutive?
(c) Qual è la probabilità condizionata che il numero 16 venga estratto l'$(n+1)$-esima giornata, dato che non è mai stato estratto nelle n giornate precedenti?

Esercizio 5.8. Non riesco ad avvertire Emilio per telefono dell'appuntamento per la cena di questa sera. Come al solito Aldo gli invierà allora un messaggio di posta elettronica, Bruno gli invierà un messaggio sul telefono cellulare, e interverrà anche Carla, cercando di avvertire di persona la sorella di Emilio. Essi avranno successo rispettivamente con probabilità $\mathbb{P}(A) = 0.8$, $\mathbb{P}(B) = 0.7$ e $\mathbb{P}(C) = 0.6$; i tre eventi, inoltre, sono completamente indipendenti.
(a) Trovare la probabilità che Emilio venga informato dell'appuntamento.
(b) Dato che Emilio si presenta effettivamente all'appuntamento, come devo valutare la probabilità che egli abbia letto la sua posta elettronica?

Esercizio 5.9. Una moneta viene lanciata n volte. Supponiamo che valga

$$\mathbb{P}(\underbrace{T, T, ..., T}_{n \text{ lanci}}) = \mathbb{P}(\underbrace{C, C, ..., C}_{n \text{ lanci}}) = \frac{1}{2}.$$

Posto, per $i = 1, \cdots, n$, $E_i = \{$ *esce testa all'i-simo lancio* $\}$,
(a) i due eventi E_1 ed E_2 sono indipendenti, correlati positivamente o correlati negativamente?
(b) i due eventi E_1 ed $\overline{E}_2$ sono indipendenti, correlati positivamente o correlati negativamente?
(c) al variare di n, qual è la probabilità che vi sia un numero pari di risultati testa sugli n lanci? (considerando 0 come un numero pari)

Esercizio 5.10. Siano E_1 ed E_2 due eventi in uno spazio campione Ω. Elencate gli eventi appartenenti all'algebra generata da $\{E_1, E_2\}$.

6 Probabilità binomiali e ipergeometriche; estrazioni casuali da urne di composizione nota

In questo paragrafo vogliamo discutere in modo più sistematico due particolari modelli probabilistici, che sono già comparsi in precedenti esempi e che portano alle probabilità binomiali e ipergeometriche.

6.1 Probabilità binomiali

Consideriamo, su uno spazio finito di probabilità, n prove bernoulliane $E_1, ..., E_n$; cioè, ricordando la Definizione 5.8, assumiamo che $E_1, ..., E_n$ siano completamente indipendenti ed equiprobabili: ponendo $\mathbb{P}(E_i) = \theta$, per $i = 1, 2..., n$ e ponendo

$$X_i = \begin{cases} 1 & \text{se si verifica } E_i \\ 0 & \text{se si verifica } \overline{E}_i \end{cases} \qquad i = 1, 2, \cdots, n, \tag{41}$$

si ha, per ogni n-upla $\mathbf{x} := (x_1, ..., x_n) \in \{0, 1\}^n$,

$$\mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\}) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, \tag{42}$$

dove $\{X_1 = x_1, ..., X_n = x_n\}$ è un modo rapido di scrivere l'evento $\{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}$.

Per capire la relazione (42) facciamo un esempio: sia $n = 5$, e $\mathbf{x} = (0, 1, 1, 0, 1)$,

$$\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1\} = \overline{E}_1 \cap E_2 \cap E_3 \cap \overline{E}_4 \cap E_5$$

e quindi

$$\mathbb{P}(\{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1\}) = \mathbb{P}(\overline{E}_1 \cap E_2 \cap E_3 \cap \overline{E}_4 \cap E_5)$$
$$= \mathbb{P}(\overline{E}_1)\, \mathbb{P}(E_2)\, \mathbb{P}(E_3)\, \mathbb{P}(\overline{E}_4)\, \mathbb{P}(E_5) = (1-\theta)\theta\theta(1-\theta)\theta = \theta^3\,(1-\theta)^2$$

Poniamo ora

$$S_n = \sum_{i=1}^n X_i$$

e consideriamo, per $k = 0, 1, ..., n$, la probabilità dell'evento composto $\{S_n = k\}$; osserviamo che potremo scrivere

$$\{S_n = k\} = \{\textstyle\sum_{i=1}^n X_i = k\} = \displaystyle\bigcup_{\mathbf{x} \in \{0,1\}^n : \sum_{i=1}^n x_i = k} \{X_1 = x_1, ..., X_n = x_n\}.$$

Due eventi del tipo

$$\{X_1 = x'_1, ..., X_n = x'_n\}, \quad \{X_1 = x''_1, ..., X_n = x''_n\}$$

sono ovviamente incompatibili nel caso $\mathbf{x}' = (x'_1, ..., x'_n) \neq \mathbf{x}'' = (x''_1, ..., x''_n)$; inoltre, nel caso in cui

$$\textstyle\sum_{i=1}^n x'_i = \sum_{i=1}^n x''_i = k,$$

essi risultano equiprobabili, entrambi di probabilità $\theta^k(1-\theta)^{n-k}$, in virtù dell'equazione (42). E dunque, dal momento che la cardinalità dell'insieme $\{\mathbf{x} \in \{0,1\}^n : \sum_{i=1}^n x_i = k\}$ è uguale a $\binom{n}{k}$, potremo scrivere, per $k = 0, 1, ..., n$,

$$\mathbb{P}(\{S_n = k\}) = \sum_{\mathbf{x} \in \{0,1\}^n : \sum_{i=1}^n x_i = k} \mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\}) = \binom{n}{k} \theta^k (1-\theta)^{n-k}. \tag{43}$$

Per capire la relazione (43), ossia $\mathbb{P}(S_n = k)$, per $k = 0, 1, ..., n$, facciamo un esempio: sia $n = 5$, e $k = 3$

$$\begin{aligned}
&\{S_5 = 3\} \\
= \quad &\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0\} \cup \{X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0\} \\
&\cup \{X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 1\} \cup \{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 0\} \\
&\cup \{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0, X_5 = 1\} \cup \{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 1, X_5 = 1\} \\
&\cup \{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 0\} \cup \{X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1\} \\
&\cup \{X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1\} \cup \{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1\}
\end{aligned}$$

ossia abbiamo considerato solo le $\binom{5}{3}$ 5-ple in cui compaiono 3 "uno" e 2 "zero", ossia le $(x_1, x_2, x_3, x_4, x_5)$ tali che, $x_i \in \{0, 1\}$, e $x_1 + x_2 + x_3 + x_4 + x_5 = 3$.
Inoltre

$$\begin{aligned}
&\mathbb{P}(\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0\}) = \mathbb{P}(\{X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0\}) \\
=&\cdots \\
=&\mathbb{P}(\{X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 1\}) = \mathbb{P}(\{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 1, X_5 = 1\}) \\
=&\theta^3 (1 - \theta)^2,
\end{aligned}$$

da cui

$$\mathbb{P}(S_5 = 3) = \binom{5}{3} \theta^3 (1 - \theta)^2$$

Una probabilità del tipo in (43) sull'insieme $\{0, 1, \ldots, n\}$, prende il nome di ***probabilità binomiale*** di parametri n e θ, ossia una misura di probabilità $\mathbf{P}$ definita sull'insieme delle parti di $\{0, 1, 2, ..., n\}$ come segue

$$\mathbf{P} : \mathscr{P}\big(\{0, 1, 2, ..., n\}\big) \to [0, 1], \quad I \mapsto \mathbf{P}(I) \overset{def}{=} \sum_{k \in I} p(k),$$

dove $p(k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}$, $k \in \{0, 1, 2, ..., n\}$.

Esempio* 6.1. *Un dado viene lanciato* 10 *volte. Sia S il numero di volte in cui si ottiene il risultato asso. Calcolare* $\mathbb{P}(\{S = 5\})$.

Soluzione: Si tratta di 10 prove bernoulliane, di probabilità $\frac{1}{6}$ (infatti in modo del tutto analogo all'Esempio 5.7 anche in questo caso eventi relativi a lanci diversi sono indipendenti); dunque

$$\mathbb{P}(\{S = 5\}) = \binom{10}{5} \cdot \left(\frac{5}{6}\right)^5 \cdot \left(\frac{1}{6}\right)^5 = \binom{10}{5} \cdot \frac{5^5}{6^{10}}.$$

Esempio* 6.2. *Ciascun viaggiatore che occupa un posto in uno scompartimento "per fumatori" in un treno EuroStar è effettivamente un fumatore (o fumatrice) con probabilità uguale al* 70%*. Se lo scompartimento contiene* 5 *posti prenotati (oltre a quello da me prenotato), qual è la probabilità che io vi incontri meno di tre fumatori?*

Soluzione: Si sta sottointendendo che lo scompartimento venga riempito e che i viaggiatori si comportino, rispetto al fumo, in modo ciascuno indipendente dall'altro. Se S indica il numero dei fumatori nei 5 posti rimanenti, si ha:

$$\begin{aligned}
\mathbb{P}(\{S < 3\}) = \quad & \mathbb{P}(\{S = 0\}) + \mathbb{P}(\{S = 1\}) + \mathbb{P}(\{S = 2\}) \\
= \quad & \binom{5}{0} \left(\frac{7}{10}\right)^0 \left(\frac{3}{10}\right)^5 + \binom{5}{1} \left(\frac{7}{10}\right)^1 \left(\frac{3}{10}\right)^4 + \binom{5}{2} \left(\frac{7}{10}\right)^2 \left(\frac{3}{10}\right)^3.
\end{aligned}$$

Esempio* 6.3. *Un testo, contenente 20 errori di stampa, viene sottoposto a due diversi correttori di bozze. Ciascun errore contenuto nel testo viene individuato da ciascun correttore con probabilità $p = 0.6$ ed indipendentemente da quello che accade per gli altri errori. Inoltre i due correttori lavorano indipendentemente uno dall'altro.*

Trovare la probabilità che il numero degli errori individuati da almeno uno dei due correttori sia superiore a 15.

Soluzione: Ciascun errore viene individuato (da almeno uno dei due correttori) con probabilità[15]

$$\theta = 2p - p^2 = \frac{84}{100}$$

(non ci interessa se viene individuato da uno dei due correttori o dall'altro o, eventualmente, da entrambi; a noi interessa che almeno uno dei due individui l'errore).

Si tratta quindi di 20 prove bernuolliane, in ognuna delle quali vi è una probabilità di successo uguale a θ. Indicando dunque con S il numero complessivo degli errori individuati, si ha

$$\mathbb{P}(\{S > 15\}) = \sum_{k=16}^{20} \binom{20}{k} \left(\frac{84}{100}\right)^k \left(\frac{16}{100}\right)^{20-k}.$$

6.2 Estrazioni casuali da urne con reiserimento

Illustriamo ora una tipica situazione in cui si incontra uno schema di prove bernoulliane.

Pensiamo ad una popolazione composta di M oggetti, diciamo $c_1, ..., c_M$, di cui m_1 di tipo A e i rimanenti $m_2 = M - m_1$ di un diverso tipo B. Supponiamo, ad esempio, di aver numerato $c_1, ..., c_M$ in modo tale che

$$c_1, ..., c_{m_1} \text{ sono di tipo A} \quad \text{e} \quad c_{m_1+1}, ..., c_M \text{ di tipo B.} \tag{44}$$

> Per fissare le idee si può pensare ad estrazioni con reinserimento da un'urna che contiene $M = 7$ palline,
> di cui $m_1 = 4$ sono rosso amaranto e sono numerate $\boxed{1}$, $\boxed{2}$, $\boxed{3}$, $\boxed{4}$,
> mentre le rimanenti $M - m_1 = 7 - 4 = 3$ sono blu e sono numerate $\boxed{5}$, $\boxed{6}$ e $\boxed{7}$.

Eseguiamo ora n estrazioni *casuali con reinserimento* da tale popolazione; con tale termine si vuole esprimere il fatto che si eseguono n estrazione successive, reinserendo ogni volta nella popolazione l'oggetto estratto ed estraendo in modo tale che **ciascun** oggetto abbia, ogni volta, la **stessa probabilità** (uguale a $\frac{1}{M}$) di essere estratto, sia esso di tipo A o di tipo B.

Lo spazio campione in tale esperimento (consistente nell'eseguire le n estrazioni) può essere identificato come l'insieme

$$\Omega := \{1, ..., M\}^n = \big\{(j_1, j_2, ..., j_n) : j_k \in \{1, 2, ..., M\}, \text{per ogni } k = 1, 2, ..., n\big\}$$

costituito delle n-uple ordinate di elementi in $\{1, ..., M\}$; esso ha dunque cardinalità M^n; il numero j_1 rappresenta il numero del primo oggetto estratto, il numero j_2 rappresenta il numero del secondo oggetto estratto, e così via fino a j_n, che rappresenta il numero dell'$n - simo$ oggetto estratto.

> Per fissare le idee, pensando ad $n = 3$ estrazioni con reinserimento da un'urna che contiene $M = 7$ palline, di cui $m_1 = 4$ sono rosso amaranto e le rimanenti 3 sono blu,
> l'elemento $(3, 7, 7) \in \Omega = \{1, 2, 3, 4, 5, 6, 7\}^3$ rappresenta la seguente estrazione:
> prima pallina estratta $\boxed{3}$, seconda pallina estratta $\boxed{7}$, terza pallina estratta $\boxed{7}$.

[15] Il ragionamento per arrivare al calcolo di $\theta = 2p - p^2$ è simile a quello usato negli Esempi 5.2 e 5.3

In tale spazio campione, consideriamo, per $i = 1, ..., n$, ora gli eventi del tipo:

$$A_i = \{l'oggetto\ estratto\ nella\ i - sima\ estrazione\ è\ di\ tipo\ \mathrm{A}\}.$$

Per fissare le idee, pensando ad $n = 3$ estrazioni con reinserimento da un'urna che contiene $M = 7$ palline, di cui $m_1 = 4$ sono rosso amaranto e le rimanenti 3 sono blu.
Allora l'evento $\{$*l'oggetto estratto nella seconda estrazione è di tipo* $\mathrm{A}\}$ è rappresentato dall'insieme

$$A_2 = \{1, 2, 3, 4, 5, 6, 7\} \times \{1, 2, 3, 4\} \times \{1, 2, 3, 4, 5, 6, 7\}$$

ossia

$$A_2 = \big\{(j_1, j_2, j_3) : \ j_1 \in \{1, 2, 3, 4, 5, 6, 7\}, j_2 \in \{1, 2, 3, 4\}, j_3 \in \{1, 2, 3, 4, 5, 6, 7\}\big\}$$

ovvero, ricordando che $\Omega = \{1, 2, 3, 4, 5, 6, 7\}^3$,

$$A_2 = \big\{(j_1, j_2, j_3) \in \Omega : \ 1 \leq j_2 \leq 4\big\}.$$

***Proposizione* 6.1.** *Gli eventi $A_1, ..., A_n$ (definiti qui sopra) costituiscono delle prove bernoulliane con $\mathbb{P}(A_i) = \frac{m_1}{M}$.*

Dimostrazione. Per vedere che gli eventi $A_1, ..., A_n$ costituiscono delle prove bernoulliane dobbiamo mostrare che hanno tutti la stessa probabilità e che sono completamente indipendenti.
La condizione che le estrazioni siano casuali con reinserimento corrisponde all'assegnazione della stessa probabilità $\frac{1}{M^n}$ a ciascuno degli eventi elementari $(j_1, ..., j_n) \in \Omega$. Inoltre, in virtù della posizione (44) possiamo scrivere gli eventi A_i come segue

$$\begin{aligned} A_i &:= \{(j_1, ..., j_n) \in \Omega : \ 1 \leq j_i \leq m_1\} \\ &= \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{i-1\, volte} \times \{1, 2, \cdots, m_1\} \times \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{n-i\, volte}, \end{aligned}$$

ed in modo analogo

$$\begin{aligned} \overline{A}_i &:= \{(j_1, ..., j_n) \in \Omega : \ m_1 + 1 \leq j_i \leq M\} \\ &= \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{i-1\, volte} \times \{m_1 + 1, m_1 + 2, \cdots, m_1 + m_2\} \times \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{n-i\, volte}. \end{aligned}$$

Dunque, considerando per $1 \leq i \leq n$, la quantità X_i come definita nella (41), si ha

$$\begin{aligned} \mathbb{P}(\{X_i = 1\}) = \mathbb{P}(A_i) &= \frac{|\{(j_1, ..., j_n) \in \Omega : \ 1 \leq j_i \leq m_1\}|}{M^n} \\ &= \frac{m_1 \cdot M^{n-1}}{M^n} = \frac{m_1}{M}, \end{aligned}$$

e, analogamente

$$\mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\}) = \frac{m_1^{\sum_{i=1}^n x_i} m_2^{n - \sum_{i=1}^n x_i}}{M^n} = \left(\frac{m_1}{M}\right)^{\sum_{i=1}^n x_i} \left(\frac{m_2}{M}\right)^{n - \sum_{i=1}^n x_i}.$$

Mettendo insieme le due relazioni dimostrate si ottiene che, per tutti gli eventi

$$\widetilde{A}_i \in \mathcal{P}_i = \{A_i, \overline{A}_i\} = \Big\{\{X_i = 1\}, \ \{X_i = 0\}\Big\}$$

si ha

$$\mathbb{P}(\widetilde{A}_1 \cap \widetilde{A}_2 \cap ... \cap \widetilde{A}_n) = \mathbb{P}(\widetilde{A}_1) \cdot \mathbb{P}(\widetilde{A}_2) \cdot ... \cdot \mathbb{P}(\widetilde{A}_n),$$

che corrisponde alla relazione (37), che come osservato in Osservazione 3 della Lezione 5, è equivalente alla relazione (35) che definisce l'indipendenza completa. □

Una precisazione sulla dimostrazione della precedente Proposizione 6.1

Al lettore più attento la dimostrazione potrebbe sembrare non soddisfacente, in quanto abbiamo dimostrato la condizione (37), senza aver dato la dimostrazione che essa è equivalente alla (35), che abbiamo preso come definizione dell'indipendenza completa. Tuttavia va osservato che, in modo assolutamente analogo, si potrebbe dimostrare direttamente la (35), ossia verificare che per ogni $2 \leq m \leq n$,

$$\begin{aligned}\mathbb{P}(\{X_{j_1} = x_{j_1}, ..., X_{j_m} = x_{j_m}\}) &= \frac{m_1^{\sum_{i=1}^m x_{j_i}} m_2^{m-\sum_{i=1}^m x_{j_i}} M^{n-m}}{M^n} \\ &= \left(\frac{m_1}{M}\right)^{\sum_{i=1}^m x_{j_i}} \left(\frac{m_2}{M}\right)^{m-\sum_{i=1}^m x_{j_i}} \\ &= \prod_{i=1}^m \mathbb{P}(\{X_{j_i} = x_{j_i}\}).\end{aligned}$$

La perfetta conoscenza della composizione della popolazione da cui vengono estratti gli oggetti è fondamentale per la proprietà di completa indipendenza. Come vedremo nella nota a pagina 81 alla fine di questa Lezione e nella Lezione 11, le estrazioni da un'urna con composizione non nota, invece non sono eventi indipendenti.

Consideriamo ora

$$S_n = \sum_{i=1}^n X_i,$$

in modo che S_n rappresenta il numero di elementi di tipo A in un campionamento casuale *con reinserimento* di n oggetti da una popolazione complessivamente costituita da M elementi, di cui m_1 di tipo A e $m_2 = M - m_1$ di tipo B. Ricordando la formula (43) ed in virtù della ***Proposizione* 6.1**, possiamo concludere scrivendo

$$\mathbb{P}(\{S_n = k\}) = \binom{n}{k} \left(\frac{m_1}{M}\right)^k \left(\frac{m_2}{M}\right)^{n-k}, \qquad \text{per } k = 0, 1, \cdots, n. \tag{45}$$

6.3 Estrazioni casuali da urne senza reiserimento e Probabilità ipergeometriche

Consideriamo ora la stessa situazione come descritta nel paragrafo precedente, cioè con una popolazione di M individui, numerati da 1 ad M, di cui gli m_1 individui di tipo A sono numerati da 1 ad m_1, e i rimanenti sono numerati da $m_1 + 1$ fino ad M, come in (44), ma con la differenza che le n estrazioni siano eseguite ***senza reinserimento***. Poniamo di nuovo

$$A_i := \{\textit{oggetto di tipo } \mathrm{A} \textit{ alla i-esima estrazione}\}, \qquad i = 1, 2, ..., n$$

$$X_i = \begin{cases} 1 & \text{se si verifica } A_i, \\ 0 & \text{se si verifica } \overline{A}_i, \end{cases} \qquad i = 1, 2, ..., n.$$

$$S_n = \sum_{i=1}^n X_i.$$

Nel caso di estrazioni senza reinserimento potremo considerare come spazio campione lo spazio costituito dalle n-uple ordinate di elementi di $\{1, ..., M\}$, senza ripetizione (ossia le disposizioni di M elementi di classe n):

$$\Omega := \{(j_1, ..., j_n) : \ 1 \leq j_i \leq M, \ j_1 \neq ... \neq j_n\} = \{(j_1, ..., j_n) : \ 1 \leq j_i \leq M, \ j_1, \cdots, j_n \text{ tutti distinti}\},$$

e dunque la cardinalità di Ω è

$$|\Omega| = M\big(M-1\big)...\big(M-(n-1)\big) = M(M-1)...(M-n+1) = \frac{M!}{(M-n)!}.$$

Lo schema di estrazioni casuali senza reinserimento si traduce nella condizione che tutti gli elementi di Ω (eventi elementari) hanno uguale probabilità

$$\frac{1}{M(M-1)...(M-n+1)} = \frac{(M-n)!}{M!}.$$

Un qualunque evento composto, della forma $\{X_1 = x_1, ..., X_n = x_n\}$, ha una cardinalità che dipende soltanto dal numero $k = \sum_{i=1}^n x_i$ ed, esattamente, è data da

$$\begin{aligned}
&\overbrace{m_1 \cdot \big(m_1-1\big) \cdot ... \cdot \big(m_1-(k-1)\big)}^{k \ fattori} \cdot \overbrace{m_2 \cdot \big(m_2-1\big) \cdot ... \cdot \big(m_2-(n-k-1)\big)}^{n-k \ fattori} \\
=& m_1 \cdot (m_1-1) \cdot ... \cdot (m_1-k+1) \cdot m_2 \cdot (m_2-1) \cdot ... \cdot (m_2-(n-k)+1) \\
=& \frac{m_1!}{(m_1-k)!} \cdot \frac{m_2!}{\big(m_2-(n-k)\big)!}
\end{aligned}$$

e dunque

$$\begin{aligned}
&\mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\}) \\
&= \frac{m_1 \cdot (m_1-1) \cdot ... \cdot (m_1-k+1) \cdot m_2 \cdot (m_2-1) \cdot ... \cdot (m_2-(n-k)+1)}{M(M-1)...(M-n+1)} \\
&= \frac{m_1!}{(m_1-k)!} \frac{m_2!}{(m_2-(n-k))!} \frac{(M-n)!}{M!}
\end{aligned}$$

da cui

$$\begin{aligned}
\mathbb{P}(\{S_n = k\}) &= \sum_{\mathbf{x} \in \{0,1\}^n; \sum_i x_i = k} \mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\}) \\
&= \binom{n}{k} \frac{m_1 \cdot (m_1-1) \cdot ... \cdot (m_1-k+1) \cdot m_2 \cdot (m_2-1) \cdot ... \cdot (m_2-(n-k)+1)}{M(M-1)...(M-n+1)} \\
&= \binom{n}{k} \frac{m_1!}{(m_1-k)!} \frac{m_2!}{(m_2-(n-k))!} \frac{(M-n)!}{M!};
\end{aligned}$$

riscrivendo in forma più compatta tale ultima frazione, attraverso la notazione dei coefficienti binomiali, possiamo concludere: per $k \in \{0, 1, \ldots, n\}$

$$\mathbb{P}(\{S_n = k\}) = \begin{cases} \dfrac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}} & \text{per } \max(0, n+m_1-M) \leq k \leq \min(n, m_1), \\ 0 & \text{altrimenti} \end{cases} \tag{46}$$

In effetti già conosciamo questo risultato dall'Esempio 3.8 (si veda anche l'Esercizio proposto 3.2) della Lezione 3.

Una probabilità del tipo (46) su $\{0,1,2,...,n\}$ prende il nome di ***probabilità ipergeometriche***, ossia una misura di probabilità $\mathbf{P}$ definita sull'insieme delle parti di $\{0,1,2,...,n\}$ come segue

$$\mathbf{P}: \mathscr{P}\big(\{0,1,2,...,n\}\big) \to [0,1], \quad I \mapsto \mathbf{P}(I) \stackrel{def}{=} \sum_{k\in I} p(k),$$

dove $p(k) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}}$, se k è tale che $0 \le k \le m_1$ e $0 \le n-k \le m_2$ (o equivalentemente se $\max(0, n+m_1-M) \le k \le \min(n, m_1)$), mentre $p(k) = 0$ per gli eventuali altri valori di k.

Dal momento che, per fissati M, m_1, n, la famiglia degli eventi $\{S_n = k\}$, per $\max(0, n+m_1-M) \le k \le \min(n, m_1)$, costituisce una partizione dello spazio campione, deve ovviamente risultare

$$\sum_{k=0\vee(n+m_1-M)}^{n\wedge m_1} \mathbb{P}(\{S_n = k\}) = 1 \quad \text{cioè} \quad \sum_{k=0\vee(n+m_1-M)}^{n\wedge m_1} \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}} = 1.$$

In effetti quest'ultima identità coincide con l'identità di Vandermonde (25).

***Osservazione* 6.1.** Si può pervenire al risultato (46) anche in un modo alternativo, senza utilizzare il calcolo combinatorio, ma applicando direttamente la formula delle probabilità composte agli eventi del tipo $\{X_1 = x_1, ..., X_n = x_n\} = \{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}$:

$$\begin{aligned}&\mathbb{P}(\{X_1 = x_1, ..., X_n = x_n\})\\ &= \mathbb{P}(\{X_1 = x_1\})\,\mathbb{P}(\{X_2 = x_2\}|\{X_1 = x_1\})\cdots\mathbb{P}(\{X_n = x_n\}|\{X_1 = x_1, ..., X_{n-1} = x_{n-1}\}).\end{aligned}$$

Possiamo infatti ***imporre direttamente***, a partire dalla descrizione del problema, che

$$\mathbb{P}(\{X_1 = 1\}) = \frac{m_1}{M}$$

e

$$\mathbb{P}(\{X_{r+1} = 1\}|\{X_1 = x_1...., X_r = x_r\}) = \frac{m_1 - \sum_{i=1}^r x_i}{M-r}, \quad \text{per } 1 \le r \le n-1$$

ovvero possiamo imporre che le precedenti probabilità condizionate agli esiti delle prime r estrazioni, siano uguali al rapporto fra il numero $m_1 - \sum_{i=1}^r x_i$ degli elementi di tipo A rimasti nella popolazione dopo le prime r estrazioni ed il numero complessivo $M - r$ degli elementi rimasti nella popolazione.

Per rendere più comprensibile il ragionamento si pensi al caso in cui siano $n = 4$, $M = 6$ $m_1 = 4$ e $k = 2$, allora, ad esempio,

$$\begin{aligned}&\mathbb{P}(\{X_1 = 1, X_2 = 0, X_3 = 1, X_4 = 0\}) = \mathbb{P}(A_1 \cap \overline{A}_2 \cap A_3 \cap \overline{A}_4)\\ =&\,\mathbb{P}(A_1)\cdot\mathbb{P}(\overline{A}_2|A_1)\cdot\mathbb{P}(A_3|A_1\cap\overline{A}_2)\cdot\mathbb{P}(\overline{A}_4|A_1\cap\overline{A}_2\cap A_3)\\ =&\,\mathbb{P}(\{X_1 = 1\})\cdot\mathbb{P}(\{X_2 = 0\}|\{X_1 = 1\})\cdot\mathbb{P}(\{X_3 = 1\}|\{X_1 = 1, X_2 = 0\})\cdot\\ &\cdot\mathbb{P}(\{X_4 = 0\}|\{X_1 = 1, X_2 = 0, X_3 = 1\})\\ =&\,\frac{4}{6}\cdot\frac{2}{6-1}\cdot\frac{4-1}{6-2}\cdot\frac{2-1}{6-3}. = \frac{4}{6}\cdot\frac{2}{5}\cdot\frac{3}{4}\cdot\frac{1}{3}.\end{aligned}$$

Notiamo dunque che tali probabilità condizionate ***non vengono calcolate*** *utilizzando la loro definizione (data in Definizione 4.1 della Lezione 4), ma* ***vengono direttamente assegnate*** *a partire dalle condizioni del problema.*

Ciò costituisce un esempio di quanto era stato già accennato più in generale in merito alla nozione di probabilità condizionata: ***si giunge cioè a calcolare delle probabilità di eventi composti non tramite un calcolo combinatorio, bensì assegnando delle probabilità condizionate e delle condizioni di simmetria fra eventi***; tali considerazioni sono analoghe a quelle già svolte nell'Esempio 6.2.

***Osservazione* 6.2.** È importante osservare che anche nel caso di estrazioni senza reinserimento, per ogni $i = 1, 2, ..., n$, per ciascuno degli eventi $A_i = \{X_i = 1\}$, ossia degli eventi $\{$*all'$i-$sima estrazione viene estratto un individuo/pallina di tipo* A$\}$ si ha

$$\mathbb{P}(A_i) = \frac{m_1}{M}, \quad \text{per ogni } i = 1, 2, ..., n,$$

ossia
nel caso di senza reinserimento le probabilità di estrarre un oggetto di tipo A alla i-sima estrazione sono tutte uguali alla probabilità di estrarre un oggetto di tipo A alla prima estrazione e quindi hanno lo stesso valore delle analoghe probabilità calcolate nel caso di estrazioni con reinserimento.
Per convincersene si può ragionare in diversi modi.

Per iniziare è ovvio che ciò sia vero per la prima estrazione, cioè per $i = 1$, ossia $\mathbb{P}(A_1) = \frac{m_1}{M}$.
Per il caso $i = 2$ possiamo procedere in modo analogo all'Esempio 4.3, con la formula delle probabilità totali rispetto alla partizione data da $H_1 = A_1$ e $H_2 = \overline{A}_1$:

$$\begin{aligned}\mathbb{P}(A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1) + \mathbb{P}(\overline{A}_1)\,\mathbb{P}(A_2|\overline{A}_1) &= \frac{m_1}{M} \cdot \frac{m_1 - 1}{M - 1} + \frac{M - m_1}{M} \cdot \frac{m_1}{M - 1} \\ = \frac{m_1}{M} \cdot \left(\frac{m_1 - 1}{M - 1} + \frac{M - m_1}{M - 1} \right) &= \frac{m_1}{M} \cdot \frac{\cancel{m_1} - 1 + M - \cancel{m_1}}{M - 1} = \frac{m_1}{M}\end{aligned}$$

Tuttavia è forse più conveniente ragionare nel seguente modo: stiamo utilizzando le probabilità classiche, e quindi per ottenere che $\mathbb{P}(A_i) \left(:= \frac{|A_i|}{|\Omega|} \right) = \frac{m_1}{M}$, basta mostrare che la cardinalità di A_i è uguale alla cardinalità di A_1, ossia che c'è una corrispondenza biunovoca tra A_i ed A_1, per ogni i in $\{1, 2, ..., n\}$.

Consideriamo di nuovo il caso $i = 2$. Ricordando che

$$A_1 := \big\{(j_1, j_2, ..., j_n) \in \Omega \text{ tali che } j_1 \in \{1, 2, ..., m_1\}\big\} \; e \; A_2 := \big\{(j_1, j_2, ..., j_n) \in \Omega \text{ tali che } j_2 \in \{1, 2, ..., m_1\}\big\},$$

è semplice convincersi che scambiando j_1 e j_2 da un elemento $(j_1, j_2, j_3, ..., j_n)$ di A_1 si ottiene un elemento di A_2 e viceversa, da un elemento di A_2 si ottiene un elemento di A_1. Gli eventi A_1 ed A_2 sono quindi in corrispondenza biunivoca ed hanno perciò la stessa cardinalità. Analogamente la corrispondenza biunivoca tra A_1 e A_i si ottiene scambiando j_1 e j_i.

> Per rendere più comprensibile il ragionamento si pensi al caso in cui si abbia $n = 4$, $M = 7$, $m_1 = 3$, e $i = 3$, ossia gli elementi di tipo A sono numerati 1, 2, 3, mentre gli elementi di tipo B sono numerati 4, 5, 6, 7. Allora, ad esempio, la disposizione $(5, 6, 2, 1)$ appartiene a A_3 in quanto $j_3 = 2 \in \{1, 2, 3\}$ e la disposizione ottenuta da $(5, 6, 2, 1)$ scambiando $j_1 = 5$ con $j_3 = 2$, ossia la disposizione $(2, 6, 5, 1)$, appartiene ad A_1.

***Osservazione* 6.3.** Il problema qui affrontato riguarda il calcolo della probabilità di avere k elementi di tipo A in un campionamento casuale di n oggetti da una popolazione complessivamente costituita da M elementi, di cui m_1 di tipo A e $m_2 = (M - m_1)$ di tipo B. Tale calcolo si applica a problemi di diverso tipo (quali *estrazioni da urne*, *sondaggio elettorale*, *analisi statistica di una popolazione*, etc...) tutti, fra di loro, sostanzialmente isomorfi. È interessante confrontare fra loro le due formule (45) e (46). Entrambe risolvono il problema detto; la prima riguarda però estrazioni con reinserimento, mentre la seconda riguarda estrazioni senza reinserimento. Intuitivamente ci possiamo aspettare che le due formule tendano a coincidere nel caso in cui M, m_1 ed m_2 siano numeri molto grandi rispetto a n; infatti in tal caso ci si può aspettare che non vi sia grande differenza fra estrazioni con o senza reinserimento. Ciò può essere formalizzato come segue: è possibile dimostrare che, per fissati valori di n, k (con $0 \le k \le n$) e θ (con $0 < \theta < 1$), mandando M ed m_1 all'infinito in modo che $\frac{m_1}{M}$ tende a θ, allora risulta

$$\lim_{\substack{M \to \infty, m_1 \to \infty \\ \frac{m_1}{M} \to \theta}} \frac{\binom{m_1}{k}\binom{M - m_1}{n-k}}{\binom{M}{n}} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

Convergenza delle probabilità ipergeometriche alle probabilità binomiali

Dimostriamo che, se il numero n è fissato, mentre M ed m_1 sono molto grandi e con $\frac{m_1}{M}$ vicino a θ, allora le probabilità ipergeometriche e le probabilità binomiali con probabilità θ sono molto vicine. Formalmente si tratta di dimostrare che, per ogni $k = 0, 1, ..., n$

$$\lim_{\substack{M\to\infty, m_1\to\infty \\ \frac{m_1}{M}\to\theta}} \frac{\binom{m_1}{k}\binom{M-m_1}{n-k}}{\binom{M}{n}} = \binom{n}{k}\theta^k(1-\theta)^{n-k}.$$

Ricordiamo che, per arrivare alla (46), abbiamo mostrato che

$$\begin{aligned}
&\binom{n}{k}\frac{m_1!}{(m_1-k)!}\frac{m_2!}{(m_2-(n-k))!}\frac{(M-n)!}{M!} \\
&=\frac{n!}{k!(n-k)!}\frac{m_1!}{(m_1-k)!}\frac{m_2!}{(m_2-(n-k))!}\frac{(M-n)!}{M!} \\
&=\frac{m_1!}{k!\,(m_1-k)!}\frac{m_2!}{(n-k)!\,(m_2-(n-k))!}\frac{(M-n)!\,n!}{M!}=\frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}}.
\end{aligned}$$

Utilizzando la precedente relazione con $m_2 = M - m_1$, basta quindi verificare che

$$\lim_{\substack{M\to\infty, m_1\to\infty \\ \frac{m_1}{M}\to\theta}} \frac{m_1!}{(m_1-k)!}\frac{(M-m_1)!}{\big(M-m_1-(n-k)\big)!}\frac{(M-n)!}{M!} = \theta^k(1-\theta)^{n-k}.$$

Infatti si ha

$$\begin{aligned}
\frac{m_1!}{(m_1-k)!} &= m_1(m_1-1)\cdots\big(m_1-(k-1)\big), \\
\frac{(M-m_1)!}{\big(M-m_1-(n-k)\big)!} &= (M-m_1)(M-m_1-1)\cdots\big(M-m_1-(n-k-1)\big) \\
\frac{(M-n)!}{M!} &= \frac{1}{M(M-1)\cdots\big(M-(n-1)\big)}
\end{aligned}$$

e quindi

$$\begin{aligned}
&\frac{m_1!}{(m_1-k)!}\frac{(M-m_1)!}{\big(M-m_1-(n-k)\big)!}\frac{(M-n)!}{M!} \\
&=\frac{m_1(m_1-1)\cdots\big(m_1-(k-1)\big)\;(M-m_1)(M-m_1-1)\cdots\big(M-m_1-(n-k-1)\big)}{M(M-1)\cdots\big(M-(n-1)\big)} \\
&=\frac{m_1}{M}\cdot\frac{m_1-1}{M-1}\cdots\frac{m_1-(k-1)}{M-(k-1)}\cdot\frac{M-m_1}{M-k}\cdot\frac{M-m_1-1}{M-k-1}\cdots\frac{M-m_1-(n-k-1)}{M-(n-1)}
\end{aligned}$$

e la tesi si ottiene considerando che, per ogni i, j,

$$\lim_{\substack{M\to\infty, m_1\to\infty \\ \frac{m_1}{M}\to\theta}} \frac{m_1-i}{M-i}=\theta, \qquad \lim_{\substack{M\to\infty, m_1\to\infty \\ \frac{m_1}{M}\to\theta}} \frac{M-m_1-i}{M-j}=1-\theta,$$

Ricordando che la parte intera $\lfloor x\rfloor$ di un numero reale x è quel numero intero k tale che $k \le x < k+1$, osserviamo che si può prendere m_1 in funzione di M come segue

$$m_1 = m_1(M) = \lfloor\theta M\rfloor \quad \text{e mandare } M \text{ all'infinito;}$$

infatti, chiaramente $m_1 = \lfloor\theta M\rfloor$ tende ad infinito se M tende all'infinito, e inoltre

$$\lim_{M\to\infty}\frac{m_1(M)}{M} = \lim_{M\to\infty}\frac{\lfloor\theta M\rfloor}{M} = \theta.$$

L'ultimo limite si ottiene tenendo conto che $\lfloor\theta M\rfloor \le \theta M < \lfloor\theta M\rfloor + 1$ e che quindi

$$0 \le \theta M - \lfloor\theta M\rfloor < 1, \quad \text{ovvero} \quad 0 \le \frac{\theta M}{M} - \frac{\lfloor\theta M\rfloor}{M} < \frac{1}{M}.$$

6.4 Esercizi di verifica

Esercizio 6.1. Un candidato ad un'elezione ha bisogno di almeno 50 voti per essere eletto. Prepara allora una lettera per informare i potenziali elettori circa la sua candidatura, il suo programma elettorale, etc.... Egli valuta che ogni persona che riceve la lettera si recherà effettivamente a votare per lui con una probabilità del 40%, indipendentemente dal comportamento degli altri (e si sottointende che egli certamente non ottiene voti da coloro ai quali non ha inviato la lettera).
(a) Qual è la probabilità che egli riceva esattamente 51 voti se invia la lettera a 200 persone? (basta trovare l'espressione)
(b) Qual è la probabilità di essere eletto se invia la lettera a 100 persone? (basta trovare l'espressione)
(c) Caratterizzare il numero minimo n_{min} di persone alle quali deve inviare copia della lettera affinché la probabilità di essere eletto sia superiore all'80%.

Riprenderemo questo esercizio per ottenere delle soluzioni approssimate con l'ausilio del Teorema Centrale del Limite (approssimazione normale)

Esercizio 6.2. Si prendono a caso $n = 5$ viti da una scatola contenente complessivamente $M = 26$ viti, di cui alcune nuove ed altre usurate.
(a) Supponendo che la scatola contiene $m_1 = 20$ viti nuove e $m_2 = 6$ viti usurate, calcolare la probabilità che almeno quattro delle cinque viti scelte siano nuove.
(b) Si supponga ora di non conoscere inizialmente il numero M_1 delle viti nuove nella scatola e si ponga

$$\mathbb{P}(\{M_1 = h\}) = \binom{26}{h} \left(\frac{4}{5}\right)^h \left(\frac{1}{5}\right)^{26-h}, \qquad h = 0, 1, 2, ..., 26$$

Dopo aver verificato che tutte le cinque viti scelte sono nuove, come va calcolata la probabilità dell'ipotesi $\{M_1 = 26\}$?

Esercizio 6.3. In un lotto di 15 lampadine, ve ne sono 5 guaste. Se ne estraggono a caso 3. Calcolare la probabilità dei seguenti eventi:
(a) nessuna lampadina difettosa fra le tre estratte
(b) esattamente una lampadina difettosa fra le tre estratte
(c) almeno una lampadina difettosa fra le tre estratte.
Si considerino separatamente i due diversi casi in cui
(i) si estraggano le tre lampadine contemporaneamente
(ii) le estrazioni sono con reimbussolamento

Esercizio 6.4. Si hanno m esemplari di un certo tipo di telecomando (TC) per televisore; ciascun TC ha bisogno di due batterie per il suo funzionamento. Si hanno a disposizione $2m$ batterie, di cui però h cariche e $(2m - h)$ scariche. Da tale gruppo di batterie vengono costituite in modo casuale m coppie, che vengono inserite negli m TC.
Calcolare la probabilità che un fissato TC abbia entrambe le batterie cariche.

Esercizio 6.5. Riottenere la formula (46) seguendo le indicazioni contenute nella precedente ***Osservazione 6.1*** di pagina 76.

Esercizio 6.6. Un'urna contiene una pallina rossa e una blu. Si effettuano 5 estrazioni con doppio reinserimento, ossia, si estrae una pallina per volta e, ogni volta la pallina viene reinserita nell'urna insieme a una seconda pallina dello stesso colore.
Calcolare la probabilità di estrarre k volte una pallina rossa, per $k = 0,\ 1,\ ...,\ 5$.

Non indipendenza nel caso di estrazioni con reinserimento da un'urna di composizione incognita

Supponiamo ora di avere due urne:
la prima urna contiene a_1 palline di tipo A e b_1 di tipo B,
le seconda urna contiene a_2 palline di tipo A e b_2 di tipo B.
Supponiamo che le due urne siano esternamente uguali e di sapere che verrano effettuate due estrazioni con reinserimento da una delle due urne, ma non sappiamo quale. Posto

$$H_1 = \{\textit{viene scelta l'urna 1}\}, \quad H_2 = \{\textit{viene scelta l'urna 2}\}(=\overline{H}_1)$$

supponiamo tuttavia di conoscere $\mathbb{P}(H_1)$ e $\mathbb{P}(H_2) = 1 - \mathbb{P}(H_1)$.
Ad esempio, la scelta dell'urna potrebbe essere effettuata lanciando un dado ben equilibrato e, ad esempio, potremmo sapere che verrà scelta l'urna 1 se esce un numero minore o uguale a 2, e verrà scelta l'urna 2 altrimenti. In questo caso si avrebbe $\mathbb{P}(H_1) = 1/3$ e $\mathbb{P}(H_2) = 1 - \mathbb{P}(H_1) = 2/3$.
Sottolineiamo che la scelta dell'urna viene fatta a nostra insaputa, sappiamo solo che verrà scelta in questo modo.
Dall'urna scelta, sempre la stessa, vengono effettuate due estrazione con reinserimento. In analogia a quanto fatto precedentemente poniamo

$$A_1 = \{\textit{la prima pallina estratta è di tipo } \text{A}\},$$

$$A_2 = \{\textit{la seconda pallina estratta è di tipo } \text{A}\}.$$

Chiaramente, se sapessimo che l'urna da cui vengono effettuate le estrazioni potremmo calcolare le probabilità di A_1, A_2 e di $A_1 \cap A_2$, ossia

$$\mathbb{P}(A_1|H_1) = \mathbb{P}(A_2|H_1) = \frac{a_1}{a_1+b_1}, \text{ e } \mathbb{P}(A_1 \cap A_2|H_1) = \frac{a_1}{a_1+b_1}\frac{a_1}{a_1+b_1} = \left(\frac{a_1}{a_1+b_1}\right)^2$$

e analogamente

$$\mathbb{P}(A_1|H_2) = \mathbb{P}(A_2|H_2) = \frac{a_2}{a_2+b_2}, \text{ e } \mathbb{P}(A_1 \cap A_2|H_1) = \frac{a_2}{a_2+b_2}\frac{a_2}{a_2+b_2} = \left(\frac{a_2}{a_2+b_2}\right)^2.$$

Di conseguenza possiamo calcolare la probabilità di A_1, con la formula delle probabilità totali come

$$\mathbb{P}(A_1) = \mathbb{P}(H_1)\,\mathbb{P}(A_1|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A_1|H_2) = \mathbb{P}(H_1)\frac{a_1}{a_1+b_1} + (1-\mathbb{P}(H_1))\frac{a_2}{a_2+b_2}$$

$$\mathbb{P}(A_2) = \mathbb{P}(H_1)\,\overbrace{\mathbb{P}(A_2|H_1)}^{=\mathbb{P}(A_1|H_1)} + \mathbb{P}(H_2)\,\overbrace{\mathbb{P}(A_2|H_2))}^{=\mathbb{P}(A_1|H_2)} = \mathbb{P}(A_1).$$

e infine

$$\begin{aligned}\mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(H_1)\,\mathbb{P}(A_1 \cap A_2|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A_1 \cap A_2|H_2)\\ &= \mathbb{P}(H_1)\left(\frac{a_1}{a_1+b_1}\right)^2 + (1-\mathbb{P}(H_1))\left(\frac{a_2}{a_2+b_2}\right)^2\end{aligned}$$

Accade che gli eventi A_1 e A_2 non sono indipendenti, ma sono correlati positivamente, ossia

$$\mathbb{P}(A_1 \cap A_2) \geq \mathbb{P}(A_1)\,\mathbb{P}(A_2).$$

Infatti posto $\lambda := \mathbb{P}(H_1)$, e $x_i := \frac{a_i}{a_i+b_i}$, per $i = 1, 2$,

$$\mathbb{P}(A_1 \cap A_2) \geq \mathbb{P}(A_1)\,\mathbb{P}(A_2) \;\Leftrightarrow\; \lambda x_1^2 + (1-\lambda)x_2^2 \geq (\lambda x_1 + (1-\lambda)x_2)^2\,,$$

e quest'ultima condizione è verificata in quanto la funzione $x \mapsto x^2$ è convessa.

Richiamo sulle funzioni convesse

Una funzione $\varphi : [a, b] \to \mathbb{R}$, $x \mapsto \varphi(x)$ è convessa nell'intervallo $[a, b]$, se e solo se

$$\forall\, x_1, x_2 \in [a, b], \text{ e } \forall \lambda \in [0, 1], \text{ si ha } \quad \varphi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\varphi(x_1) + (1-\lambda)\varphi(x_2)$$

La funzione $x \mapsto \varphi(x) := x^2$ è il prototipo delle funzioni convesse.
Per controllare che è convessa basta osservare che

$$\varphi(\lambda x_1 + (1-\lambda)x_2) = (\lambda x_1 + (1-\lambda)x_2)^2 = \lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1 x_2,$$

mentre

$$\lambda\varphi(x_1) + (1-\lambda)\varphi(x_2) = \lambda x_1^2 + (1-\lambda)x_2^2$$

e quindi

$$\begin{gathered}
\varphi(\lambda x_1 + (1-\lambda)x_2) \leq \lambda\varphi(x_1) + (1-\lambda)\varphi(x_2) \\
\Updownarrow \\
\lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1 x_2 \leq \lambda x_1^2 + (1-\lambda)x_2^2 \\
\Updownarrow \\
\lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1 x_2 - \lambda x_1^2 - (1-\lambda)x_2^2 \leq 0 \\
\Updownarrow \\
\big(\lambda^2 - \lambda\big)x_1^2 + \big((1-\lambda)^2 - (1-\lambda)\big)x_2^2 + 2\lambda(1-\lambda)x_1 x_2 \leq 0 \\
\Updownarrow \\
-\lambda(1-\lambda)x_1^2 + (1-\lambda)\big((1-\lambda) - 1\big)x_2^2 + 2\lambda(1-\lambda)x_1 x_2 \leq 0 \\
\Updownarrow \\
-\lambda(1-\lambda)\big[x_1^2 + x_2^2 - 2x_1 x_2\big] \leq 0 \;\Leftrightarrow\; -\underbrace{\lambda(1-\lambda)}_{\geq 0}\underbrace{(x_1 - x_2)^2}_{\geq 0} \leq 0
\end{gathered}$$

7 Variabili aleatorie e distribuzioni di probabilità

7.1 Variabili aleatorie in spazi finiti di probabilità

Varie questioni incontrate nelle precedenti lezioni trovano una adeguata formalizzazione tramite l'introduzione della nozione di ***variabile aleatoria***. Nei precedenti esempi, infatti, ci siamo ripetutamente imbattuti in oggetti quali: somma dei punteggi nel lancio di due dadi, numero di votanti per uno schieramento in un sondaggio elettorale, numero di successi su n prove bernoulliane, massimo fra i cinque numeri risultanti da un'estrazione del lotto, etc....

Sarebbe forse più evocativo parlare di *numeri aleatori*, ma dal punto di vista storico è ormai entrato nell'uso parlare di variabili aleatorie invece che di numeri aleatori, e non possiamo modificare la sua denominazione

A parte la diversa natura dei problemi considerati, notiamo che si è trattato in ogni caso di situazioni in cui si considera una grandezza aleatoria X e valgono le seguenti condizioni:
- il valore che X assumerà sarà connesso (in qualche preciso modo) al risultato elementare di un qualche esperimento aleatorio;

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE:
X_1 rappresenta il valore/punteggio del primo dado, X_2 il valore/punteggio del secondo dado, $X = X_1 + X_2$ la somma dei punteggi ottenuti
QUI $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} = \big\{(i,j),\ i,j \in \{1,2,3,4,5,6\}\big\}$
e se esce il risultato (i,j) allora X_1 assume il valore i, X_2 assume il valore j ed $X = X_1 + X_2$ assume il valore $i+j$.
Dal punto di vista matematico si tratta quindi di funzioni:

$$
\begin{aligned}
(i,j) &\mapsto X_1(i,j) = i \\
(i,j) &\mapsto X_2(i,j) = j \\
(i,j) &\mapsto X(i,j) = i+j
\end{aligned}
$$

- possiamo elencare i valori che possono essere assunti da X;

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE:
L'insieme dei valori che può assumere X_1 è l'insieme $\{1,2,3,4,5,6\}$, lo stesso vale per X_2, invece l'insieme dei valori che può assumere $X = X_1 + X_2$ è l'insieme $\{2,3,4,5,6,7,8,9,10,11,12\}$.
In altre parole $X_1(\Omega) = \{1,2,3,4,5,6\}$, dove $X_1(\Omega)$ denota l'immagine di X_1, anche $X_2(\Omega) = \{1,2,3,4,5,6\}$, e invece $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$, e potremmo anche considerare, sempre con $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$,

$$
\begin{aligned}
X_1 :\ \Omega &\to X_1(\Omega) = \{1,2,3,4,5,6\} \subset \mathbb{R} \\
(i,j) &\mapsto X_1(i,j) = i \\
X_2 :\ \Omega &\to X_2(\Omega) = \{1,2,3,4,5,6\} \subset \mathbb{R} \\
(i,j) &\mapsto X_2(i,j) = j \\
X :\ \Omega &\to X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\} \subset \mathbb{R} \\
(i,j) &\mapsto X(i,j) = i+j
\end{aligned}
$$

- sussiste una situazione di incertezza relativamente allo specifico valore che X effettivamente assume.

In base alla misura di probabilità assegnata sullo spazio campione in tale esperimento, potremo valutare la probabilità che si presentino i vari possibili valori per la grandezza X.

ESEMPIO: LANCIO DI DUE DADI A 6 FACCE: Se i dadi sono ben equilibrati e prendiamo come misura di probabilità $\mathbb{P}_c$, la probabilità classica, e valutiamo quindi $\mathbb{P}_c(X_1 = k) = \frac{1}{6}$, per ogni $k \in \{1, 2, 3, 4, 5, 6\}$, similmente $\mathbb{P}_c(X_2 = k) = \frac{1}{6}$, per ogni $k \in \{1, 2, 3, 4, 5, 6\}$, e, tenendo conto che

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

otteniamo

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

Se invece i due dadi fossero truccati e si supponesse che la probabilità che il risultato (i, j) abbia probabilità proporzionale a $i \cdot j$, ossia si usasse la probabilità $\mathbb{Q}$ tale che $\mathbb{Q}(\{(i, j)\}) = \frac{i \cdot j}{21^2}$ si avrebbe, come visto in precedenza $\mathbb{Q}(X_1 = k) = \frac{k}{21}$, per $k \in \{1, 2, 3, 4, 5, 6\}$, similmente $\mathbb{Q}(X_2 = k) = \frac{k}{21}$, per $k \in \{1, 2, 3, 4, 5, 6\}$, e, sempre tenendo conto che

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

otteniamo

$$\begin{aligned}
\mathbb{Q}(X = 2) &= \frac{1 \cdot 1}{21^2} = \frac{1}{21^2},\\
\mathbb{Q}(X = 3) &= \frac{1 \cdot 2 + 2 \cdot 1}{21^2} = \frac{4}{21^2},\\
\mathbb{Q}(X = 4) &= \frac{1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1}{21^2} = \frac{10}{21^2},\\
\mathbb{Q}(X = 5) &= \frac{1 \cdot 4 + 2 \cdot 3 + 3 \cdot 2 + 4 \cdot 1}{21^2} = \frac{20}{21^2},\\
\mathbb{Q}(X = 6) &= \frac{1 \cdot 5 + 2 \cdot 4 + 3 \cdot 3 + 4 \cdot 2 + 5 \cdot 1}{21^2} = \frac{35}{21^2},\\
\mathbb{Q}(X = 7) &= \frac{1 \cdot 6 + 2 \cdot 5 + 3 \cdot 4 + 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 1}{21^2} = \frac{56}{21^2},\\
\mathbb{Q}(X = 8) &= \frac{2 \cdot 6 + 3 \cdot 5 + 4 \cdot 4 + 5 \cdot 3 + 6 \cdot 2}{21^2} = \frac{70}{21^2},\\
\mathbb{Q}(X = 9) &= \frac{3 \cdot 6 + 4 \cdot 5 + 5 \cdot 4 + 6 \cdot 3}{21^2} = \frac{76}{21^2},\\
\mathbb{Q}(X = 10) &= \frac{4 \cdot 6 + 5 \cdot 5 + 6 \cdot 4}{21^2} = \frac{73}{21^2},\\
\mathbb{Q}(X = 11) &= \frac{5 \cdot 6 + 6 \cdot 5}{21^2} = \frac{60}{21^2},\\
\mathbb{Q}(X = 12) &= \frac{6 \cdot 6}{21^2} = \frac{36}{21^2}
\end{aligned}$$

Osserviamo che la famiglia degli eventi $H_k^X := \{X = k\}$, per $k \in X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$ è una partizione, che chiameremo la **partizione generata da X**, e quindi, sia per $\mathbb{P} = \mathbb{P}_c$ che per $\mathbb{P} = \mathbb{Q}$ si ha

$$\mathbb{P}(H_2^X) + \mathbb{P}(H_3^X) + \cdots + \mathbb{P}(H_{11}^X) + \mathbb{P}(H_{12}^X) = \sum_{k=2}^{12} \mathbb{P}(H_k^X) = 1$$

ovvero $\sum_{k=2}^{12} \mathbb{P}(\{X = k\}) = 1$. Quindi la funzione definita da:

$$p_X : \ X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\} \to [0,1]; \quad k \mapsto p_X(k) := \mathbb{P}(\{X = k\})$$

definisce, nel solito modo, una probabilità sull'insieme delle parti di $X(\Omega) = \{2,3,4,5,6,7,8,9,10,11,12\}$:

$$\mathbf{P}_X : \ \mathscr{P}\big(X(\Omega)\big) = \mathscr{P}\big(\{2,3,4,5,6,7,8,9,10,11,12\}\big) \to [0,1];$$

$$I \mapsto \mathbf{P}_X(I) := \sum_{k \in I} p_X(k) \Big(= \sum_{k \in I} \mathbb{P}(\{X = k\}) \Big)$$

Tale probabilità prende il nome di **distribuzione (di probabilità) di X** e invece la funzione $k \mapsto p_X(k)$, viene detta **densità discreta di X**.
Osserviamo infine che $\mathbf{P}_X(I)$ coincide con $\mathbb{P}(X \in I)$, infatti

$$\mathbf{P}_X(I) := \sum_{k \in I} p_X(k) = \sum_{k \in I} \mathbb{P}(\{X = k\}) = \mathbb{P}\left(\cup_{k \in I} \{X = k\}\right) = \mathbb{P}(X \in I)$$

Ad esempio, se $I = \{2,3,4,5\}$ allora possiamo scrivere l'evento $\{X \in I\}$

$$\big\{X \in \{2,3,4,5\}\big\} = \{X = 2\} \cup \{X = 3\} \cup \{X = 4\} \cup \{X = 5\}$$

e quindi

$$\begin{aligned} \mathbb{P}(X \in \{2,3,4,5\}) &= \mathbb{P}(\{X = 2\} \cup \{X = 3\} \cup \{X = 4\} \cup \{X = 5\}) \\ &= \mathbb{P}(\{X = 2\}) + \mathbb{P}(\{X = 3\}) + \mathbb{P}(\{X = 4\}) + \mathbb{P}(\{X = 5\}) \\ &= p_X(2) + p_X(3) + p_X(4) + p_X(5). \end{aligned}$$

Nel caso in cui $X(\Omega)$ abbia cardinalità grande, non è possibile elencare tutti i valori di $p(x_k)$, per ogni $x_k \in X(\Omega)$: ad esempio nel caso di X uguale alla somma ottenuta lanciando due dadi ben equilibrati abbiamo trovato che

$$p_X(2) = 1/36, \ p_X(3) = 2/36, \ p_X(3) = 4/36, \ p_X(4) = 3/36,$$

$$p_X(5) = 4/36, \ p_X(6) = 5/36, \ p_X(7) = 6/36,$$

$$p_X(8) = 5/36, \ p_X(9) = 4/36, \ p_X(10) = 3/36, \ p_X(11) = 2/36, \ p_X(12) = 1/36,$$

e, per $k = 2,3,4,5,6,7$, è immediato osservare che il numeratore di $p_X(k)$ cresce di volta in volta di 1 e coincide con $k - 1$, mentre, per $k = 8,9,10,11,12$, il numeratore decresce di volta in volta di 1 e quindi potremo sicuramente scriverlo come $a - 1$, per un opportuno intero a: non è difficile convincersi che $a = 13$, osservando ad esempio che $8 + 5 = 9 + 4 = 10 + 3 = 11 + 2 = 12 + 1 = 13$. Potremo quindi scrivere sinteticamente

$$p_X(k) = \begin{cases} \dfrac{k-1}{36} & \text{per } k = 2,3,...,7 \\[2ex] \dfrac{13-k}{36} & \text{per } k = 8,9,...,12 \end{cases}$$

Tali considerazioni motivano le definizioni seguenti.

Definizione 7.1 (provvisoria). *Sia Ω un insieme finito. Un'applicazione $X : \Omega \to X(\Omega) \subseteq \mathbb{R}$ viene detta* **variabile aleatoria** *(definita su $(\Omega, \mathscr{P}(\Omega))$).*

***Osservazione* 7.1.** Essendo $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ un insieme finito, l'immagine di X, ovvero $X(\Omega)$, è un insieme del tipo $\{x_1, ..., x_n\}$, con $x_\ell \in \mathbb{R}$, per $\ell = 1, ..., n$, e ovviamente con $n \leq N$.

Consideriamo ora gli eventi $\{X = x_\ell\}$ come sottoinsiemi di Ω: un tale evento si verifica se e solo se $\omega \in \{X = x_\ell\}$, in altre parole

$$\{X = x_\ell\} = \{\omega \in \Omega : \ X(\omega) = x_\ell\} = X^{-1}(\{x_\ell\}), \quad \ell = 1, 2, ..., n.$$

Indicheremo brevemente tali eventi anche con i simboli

$$H_1^X := \{X = x_1\}, ..., \ H_n^X = \{X = x_n\}.$$

È immediato verificare che la famiglia degli eventi

$$\mathcal{H}^X = \left\{H_1^X, ..., H_n^X\right\}$$

costituisce una partizione di Ω, detta anche la ***partizione generata da*** X.

Va osservato che, così come per un evento, una variabile aleatoria X è tale a prescindere dalla probabilità $\mathbb{P}$ definita su $(\Omega, \mathscr{P}(\Omega))$, e lo stesso vale per la partizione $\mathcal{H}^X = \left\{H_1^X, ..., H_n^X\right\}$ generata da X.
Inoltre va detto che di solito, per indicare le variabili aleatorie si usano le lettere maiuscole stampatello, così come si usa per gli eventi, ma in genere per le variabili aleatorie si usano le ultime lettere dell'alfabeto, mentre per gli eventi si preferiscono le prime lettere dell'alfabeto.

Proprio grazie al fatto che $\mathcal{H}^X = \left\{H_1^X, ..., H_n^X\right\}$ è una partizione, se $\mathbb{P} : \mathscr{P}(\Omega) \to [0,1]$ è una probabilità, ponendo

$$p_X(x_\ell) := \mathbb{P}(\{X = x_\ell\}), \quad \ell = 1, ..., n,$$

risulta

$$\sum_{\ell=1}^{n} p_X(x_\ell) = \sum_{\ell=1}^{n} \mathbb{P}(\{X = x_\ell\}) = 1, \qquad p_X(x_\ell) \geq 0, \qquad \ell = 1, ..., n. \tag{47}$$

Per semplicità di notazione, a volte scriveremo $\mathbb{P}(X = x_\ell)$ invece di $\mathbb{P}(\{X = x_\ell\})$.

Osserviamo che, se $\mathbb{P}$ è associato a $p(\omega_i) = \mathbb{P}(\{\omega_i\})$ come in (2.3), allora, essendo banalmente

$$H_\ell^X = \{X = x_\ell\} = \{\omega \in \Omega : \ X(\omega_i) = x_\ell\} = \cup_{i: X(\omega_i) = x_\ell} \{\omega_i\},$$

si ha

$$\boxed{p_X(x_\ell) = \sum_{i: X(\omega_i) = x_\ell} p(\omega_i).}$$

Definizione 7.2. *La funzione*

$$p_X : \ X(\Omega) \to [0,1] \subset \mathbb{R}; \qquad x \mapsto p_X(x) := \mathbb{P}(\{X = x\})$$

viene detta **densità discreta** *di X.*

La precedente proprietà (47) ci permette di definire un nuovo spazio di probabilità

$$(X(\Omega), \mathcal{P}(X(\Omega)), \mathbf{P}_X)$$

dove, per $A \in \mathcal{P}(X(\Omega))$, la probabilità $\mathbf{P}_X(A)$ è definita da

$$\mathbf{P}_X(A) := \sum_{\ell : x_\ell \in A} p_\ell = \sum_{\ell : x_\ell \in A} p_X(x_\ell). \tag{48}$$

Come vedremo tra poco, vale la seguente relazione:

$$\mathbf{P}_X(A) = \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}(X \in A), \tag{49}$$

dove $X^{-1}(A)$ indica la controimmagine dell'insieme A tramite la funzione X, ovvero

$$X^{-1}(A) := \{\omega \in \Omega : \; X(\omega) \in A\}.$$

A volte, per comodità di notazione, si scrive più brevemente $\{X \in A\}$ invece di $\{\omega \in \Omega : \; X(\omega) \in A\}$.

Riassumendo, possiamo interpretare $\mathbf{P}_X(A)$ come $\mathbb{P}(\{X \in A\})$. Tale interpretazione e l'equivalenza tra (48) e (49) sono basate sul fatto che l'evento $X^{-1}(A) = \{X \in A\}$ si può scrivere come

$$\{X \in A\} = \bigcup_{\ell : x_\ell \in A} \{X = x_\ell\},$$

e che, di conseguenza,

$$\mathbb{P}(X^{-1}(A)) := \mathbb{P}(\{X \in A\}) = \sum_{\ell : x_\ell \in A} \mathbb{P}(\{X = x_\ell\}) = \sum_{\ell : x_\ell \in A} p_X(x_\ell) = \sum_{\ell : x_\ell \in A} p_\ell = \mathbf{P}_X(A).$$

Definizione 7.3 (provvisoria). *La misura di probabilità $\mathbf{P}_X(\cdot)$ su $\mathcal{P}(X(\Omega))$ definita da (49) o equivalentemente da (48) prende il nome di* **distribuzione di probabilità** *della variabile aleatoria X.*

Osserviamo che $\mathbf{P}_X$ si può estendere a qualunque sottoinsieme J di $\mathbb{R}$ e non solo a sottoinsiemi di $X(\Omega)$ (ed in particolare ad ogni intervallo J, limitato o illimitato) defindendo $\mathbf{P}_X(J) := \mathbb{P}(X \in J) = \sum_{\ell : x_\ell \in J} \mathbb{P}(X = x_\ell) = \sum_{\ell : x_\ell \in J} p_X(x_\ell)$.
Questa idea è alla base della definizione di distribuzione per variabili aleatorie più generali.

***Osservazione* 7.2.** Ovviamente a qualunque variabile aleatoria, definita su uno spazio di probabilità, possiamo associare la sua distribuzione di probabilità.
Grazie alla relazione (48), come per qualunque misura di probabilità, per individuare la distribuzione di probabilità $\mathbf{P}_X$ di una variabile aleatoria X, basta specificare
- la sua immagine $X(\Omega)$, cioè l'insieme dei valori che può assumere X,
- la densità discreta p_X, ossia i valori

$$p_X(x_\ell) = \mathbf{P}_X(\{x_\ell\}) = \mathbb{P}(X = x_\ell) \quad \text{per ogni } x_\ell \in X(\Omega).$$

Due diverse variabili aleatorie, definite o meno su uno stesso spazio di probabilità, possono dar luogo ad una stessa distribuzione di probabilità (vedere i due successivi Esempi 7.1 e 7.2).

7.2 Primi esempi notevoli di variabili aleatorie

È opportuno innanzitutto richiamare l'attenzione sui due particolari tipi di variabili aleatorie: le variabili aleatorie ***degeneri*** e le variabili aleatorie ***binarie***.

Definizione 7.4 (variabili aleatorie degeneri)**.** *Diciamo che X, variabile aleatoria definita su Ω, è una variabile aleatoria* **degenere** *se esiste $\overline{x} \in \mathbb{R}$, tale che $X(\Omega) = \{\overline{x}\}$, cioè se X è costante su Ω.*

La distribuzione di una variabile aleatoria degenere è banale: $X(\Omega) = \{\overline{x}\}$, $\mathscr{P}(X(\Omega)) = \{\emptyset, \{\overline{x}\}\}$, con $\mathbf{P}_X(\emptyset) = 0$ e $\mathbf{P}_X(\{\overline{x}\}) = 1$.

Definizione 7.5 (variabili aleatorie binarie)**.** *Diciamo che X, variabile aleatoria definita su Ω, è una variabile aleatoria* **binaria** *se $X(\Omega) = \{0, 1\}$.*

Definizione 7.6 (indicatore o funzione indicatrice di un evento)**.** *Sia $E \in \mathscr{P}(\Omega)$ un evento e sia $\mathbf{1}_E$ la funzione definita da*[16]*:*

$$\mathbf{1}_E(\omega) = \begin{cases} 1 & \text{per } \omega \in E, \qquad \text{(cioè se si è verificato l'evento } E\text{);} \\ 0 & \text{per } \omega \notin E, \qquad \text{(cioè se \textbf{non} si è verificato l'evento } E\text{).} \end{cases}$$

La funzione $\mathbf{1}_E$ è dunque una variabile aleatoria binaria, che viene chiamata con il termine **indicatore** *di E (o anche* **funzione indicatrice** *di E).*

Ricordiamo infine che si usa il simbolo $\mathbf{1}_E$, per mettere in evidenza il fatto che tale variabilea aleatoria vale 1 su E.

***Osservazione* 7.3.** Per qualunque v.a. (variabile aleatoria) binaria X esiste $E \subseteq \Omega = \{\omega_1, \cdots, \omega_N\}$ tale che

$$X(\omega_i) = \mathbf{1}_E(\omega_i), \qquad i = 1, ..., N.$$

Poniamo infatti

$$E = \{\omega_i \in \Omega : \ X(\omega_i) = 1\}$$

Si ha allora

$$\overline{E} = \{\omega_i \in \Omega : \ X(\omega_i) = 0\}$$

e dunque possiamo scrivere $X(\omega_i) = \mathbf{1}_E(\omega_i), \qquad i = 1, ..., N$.

La distribuzione di una variabile binaria $X = \mathbf{1}_E$, con $p = \mathbb{P}(E)$, è individuata ovviamente dal fatto che $X(\Omega) = \{0, 1\}$ e da $p_0 = p_X(0) = 1 - p$ e $p_1 = p_X(1) = p$, è detta ***distribuzione di Bernoulli di parametro*** p, e si scrive $X \sim Bern(p)$.

Il seguente risultato mette in luce l'importanza delle funzioni indicatrici e risulterà particolarmente interessante quando parleremo del valore atteso di una variabile aleatoria (vedere la successiva Lezione 9)

***Proposizione* 7.1.** *Una qualunque variabile aleatoria X si può scrivere come combinazione lineare delle variabili aleatorie binarie $\mathbf{1}_{H_k^X}$: Posto $X(\Omega) = \{x_1, ..., x_n\}$ l'insieme dei valori assumibili da X, e la partizione generata da X, ossia*

$$H_\ell^X := \{X = x_\ell\}, \quad \ell = 1, ..., n$$

allora

$$X(\omega) = \sum_{k=1}^{n} x_k \, \mathbf{1}_{H_k^X}(\omega).$$

[16] In altre discipline della Matematica tale funzione viene usualmente chiamata ***funzione caratteristica*** di E, ma in Probabilità si usa un termine diverso, in quanto il termine ***funzione caratteristica*** è utilizzato per un altro concetto, che verrà studiato in corsi successivi.

Dimostrazione. Siano $X_1, ..., X_n$ le variabili aleatorie binarie definite come indicatori degli eventi $H_\ell^X := \{X = x_\ell\}$, $\ell = 1, ..., n$, ovvero

$$X_\ell = \mathbf{1}_{H_\ell^X}$$

È facile convincersi che allora possiamo scrivere

$$X(\omega_i) = \sum_{\ell=1}^{n} x_\ell \cdot X_\ell(\omega_i), \qquad i = 1, ..., N.$$

Infatti basta mostrare che la funzione $Y(\omega)$, definita su Ω da

$$Y(\omega) := \sum_{\ell=1}^{n} x_\ell \cdot X_\ell(\omega)$$

coincide con $X(\omega)$ per ogni ω. A questo scopo basta ricordare che la famiglia di eventi H_ℓ^X forma una partizione dell'evento certo Ω e di conseguenza basta mostrare che

$$Y(\omega) := \sum_{\ell=1}^{n} x_\ell \cdot X_\ell(\omega) = X(\omega) \qquad \text{per ogni } \omega \in H_\kappa^X, \qquad \text{e per ogni } \kappa = 1, ..., n.$$

E infatti per ogni $\omega \in H_\kappa^X$ ovviamente $X(\omega) = x_\kappa$, e anche $Y(\omega) = x_\kappa$, come si vede subito, tenendo conto che $\mathbf{1}_{H_\ell^X}(\omega) = 0$ se $\ell \neq \kappa$ e che ovviamente $\mathbf{1}_{H_\kappa^X}(\omega) = 1$:

$$Y(\omega) = x_\kappa \cdot \mathbf{1}_{H_\kappa^X}(\omega) + \sum_{\ell \neq \kappa} x_k \cdot \mathbf{1}_{H_\ell^X}(\omega) = x_\kappa \cdot 1 + 0 = x_\kappa.$$

□

***Osservazione* 7.4.** Spesso si mira a determinare direttamente la distribuzione di probabilità di una variabile aleatoria, sulla base di considerazioni circa l'esperimento consistente nell'osservare il valore della variabile stessa. In tali casi non teniamo conto dello spazio di probabilità Ω su cui la variabile può essere definita, ne' come tale variabile vi possa essere definita, ne' quale sia la misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$.

Vediamo ora qualche esempio di distribuzione di probabilità.

Esempio 7.1. *Consideriamo ancora una volta l'esperimento legato al lancio di due dadi, in cui*

$$\Omega := \{(h,k) : \ 1 \leq h \leq 6, \ 1 \leq k \leq 6\}.$$

$$\mathbb{P}(\{(h,k)\}) = \frac{1}{36}, \qquad 1 \leq h \leq 6, \quad 1 \leq k \leq 6$$

Su questo spazio possiamo definire diverse variabili aleatorie, ad esempio:

$X_1 : (h,k) \to h$ *("punteggio del primo dado"),* $X_2 : (h,k) \to k$ *("punteggio del secondo dado"),* $X : (h,k) \to h + k$ *("somma dei due punteggi"),* $W : (h,k) \to \frac{h}{k}$*, etc...*

Le variabili aleatorie X_1 e X_2 hanno la stessa distribuzione di probabilità, data da :

$$p_{X_1}(x) = \mathbf{P}_{X_1}(\{x\}) = \mathbb{P}(\{X_1 = x\}) = p_{X_2}(x) = \mathbf{P}_{X_2}(\{x\}) = \mathbb{P}(\{X_2 = x\}) = \frac{1}{6}, \qquad x = 1, 2, ...6;$$

questa è la **distribuzione uniforme** *su {1,2,...,6}.*

La distribuzione di probabilità di $X = X_1 + X_2$ è invece data da

$$p_X(x) = \mathbf{P}_X(\{x\}) = \mathbb{P}(\{X = x\}) = \frac{x-1}{36}, \quad x = 2, 3, ..., 7,$$
$$p_X(x) = \mathbf{P}_X(\{x\}) = \mathbb{P}(\{X = x\}) = \frac{13-x}{36}, \quad x = 8, 9, ..., 12.$$

Esempio 7.2. *Riprendiamo l'Esempio 4.2, considerando il caso $n = 6$: un uomo con 6 chiavi di cui una sola apre la porta, le prova una alla volta fino a quando trova quella giusta.. La variabile aleatoria*

$$T := \textit{numero dei tentativi fino a trovare la chiave giusta}$$

può prendere i valori $1, 2, ..., 6$ e risulta, grazie a quanto avevamo visto, (ma anche perché il problema equivale al problema di estrazioni senza reinserimento da un'urna con 1 palline di tipo A e 5 di tipo B) e, per $k = 1, 2, ..., 6$, l'evento $\{T = k\}$ corrisponde all'evento $A_k :=$ {La palline di tipo A esce alla k-sima estrazione})

$$p_T(x) = \mathbf{P}_T(\{x\}) = \mathbb{P}(\{T = x\}) = \frac{1}{6}, \quad x = 1, 2, ...6.$$

Confrontando tale risultato con quanto visto prima, troviamo l'esempio di due variabili aleatorie (cioè X_1 dell'Esempio 7.1 e T dell'Esempio 7.2), definite su spazi diversi, che hanno la stessa distribuzione di probabilità.

Prima di considerare il successivo esempio è utile fare mente locale sulla seguente semplice osservazione

***Osservazione* 7.5.** Siano $E_1, ..., E_n$ degli eventi in uno spazio $(\Omega, \mathscr{P}(\Omega))$ e indichiamo rispettivamente con $X_1, ..., X_n$ i loro indicatori. Ovviamente $X_1, ..., X_n$ sono delle variabili aleatorie definite su $(\Omega, \mathscr{P}(\Omega))$. Su $(\Omega, \mathscr{P}(\Omega))$ possiamo definire anche la variabile aleatoria $S_n := \sum_{h=1}^{n} X_h$; poniamo cioè

$$S_n(\omega_i) := \sum_{h=1}^{n} X_h(\omega_i) = \sum_{h=1}^{n} \mathbf{1}_{E_h}(\omega_i).$$

Ovviamente S_n ha il significato di ***numero di successi fra gli eventi*** $E_1, ..., E_n$.

Prima di passare ad esaminare i due importanti Esempi 7.3 e 7.4, si noti che la variabile aleatoria S_n può assumere $n + 1$ valori, ossia $S_n(\Omega) = \{0, 1, ..., n\}$. Inoltre la famiglia degli n eventi E_h, $h = 1, ..., n$, ***non coincide*** con la partizione $\big\{\{S_n = 0\}, \{S_n = 1\}, ..., \{S_n = n\}\big\}$, che invece è costituita da $n + 1$ eventi. Questa osservazione mostra anche che la rappresentazione di una variabile aleatoria come combinazione lineare di variabili aleatorie binarie non è unica: posto

$$H_\ell^{S_n} := \{S_n = \ell\}, \quad \text{per } \ell = 0, 1, ..., n,$$

si ha che

$$S_n = \sum_{\ell=0}^{n} \ell\, \mathbf{1}_{H_\ell^{S_n}}$$

corrisponde alla rappresentazione usata nella dimostrazione della ***Proposizione* 7.1**, che è diversa dalla precedente rappresentazione

$$S_n = \sum_{h=1}^{n} \mathbf{1}_{E_h}.$$

Esempio 7.3 (Variabili aleatorie binomiali)**.** *Consideriamo n prove bernoulliane, cioè n eventi completamente indipendenti, ciascuno di probabilità θ $(0 < \theta < 1)$ e consideriamo la variabile aleatoria $S_n :=$ numero di successi sulle n prove. I valori possibili per tale variabile sono ovviamente $0, 1, ..., n$ e, come abbiamo visto nella lezione precedente, si ha,*

$$p_{S_n}(k) = \mathbb{P}(\{S_n = k\}) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \qquad k = 0, 1, ..., n.$$

Si dice che S_n segue una **distribuzione binomiale di parametri *n* e *θ***; *ciò si indica con il simbolo $S_n \sim Bin(n, \theta)$. Le variabili aleatorie S_n, al variare di n e θ, sono il prototipo delle variabili aleatorie binomiali, ma va ricordato che si dice che $X \sim Bin(n, \theta)$ anche per ogni una variabile aletoria X con $X(\Omega) = \{0, 1, ..., n\}$ e per la quale $\mathbb{P}(\{X = k\}) = \binom{n}{k}\theta^k (1 - \theta)^{n-k}$, per ogni $k = 0, 1, ..., n$.*

Esempio 7.4 (Variabili aleatorie ipergeometriche). *Vengono eseguite n estrazioni casuali senza reinserimento da una popolazione che contiene complessivamente M elementi, di cui m_1 elementi di tipo A e $m_2 = M - m_1$ elementi di tipo B. Consideriamo la variabile aleatoria $S_n :=$ numero di elementi di tipo A fra gli n elementi estratti. Sappiamo che vale*

$$p_{S_n}(k) = \mathbb{P}(\{S_n = k\}) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}}, \quad \textit{per } k \textit{ tale che } \max(0, n + m_1 - M) \le k \le \min(n, m_1),$$

in quanto $0 \le k \le m_1,\ 0 \le n - k \le m_2 (= M - m_1)$ equivale a $0 \le k \le m_1,\ n - m_2 \le k \le n$.
Si dice che S_n segue una **distribuzione ipergeometrica di parametri** *M, m_1, n e nel seguito indicheremo questa proprietà con il simbolo $S_n \sim Hyp(M, m_1, n)$.*
Le variabili aleatorie precedenti sono il prototipo delle variabili aleatorie ipergeometriche, ma si dice ugualmente che una variabile aleatoria X ha distribuzione ipergeometrica di parametri M, m_1, n, in simboli $X \sim Hyp(M, m_1, n)$, se $X(\Omega) = \{k \text{ interi: } 0 \le k \le m_1,\ 0 \le n - k \le M - m_1\}$ e $\mathbb{P}(X = k) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}}$, per $k \in X(\Omega)$.

Finora abbiamo quasi esclusivamente considerato variabili aleatorie ***a valori interi*** (cioè tali che $X(\Omega) \subseteq \mathbb{Z}$); ma questi non sono gli unici casi di possibile interesse; nel caso considerato nel precedente Esempio 7.3, è interessante considerare anche la variabile aleatoria Y_n ("frequenza dei successi") definita dalla relazione

$$Y_n = \frac{S_n}{n} = \frac{\sum_{h=1}^n \mathbf{1}_{E_h}}{n}.$$

A questo proposito è interessante più in generale, date n variabili aleatorie $X_1, ..., X_n$, studiare il comportamento probabilistico della media aritmetica

$$Y_n = \frac{\sum_{h=1}^n X_h}{n}.$$

A tale tipo di variabile aleatoria, daremo particolare attenzione nel seguito, in particolare nella Lezione 10, dove si trova una prima versione della Legge dei Grandi Numeri. Inoltre, successivamente ce ne occuperemo anche nel caso di variabili aleatorie più generali, e che possono prendono valori in intervalli dell'insieme dei numeri reali.

7.3 Ulteriori esempi di variabili aleatorie a valori interi

Torniamo al caso di variabili aleatorie a valori interi e notiamo quanto segue.

***Osservazione* 7.6.** Per una variabile aleatoria X, a valori interi, cioè $X(\Omega) \subseteq \mathbb{Z}$, può essere spesso conveniente calcolare la distribuzione di probabilità tenendo conto della relazione

$$\mathbb{P}(X = k) = \mathbb{P}(X \le k) - \mathbb{P}(X \le k - 1), \qquad \text{per } k \in \mathbb{Z}.$$

Altre volte può essere conveniente tenere conto invece che

$$\begin{aligned}\mathbb{P}(X = k) &= \mathbb{P}(X \ge k) - \mathbb{P}(X \ge k + 1)\\ &= \mathbb{P}(X > k - 1) - \mathbb{P}(X > k), \qquad \text{per } k \in \mathbb{Z}.\end{aligned}$$

La dimostrazione delle precedenti relazioni è basato sul fatto che

$$\{X \le k\} = \{X = k\} \cup \{X \le k - 1\} \quad \text{e} \quad \{X \ge k\} = \{X = k\} \cup \{X \ge k + 1\},$$

da cui

$$\mathbb{P}(X \le k) = \mathbb{P}(X = k) + \mathbb{P}(X \le k - 1) \quad \text{e} \quad \mathbb{P}(X \ge k) = \mathbb{P}(X = k) + \mathbb{P}(X \ge k + 1)$$

e infine che $\{X \ge h\} = \{X > h - 1\}$.

Esempio* 7.5. *Siano X_1 e X_2 i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria Z definita come il massimo dei due punteggi. Individuare i valori che può assumere Z e con quali probabilità.*

Soluzione: I valori possibili per Z sono ovviamente $\{1, 2, ..., 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} := \big\{\{X_1 = i\},\ i = 1, \cdots, 6\big\}$ e $\mathcal{B} := \big\{\{X_2 = j\},\ j = 1, \cdots, 6\big\}$ sono indipendenti[17], (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti[18]), risulta

$$\begin{aligned}\mathbb{P}(Z = k) &= \mathbb{P}(Z \le k) - \mathbb{P}(Z \le k-1)\\ &= \mathbb{P}(\{X_1 \le k\} \cap \{X_2 \le k\}) - \mathbb{P}(\{X_1 \le k-1\} \cap \{X_2 \le k-1\})\end{aligned}$$

per l'indipendenza degli eventi $\{X_1 \le h_1\}$ e $\{X_2 \le h_2\}$,

$$\begin{aligned}&= \mathbb{P}(X_1 \le k) \cdot \mathbb{P}(X_2 \le k) - \mathbb{P}(X_1 \le k-1) \cdot \mathbb{P}(X_2 \le k-1)\\ &= \left(\frac{k}{6}\right)^2 - \left(\frac{k-1}{6}\right)^2 = \frac{2k-1}{36}, \qquad k = 1, 2, ..., 6.\end{aligned}$$

Esempio* 7.6. *Siano X_1 e X_2 i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria W definita come il minimo dei due punteggi. Individuare i valori che può assumere W e con quali probabilità.*

Soluzione: I valori possibili per W sono ovviamente $\{1, 2, ..., 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} := \big\{\{X_1 = i\},\ i = 1, \cdots, 6\big\}$ e $\mathcal{B} := \big\{\{X_2 = j\},\ j = 1, \cdots, 6\big\}$ sono indipendenti (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti), risulta

$$\begin{aligned}\mathbb{P}(W = k) &= \mathbb{P}(W \ge k) - \mathbb{P}(W \ge k+1)\\ &= \mathbb{P}(\{X_1 \ge k\} \cap \{X_2 \ge k\}) - \mathbb{P}(\{X_1 \ge k+1\} \cap \{X_2 \ge k+1\})\\ &= \mathbb{P}(X_1 \ge k) \cdot \mathbb{P}(X_2 \ge k) - \mathbb{P}(X_1 \ge k+1) \cdot \mathbb{P}(X_2 \ge k+1)\\ &= \mathbb{P}(X_1 > k-1) \cdot \mathbb{P}(X_2 > k-1) - \mathbb{P}(X_1 > k) \cdot \mathbb{P}(X_2 > k)\\ &= (1 - \mathbb{P}(X_1 \le k-1)) \cdot (1 - \mathbb{P}(X_2 \le k-1)) - (1 - \mathbb{P}(X_1 \le k)) \cdot (1 - \mathbb{P}(X_2 \le k))\\ &= \left(1 - \frac{k-1}{6}\right)^2 - \left(1 - \frac{k}{6}\right)^2 = \frac{49 - 14k + k^2 - (36 - 12k + k^2)}{36}\\ &= \frac{13 - 2k}{36}, \qquad k = 1, 2, ..., 6.\end{aligned}$$

Esercizio proposto 7.1. *Ripetere gli esempi precedenti nel caso in cui i due dadi sono truccati in modo che $\mathbb{P}(X_\ell = i) = Ki$, per $i = 1, \cdots, 6$, ed $\ell = 1, 2$, e si assuma l'indipendenza delle partizioni $\mathcal{A} := \big\{\{X_1 = i\},\ i = 1, \cdots, 6\big\}$ e $\mathcal{B} := \big\{\{X_2 = j\},\ j = 1, \cdots, 6\big\}$, generate rispettivamente da X_1 e da X_2 .*

***Osservazione* 7.7.** Il concetto di variabile aleatoria così come introdotto in questa lezione (vedere la Definizione 7.1), può talvolta apparire un po' forzato o artificiale a chi sia all'inizio dello studio della teoria assiomatica della probabilità. Di fatto, invece, esso si rivela di importanza fondamentale sia nella formalizzazione rigorosa che nella comprensione di numerose questioni specifiche del calcolo delle probabilità.

In particolare esso permette di dare un chiaro significato alle operazioni fra variabili aleatorie, estendendo in modo diretto a questi oggetti (essendo funzioni a valori reali) le operazioni definite nel campo dei numeri reali.

Ad esempio, come abbiamo precedentemente visto, la somma $X = X_1 + X_2$ di due variabili aleatorie X_1, X_2 (definite su uno stesso spazio Ω) altro non è che la funzione definita dalla relazione

$$X(\omega) = X_1(\omega) + X_2(\omega), \text{ per } \omega \in \Omega.$$

[17] Questa proprietà è alla base della successiva definizione di indipendenza stocastica per le v.a. X_1 e X_2

[18] Si veda a questo proposito l'Esercizio proposto 5.6.

Avevamo già avvertito comunque che la Definizione 7.1 di variabile aleatoria, così come è stata formulata, è provvisoria. Come si vedrà, essa va infatti adeguatamente modificata e completata quando si passi a trattare il caso in cui Ω non è un insieme finito (si veda la Lezione 14 ed in particolare la Definizione 14.5).

7.4 Trasformazioni di una variabile aleatoria in spazi finiti

Data una variabile aleatoria X, e una funzione h, si può definire una seconda variabile aleatoria come la trasformata di X tramite la funzione h, ossia

$$Z: \ \Omega \to \mathbb{R}, \quad \omega \mapsto Z(\omega) := h(X(\omega)).$$

Supponendo nota la distribuzione di X, ovvero l'insieme $X(\Omega) = \{x_1, ..., x_n\}$ dei valori che può assumere X, e la sua densità discreta $p_X(x_k)$, $k = 1, ..., n$, nasce immediatamente il problema di come ottenere la distribuzione di Z, ovvero l'insieme $Z(\Omega) = \{z_1, ..., z_m\}$ dei valori che può assumere Z e la sua densità discreta, $p_Z(z_h) = \mathbb{P}(Z = z_h)$.
Ad esempio se X assume i valori ± 2, ± 1 e 0, con $p_X(x) = \frac{1}{5}$ per ogni $x \in X(\Omega)$ ed $h(x) = x^2$, allora $Z = X^2$ può assumere solo i valori 0, 1 e 4, ed inoltre, essendo

$$\{Z = 0\} = \{X = 0\}, \quad \{Z = 1\} = \{X = 1\} \cup \{X = -1\}, \quad \{Z = 4\} = \{X = 2\} \cup \{X = -2\}$$

si ottiene che

$$\begin{aligned}
\mathbb{P}(Z = 0) &= \mathbb{P}(X = 0) = \frac{1}{5}, \\
\mathbb{P}(Z = 1) &= \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = \frac{2}{5}, \\
\mathbb{P}(Z = 4) &= \mathbb{P}(X = 2) + \mathbb{P}(X = -2) = \frac{2}{5}.
\end{aligned}$$

Torniamo ora al caso generale: ovviamente $Z(\Omega) = h\big(X(\Omega)\big)$, e, tenendo conto che

$$\{Z = z_j\} = \bigcup_{\ell : h(x_\ell) = z_j} \{X = x_\ell\},$$

si ottiene immediatamente che, per ogni $z_j \in h(X(\Omega))$,

$$p_Z(z_j) = \mathbb{P}(Z = z_j) = \sum_{\ell : h(x_\ell) = z_j} \mathbb{P}(X = x_\ell) = \sum_{\ell : h(x_\ell) = z_j} p_X(x_\ell).$$

7.5 Esercizi di verifica

Esercizio 7.1. Indichiamo con $X_1, ..., X_5$ i 5 numeri estratti su una ruota del lotto (si estrae senza reinserimento da un'urna contenente i numeri $\{1,2,..., 90\}$) e sia inoltre X il valore più alto fra $X_1, ..., X_5$. Calcolate $\mathbb{P}(X \leq k)$ e $\mathbb{P}(X = k)$ per $k = 1, 2...., 90$.

Esercizio 7.2. Supponiamo che X sia una variabile aleatoria a valori nell'insieme $\{0, 1, ..., n\}$ e che, per una coppia di costanti positive A e ρ, risulti

$$\mathbb{P}(X = k) = A \cdot \frac{\rho^k}{k!\,(n-k)!}, \qquad k = 0, 1, ..., n,$$

Dimostrare che X segue una distribuzione binomiale ed individuarne i parametri.

Esercizio 7.3. Individuare una distribuzione di probabilità (non degenere) per una variabile aleatoria X in modo tale che risulti degenere la distribuzione della variabile aleatoria $Y = X^2$.

Esercizio 7.4. Individuare una distribuzione di probabilità per una variabile aleatoria X (non binaria) in modo tale che $Y = X^2$ risulti una variabile aleatoria binaria.

Esercizio 7.5. Sia X una variabile aleatoria con distribuzione di probabilità binomiale $Bin(6, \frac{1}{3})$. Trovare qual è il valore più probabile per X. (confrontare il risultato con il successivo Esercizio 7.8).

Esercizio 7.6. Consideriamo una variabile aleatoria con distribuzione di probabilità ipergeometrica

$$X \ \propto Hyp(6, 3; 3).$$

Qual è il più probabile fra i due eventi $\{X \leq 1\}, \{X > 1\}$?

Esercizio 7.7. In una lotteria sono stati emessi 1000 biglietti e vengono distribuiti 2 primi premi del valore di 1000 Euro, 4 secondi premi del valore di 500 Euro e 20 terzi premi del valore di 100 Euro.
(a) Trovare la distribuzione di probabilità della variabile aleatoria X che indica il valore del premio associato ad un singolo biglietto.
(b) Scrivere la distribuzione di probabilità della variabile aleatoria $2X$.
Un tizio ha acquistato 2 biglietti della lotteria ed indichiamo con Z la variabile aleatoria che indica il valore complessivo dei premi che potrebbe vincere alla lotteria.
(c) Trovare la distribuzione di probabilità di Z.

Esercizio 7.8. Sia S_n una variabile aleatoria binomiale di parametri n e θ.
(a) Verificate che, per $k = 0, 1, ..., n - 1$

$$\mathbb{P}(S_n = k + 1) = \frac{n - k}{k + 1} \frac{\theta}{1 - \theta} \, \mathbb{P}(S_n = k).$$

(b) Utilizzando la proprietà precedente, verificate che esiste un $\overline{k}$ tale che

$$\begin{aligned} \mathbb{P}(S_n = k + 1) \geq \mathbb{P}(S_n = k) \quad & k \leq \overline{k} \\ \mathbb{P}(S_n = k + 1) \leq \mathbb{P}(S_n = k) \quad & k > \overline{k} \end{aligned}$$

8 Distribuzioni congiunte di più variabili aleatorie

In questa lezione esaminiamo alcune definizioni relative al caso in cui si considerino contemporaneamente, su uno stesso spazio di probabilità, due variabili aleatorie. I concetti che verranno introdotti si possono estendere senza difficoltà al caso di un numero di variabili maggiore di due.

Sia dunque $(\Omega, \mathcal{P}(\Omega), P)$ uno spazio di probabilità e X, Y una coppia di variabili aleatorie definite su di esso, con $X : \Omega \to X(\Omega) := \{x_1, ..., x_n\}, Y : \Omega \to Y(\Omega) := \{y_1, ..., y_m\}$. Possiamo considerare, su Ω, la partizione costituita dagli eventi del tipo

$$\{X = x_i\} \cap \{Y = y_j\} := \{\omega \in \Omega : \ X(\omega) = x_i, Y(\omega) = y_j\}, \qquad i = 1, ..., n; \quad j = 1, ..., m.$$

Spesso, d'ora in poi, la scrittura $\{X = x_i\} \cap \{Y = y_j\}$ verrà più semplicemente sostituita da $\{X = x_i, Y = y_j\}$, come del resto già fatto nella precedente Lezione 7.

In analogia con il caso di una sola variabile aleatoria utilizzeremo la notazione

$$p_{X,Y}(x_i, y_j) := \mathbb{P}(\{X = x_i, Y = y_j\}), \qquad i = 1, ..., n; \quad j = 1, ..., m$$

per mettere in evidenza le variabili aleatorie X ed Y coinvolte, ed i valori $\{x_1, \cdots, x_n\}$ e $\{y_1, \cdots, y_m\}$ che possono assumere.

Definizione 8.1. *La funzione*

$$X(\Omega) \times Y(\Omega) \to [0, 1], \qquad (x_i, y_j) \mapsto p_{X,Y}(x_i, y_j)$$

assume il nome di **densità discreta congiunta** *di X e Y.*

Ovviamente, essendo $\{H^{X,Y}_{i,j} := \{X = x_i, Y = y_j\},\ i = 1, ..., n;\ j = 1, ..., m;\}$ una partizione, risulta

$$p_{X,Y}(x_i, y_j) \geq 0, \quad i = 1, ..., n;\ j = 1, ..., m; \qquad \sum_{i=1}^{n}\sum_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1.$$

Possiamo dunque considerare, analogamente a quanto visto sopra per il caso di una singola variabile aleatoria, il nuovo spazio di probabilità, indotto da X, Y, definito come

$$(X(\Omega) \times Y(\Omega), \mathcal{P}(X(\Omega) \times Y(\Omega)), \mathbf{P}_{X,Y}),$$

dove, per $E \subseteq X(\Omega) \times Y(\Omega)$, si pone

$$\mathbf{P}_{X,Y}(E) = \sum_{(i,j):(x_i,y_j)\in E} p_{X,Y}(x_i, y_j); \tag{50}$$

È importante osservare che

$$\mathbf{P}_{X,Y}(E) = \mathbb{P}(\{(X, Y) \in E\}),$$

dove, analogamente al caso di una variabile aleatoria unidimensionale si usa la notazione abbreviata, ovvero, per ogni sottoinsieme $E \subseteq X(\Omega) \times Y(\Omega)$,

$$\{(X, Y) \in E\} = \{\omega \in \Omega : \ (X(\omega), Y(\omega)) \in E\}.$$

Il fatto che $\mathbf{P}_{X,Y}(E)$ definito in (50) coincida con $\mathbb{P}(\{(X, Y) \in E\})$ è dovuto all'osservazione che

$$\{(X, Y) \in E\} = \bigcup_{i,j:(x_i,y_j)\in E} \{(X, Y) = (x_i, y_j)\} = \bigcup_{i,j:(x_i,y_j)\in E} \{X = x_i, Y = y_j\},$$

da cui

$$\mathbb{P}(\{(X, Y) \in E\}) = \sum_{i,j:(x_i,y_j)\in E} \mathbb{P}(X = x_i, Y = y_j).$$

Definizione 8.2. *La misura di probabilità*

$$\mathbf{P}_{X,Y}:\ (X(\Omega)\times Y(\Omega),\mathscr{P}(X(\Omega)\times Y(\Omega)))\,;\qquad E\mapsto \mathbf{P}_{X,Y}(E)=\mathbb{P}(\{(X,Y)\in E\})=\sum_{(i,j):(x_i,y_j)\in E} p_{X,Y}(x_i,y_j)$$

prende il nome di **distribuzione di probabilità congiunta di** X, Y.

Siano ora date due variabili aleatorie X, Y i cui insiemi di valori possibili siano $\{x_1, ..., x_n\}$, $\{y_1, ..., y_m\}$ rispettivamente e sia la loro distribuzione di probabilità congiunta individuata dalla densità discreta congiunta, ossia dalle probabilità

$$p_{X,Y}(x_i, y_j) := \mathbb{P}(\{X = x_i, Y = y_j\}), \qquad i = 1, ..., n; \quad j = 1, ..., m.$$

Quando la distribuzione congiunta di (X, Y) è nota, possiamo ottenere la distribuzione della variabile aleatoria X, che in questo contesto è detta la ***distribuzione marginale*** per X. Tale distribuzione di probabilità è concentrata sui valori $\{x_1, ..., x_n\}$ e attribuisce loro le probabilità

$$p_X(x_i) = \mathbb{P}(X = x_i) = \sum_{j=1}^{m} \mathbb{P}(\{X = x_i, Y = y_j\}) = \sum_{j=1}^{m} p_{X,Y}(x_i, y_j), \quad i = 1, ..., n, \tag{51}$$

come segue immediatamente dal fatto che $\{X = x_i\} = \bigcup_{j=1}^{m}\{X = x_i, Y = y_j\}$.
In questo contesto, la funzione $x_i \mapsto p_X(x_i)$ viene detta anche ***densità discreta marginale di*** X, e coincide con la densità discreta di X.

Analogamente la ***distribuzione marginale*** per la v.a. Y è la distribuzione di probabilità concentrata sui valori $\{y_1, ..., y_m\}$ e che loro attribuisce le probabilità

$$p_Y(y_j) = \mathbb{P}(Y = y_j) = \sum_{i=1}^{n} \mathbb{P}(\{X = x_i, Y = y_j\}) = \sum_{i=1}^{n} p_{X,Y}(x_i, y_j), \quad j = 1, ..., m, \tag{52}$$

Esempio 8.1. *Siano* X *ed* Y *due variabili aleatorie che possono rispettivamente assumere i valori* $\{-1, 0, 1\}$ *e* $\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$, *con probabilità congiunte*

$$\begin{array}{lll}
\mathbb{P}(X=-1,Y=1/4)=0.1 & \mathbb{P}(X=0,Y=1/4)=0.2 & \mathbb{P}(X=+1,Y=1/4)=0\\
\mathbb{P}(X=-1,Y=1/2)=0 & \mathbb{P}(X=0,Y=1/2)=0.12 & \mathbb{P}(X=+1,Y=1/2)=0.05\\
\mathbb{P}(X=-1,Y=3/4)=0.05 & \mathbb{P}(X=0,Y=3/4)=0.1 & \mathbb{P}(X=+1,Y=3/4)=0.04\\
\mathbb{P}(X=-1,Y=1\quad)=0.1 & \mathbb{P}(X=0,Y=1\quad)=0.04 & \mathbb{P}(X=+1,Y=1\quad)=0.2
\end{array}$$

indicate sinteticamente nella seguente tabella[19]

$Y \diagdown X$	**−1**	**0**	**1**
1/4	0.1	0.2	0
1/2	0	0.12	0.05
3/4	0.05	0.1	0.04
1	0.1	0.04	0.2

Applichiamo la formula (51) *per trovare la distribuzione di probabilità marginale della variabile* X*; otteniamo allora*

$$\begin{aligned}\mathbb{P}(X=-1) &= \mathbb{P}(X=-1,Y=1/4)+\mathbb{P}(X=-1,Y=1/2)+\mathbb{P}(X=-1,Y=3/4)+\mathbb{P}(X=-1,Y=1)\\ &= 0.1+0+0.05+0.1=0.25;\end{aligned}$$

[19] Si noti che tutti i calcoli che seguono non dipendono dallo spazio di probabilità $(\Omega, \mathscr{P}(\Omega), \mathbb{P})$ su cui le due variabili aleatorie X ed Y sono definite, ma possono essere effettuati utilizzando solamente la distribuzione congiunta.

analogamente risulta

$$\begin{aligned}\mathbb{P}(X=0) &= \mathbb{P}(X=0,Y=1/4)+\mathbb{P}(X=0,Y=1/2)+\mathbb{P}(X=0,Y=3/4)+\mathbb{P}(X=0,Y=1)\\ &= 0.2+0.12+0.1+0.04=0.46;\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X=+1) &= \mathbb{P}(X=+1,Y=1/4)+\mathbb{P}(X=+1,Y=1/2)+\mathbb{P}(X=+1,Y=3/4)+\mathbb{P}(X=+1,Y=1)\\ &= 0+0.05+0.04+0.2=0.29;\end{aligned}$$

oppure, più semplicemente $\mathbb{P}(X=+1)=1-\mathbb{P}(X=-1)-\mathbb{P}(X=0)=1-0.25-0.46=0.29$ *Abbiamo cioè ottenuto le probabilità della distribuzione marginale di* X *calcolando le somme degli elementi nelle diverse colonne della tabella. Analogamente, calcolando le somme degli elementi sulle righe, otteniamo la distribuzione marginale della variabile* Y *:*

$$\mathbb{P}\left(Y=\tfrac{1}{4}\right)=0.3,\quad \mathbb{P}\left(Y=\tfrac{1}{2}\right)=0.17,\quad \mathbb{P}\left(Y=\tfrac{3}{4}\right)=0.19,\quad \mathbb{P}\left(Y=1\right)=0.34.$$

Riportiamo allora tali distribuzioni marginali, inserendole in una riga ed in una colonna aggiunte rispettivamente nei margini in basso e a destra della tabella. Cioè completiamo la tabella precedente riportandovi anche le somme di riga e le somme di colonna; otteniamo dunque

$Y\diagdown X$	-1	0	1	p_Y
$1/4$	0.1	0.2	0	0.30
$1/2$	0	0.12	0.05	0.17
$3/4$	0.05	0.1	0.04	0.19
1	0.1	0.04	0.2	0.34
p_X	0.25	0.46	0.29	1

(53)

Passiamo ora dal precedente Esempio 8.1 al caso generale di una coppia di variabili aleatorie (indicate come al solito con X, Y), risulta naturale descrivere la loro distribuzione di probabilità congiunta attraverso una *tabella a doppia entrata* (cioè, insomma, una matrice), come segue:

$Y\diagdown X$	x_1	$\cdots$	x_i	$\cdots$	x_n	$\mathbf{P}_Y$
y_1	$p_{1,1}$	$\cdots$	$p_{i,1}$	$\cdots$	$p_{n,1}$	p''_1
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
y_j	$p_{1,j}$	$\cdots$	$p_{i,j}=\mathbb{P}(X=x_i,Y=y_j)$	$\cdots$	$p_{n,j}$	$p''_j=\mathbb{P}(Y=y_j)$
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
y_m	$p_{1,m}$	$\cdots$	$p_{i,m}$	$\cdots$	$p_{n,m}$	p''_m
$\mathbf{P}_X$	p'_1	$\cdots$	$p'_i=\mathbb{P}(X=x_i)$	$\cdots$	p'_n	1

In questa tabella:

• la prima riga indica i valori $\{x_i, i=1,...,n\}$ che può assumere la variabile aleatorie X, e la prima colonna indica i valori possibili $\{y_j, j=1,...,m\}$ per Y;

• gli elementi nelle righe e colonne interne indicano le probabilità dei corrispondenti eventi, ovvero le probabilità $p_{ij}=\mathbb{P}(\{X=x_i,Y=y_j\})$;

• l'ultima riga e l'ultima colonna, dunque ai *margini* della tabella, vengono riportate rispettivamente le somme di colonna e le somme di riga, cioè le probabilità delle distribuzioni *marginali* (di qui il nome) delle due variabili, ovvero $p'_i=\sum_{j=1}^{m} p_{ij}$ e $p''_j=\sum_{i=1}^{n} p_{ij}$, rispettivamente.

***Osservazione* 8.1.** È interessante osservare che, dati due insiemi $\{x_1, ..., x_n\}$ ed $\{y_1, ..., y_m\}$, e data una tabella a doppia entrata p_{ij}, $i = 1, .., n$, $j = 1, ..., m$, se p_{ij} soddisfa le seguenti condizioni

$$p_{ij} \geq 0, \quad i = 1, ..., n; \; j = 1, ..., m; \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1. \tag{54}$$

Allora possiamo costruire, a partire dall'insieme $\{x_1, ..., x_n\} \times \{y_1, ..., y_m\}$ uno spazio di probabilità, dove la probabilità è definita a partire dalla densità

$$p: \; \{x_1, ..., x_n\} \times \{y_1, ..., y_m\} \to [0, 1]; \quad (x_i, y_j) \mapsto p(x_i, y_j) := p_{ij}.$$

Inoltre se p'_i è definito da

$$p'_i = \sum_{j=1}^{m} p_{ij}, \quad i = 1, ..., n, \tag{55}$$

allora ovviamente p'_i, $i = 1, ..., n$, è effettivamente una densità su $\{x_1, ..., x_n\}$, in quanto risulta

$$p'_i \geq 0, \quad i = 1, ..., n; \qquad \sum_{i=1}^{n} p'_i = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} p_{ij} \right) = 1.$$

Una considerazione analoga vale per

$$p''_j = \sum_{i=1}^{n} p_{ij}, \qquad j = 1, ..., m. \tag{56}$$

Ciò è legato al fatto che nello spazio di probabilità appena costruito è possibile definire delle variabili aleatorie X e Y che hanno come densità congiunta proprio tale densità: basta infatti prendere

$$X: \; \{x_1, ..., x_n\} \times \{y_1, ..., y_m\} \to \{x_1, ..., x_n\}, \quad \omega = (x_i, y_j) \mapsto X\big((x_i, y_j)\big) = x_i$$

$$Y: \; \{x_1, ..., x_n\} \times \{y_1, ..., y_m\} \to \{y_1, ..., y_m\}, \quad \omega = (x_i, y_j) \mapsto Y\big((x_i, y_j)\big) = y_j.$$

Chiaramente la densità discreta congiunta di X, Y è data da

$$p_{X,Y}(x : i, y_j) = p_{ij}, \quad (i, j) \in \{1, ..., n\} \times \{1, ..., m\}$$

e le densità discrete marginali sono date rispettivamente da

$$p_X(x_i) = p'_i, \text{ con } i \in \{1, ..., n\}, \quad \text{e da} \quad p_Y(y_j) = p''_j, \text{ con } j \in \{1, ..., m\}.$$

8.1 Distribuzioni condizionate

In una situazione quale quella descritta nella precedente sezione, ossia con due variabili aleatorie X e Y definite su uno stesso spazio Ω, fissiamo ora $1 \leq j \leq m$ e consideriamo l'evento $\{Y = y_j\}$, che assumiamo avere probabilità strettamente positiva.

Supponiamo ora che sia stato osservato questo evento, mentre non è stato osservato il valore assunto dalla variabile X.

Ci possiamo allora domandare quale sia, data questa informazione, la distribuzione di probabilità che esprime lo stato di informazione parziale circa X.

A questo quesito, risulta naturale rispondere con la seguente definizione:

Definizione 8.3. *Fissato $y_j \in Y(\Omega)$, con $\mathbb{P}(Y = y_j) > 0$, la* **distribuzione di probabilità condizionata della variabile** X, **dato l'evento** $\{Y = y_j\}$, *è la distribuzione di probabilità che concentra, sui valori x_i ($1 \leq i \leq n$), le probabilità condizionate*

$$\mathbb{P}(X = x_i | Y = y_j) := \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}\,(Y = y_j)}, \quad 1 \leq i \leq n.$$

Useremo anche la seguente notazione

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

e infine useremo il termine **densità discreta di** X **condizionata a** $Y = y_j$ *per denotare la funzione*

$$p_{X|Y}(\cdot|y_j) : \ X(\Omega) \to [0, 1], \quad x_i \mapsto p_{X|Y}(x_i|y_j).$$

È ovvio che i valori di $p_{X|Y}(x_i|y_j)$ si ricavano dalle probabilità congiunte tramite la formula

$$p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{i=1}^{n} p_{X,Y}(x_i, y_j)}, \tag{57}$$

e che, qualunque sia $j = 1, \cdots, m$,

$$p_{X|Y}(x_i|y_j) \geq 0, \quad i = 1, \cdots n, \qquad \sum_{i=1}^{n} p_{X|Y}(x_i|y_j) = 1.$$

Esempio 8.2. *Siano X ed Y come nell'Esempio 8.1 e fissiamo l'evento $\{Y = 1/2\}$. Allora la distribuzione di probabilità della variabile X, condizionata all'evento $\{Y = 1/2\}$ è individuata dalla densità discreta $p_{X|Y}(i|1/2) = \mathbb{P}(X = i | \{Y = 1/2\})$ condizionata a $Y = 1/2$:*

$$p_{X|Y}(-1|1/2) = \mathbb{P}(X = -1 | \{Y = 1/2\}) = \frac{\mathbb{P}(X = -1, Y = 1/2)}{\mathbb{P}(Y = 1/2)} = \frac{0}{0.17} = 0$$
$$p_{X|Y}(0|1/2) = \mathbb{P}(X = 0 | \{Y = 1/2\}) = \frac{\mathbb{P}(X = 0, Y = 1/2)}{\mathbb{P}(Y = 1/2)} = \frac{0.12}{0.17} = \frac{12}{17} \approx 0,706$$
$$p_{X|Y}(1|1/2) = \mathbb{P}(X = 1 | \{Y = 1/2\}) = \frac{\mathbb{P}(X = 1, Y = 1/2)}{\mathbb{P}(Y = 1/2)} = \frac{0.05}{0.17} = \frac{5}{17} \approx 0,294.$$

Ovviamente $p_{X|Y}(-1|1/2) + p_{X|Y}(0|1/2) + p_{X|Y}(1|1/2) = 0 + \frac{12}{17} + \frac{5}{17} = 1$. D'altra parte la densità discreta condizionale di X dato $Y = y$ si ricava prendendo la riga relativa alla densità discreta congiunta $p_{X,Y}(x_i, 1/2)$, cioè, in questo caso, la riga

$Y \diagdown X$	-1	0	1	p_Y
$1/2$	0	0.12	0.05	0.17

ed effettuandone la sua normalizzazione.

Analogamente definiamo, per un fissato indice $1 \leq i \leq n$, la ***distribuzione condizionata di*** Y, ***dato l'evento*** $\{X = x_i\}$, con $\mathbb{P}(X = x_i) > 0$, come la distribuzione di probabilità che concentra, sui valori y_j ($1 \leq j \leq m$), le probabilità condizionate date da

$$p_{Y|X}(y_j|x_i) := \mathbb{P}(Y = y_j | X = x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{j=1}^{m} p_{X,Y}(x_i, y_j)}. \tag{58}$$

Anche in questo caso, qualunque sia $i = 1, \cdots, n$

$$p_{Y|X}(y_j|x_i) \geq 0, \quad j = 1, \cdots m, \qquad \sum_{j=1}^{m} p_{Y|X}(y_j|x_i) = 1$$

***Osservazione* 8.2.** Consideriamo una coppia di variabili aleatorie X ed Y, per le quali gli insiemi di valori possibili siano $X(\Omega) := \{x_1, ..., x_n\}$ e $Y(\Omega) := \{y_1, ..., y_m\}$. La distribuzione di probabilità congiunta è allora individuata dall'insieme delle probabilità congiunte, che per brevità scriveremo come

$$\{p_{ij} = p_{X,Y}(x_i, y_j);\ i = 1, ..., n;\ j = 1, ..., m\}.$$

In base a $\{p_{ij};\ i = 1, ..., n;\ \ j = 1, ..., m\}$, attraverso le formule (51), (52), (57) , (58), si determinano univocamente
• le probabilità marginali

$$\{p'_i = p_X(x_i);\ i = 1, ..., n\}, \qquad \{p''_j = p_Y(y_j);\ j = 1, ..., m\}$$

• e le probabilità condizionate

$$\{p'_{i|j} = p_{X|Y}(x_i|y_j);\ j = 1, ..., m,\ i = 1, ..., n\}, \qquad \{p''_{j|i} = p_{Y|X}(y_j|x_i);\ i = 1, ..., n,\ j = 1, ..., m\}.$$

Supponiamo ora invece di assegnare la coppia $\{p'_i;\ i = 1, ..., n\}$, $\{p''_j;\ j = 1, ..., m\}$, (con $p'_i \geq 0$ e $p''_j \geq 0$) delle densità discrete marginali; si ricordi che vanno rispettati i vincoli dati dalle condizioni di normalizzazione

$$\sum_{i=1}^{n} p'_i = 1; \qquad \sum_{j=1}^{m} p''_j = 1.$$

Per ogni tabella p_{ij} a doppia entrata che ammette come probabilità marginali $\{p'_i;\ i = 1, ..., n\}$ e $\{p''_j;\ j = 1, ..., m\}$, deve necessariamente valere

$$\begin{cases} \sum_{j=1}^{m} p_{ij} = p'_i, & i = 1, ..., n \\ \sum_{i=1}^{n} p_{ij} = p''_j, & j = 1, ..., m \\ \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1. \end{cases} \tag{59}$$

Osserviamo che (59) può anche essere interpretato come un sistema lineare nelle incognite p_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, in cui le probabilità marginali p'_i e p''_j sono termini noti: si tratta di un sistema di $(n + m + 1)$ equazioni in $n \cdot m$ incognite, che risulta indeterminato, nel caso in cui $n \geq 2$, $m \geq 2$.
Vi sono infatti diverse distribuzioni congiunte che ammettono $\{p'_i;\ i = 1, ..., n\}$, $\{p''_j;\ j = 1, ..., m\}$ come probabilità marginali. Ad esempio, come è facile verificare, c'è sempre la soluzione

$$p_{ij} = p'_i \cdot p''_j, \qquad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m.$$

Tale soluzione equivale a considerare la densità discreta congiunta come prodotto delle densità discrete marginali:

$$p_{X,Y}(x_i, y_j) = p_X(x_i) p_Y(y_j), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, m.$$

Come vedremo nella sezione successiva, corrisponde all'*indipendenza delle due variabili aleatorie* X, Y (vedere la Definizione 8.4).

Tuttavia $p'_i \cdot p''_j$ non è l'unica soluzione. Per convincersene basta prendere le probabilità marginali della tabella (53) dell'Esempio 8.1 e osservare che la distribuzione congiunta chiaramente differisce da $p'_i \cdot p''_j$: ad esempio $\mathbb{P}(X = -1, Y = \frac{1}{2}) = 0$ mentre $\mathbb{P}(X = -1)\,\mathbb{P}(Y = \frac{1}{2}) \neq 0$.

Sulle distribuzioni congiunte con marginali date

In realtà, date $\{p'_i;\ i = 1,...,n\}$ e $\{p''_j;\ j = 1,...,m\}$, con le condizioni di normalizzazione

$$\sum_{i=1}^{n} p'_i = 1; \qquad \sum_{j=1}^{m} p''_j = 1,$$

il sistema che devono soddisfare le $n \cdot m$ incognite p_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, è

$$\begin{cases} \sum_{j=1}^{m} p_{ij} = p'_i, & i = 1,...,n \\ \sum_{i=1}^{n} p_{ij} = p''_j, & j = 1,...,m \\ p_{ij} \geq 0, & i = 1,...,n \quad j = 1,...,m. \end{cases}$$

Infatti, grazie alle condizioni di normalizzazione per $\{p'_i;\ i = 1,...,n\}$ e $\{p''_j;\ j = 1,...,m\}$, la condizione $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$ è automaticamente soddisfatta.
Si tratta di un sistema con $(n-1) \cdot (m-1)$ gradi di libertá.

Per convincersi che i gradi di libertà del sistema

$$\begin{cases} \sum_{j=1}^{m} p_{ij} = p'_i, & i = 1,...,n \\ \sum_{i=1}^{n} p_{ij} = p''_j, & j = 1,...,m \\ p_{ij} \geq 0, & i = 1,...,n \quad j = 1,...,m. \end{cases}$$

sono $(n-1) \cdot (m-1)$, si cominci con il caso $n = m = 2$. Chiaramente in questo caso basta fissare, ad esempio p_{11}, con $0 \leq p_{11} \leq p'_1$ e $p_{11} \leq p''_1$, ovvero $p_{11} \leq \min(p'_1, p''_1)$, per ottenere automaticamente i valori $p_{12} = p'_1 - p_{11}$ e $p_{21} = p''_1 - p_{11}$. Infine $p_{22} = p'_2 - p_{21} = p''_2 - p_{12} = 1 - (p_{11} + p_{12} + p_{21})$.

Lasciamo al lettore più interessato la verifica che le condizioni di normalizzazione $p'_1 + p'_2 = 1$ e $p''_1 + p''_2 = 1$ garantiscono che $p'_2 - p_{21} = p''_2 - p_{12}$ e che, per ottenere $p_{22} \geq 0$, bisogna imporre anche la condizione che $p_{11} \geq p''_2 - p'_1$ e $p_{11} \geq p'_2 - p''_1$.

Il caso generale è analogo: ad esempio, si possono fissare i valori di $p_{i,j}$ per $i = 1,...,n-1$ e $j = 1,...,m-1$, in modo che

$$p_{i,j} \geq 0, \ i = 1,...,n-1, \ j = 1,...,m-1,$$

$$\sum_{j=1}^{m-1} p_{ij} \leq p'_i, \quad i = 1,...,n-1 \quad e \quad \sum_{i=1}^{n-1} p_{ij} \leq p''_j, \quad j = 1,...,m-1.$$

I rimanenti $n - 1 + m - 1 + 1 = n + m - 1$ valori

$$p_{n,j} \text{ per } j = 1,...,m-1,\ p_{i,m}, \text{ per } i = 1,...,n-1, \text{ e } p_{n,m}$$

sono automaticamente ricavati dalle equazioni del sistema. Bisogna infine imporre la condizione che anche questi ultimi valori siano non negativi, altrimenti la soluzione trovata non è una distribuzione congiunta.

***Osservazione* 8.3.** Siano date due variabili aleatorie X e Y. Sappiamo che la loro densità discreta congiunta (e quindi la loro distribuzione congiunta) è determinata dagli insiemi $\{x_1, ..., x_n\}$ e $\{y_1, ..., y_m\}$ dei valori che possono assumere rispettivamente, e dalla tabella $\{p_{ij} = p_{X,Y}(x_i, y_j);\ i = 1, ..., n; j = 1, ..., m\}$. Questa tabella risulta univocamente determinata quando si impongano, ad esempio, sia le probabilità marginali $\{p'_i = p_X(x_i);\ i = 1, ..., n\}$ che le probabilità condizionate $\{p''_{j|i} = p_{Y|X}(y_j|x_i);\ i = 1, ..., n,\ j = 1, ..., m\}$. Infatti si ha

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j|X = x_i) \qquad \Leftrightarrow \qquad p_{ij} = p'_i \cdot p''_{j|i}, \qquad i = 1, ..., n, \quad j = 1, ..., m.$$

Ciò mostra anche che la conoscenza di $\{p'_i = p_X(x_i);\ i = 1, ..., n\}$ e $\{p''_{j|i} = p_{Y|X}(y_j|x_i);\ i = 1, ..., n,\ j = 1, ..., m\}$ determina la coppia $\{p''_j = p_Y(y_j);\ j = 1, ..., m\}$ $\{p'_{i|j} = p_{X|Y}(x_i|y_j);\ i = 1, ..., n,\ j = 1, ..., m\}$.

Ovviamente, scambiando il ruolo di X ed Y si ottiene anche il viceversa.

Notiamo anche che la relazione che lega fra loro tali probabilità non è nient'altro che la Formula di Bayes

$$\mathbb{P}(X = x_i|Y = y_j) = \frac{\mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j|X = x_i)}{\mathbb{P}(Y = y_j)} = \frac{\mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j|X = x_i)}{\sum_\ell \mathbb{P}(X = x_\ell)\,\mathbb{P}(Y = y_j|X = x_\ell)},$$

oppure, equivalentemente,

$$\mathbb{P}(Y = y_j|X = x_i) = \frac{\mathbb{P}(Y = y_j)\,\mathbb{P}(X = x_i|Y = y_j)}{\mathbb{P}(X = x_i)} = \frac{\mathbb{P}(Y = y_j)\,\mathbb{P}(X = x_i|Y = y_j)}{\sum_k \mathbb{P}(Y = y_k)\,\mathbb{P}(X = x_i|Y = y_k)},$$

Tali formule, riadattate al simbolismo breve introdotto, possono essere riscritte, rispettivamente, nella forma:

$$p'_{i|j} = \frac{p'_i \cdot p''_{j|i}}{p''_j} = \frac{p'_i \cdot p''_{j|i}}{\sum_\ell p'_\ell \cdot p''_{j|\ell}}, \qquad \text{e} \qquad p''_{j|i} = \frac{p''_j \cdot p'_{i|j}}{p'_i} = \frac{p''_j \cdot p'_{i|j}}{\sum_k p''_k \cdot p'_{i|k}}.$$

Esempio 8.3. *Sia X una variabile aleatoria con distribuzione binomiale di parametri $n = 3$ e $\theta = 1/4$, ossia*

$$p_X(0) = \mathbb{P}(X = 0) = \binom{3}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^3 = \frac{27}{64}, \quad p_X(1) = \mathbb{P}(X = 1) = \binom{3}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^2 = \frac{27}{64},$$
$$p_X(2) = \mathbb{P}(X = 2) = \binom{3}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^1 = \frac{9}{64}, \quad p_X(3) = \mathbb{P}(X = 3) = \binom{3}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^0 = \frac{1}{64}.$$

Supponiamo inoltre che, per ogni $h \in X(\Omega)$, la distribuzione condizionata di Y dato $\{X = h\}$ sia uniforme in $\{k$ interi: $0 \le k \le h\}$, ossia

$$p_{Y|X}(0|0) = \mathbb{P}(Y = 0|\{X = 0\}) = 1, \qquad p_{Y|X}(k|0) = \mathbb{P}(Y = k|\{X = 0\}) = 0,\ k = 1, 2, 3;$$
$$p_{Y|X}(k|1) = \mathbb{P}(Y = k|\{X = 1\}) = \frac{1}{2}, \quad k = 0, 1, \qquad p_{Y|X}(k|1) = \mathbb{P}(Y = k|\{X = 1\}) = 0,\ k = 2, 3;$$
$$p_{Y|X}(k|2) = \mathbb{P}(Y = k|\{X = 2\}) = \frac{1}{3}, \quad k = 0, 1, 2, \qquad p_{Y|X}(k|2) = \mathbb{P}(Y = k|\{X = 2\}) = 0,\ k = 3;$$
$$p_{Y|X}(k|3) = \mathbb{P}(Y = k|\{X = 3\}) = \frac{1}{4}, \quad k = 0, 1, 2, 3.$$

Possiamo quindi facilmente calcolare la densità discreta congiunta di X, Y: $p_{X,Y}(h, k) = \mathbb{P}(X = h, Y = k) = \mathbb{P}(Y = k|X = h)\,\mathbb{P}(X = h)$, da cui

$$p_{X,Y}(h, k) = \begin{cases} \frac{1}{h+1}\binom{3}{h}\left(\frac{1}{4}\right)^h\left(\frac{3}{4}\right)^{3-h}, & \textit{per } 0 \le k \le h, \\ 0 & \textit{altrimenti} \end{cases}$$

Inoltre è possibile calcolare la densità marginale di Y

$$\begin{aligned} p_Y(0) &= \sum_{h=0}^{3} \mathbb{P}(X=h, Y=0) = \sum_{h=0}^{3} \mathbb{P}(X=h)\, \mathbb{P}(Y=0|X=h) \\ &= \frac{27}{64}\,1 + \frac{27}{64}\,\frac{1}{2} + \frac{9}{64}\,\frac{1}{3} + \frac{1}{64}\,\frac{1}{4} = \frac{108+54+12+1}{256} = \frac{175}{256} \end{aligned}$$

tenendo conto che $\mathbb{P}(X=0, Y=1) = 0$

$$\begin{aligned} p_Y(1) &= \sum_{h=0}^{3} \mathbb{P}(X=h, Y=1) = \sum_{h=1}^{3} \mathbb{P}(X=h)\, \mathbb{P}(Y=1|X=h) \\ &= \frac{27}{64}\,\frac{1}{2} + \frac{9}{64}\,\frac{1}{3} + \frac{1}{64}\,\frac{1}{4} = \frac{54+12+1}{256} = \frac{67}{256} \end{aligned}$$

tenendo conto che $\mathbb{P}(X=0, Y=2) = \mathbb{P}(X=1, Y=2) = 0$

$$p_Y(2) = \sum_{h=0}^{3} \mathbb{P}(X=h, Y=2) = \sum_{h=2}^{3} \mathbb{P}(X=h)\, \mathbb{P}(Y=2|X=h) = \frac{9}{64}\,\frac{1}{3} + \frac{1}{64}\,\frac{1}{4} = \frac{12+1}{256} = \frac{13}{256}$$

tenendo conto che $\mathbb{P}(X=0, Y=3) = \mathbb{P}(X=1, Y=3) = \mathbb{P}(X=2, Y=3) = 0$

$$p_Y(3) = \sum_{h=0}^{3} \mathbb{P}(X=h, Y=3) = \mathbb{P}(X=3, Y=3) = \mathbb{P}(X=3)\, \mathbb{P}(Y=3|X=3) = \frac{1}{64}\,\frac{1}{4} = \frac{1}{256}$$

Possiamo riassumere i risultati ottenuti nella tabella

$Y \diagdown X$	**0**	**1**	**2**	**3**	p_Y
0	$\frac{108}{256}$	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$	$\frac{175}{256}$
1	0	$\frac{54}{256}$	$\frac{12}{256}$	$\frac{1}{256}$	$\frac{67}{256}$
2	0	0	$\frac{12}{256}$	$\frac{1}{256}$	$\frac{13}{256}$
3	0	0	0	$\frac{1}{256}$	$\frac{1}{256}$
p_X	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$	1

Infine è possibile anche calcolare la densità discreta condizionale di X dato $Y = k$: prima di tutto osserviamo che

$$p_{X|Y}(k|h) = \mathbb{P}(X=h|\{Y=k\}) = 0 \ \text{ per tutti i valori di } h \text{ e } k \text{ per i quali } h < k.$$

in quanto $X \geq Y$ sempre, e successivamente calcoliamo

$$\begin{aligned} p_{X|Y}(0|0) &= \mathbb{P}(X=0|\{Y=0\}) \propto \mathbb{P}(X=0, Y=0) = \frac{108}{256}, \quad \Leftrightarrow \quad p_{X|Y}(0|0) \propto 108, \\ p_{X|Y}(1|0) &= \mathbb{P}(X=1|\{Y=0\}) \propto \mathbb{P}(X=1, Y=0) = \frac{54}{256}, \quad \Leftrightarrow \quad p_{X|Y}(1|0) \propto 54, \\ p_{X|Y}(2|0) &= \mathbb{P}(X=2|\{Y=0\}) \propto \mathbb{P}(X=2, Y=0) = \frac{12}{256}, \quad \Leftrightarrow \quad p_{X|Y}(1|0) \propto 12, \\ p_{X|Y}(3|0) &= \mathbb{P}(X=3|\{Y=0\}) \propto \mathbb{P}(X=2, Y=0) = \frac{1}{256}, \quad \Leftrightarrow \quad p_{X|Y}(1|0) \propto 1, \end{aligned}$$

da cui, tenendo conto che $\sum_{h=0}^{3} p_{X,Y}(h|0) = 1$,

$$p_{X|Y}(0|0) = \tfrac{108}{175}, \quad p_{X|Y}(1|0) = \tfrac{54}{175}, \quad p_{X|Y}(2|0) = \tfrac{12}{175}, \quad p_{X|Y}(3|0) = \tfrac{1}{175}.$$

In modo analogo otteniamo:

$$p_{X|Y}(0|1) \propto 0, \quad p_{X|Y}(1|1) \propto 54, \quad p_{X|Y}(2|1) \propto 12, \quad p_{X|Y}(3|1) \propto 1,$$

$$p_{X|Y}(0|2) \propto 0, \quad p_{X|Y}(1|2) \propto 0, \quad p_{X|Y}(2|2) \propto 12, \quad p_{X|Y}(3|2) \propto 1,$$

$$p_{X|Y}(0|3) \propto 0, \quad p_{X|Y}(1|3) \propto 0, \quad p_{X|Y}(2|1) \propto 0, \quad p_{X|Y}(3|1) \propto 1.$$

e quindi

$$p_{X|Y}(0|1) = 0, \quad p_{X|Y}(1|1) = \tfrac{54}{67}, \quad p_{X|Y}(2|1) = \tfrac{12}{67}, \quad p_{X|Y}(3|1) = \tfrac{1}{67},$$

$$p_{X|Y}(0|2) = 0, \quad p_{X|Y}(1|2) = 0, \quad p_{X|Y}(2|2) = \tfrac{12}{13}, \quad p_{X|Y}(3|2) = \tfrac{1}{13},$$

$$p_{X|Y}(0|3) = 0, \quad p_{X|Y}(1|3) = 0, \quad p_{X|Y}(2|1) = 0, \quad p_{X|Y}(3|1) = 1.$$

8.2 Trasformazioni di coppie di variabili aleatorie in spazi finiti

Siano X ed Y due variabili aleatorie e $g(x,y)$ una fissata funzione di due variabili a valori reali; sia infine U come la variabile aleatoria definita da

$$U = g(X,Y).$$

In modo del tutto analogo a quanto visto nella Sezione 7.4 sulle trasformazioni di una variabile aleatoria, c'è un procedimento generale per ottenere la distribuzione della variabile aleatoria U:

$$p_U(u) = \mathbb{P}(U = u) = \sum\sum_{(x_i,y_j)\in X(\Omega)\times Y(\Omega):\, g(x_i,y_j)=u} p_{X,Y}(x_i,y_j).$$

Per verificarlo basta osservare che

$$\begin{aligned}
\{U = u\} &= \{g(X,Y) = u\} = \big\{(X,Y) \in \{(x,y) :\ g(x,y) = u\}\big\} \\
&= \big\{(X,Y) \in \{(x,y) \in X(\Omega)\times Y(\Omega) :\ g(x,y) = u\}\big\} \\
&= \bigcup_{\substack{(x_i,y_j)\,\in\, X(\Omega)\times Y(\Omega) \\ \text{tali che } g(x_i,y_j)=u}} \{(X,Y) = (x_i,y_j)\} = \bigcup_{\substack{(x_i,y_j)\,\in\, X(\Omega)\times Y(\Omega) \\ \text{tali che } g(x_i,y_j)=u}} \{X = x_i, Y = y_j\}.
\end{aligned}$$

In casi specifici, il calcolo della densità discreta di U può essere effettuato utilizzando la tabella della densità discreta congiunta di X, Y, come spiegato nella seguente nota (a pagina 104).

Densità discreta di U=$g(X,Y)$ con la tabella di $p_{X,Y}$

Siano X ed Y come nell'Esempio 8.1, e sia $g(x,y) = |xy|$. Oltre alla tabella della distribuzione congiunta di X ed Y aggiungiamo una tabella in cui poniamo in ogni casella relativa alla coppia (x_i, y_j) il valore della funzione $g(x_i, y_j)$, ossia, in questo caso, il valore di $|x_i y_j|$,

$Y \diagdown X$	-1	0	1
$1/4$	0.1	0.2	0
$1/2$	0	0.12	0.05
$3/4$	0.05	0.1	0.04
1	0.1	0.04	0.2

$Y \diagdown X$	-1	0	1
$1/4$	$1/4$	0	$1/4$
$1/2$	$1/2$	0	$1/2$
$3/4$	$3/4$	0	$3/4$
1	1	0	1

Dalla seconda tabella vediamo immediatamente che la variabile aleatoria U assume valori nell'insieme $\{0,\, 1/4,\, 1/2,\, 3/4,\, 1\}$. Per calcolare la sua distribuzione basta sommare le probabilità corrispondenti, ossia, ad esempio, essendo

$$\{U = 0\} = \bigcup_{y_j \in Y(\Omega)} \{X = 0, Y = y_j\}, \quad Y(\Omega) = \{1/4,\, 1/2,\, 3/4,\, 1\}$$

si ottiene

$$\begin{aligned} \mathbb{P}(U = 0) \; &= \mathbb{P}(X = 0, Y = 1/4) + \mathbb{P}(X = 0, Y = 1/2) \\ &\quad + \mathbb{P}(X = 0, Y = 3/4) + \mathbb{P}(X = 0, Y = 1), \\ = \; &0.2 + 0.12 + 0.1 + 0.04 = 0.46. \end{aligned}$$

In modo del tutto analogo possiamo calcolare

$$\begin{aligned} \mathbb{P}(U = 1/4) \; &= \mathbb{P}(X = -1, Y = 1/4) + \mathbb{P}(X = 1, Y = 1/4) = 0.1 + 0 = 0.1 \\ \mathbb{P}(U = 1/2) \; &= \mathbb{P}(X = -1, Y = 1/2) + \mathbb{P}(X = 1, Y = 1/2) = 0 + 0.05 = 0.05 \\ \mathbb{P}(U = 3/4) \; &= \mathbb{P}(X = -1, Y = 3/4) + \mathbb{P}(X = 1, Y = 3/4) = 0.05 + 0.04 = 0.09 \\ \mathbb{P}(U = 1) \; &= \mathbb{P}(X = -1,\, Y = 1\,) + \mathbb{P}(X = 1,\, Y = 1\,) = 0.1 + 0.2 = 0.3 \end{aligned}$$

Vale la pena osservare che può bastare anche una sola tabella, in cui mettiamo da una parte (e in grassetto) i valori che assume $g(x_i, y_j)$ e dall'altra mettiamo la densità discreta congiunta di X, Y, ossia $p_{X,Y}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$:

$Y \diagdown X$	-1		0		1	
$1/4$	$1/4$	0.1	0	0.2	$1/4$	0
$1/2$	$1/2$	0	0	0.12	$1/2$	0.05
$3/4$	$3/4$	0.05	0	0.1	$3/4$	0.04
1	1	0.1	0	0.04	1	0.2

Per calcolare $\mathbb{P}(U = u)$, si tratta solo di sommare i valori delle probabilità (a destra) per tutte le coppie di valori (x_i, y_j) in cui la funzione $g(x_i, y_j)$ assume il valore u.

Tuttavia si può procedere anche senza l'aiuto della tabella: ad esempio possiamo usare il fatto che la famiglia $H_i^X = \{X = x_i\}$, $i = 1, \ldots, n$ è una partizione, e scrivere

$$\{U = u\} = \{U = u\} \cap \Omega = \{U = u\} \cap \cup_{i=1}^n H_i^X = \cup_{i=1}^n \big(\{U = u\} \cap H_i^X\big)$$

e osservare che, per ogni $i = 1, 2, \ldots, n$,

$$H_i^X \cap \{U = u\} = \{X = x_i\} \cap \{g(X,Y) = u\} = \{X = x_i\} \cap \{g(x_i, Y) = u\},$$

da cui

$$p_U(u) = \mathbb{P}(U = u) = \sum_{x_i \in X(\Omega)} \mathbb{P}(H_i^X \cap \{g(X,Y) = u\})$$
$$= \sum_{x_i \in X(\Omega)} \mathbb{P}(X = x_i, g(X,Y) = u) = \sum_{x_i \in X(\Omega)} \mathbb{P}(X = x_i, g(x_i, Y) = u)$$

da cui, ancora

$$p_U(u) = \sum_{x_i \in X(\Omega)} \sum_{\substack{y_j \in Y(\Omega): \\ g(x_i, y_j) = u}} \mathbb{P}(X = x_i, Y = y_j),$$

e ritrovare così la formula generale.
Analogamente si può prendere invece la partizione $H_j^Y = \{Y = y_j\}$, $j = 1, \ldots, m$, e ottenere

$$p_U(u) = \mathbb{P}(U = u) = \sum_{y_j \in Y(\Omega)} \mathbb{P}(\{g(X,Y) = u\} \cap H_j^Y)$$
$$= \sum_{y_j \in Y(\Omega)} \mathbb{P}(g(X,Y) = u, y_j) = \sum_{y_j \in Y(\Omega)} \mathbb{P}(g(X, y_j) = u, Y = y_j)$$
$$= \sum_{y_j \in Y(\Omega)} \sum_{\substack{x_i \in X(\Omega): \\ g(x_i, y_j) = u}} \mathbb{P}(X = x_i, Y = y_j).$$

Trasformazioni bidimensionali Si procede in modo del tutto analogo anche nel caso in cui si abbia una coppia di funzioni $(x,y) \mapsto g_1(x,y)$ e $(x,y) \mapsto g_2(x,y)$, che definiscono le variabili aleatorie

$$U = g_1(X,Y) \quad V = g_2(X,Y).$$

In formule possiamo scrivere

$$p_{U,V}(u,v) = \mathbb{P}(U = u, V = v) = \sum_{\substack{(x_i,y_j): \\ g_1(x_i,y_j) = u \\ g_2(x_i,y_j) = v}} \sum p_{X,Y}(x_i, y_j)$$

Se, ad esempio, sempre nelle ipotesi dell'Esempio 8.1 $g_1(x,y) = |xy|$ e $g_2(x,y) = |x/(y - \frac{5}{8})|$, possiamo anche procedere con la tabella nel seguente modo e ottenere

$Y \diagdown X$	-1		0		1	
$1/4$	$(1/4, 8/3)$	0.1	$(0,0)$	0.2	$(1/4, 8/3)$	0
$1/2$	$(1/2, 8)$	0	$(0,0)$	0.12	$(1/2, 8)$	0.05
$3/4$	$(3/4, 8)$	0.05	$(0,0)$	0.1	$(3/4, 8)$	0.04
1	$(1, 8/3)$	0.1	$(0,0)$	0.04	$(1, 8/3)$	0.2

da cui otteniamo immediatamente la tabella della densità discreta congiunta di U, V:

$V \diagdown U$	0	$1/4$	$1/2$	$3/4$	1
0	0.46	0	0	0	0
$8/3$	0	0.1	0	0	0.3
8	0	0	0.05	0.09	0

Esercizio proposto 8.1. *Consideriamo i due punteggi X_1 e X_2 derivanti dal lancio di due dadi e definiamo le variabili aleatorie*

$$X = X_1 + X_2, \quad Y = \max(X_1, X_2).$$

Per tali variabili sono già state calcolate (nell'Esempio 7.1 e nell'Esempio 7.5 della precedente lezione) le distribuzioni marginali. Costruite ora la tabella delle probabilità congiunte e calcolate la distribuzione condizionata di Y, dato l'evento $\{X = 9\}$.

Esercizio proposto 8.2. *Consideriamo i due punteggi X_1, X_2, X_3 e X_4 derivanti dal lancio di quattro dadi e definiamo le variabili aleatorie*

$$X = X_1 + X_2, \quad Y = \max(X_1, X_2).$$

$$X' = X_3 + X_4, \quad Y' = Y = \max(X_1, X_2).$$

Dopo aver osservato che le variabili aleatorie X ed X' hanno la stessa distribuzione marginale, e che lo stesso vale ovviamente per Y ed Y', costruite tabella delle probabilità congiunte di X' e Y' e calcolate la distribuzione condizionata di X, dato l'evento $\{Y = 5\}$, e quella di X' dato l'evento $\{Y' = 5\}$.

8.3 Indipendenza stocastica fra variabili aleatorie.

Vogliamo ora definire il concetto di indipendenza stocastica fra due variabili aleatorie X, Y.

Ricordando quanto discusso nel caso di due eventi potremo dire, dal punto di vista euristico, che due variabili aleatorie sono indipendenti se, qualunque informazione raccolta su una delle due variabili, ad esempio X, non porta a modificare lo stato di informazione su Y. Arriveremo subito, in effetti, a formulare una definizione rigorosa proprio partendo da tali considerazioni.

Cominciamo con il seguente, semplice, ma ***fondamentale*** esercizio:

Esercizio proposto 8.3. *Siano X ed Y due variabili aleatorie. Verificare*[20]*che, le seguenti condizioni sono fra di loro equivalenti:*

(i) le probabilità condizionate $y_j \mapsto p''_{j|i} = p_{Y|X}(y_j|x_i) = \mathbb{P}(Y = y_j|X = x_i)$ non dipendono dall'indice i, ossia dal valore x_i;

(ii) $\forall\ 1 \leq i \leq n, 1 \leq j \leq m$ *risulta*

$$p''_{j|i} = p''_j \quad \Leftrightarrow \quad p_{Y|X}(y_j|x_i) = p_Y(y_j) \quad \Leftrightarrow \quad \mathbb{P}(Y = y_j|X = x_i) = \mathbb{P}(Y = y_j);$$

(iii) $\forall\ 1 \leq i \leq n, 1 \leq j \leq m$ *risulta*

$$p_{ij} = p'_i \cdot p''_j \quad \Leftrightarrow \quad p_{X,Y}(x_i, y_j) = p_X(x_i)\, p_Y(y_j) \quad \Leftrightarrow \quad \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\,\mathbb{P}(Y = y_j);.$$

[20]Le implicazioni $(iii) \Rightarrow (ii)$, $(ii) \Rightarrow (i)$ sono ovvie. Basta a questo punto dimostrare l'implicazione $(i) \Rightarrow (iii)$: si osservi che, se vale (i) e se si pone $q_j := p''_{j|1} := p''_{j|i}$, allora

$$p_{ij} = p'_i \cdot p''_{j|i} = p'_i \cdot q_j.$$

Da ciò, sommando sugli indici $i = 1, \ldots, n$, si ottiene che

$$p''_j = \sum_{i=1}^{n} p_{ij} = \sum_{i=1}^{n} p'_i \cdot q_j = q_j \cdot \sum_{i=1}^{n} p'_i = q_j \cdot 1 = q_j.$$

Confrontando tra loro le ultime due relazioni si ottiene immediatamente la (iii).

Potremo dire allora che X ed Y sono stocasticamente indipendenti qualora si verifichi una (e quindi tutte) delle condizioni (i), (ii), o (iii) del precedente esercizio.

A partire, in particolare, da (iii) e, ricordando la Definizione 5.4 della Lezione 5 di partizioni indipendenti, potremo allora giungere alla seguente

Definizione 8.4. *Le variabili aleatorie X ed Y si dicono* **stocasticamente indipendenti** *se*

$$\mathbb{P}(X \in I,\, Y \in J) = \mathbb{P}(X \in I)\,\mathbb{P}(Y \in J), \qquad \textit{qualunque siano } I \subseteq X(\Omega) \textit{ e } J \subseteq Y(\Omega) \tag{60}$$

o equivalentemente se sono fra loro indipendenti le due partizioni

$$\mathcal{A} := \big\{\{X = x_1\}, ..., \{X = x_n\}\big\} \quad e \quad \mathcal{B} := \big\{\{Y = y_1\}, ..., \{Y = y_m\}\big\}.$$

In altre parole le variabili aleatorie X ed Y sono stocasticamente indipendenti se e solo se la densità discreta congiunta è il prodotto delle densità discrete marginali, ossia se e solo se vale[21]

$$p_{ij} = p'_i \cdot p''_j, \quad \Leftrightarrow \quad p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j) \qquad \forall\, 1 \le i \le n, \quad 1 \le j \le m,$$

ovvero

$$\mathbb{P}\left(\{X = x_i\} \cap \{Y = y_j\}\right) = \mathbb{P}\left(\{X = x_i\}\right) \cdot \mathbb{P}\left(\{Y = y_j\}\right), \qquad \forall\, 1 \le i \le n, \quad 1 \le j \le m \tag{60$'$}$$

L'equivalenza delle due condizioni nella definizione di indipendenza si ottiene grazie alla ***Proposizione* 5.2** della Lezione 5. Infatti, posto $\mathcal{A} := \big\{\{X = x_1\}, ..., \{X = x_n\}\big\}$, la partizione generata dalla variabile aleatoria X, e $\mathcal{B} := \big\{\{Y = y_1\}, ..., \{Y = y_m\}\big\}$ la partizione generata dalla variabile aleatoria Y, e denotando con $\mathcal{G}_X = \mathcal{G}(\mathcal{A})$, l'algebra generata dalla partizione $\mathcal{A}$, e con $\mathcal{G}_Y = \mathcal{G}(\mathcal{B})$, l'algebra generata dalla partizione $\mathcal{B}$, la ***Proposizione* 5.2** garantisce che

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B), \qquad \text{qualunque siano } A \in \mathcal{G}_X \text{ e } B \in \mathcal{G}_Y,$$

ovvero la (60). Per capire questa equivalenza con la (60) basta ricordare che
• l'algebra generata da $\mathcal{A}$ è la famiglia degli eventi A che si possono scrivere come unioni di elementi della partizione ossia

$$A \in \mathcal{G}_X = \mathcal{G}(\mathcal{A}) \quad \Leftrightarrow A = \cup_{i: x_i \in I}\{X = x_i\}, \text{ per un } I \subset \{1, ..., n\};$$

• gli eventi $\{X \in I\}$ sono appunto elementi di $\mathcal{G}(\mathcal{A})$, in quanto

$$\{X \in I\} = \cup_{i: x_i \in I}\{X = x_i\}$$

e analogamente l'algebra generata da $\mathcal{B}$ è la famiglia degli eventi

$$\{Y \in J\} = \cup_{j: y_j \in J}\{Y = y_j\}, \quad J \subset \{1, 2, .., m\}.$$

In realtà, come già detto e illustrato nella nota precedente, basta solo applicare la ***Proposizione* 5.2** della Lezione 5. Tuttavia ci sembra utile riportarne la dimostrazione diretta:

***Proposizione* 8.1.** *Le condizioni (60) e (60)$'$ per l'indipendenza di X e Y nella precedente Definizione 8.4 sono equivalenti, ossia*

$$\mathbb{P}(X \in I,\, Y \in J) = \mathbb{P}(X \in I)\,\mathbb{P}(Y \in J), \qquad \textit{qualunque siano } I \subseteq X(\Omega) \textit{ e } J \subseteq Y(\Omega)$$
$$\Updownarrow$$
$$\mathbb{P}\left(\{X = x_i\} \cap \{Y = y_j\}\right) = \mathbb{P}\left(\{X = x_i\}\right) \cdot \mathbb{P}\left(\{Y = y_j\}\right), \qquad \forall\, 1 \le i \le n, \quad 1 \le j \le m.$$

[21] Si noti che si tratta della condizione (iii) dell'Esercizio proposto 8.3.

Dimostrazione. Iniziamo prendendo come ipotesi la seconda condizione, ossia la condizione che $\{X = x_i\} \perp\!\!\!\perp \{Y = y_j\}$ per ogni $i = 1, ...m,\ j = 1, ..., m$, e mostriamo che $\{X \in I\} \perp\!\!\!\perp \{Y \in J\}$ per ogni $I \subseteq X(\Omega)$ e $J \subseteq Y(\Omega)$.
Gli eventi $\{X \in I\}$ ed $\{Y \in J\}$ si possono scrivere rispettivamente come

$$\{X \in I\} = \bigcup_{i\colon x_i \in I} \{X = x_i\}, \qquad \{Y \in J\} = \bigcup_{j\colon y_j \in J} \{Y = y_j\},$$

e quindi

$$\{X \in I\} \cap \{Y \in J\} = \left(\bigcup_{i\colon x_i \in I} \{X = x_i\} \right) \bigcap \left(\bigcup_{j\colon y_j \in J} \{Y = y_j\} \right) = \bigcup_{i\colon x_i \in I} \bigcup_{j\colon y_j \in J} \{X = x_i\} \cap \{Y = y_j\}.$$

Di conseguenza allora

$$\begin{aligned} \mathbb{P}(\{X \in I\} \cap \{Y \in J\}) &= \sum_{i\colon x_i \in I} \sum_{j\colon y_j \in J} \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) \\ &= \sum_{i\colon x_i \in I} \sum_{j\colon y_j \in J} \mathbb{P}(\{X = x_i\}) \cdot \mathbb{P}(\{Y = y_j\}) = \sum_{i\colon x_i \in I} \mathbb{P}(\{X = x_i\}) \cdot \sum_{j\colon y_j \in J} \mathbb{P}(\{Y = y_j\}) \\ &= \sum_{i\colon x_i \in I} \mathbb{P}(\{X = x_i\}) \cdot \mathbb{P}(\{Y \in J\}) = \mathbb{P}(X \in I)\, \mathbb{P}(Y \in J). \end{aligned}$$

Ciò dimostra che se $\{X = x_i\}$ e $\{Y = y_j\}$ sono indipendenti per ogni $i \in \{1, 2, ..., n\}$ e per ogni $\jmath \in \{1, 2, ..., m\}$ allora anche tutti gli eventi del tipo $\{X \in I\}$ ed $\{Y \in J\}$ sono indipendenti.

La dimostrazione dell'implicazione inversa è banale: basta prendere $I = \{x_i\}$ e $J = \{y_j\}$.

□

8.3.1 Distribuzione della somma di X e Y e distribuzione condizionata di X dato il valore della somma

In questa sezione proponiamo al lettore i seguenti due importanti esercizi:

Esercizio proposto 8.4. *Siano X ed Y variabili aleatorie stocasticamente indipendenti con*

$$X\,(\Omega) := \{x_1, ..., x_n\} \quad e \quad Y\,(\Omega) := \{y_1, ..., y_m\}$$

e siano note le relative densità discrete marginali.
(i) Trovare l'espressione della distribuzione di probabilità della variabile $Z := X + Y$.
(ii) Trovare l'espressione della distribuzione condizionata di probabilità di X, dato $\{Z = z\}$.

Esercizio proposto 8.5. *Siano X ed Y variabili aleatorie stocasticamente indipendenti, con distribuzioni binomiali di parametri (r, θ) e (s, θ), rispettivamente.*
(i) Determinare la distribuzione di probabilità di $Z := X + Y$.
(ii) Determinare la distribuzione di probabilità condizionata di X, dato $\{Z = k\}$, con $0 \leq k \leq r + s$.

Vista l'importanza dei problemi proposti, li risolviamo in modo rapido.

Soluzione del punto (i) dell'Esercizio-proposto 8.4.

Ricordiamo che X e Y sono due variabili aleatorie e che stiamo cercando la distribuzione di $Z = X + Y$.

Iniziamo notando immediatamente che $Z(\Omega) = \{z \in \mathbb{R} : \exists x_k \in X(\Omega),\, y_h \in Y(\Omega) :\ x_k + y_h = z\}$, e

$$\begin{aligned}\{Z = z\} = \{X + Y = z\} &= \bigcup_{x_k,\, y_h \colon x_k + y_h = z} \{X = x_k,\, Y = y_h\} \\ &= \bigcup_k \{X = x_k,\, Y = z - x_k\} = \bigcup_h \{X = z - y_h,\, Y = y_h\}\end{aligned}$$

dove gli eventi $\{X = x_k,\, Y = y_h\}$ sono disgiunti a due a due, e, in modo del tutto simile, lo stesso vale per gli eventi $\{X = x_k,\, Y = z - x_k\}$ (o per gli eventi $\{X = z - y_h,\, Y = y_h\}$), e dove alcuni degli eventi $\{X = x_k,\, Y = z - x_k\}$ (o $\{X = z - y_h,\, Y = y_h\}$) possono essere vuoti.

Quindi, ***anche senza l'ipotesi di indipendenza*** tra X ed Y, si ha

$$\begin{aligned}\mathbb{P}(\{Z = z\}) = \boxed{\mathbb{P}(\{X + Y = z\}) =} &\sum_{x_k,\, y_h \colon x_k + y_h = z} \mathbb{P}(\{X = x_k,\, Y = y_h\}) \\ &\boxed{= \sum_k \mathbb{P}(\{X = x_k,\, Y = z - x_k\}) = \sum_h \mathbb{P}(\{X = z - y_h,\, Y = y_h\}).}\end{aligned}$$

Inoltre, ***se si suppone l'indipendenza*** tra X ed Y, allora

$$\begin{aligned}\mathbb{P}(\{Z = z\}) = \boxed{\mathbb{P}(X + Y = z) =} &\sum_{x_k,\, y_h \colon x_k + y_h = z} \mathbb{P}(X = x_k)\, \mathbb{P}(Y = y_h) \\ &\boxed{= \sum_k \mathbb{P}(X = x_k)\, \mathbb{P}(Y = z - x_k) = \sum_h \mathbb{P}(X = z - y_h)\, \mathbb{P}(Y = y_h)}.\end{aligned}$$

Soluzione del punto (ii) dell'Esercizio-proposto 8.4.

Si tratta di trovare, per i valori di z tali che $\mathbb{P}(Z = z) \neq 0$, l'espressione di

$$\mathbb{P}(X = x_k | \{Z = z\}) = \frac{\mathbb{P}(\{X = x_k\} \cap \{Z = z\})}{\mathbb{P}(Z = z)}.$$

Basta allora osservare che $\{X = x_k\} \cap \{Z = z\} = \{X = x_k,\, Z = z\}$ coincide con l'evento

$$\{X = x_k, X + Y = z\} = \{X = x_k, x_k + Y = z\} = \{X = x_k,\, Y = z - x_k\},$$

e utilizzare il precedente punto *(i)*, ed ottenere

$$\boxed{\mathbb{P}(\{X = x_k\} | \{X + Y = z\}) = \frac{\mathbb{P}(X = x_k,\, Y = z - x_k)}{\sum_{k'} \mathbb{P}(X = x_{k'},\, Y = z - x_{k'})}}$$

Soluzione del punto (i) dell'Esercizio-proposto 8.5:
Ricordiamo che $X \sim Bin(r,\theta)$ ed $Y \sim Bin(s,\theta)$ sono due variabili aleatorie indipendenti e che $Z = X + Y$. Si ottiene che $Z = X + Y \sim Bin(r + s, \theta)$.

Infatti, utilizzando il risultato del precedente Esercizio 8.4, essendo $X \sim Bin(r,\theta)$ ed $Y \sim Bin(s,\theta)$, indipendenti, si ha che $Z(\Omega) = 0, 1, 2, \ldots r + s$,

$$\begin{aligned}
\mathbb{P}(Z=\ell) = \boxed{\mathbb{P}(X+Y=\ell) =} &\sum_{\substack{k+h=\ell\\ 0\le k\le r,\ 0\le h\le s}} \binom{r}{k}\theta^k(1-\theta)^{r-k}\binom{s}{h}\theta^h(1-\theta)^{s-h}\\
&= \sum_{\substack{k+h=\ell\\ 0\le k\le r,\ 0\le h\le s}} \theta^\ell(1-\theta)^{r+s-\ell}\binom{r}{k}\binom{s}{h}\\
&= \theta^\ell(1-\theta)^{r+s-\ell}\sum_{0\le k\le r,\ 0\le \ell-k\le s}\binom{r}{k}\binom{s}{\ell-k},
\end{aligned}$$

da cui, ricordando l'dentità di Vandermonde (25),

$$\boxed{\mathbb{P}(Z=\ell) = \theta^\ell(1-\theta)^{r+s-\ell}\binom{r+s}{\ell} \qquad \ell = 0,1,\ldots,r+s.}$$

Va notato che è possibile arrivare allo stesso risultato con un ragionamento sintetico (si veda la succcessiva ***Osservazione 8.4***)

Soluzione del punto (ii) dell'Esercizio 8.5:

Ricordiamo che $X \sim Bin(r,\theta)$ ed $Y \sim Bin(s,\theta)$ sono due variabili aleatorie indipendenti e che $Z = X + Y$. Si ottiene che la distribuzione condizionata di X dato $\{Z = k\}$ è di tipo ipergeometrico, e precisamente $Hyp(r + s, r; k)$.

Infatti, utilizzando le definizioni di probabilità condizionata si ha:

$$\mathbb{P}(X=i|Z=k) = \frac{\mathbb{P}(X=i,\ Z=k)}{\mathbb{P}(Z=k)},$$

considerando che $\{X = i, Z = k\} = \{X = i, Y = k - i\}$, e che $\{X = i, Y = k - i\} \neq \emptyset$, solo se $0 \le i \le r$ e $0 \le k - i \le s$, e tenendo conto che $\theta^i\theta^{k-i} = \theta^k$ e $(1-\theta)^{r-(k-i)}(1-\theta)^{s-(k-i)} = (1-\theta)^{r+s-k}$, si ottiene

$$\begin{aligned}
\boxed{\mathbb{P}(X=i|Z=k)} &= \frac{\mathbb{P}(X=i,\ Y=k-i)}{\mathbb{P}(Z=k)} = \frac{\mathbb{P}(X=i)\,\mathbb{P}(Y=k-i)}{\mathbb{P}(Z=k)}\\
&= \frac{\binom{r}{i}\theta^i(1-\theta)^{r-i}\binom{s}{k-i}\theta^{k-i}(1-\theta)^{s-(k-i)}}{\binom{r+s}{k}\theta^k(1-\theta)^{r+s-k}}\\
&\boxed{= \frac{\binom{r}{i}\binom{s}{k-i}}{\binom{r+s}{k}} \qquad 0\le i\le r \text{ e } 0\le k-i\le s.}
\end{aligned}$$

***Osservazione* 8.4.** Questi ultimi due risultati non sono sorprendenti, infatti si potrebbe anche ragionare come segue:

siano E_j, per $j = 1, 2, \ldots, r+s$, eventi globalmente indipendenti di probabilità θ, cioè tale da formare uno schema di Bernoulli.

Consideriamo le variabili $X' = \sum_{j=1}^{r} \mathbf{1}_{E_j}$ e $Y' = \sum_{j=r+1}^{r+s} \mathbf{1}_{E_j}$. Per tali variabili aleatorie valgono le seguenti proprietà:

(a) X' ha la stessa distribuzione di X, cioè $X' \sim Bin(r, \theta)$,
(b) Y' ha la stessa distribuzione di Y, cioè $Y' \sim Bin(s, \theta)$,
(c) X' ed Y' sono indipendenti[22],

e quindi (X', Y') ***ha la stessa distribuzione congiunta di*** (X, Y). Di conseguenza,

- da una parte $Z' := X' + Y' = \sum_{i=1}^{r+s} \mathbf{1}_{E_j}$ ha la stessa distribuzione di $Z = X + Y$,
- mentre dall'altra parte Z' ha chiaramente distribuzione binomiale $Bin(r+s, \theta)$.

E con ciò si ottiene che anche Z ha distribuzione binomiale $Bin(r+s, \theta)$.

Inoltre anche la distribuzione condizionata di X dato $\{Z = k\}$ coincide con la distribuzione condizionata di X' dato $\{Z' = k\}$. Il fatto di sapere che si è verificato l'evento $\{Z' = k\}$, permette di affermare che si sono verificate solo le sequenze di

$$(E_1^* \cap E_2^* \cap \ldots \cap E_r^*) \cap (E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*)$$

nelle quali per esattamente k tra gli E_ℓ^* si ha $E_\ell^* = E_\ell$, mentre per i restanti $r+s-k$ si ha $E_\ell^* = \overline{E}_\ell$.

Per individuare uno di questi eventi basta specificare quali sono i k indici ℓ per cui $E_\ell^* = E_\ell$, ovvero basta specificare un sottoinsieme $K \subseteq \{1, 2, \ldots, r+s\}$ di cardinalità k. Formalmente infatti possiamo scrivere che

$$\{Z' = k\} = \bigcup_{K \subseteq \{1,2,\ldots,r+s\}: |K| = k} \Big(\big(\cap_{i \in K} E_i \big) \cap \big(\cap_{j \notin K} \overline{E}_j \big) \Big)$$

Ciascuno degli $\binom{r+s}{k}$ eventi di questo tipo ha la stessa probabilità (ed esattamente $\theta^k \cdot (1-\theta)^{r+s-k}$).
L'evento $X' = i$, per $i = 0, 1, \ldots r$ corrisponde al caso in cui la cardinalità di $K_A := K \cap \{1, 2, \ldots, r\}$ è uguale ad i (e quindi la cardinalità di $K_B := K \cap \{r+1, r+2, \ldots, r+s\}$ è uguale ad $k-i$). Tenendo conto di ciò ci si può convincere facilmente che

$$\mathbb{P}(X' = i | Z' = k) = \frac{\binom{r}{i}\binom{s}{k-i}}{\binom{r+s}{k}} \qquad \text{per } 0 \leq i \leq r, \quad 0 \leq k-i \leq s,$$

cioè la distribuzione condizionata di X' dato $\{Z' = k\}$ è una distribuzione $Hyp(r+s, r; k)$.

---

[22] La proprietà (c) è intuitiva. Per una dimostrazione formale della la proprietà (c) bisogna utilizzare la ***Proposizione* 5.2** considerando che
i) la partizione $\mathcal{A}$ generata dagli eventi $\{E_1, E_2, \ldots, E_r\}$ e la partizione $\mathcal{B}$ generata dagli eventi $\{E_{r+1}, E_{r+2}, \ldots, E_{r+s}\}$ sono indipendenti:
infatti un evento $A \in \mathcal{A}$ se e solo se $A = E_1^* \cap E_2^* \cap \ldots \cap E_r^*$ dove $E_i^* = E_i$ oppure $E_i^* = \overline{E}_i$, e analogamente un evento $B \in \mathcal{B}$ se e solo se $B = E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*$ dove $E_j^* = E_j$ oppure $E_j^* = \overline{E}_j$, e quindi $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$, in quanto

$$\mathbb{P}\big(\underbrace{(E_1^* \cap E_2^* \cap \ldots \cap E_r^*)}_{A} \cap \underbrace{(E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*)}_{B} \big) = \underbrace{\mathbb{P}(E_1^*) \cdot \mathbb{P}(E_2^*) \cdot \cdots \cdot \mathbb{P}(E_r^*)}_{\mathbb{P}(A)} \cdot \underbrace{\mathbb{P}(E_{r+1}^*) \cdot \mathbb{P}(E_{r+2}^*) \cdot \cdots \cdot \mathbb{P}(E_{r+s}^*)}_{\mathbb{P}(B)};$$

ii) gli eventi del tipo $E = \{X' = k\}$ e gli eventi del tipo $F = \{Y' = k\}$ appartengono rispettivamente a $\mathcal{G}(\mathcal{A})$ e a $\mathcal{G}(\mathcal{B})$, le algebre generate dalle partizioni $\mathcal{A}$ e $\mathcal{B}$.

8.4 Esercizi di verifica

Esercizio 8.1. Siano X ed Y variabili aleatorie, entrambe a valori nell'insieme $\{-1, 0, 1\}$ e con distribuzione congiunta data dalla tabella

$Y \diagdown X$	$\mathbf{-1}$	$\mathbf{0}$	$\mathbf{1}$
$\mathbf{-1}$	0.1	0.1	0
$\mathbf{0}$	0	0.1	0.3
$\mathbf{1}$	0.1	0.15	θ

essendo θ un opportuno valore $0 < \theta < 1$.
(a) Determinare il valore di θ.
(b) Determinare sia la distribuzione marginale di X che quella di Y.
(c) Le variabili aleatorie X ed Y sono indipendenti?
(d) Determinare la distribuzione di probabilità condizionata di X, dato $\{Y = 1\}$, ossia calcolare la funzione di densità discreta condizionata $p_{X|Y}(k|1) = \mathbb{P}(X = k|\{Y = 1\})$ di X dato $Y = 1$.
(e) Calcolare la distribuzione di $U = |XY|$ e calcolare $\mathbb{P}(Y = 1|U = 1)$.

Esercizio 8.2. Siano X ed Y le variabili aleatorie definite nel precedente Esercizio 8.1, e sia $Z := X \cdot Y$.
(a) Determinare la distribuzione di probabilità congiunta della coppia (X, Z), dove $Z := X \cdot Y$.
(b) Determinare la distribuzione di probabilità marginale di Z.
(c) Le variabili aleatorie X ed Z sono indipendenti?
(d) Determinare la distribuzione di probabilità condizionata di X, dato $\{Z = 1\}$, ossia calcolare la funzione di densità discreta condizionata $p_{X|Z}(x|1) = \mathbb{P}(X = x|\{Z = 1\})$ di X dato $Z = 1$.

Esercizio 8.3. Verificare se sono stocasticamente indipendenti le variabili aleatorie X ed Y, la cui distribuzione di probabilità congiunta è stata considerata nel precedente Esempio 8.1.

Esercizio 8.4. Siano X ed Y le variabili aleatorie considerate nell'Esercizio proposto 8.1 di questa lezione, ossia X ed Y sono rispettivamente somma e massimo dei punteggi X_1 e X_2 risultanti dal lancio di due dadi ben equilibrati.
(a) Verificare se le variabili aleatorie X ed Y sono stocasticamente indipendenti o no.
(b) Determinare la densità discreta condizionata $p_{X|Y}(k|h)$ della somma X dato che il valore massimo $Y = h$, per $h = 1, 2, ..., 6$.

9 Valore atteso di una variabile aleatoria e relative proprietà

In questa lezione verrà introdotta la nozione di "valore atteso"di una variabile aleatoria e ne verranno messe in evidenza proprietà ed aspetti fondamentali.

In molti problemi di tipo probabilistico, data una variabile aleatoria X, sorge la necessità di individuare una quantità deterministica che, in qualche senso e entro certi fini, sia *equivalente* ad X.

Ad esempio se X rappresenta il valore (aleatorio) del ricavo derivante da una operazione finanziaria[23], nasce spesso l'esigenza di valutare una cifra deterministica, che risulti equa quale importo da pagare per avere il diritto di godere di tale ricavo[24]. Similmente, in un gioco d'azzardo vi è la necessità di verificare se il gioco sia, o meno, equo, e così via...

Come si comincerà a vedere qui di seguito, il concetto di valore atteso ha un ruolo fondamentale in tali problematiche e risulta anche ugualmente importante sia sul piano teorico, sia in altre applicazioni, ad esempio di tipo fisico, di tipo statistico, etc...

Cominciamo intanto con una definizione rigorosa di tale concetto.

Definizione 9.1 (valore atteso). *Sia* $X : \Omega := \{\omega_1, ..., \omega_N\} \to X(\Omega) \subseteq \mathbb{R}$ *una variabile aleatoria definita su* $(\Omega, \mathcal{P}(\Omega))$. *Se inoltre su* $(\Omega, \mathcal{P}(\Omega))$ *è definita una probabilità* $\mathbb{P}$, *si definisce* **valore atteso** *di* X *(rispetto a* $\mathbb{P}$*) il numero*

$$\mathbb{E}(X) := \sum_{i=1}^{N} p(\omega_i) X(\omega_i),$$

dove, come al solito, $p(\omega_i) = \mathbb{P}(\{\omega_i\})$.

Come sappiamo dalla Lezione 7, una variabile aleatoria X è una funzione definita su Ω, e quindi ovviamente **non dipende dalla probabilità** $\mathbb{P}$. Al contrario, il suo **valore atteso dipende dalla probabilità** $\mathbb{P}$.

Inoltre, come vedremo nella successiva ***Proposizione* 9.7**, il valore atteso dipende SOLO dalla distribuzione di X, e precisamente, se $X(\Omega) = \{x_1, x_2, ..., x_n\}$ e $x_k \mapsto p_X(x_k) = \mathbb{P}(X = x_k)$ è la sua densità discreta, si ha che

$$\boxed{\mathbb{E}(X) := \sum_{i=1}^{N} p(\omega_i) X(\omega_i) = \sum_{k=1}^{n} x_k p_X(x_k) = \sum_{k=1}^{n} x_k \, \mathbb{P}(X = x_k).}$$

Oltre a valore atteso si usano anche i termini ***valore medio***, ***aspettazione***, ***speranza matematica***, e a volte anche ***media***, che però è meglio evitare, perché spesso quando si parla di media ci si riferisce alla media aritmetica.

Prima di presentare alcuni esempi illustrativi è bene elencare alcune proprietà immediate di tale definizione; in quanto segue $X, Y, Z, ...$ sono variabili aleatorie definite sullo stesso spazio $\Omega = \{\omega_1, ..., \omega_N\}$.

[23] Ad esempio nel caso di un'***opzione call*** l'operazione consiste nel comprare la possibilità, ma non l'obbligo, di pagare una azione ad un prezzo prefissato K ad un istante prefissato T, invece che al prezzo (aleatorio) di mercato dell'azione. Ovviamente tale possibilità viene esercitata se il prezzo (aleatorio) al tempo T è maggiore di K, altrimenti non viene esercitata: non è conveniente pagare K quello che si può ottenere sul mercato ad un prezzo minore.

[24] Più in generale quando ci si assicura contro eventuali danni o furti, quando si scommette o si gioca in borsa, si deve pagare una quantità di denaro certa (il premio di un'assicurazione, l'ammontare della scommessa) in cambio di una quantità aleatoria (il denaro che si potrebbe ottenere in caso di danno o furto, l'importo della scommessa, in caso di vincita).

È interessante considerare anche il caso del gioco in borsa, in cui la quantità certa è, ad esempio, il prezzo di un'***opzione call***, ovvero il prezzo per ottenere la possibilità di pagare una azione ad un prezzo prefissato K, nell'istante T, mentre la quantità aleatoria è il ricavo tra il prezzo dell'azione al tempo T e K, se questa differenza è positiva.

La prima proprietà è immediata[25](ricordando che deve essere, ovviamente, $\sum_{i=1}^{N} p(\omega_i) = 1$):

***Proposizione* 9.1.** *Se X è una variabile aleatoria degenere, cioè tale che, per un valore $\widehat{x} \in \mathbb{R}$, vale $X(\omega_i) = \widehat{x}$, $i = 1, ..., N$, allora si ha*

$$\mathbb{E}(X) = \widehat{x}.$$

Anche le dimostrazioni delle seguenti due proprietà seguono banalmente dalla definizione.

***Proposizione* 9.2.** *Sia $E \subseteq \Omega$ un evento e $X = \mathbf{1}_E$ la variabile aleatoria indicatore di E. Si ha*

$$\mathbb{E}(X) = \mathbb{E}(\mathbf{1}_E) = \mathbb{P}(E).$$

La variabile aleatoria $X(\omega) = \mathbf{1}_E(\omega)$ assume solo i due valori 0 e 1:

$X(\omega) = \mathbf{1}_E(\omega) = 1$ SE E SOLO SE $\omega \in E$

mentre

$X(\omega) = \mathbf{1}_E(\omega) = 0$ SE E SOLO SE $\omega \in \overline{E}$

quindi, sommando prima sugli $\omega_i \in E$ e poi sugli $\omega_i \in \overline{E}$ si ottiene

$$\mathbb{E}(X) = \mathbb{E}(\mathbf{1}_E) = \sum_{i=1}^{N} \mathbf{1}_E(\omega_i)\, p(\omega_i) = \sum_{i:\,\omega_i \in E} \overbrace{\mathbf{1}_E(\omega_i)}^{=1}\, p(\omega_i) + \sum_{i:\,\omega_i \in \overline{E}} \overbrace{\mathbf{1}_E(\omega_i)}^{=0}\, p(\omega_i)$$
$$= \sum_{i:\,\omega_i \in E} p(\omega_i) = \mathbb{P}(E)$$

***Proposizione* 9.3.** *Siano $a \leq b$ due numeri reali tali che $a \leq X(\omega_i) \leq b$, $\forall 1 \leq i \leq N$. Allora*

$$a \leq \mathbb{E}(X) \leq b.$$

Più in generale vale la **proprietà di monotonia**: *se X ed Y sono due variabili aleatorie con la proprietà che $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$, allora*

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

La verifica è molto semplice: Essendo $p(\omega) \geq 0$ per ogni $\omega \in \Omega$, si ha che

$$X(\omega) \leq Y(\omega) \text{ per ogni } \omega \in \Omega, \quad \Rightarrow \quad X(\omega)\, p(\omega) \leq Y(\omega) p(\omega) \text{ per ogni } \omega \in \Omega$$

e di conseguenza

$$\overbrace{\sum_{i=1}^{N} X(\omega_i)\, p(\omega_i)}^{=\mathbb{E}(X)} \leq \overbrace{\sum_{i=1}^{N} Y(\omega_i)\, p(\omega_i)}^{=\mathbb{E}(Y)}.$$

Inoltre, se $a \leq X(\omega) \leq b$ per ogni $\omega \in \Omega$, prendendo poi le variabili aleatorie degenri $Y_1 = a$ ed $Y_2 = b$, l'ipotesi diviene $Y_1(\omega) \leq X(\omega) \leq Y_2(\omega)$ per ogni $\omega \in \Omega$, e quindi, per la proprietà di monotonia appena verificata si ha

$$a = \mathbb{E}(Y_1) \leq \mathbb{E}(X) \leq \mathbb{E}(Y_2) = b.$$

[25] E del resto, pensando all'interpretazione di $\mathbb{E}(X)$ come quantità certa che si è disposti a scambiare con X, si capisce che se X è a sua volta una quantità certa $\widehat{x}$, ci si aspetta che $\mathbb{E}(X)$ sia uguale a $\widehat{x}$.

Una proprietà assolutamente fondamentale del valore atteso è la ***proprietà di linearità***, che verrà enunciata e verificata qui di seguito.

***Proposizione* 9.4** (Proprietà di linearità del valore atteso)**.** *Siano a, b due numeri reali arbitrari e X, Y due variabili aleatorie definite su Ω. Consideriamo la variabile aleatoria Z, combinazione lineare di X, Y, con coefficienti a, b, cioè*

$$Z\left(\omega_i\right) := a \cdot X\left(\omega_i\right) + b \cdot Y\left(\omega_i\right), \qquad \forall 1 \leq i \leq N.$$

Si ha

$$\mathbb{E}(Z) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y),$$

ovvero

$$\boxed{\mathbb{E}(aX + bY) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).}$$

Dimostrazione. Si tratta in effetti di una semplice verifica. In virtù della definizione, abbiamo

$$\begin{aligned}\mathbb{E}(Z) = \mathbb{E}(aX + bY) &= \sum_{i=1}^{N} p(\omega_i)\left[a \cdot X\left(\omega_i\right) + b \cdot Y\left(\omega_i\right)\right] = \sum_{i=1}^{N}\left[a \cdot X\left(\omega_i\right) p(\omega_i) + b \cdot Y\left(\omega_i\right) p(\omega_i)\right] \\ &= a \cdot \sum_{i=1}^{N} p(\omega_i) X\left(\omega_i\right) + b \cdot \sum_{i=1}^{N} p(\omega_i) Y\left(\omega_i\right) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).\end{aligned}$$

□

Come immediati corollari di quanto sopra, otteniamo

***Proposizione* 9.5.** *Sia a un numero reale fissato e poniamo $Z = a \cdot X$. Allora risulta*

$$\mathbb{E}\left(Z\right) = a \cdot \mathbb{E}(X).$$

(Basta prendere $b = 0$ in ***Proposizione*** 9.4)

***Proposizione* 9.6.** *Sia b un numero reale fissato e poniamo $Z = X + b$. Allora risulta*

$$\mathbb{E}\left(Z\right) = \mathbb{E}(X) + b.$$

(Basta prendere, in ***Proposizione*** 9.4, $a = 1$ ed Y la variabile aleatoria degenere con $Y(\omega_i) = 1$ per ogni $\omega_i \in \Omega$.)

Infine va ricordato che la ***proprietà di linearità*** della ***Proposizione*** 9.4 si estende immediatamente al caso di ℓ variabili aleatorie: date ℓ variabili aleatorie $X_1, X_2, \cdots, X_\ell$ ed ℓ numeri reali $a_1, a_2, \cdots, a_\ell$ si ha

$$\boxed{\mathbb{E}\left(\sum_{k=1}^{\ell} a_k X_k\right) = \sum_{k=1}^{\ell} a_k \mathbb{E}(X_k)},$$

che si legge il valore atteso di una combinazione lineare di variabili aleatorie è la combinazione lineare dei rispettivi valori attesi.

Di fatto, per calcolare il valore atteso $\mathbb{E}(X)$ di una variabile aleatoria X, **occorre conoscere la distribuzione di probabilità di X, ma non è necessariamente richiesto di conoscere quale sia lo spazio di probabilità Ω su cui X è definita**, ne' come X vi possa essere definita, ne' quale sia la misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$.

Il valore atteso $\mathbb{E}(X)$ dipende infatti soltanto dalla distribuzione di probabilità di X, come mostrato nel seguente risultato.

***Proposizione* 9.7.** *Supponiamo che X sia una variabile aleatoria definita su un arbitrario spazio finito Ω e tale che $X(\Omega) := \{x_1, ..., x_n\}$. Risulta allora*

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j \cdot \mathbb{P}\left(\{X = x_j\}\right) = \sum_{j=1}^{n} x_j \cdot p_X(x_j). \tag{61}$$

Dimostrazione.
I modo Dal momento che $\{\{X = x_1\}, ..., \{X = x_n\}\}$ costituisce una partizione di Ω, potremo scrivere

$$\mathbb{E}(X) = \sum_{i=1}^{N} p(\omega_i) \cdot X(\omega_i) = \sum_{j=1}^{n} \left[\sum_{i:X(\omega_i)=x_j} p(\omega_i) \cdot X(\omega_i) \right]$$

in quanto l'ordine in cui si fanno le somme non ha importanza (proprietà commutativa della somma) possiamo effettuare la somma cambiando l'ordine: prima si sommano tutti i prodotti $p(\omega_i) \cdot X(\omega_i)$ per gli ω_i che appartengono a $\{X = x_1\}$, poi si sommano tutti i prodotti $p(\omega_i) \cdot X(\omega_i)$ per gli ω_i che appartengono a $\{X = x_2\}$, e così via. D'altra parte, per ciascun $1 \le j \le n$ fissato, si ha ovviamente:

$$\sum_{i:X(\omega_i)=x_j} p(\omega_i) \cdot X(\omega_i) = \sum_{i:X(\omega_i)=x_j} p(\omega_i) \cdot x_j = x_j \left(\sum_{i:X(\omega_i)=x_j} p(\omega_i) \right) = x_j \cdot \mathbb{P}\left(\{X = x_j\}\right).$$

La dimostrazione è quindi completata.

II modo La precedente dimostrazione è autocontenuta. Tuttavia è interessante notare che c'è una dimostrazione alternativa: posto, come nella ***Proposizione* 7.1**, $\mathcal{H}^X = \{H_j^X = \{X = x_j\}, j = 1, 2, .., n\}$ la partizione generata da X, si ha $X = \sum_{j=1}^{n} x_j \, \mathbf{1}_{H_j^X}$ e quindi per la proprietà di linearità del valore atteso e per il fatto che per ogni evento A si ha $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$, otteniamo

$$\mathbb{E}(X) = \mathbb{E}\Big(\sum_{j=1}^{n} x_j \mathbf{1}_{H_j^X} \Big) = \sum_{j=1}^{n} x_j \mathbb{E}(\mathbf{1}_{H_j^X}) = \sum_{j=1}^{n} x_j \, \mathbb{P}(H_j^X) = \sum_{j=1}^{n} x_j \, \mathbb{P}(\{X = x_j\}).$$

□

Per comodità del lettore, nella nota alla pagina seguente, riportiamo l'idea della dimostrazione della rappresentazione di X come combinazione lineare di funzioni indicatrici e una sua immediata conseguenza per il calcolo del valore atteso di $h(X)$, dove h è una funzione data.

Esempio 9.1 (valore atteso di variabili aleatorie uniformi in $\{1, 2, ..., n\}$). *Negli Esempi 7.1 e 7.2 della Lezione 7 abbiamo visto due variabili aleatorie, X_1 e T, definite su spazi diversi, che hanno la stessa distribuzione di probabilità, uniforme su $\{1, 2, ..., 6\}$. Il loro comune valore atteso è dato da*

$$\mathbb{E}(X_1) = \sum_{j=1}^{6} \frac{j}{6} = 3.5\big(= \mathbb{E}(T)\big).$$

Questo esempio si generalizza immediatamente al caso in cui X è una **variabile aleatoria uniforme** *su $\{1, 2, \ldots, n\}$, ovvero con $\mathbb{P}(X = j) = \frac{1}{n}$, per $j = 1, 2, \ldots, n$ per cui, ricordando che $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$, si ottiene*

$$\boxed{\mathbb{E}(X) = \sum_{j=1}^{n} j \frac{1}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}, \qquad X \sim Unif(\{1, 2, ..., n\})}$$

Come preannunciato, ricordiamo qui l'idea della dimostrazione della rappresentazione di X come combinazione lineare di funzioni indicatrici, ossia (posto, come al solito, $H_j^X = \{X = x_j\}$, per $j = 1, 2, ..., n$)

$$X = \sum_{j=1}^{n} x_j \, \mathbf{1}_{H_j^X},$$

da cui, per la proprietà di linearità e il fatto che $\mathbb{E}(\mathbf{1}_{H_j}) = \mathbb{P}(H_j^X) = \mathbb{P}(X = x_j)$, si ottiene che il valore atteso di X si può calcolare come

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j \, \mathbb{P}(X = x_j).$$

Chiaramente occorre e basta mostrare che la variabile aleatoria X e la variabile aleatoria $\sum_{j=1}^{n} x_j \, \mathbf{1}_{H_j^X}$ assumono lo stesso valore per ogni $\omega_i \in \Omega$.
Grazie al fatto che $\mathcal{H}^X = \{H_1^X, ..., H_n^X\}$ è una partizione, è immediato osservare che, per ogni ω_i esiste ed è unico un $k \in \{1, 2, ..., n\}$, tale che $\omega_i \in H_k^X$, e allora non solo $X(\omega_i) = x_k$, ma anche $\sum_{j=1}^{n} x_j \, \mathbf{1}_{H_j^X}(\omega_i) = x_k$: infatti $\mathbf{1}_{H_k^X}(\omega_i) = 1$, mentre, per ogni $j \neq k$, $\mathbf{1}_{H_j^X}(\omega_i) = 0$, e quindi

$$\sum_{j=1}^{n} x_j \, \mathbf{1}_{H_j^X}(\omega_i) = x_k \, \overbrace{\mathbf{1}_{H_k^X}(\omega_i)}^{=1} + \sum_{j \neq k, j=1}^{n} x_j \, \overbrace{\mathbf{1}_{H_j^X}(\omega_i)}^{=0} = x_k + 0 = x_k = X(\omega_i).$$

In modo del tutto analogo si ottiene che, per ogni funzione $h : \mathbb{R} \to \mathbb{R}$,

$$h(X) = \sum_{j=1}^{n} h(x_j) \, \mathbf{1}_{H_j^X}$$

da cui, come verifichiamo qui sotto, si ottiene anche che

$$\mathbb{E}\big(h(X)\big) = \sum_{j=1}^{n} h(x_j) p_X(x_j).$$

Infatti basta osservare che, per ogni $\omega_i \in \Omega$,

$$h(X(\omega_i)) = \sum_{j=1}^{n} h(x_j) \, \mathbf{1}_{H_j^X}(\omega_i)$$

in quanto se $\omega_i \in H_k^X = \{X = x_k\}$, allora $h(X(\omega_i)) = h(x_k)$, e

$$\sum_{j=1}^{n} h(x_j) \, \mathbf{1}_{H_j^X}(\omega_i) = h(x_k) \overbrace{\mathbf{1}_{H_k^X}(\omega_i)}^{=1} + \sum_{j \neq k, j=1}^{n} h(x_j) \, \overbrace{\mathbf{1}_{H_j^X}(\omega_i)}^{=0} = h(x_k) = h(X(\omega_i)).$$

Quindi, applicando di nuovo la proprietà di linearità del valore atteso, otteniamo

$$\mathbb{E}\big(h(X)\big) = \mathbb{E}\Big(\sum_{j=1}^{n} h(x_j) \, \mathbf{1}_{H_j^X}\Big) = \sum_{j=1}^{n} h(x_j) \; \mathbb{P}(H_j^X) = \sum_{j=1}^{n} h(x_j) \; \mathbb{P}(X = x_j).$$

In sostanza abbiamo dimostrato la successiva ***Proposizione*** 9.11

9.1 Variabili aleatorie a valori interi, non negativi

Nel caso specifico di una variabile aleatoria X a valori in $\{0,1,2,\dots,n\}$ o in $\{1,2,\dots,n\}$, il calcolo del valore atteso può anche essere convenientemente eseguito in termini della funzione

$$j \mapsto \mathbb{P}(\{X > j\});$$

vale infatti il seguente risultato.

***Proposizione* 9.8.** *Sia $X(\Omega) \subseteq \{0,1,\dots,n\}$. Allora si ha*

$$\mathbb{E}(X) = \sum_{j=0}^{n-1} \mathbb{P}(\{X > j\});$$

In particolare se $X(\Omega) = \{1,\dots,n\}$, allora si ha

$$\mathbb{E}(X) = 1 + \sum_{j=1}^{n-1} \mathbb{P}(\{X > j\})$$

Dimostrazione. Come già notato nel corso della Lezione 7, (precisamente nell'***Osservazione*** 7.6 di tale lezione) si può scrivere, per $1 \le j \le n$

$$\mathbb{P}(\{X = j\}) = \mathbb{P}(\{X > j-1\}) - \mathbb{P}(\{X > j\}),$$

per cui

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{j=0}^{n} j \cdot \mathbb{P}(\{X = j\}) = \sum_{j=1}^{n} j \cdot \mathbb{P}(\{X = j\}) = \\
&= \sum_{j=1}^{n} j \cdot \left[\mathbb{P}(\{X > j-1\}) - \mathbb{P}(\{X > j\})\right] = \sum_{j=1}^{n} \left[j \cdot \mathbb{P}(\{X > j-1\}) - j \cdot \mathbb{P}(\{X > j\})\right] \\
&= \sum_{j=1}^{n} j \cdot \mathbb{P}(\{X > j-1\}) - \sum_{j=1}^{n} j \cdot \mathbb{P}(\{X > j\})
\end{aligned}$$

ponendo $h = j-1$ e ricordando che $X(\Omega) \subseteq \{0,1,\cdots,n\}$, per cui $\mathbb{P}(\{X > n\}) = 0$,

$$\begin{aligned}
&= \sum_{h=0}^{n-1} (h+1) \cdot \mathbb{P}(\{X > h\}) - \sum_{j=1}^{n-1} j \cdot \mathbb{P}(\{X > j\} - \overbrace{\mathbb{P}(\{X > n\})}^{=0} \\
&= 1 \cdot \mathbb{P}(\{X > 0\}) + 2 \cdot \mathbb{P}(\{X > 1\}) + 3 \cdot \mathbb{P}(\{X > 2\}) + \cdots + n\,\mathbb{P}(\{X > n-1\}) \\
&\qquad - 1 \cdot \mathbb{P}(\{X > 1\}) - 2 \cdot \mathbb{P}(\{X > 2\}) - \cdots - (n-1)\,\mathbb{P}(\{X > n-1\}) \\
&= \mathbb{P}(\{X > 0\}) + (2-1)\,\mathbb{P}(\{X > 1\}) + \cdots + \big(n-(n-1)\big)\,\mathbb{P}(\{X > n-1\}) \\
&= \mathbb{P}(\{X > 0\}) + \sum_{h=1}^{n-1} \big((h+1)-h\big) \cdot \mathbb{P}(\{X > h\}) \\
&= \sum_{h=0}^{n-1} \mathbb{P}(\{X > h\}),
\end{aligned}$$

il che prova la prima affermazione.
La seconda affermazione dipende solo dal fatto che se $X(\Omega) = \{1,\dots,n\}$, allora $\mathbb{P}(\{X > 0\}) = 1$. □

Applichiamo ora questo modo di calcolare il valore atteso, nel caso del lancio del massimo tra i risultati del lancio di due dadi e poi lo generalizziamo al caso di variabili aleatorie indipendenti e con distribuzione uniforme in $\{1, 2, ..., n\}$.

Esempio* 9.2 (valore atteso del massimo di due v.a. uniformi in $\{1, 2, ...n\}$ e indipendenti)**.** *Siano X e Y i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria definita da $Z = X \vee Y$ (cioè $Z = \max(X, Y)$, in quanto per due numeri reali a e b si usa scrivere $\max(a, b) = a \vee b$). Calcolare $\mathbb{E}(Z)$.*

Soluzione: Nella precedente Lezione 7, e precisamente nell'Esempio 7.5, avevamo visto che

$$\mathbb{P}(\{Z \le x\}) = \left(\frac{x}{6}\right)^2, \quad x = 1, 2, ..., 6,$$

dunque $\mathbb{P}(\{Z > x\}) = 1 - \mathbb{P}(\{Z \le x\}) = 1 - \left(\frac{x}{6}\right)^2$. Da cui

$$\mathbb{E}(Z) = 1 + \sum_{j=1}^{6-1} \mathbb{P}(\{Z > j\}) = 1 + \sum_{j=1}^{5} \big(1 - \mathbb{P}(\{Z \le j\})\big) = 1 + 5 - \frac{1 + 4 + 9 + 16 + 25}{36} = 5 - \frac{19}{36} \cong 4.472$$

Anche questo esempio si generalizza immediatamente al caso in cui X ed Y sono ***variabili aleatorie indipendenti, ciascuna uniforme in*** $\{1, 2, \dots, n\}$:

$$\begin{aligned}\mathbb{E}(X \vee Y) &= \sum_{j=0}^{n-1} \mathbb{P}(\{X \vee Y > j\}) = \sum_{j=0}^{n-1} \big(1 - \mathbb{P}(\{X \vee Y \le j\})\big) \\ &= \sum_{j=0}^{n-1} 1 \quad - \quad \sum_{j=0}^{n-1} \mathbb{P}(\{X \le j,\, Y \le j\}) = n \; - \; \sum_{j=0}^{n-1} \mathbb{P}(\{X \le j,\, Y \le j\}).\end{aligned}$$

Si noti che finora si è usato solo il fatto che le due variabili aleatorie sono a valori in $\{1, 2, \dots, n\} \subseteq \{0, 1, 2, \dots, n\}$.

A questo punto per l'indipendenza delle due variabili aleatorie X ed Y si ha[26]

$$\mathbb{P}(\{X \le j,\, Y \le j\}) = \mathbb{P}(\{X \le j\})\, \mathbb{P}(\{Y \le j\}),$$

da cui

$$\mathbb{E}(X \vee Y) = n - \sum_{j=0}^{n-1} \mathbb{P}(\{X \le j\})\, \mathbb{P}(\{Y \le j\}). \tag{62}$$

Essendo sia X che Y uniformi in $\{1, 2, \dots, n\}$ si ha

$$\mathbb{P}(\{X \le j\})\, \mathbb{P}(\{Y \le j\}) = \left(\frac{j}{n}\right)^2, \qquad \text{per } j = 0, 1, \dots n,$$

e quindi, ricordando che $\sum_{k=1}^{n} k^2 = \frac{n(n+\frac{1}{2})(n+1)}{3}$,

$$\begin{aligned}\mathbb{E}(X \vee Y) &= n - \sum_{j=0}^{n-1} \left(\frac{j}{n}\right)^2 = n - \frac{1}{n^2} \frac{(n-1)(n-1+\frac{1}{2})(n-1+1)}{3} \\ &= n - \frac{(n-1)(n-\frac{1}{2})}{3n} = \frac{6n^2 - (n-1)(2n-1)}{6n} = \frac{4n^2 + 3n - 1}{6n}.\end{aligned}$$

[26] Si noti che la formula (62) vale in generale per variabili aleatorie indipendenti X e Y, a valori in $\{0, 1, 2, \dots, n\}$.

Esercizio proposto 9.1. *Si ripeta il procedimento del precedente Esempio 9.2 considerando il minimo al posto del massimo. Si ripeta il procedimento del precedente Esempio 9.2, sia per il minimo che per il massimo, considerando X ed Y sempre indipendenti, ma uniformi in $\{0, 1, \ldots, n\}$ invece che uniformi in $\{1, 2, \ldots, n\}$.*

Per calcolare il valore atteso di $X \wedge Y = \min(X, Y)$ ricordiamo che

$$\{\min(X, Y) > k\} = \{X > k\} \cap \{Y > k\}$$

per cui, ricordando che $\mathbb{P}(\{X > k\}) = \mathbb{P}(\{Y > k\}) = 1 - \mathbb{P}(\{X \leq k\}) = 1 - \frac{k}{n} = \frac{n-k}{n}$

$$\begin{aligned}
\mathbb{E}(X \wedge Y) &= \sum_{k=0}^{n-1} \mathbb{P}(\{X > k\} \cap \{Y > k\}) \overset{X \perp\!\!\!\perp Y}{=} \sum_{k=0}^{n-1} \mathbb{P}(\{X > k\}) \, \mathbb{P}(\{Y > k\}) \\
&= \sum_{k=0}^{n-1} \left(\frac{n-k}{n}\right)^2 = \sum_{h=1}^{n} \left(\frac{h}{n}\right)^2 = \frac{1}{n^2} \frac{n(2n+1)(n+1)}{6} = \frac{2n^2 + 3n + 1}{6n}
\end{aligned}$$

Tuttavia va notato che non è necessario fare tutti questi conti, se si pensa al fatto che

$$X + Y = X \wedge Y + X \vee Y \quad \Rightarrow \mathbb{E}(X) + \mathbb{E}(Y) = \mathbb{E}(X \wedge Y) + \mathbb{E}(X \vee Y)$$

da cui immediatamente

$$\begin{aligned}
\mathbb{E}(X \wedge Y) = \mathbb{E}(X) + \mathbb{E}(Y) - \mathbb{E}(X \vee Y) &= 2\,\frac{n+1}{2} - \frac{4n^2 + 3n - 1}{6n} \\
&= \frac{6n(n+1) - 4n^2 - 3n + 1}{6n} = \frac{2n^2 + 3n + 1}{6n}.
\end{aligned}$$

Anche il calcolo del valore atteso del massimo e del minimo nel caso di due v.a. indipendenti ed entrambe uniformi in $\{0, 1, ..., n\}$ si può ottenere senza troppi conti, osservando che $X_0,\, Y_0 \sim Unif(\{0, 1, ..., n\})$ se e solo se $X' := X_0{+}1,\ Y' := Y_0{+}1 \sim Unif(\{1, ..., n+1\})$ e che $\max(X_0, Y_0) = \max(X', Y') - 1$ e che $\min(X_0, Y_0) = \min(X', Y') - 1$ da cui imediatamente si ottiene

$$\mathbb{E}\big(\max(X_0, Y_0)\big) = \mathbb{E}\big(\max(X', Y')\big) - 1, \qquad \mathbb{E}\big(\min(X_0, Y_0)\big) = \mathbb{E}\big(\min(X', Y')\big) - 1$$

da cui

$$\begin{aligned}
\mathbb{E}\big(\max(X_0, Y_0)\big) &= \frac{4(n+1)^2 + 3(n+1) - 1}{6(n+1)} - 1, \\
\mathbb{E}\big(\min(X_0, Y_0)\big) &= \frac{2(n+1)^2 + 3(n+1) + 1}{6(n+1)} - 1
\end{aligned}$$

C'è da notare comunque che, ***in molti casi, il valore atteso di una variabile aleatoria può essere ottenuto in modo semplice, senza neanche calcolare la distribuzione di probabilità***, ovvero non è necessario calcolare la sua densità discreta, e, nel caso di variabili a valori interi positivi, non è neanche necessario calcolare $\mathbb{P}(X > j)$.

Piuttosto si tratta di sfruttare adeguatamente la proprietà di linearità. Il ruolo di tale proprietà verrà in parte illustrato nella seguente sezione, dove oltre a calcolare i valori attesi di alcune variabili aleatorie già incontrate, come le variabili aleatorie binomiali o ipergeometriche, analizzeremo il valore atteso della media aritmetica di n variabili aleatorie, delle intepretazioni del valore atteso come prezzo equo, o come baricentro.

9.2 Valori attesi notevoli

Esempio* 9.3. *Calcolare il valore atteso di una variabile aleatoria X con una distribuzione binomiale di parametri n, θ.*

Soluzione: Specializzando al nostro caso la formula (61), potremo scrivere:

$$\mathbb{E}(X) = \sum_{k=0}^{n} k\, p_X(k) = \sum_{k=0}^{n} k \binom{n}{k} \theta^k \,(1-\theta)^{n-k},$$

e fare i conti relativi (si veda la nota a pagine 126). Ma possiamo anche ottenere il valore di $\mathbb{E}(X)$ senza calcoli, ricordando l'Esempio 7.3 della precedente Lezione 7 e ragionando come segue.

Consideriamo n variabili aleatorie binarie $X_1, ..., X_n$, con $\mathbb{P}(\{X_j = 1\}) = \theta$, per ogni $j = 1, 2, ..., n$; è immediato calcolare il valore atteso di ognuna delle X_j e ottenere che

$$\mathbb{E}(X_j) = 0\,\mathbb{P}(X_j = 0) + 1\,\mathbb{P}(X_j = 1) = \theta.$$

Per la proprietà di linearità del valore atteso si ha dunque

$$\mathbb{E}(X_j) = \theta, \qquad \mathbb{E}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathbb{E}(X_j) = n\theta.$$

Sappiamo d'altra parte che, nel caso particolare in cui $X_1, ..., X_n$ sono indicatori di eventi completamente indipendenti, la variabile aleatorie $S := \sum_{j=1}^{n} X_j$ segue appunto una distribuzione binomiale di parametri n, θ, la stessa di X. E dunque, visto che il valore atteso di una variabile aleatoria dipende soltanto dalla sua distribuzione di probabilità, abbiamo che, anche per la nostra variabile aleatoria X, risulta $\mathbb{E}(X) = \mathbb{E}(S) = n\theta$.

Esercizio proposto 9.2. *Una prova di esonero è sostenuta da 25 studenti. Supponendo che, per ognuno di loro, la probabilità di successo sia uguale a 0.80, qual è il valore atteso del numero di studenti che passano la prova?*

Naturalmente possiamo avere distribuzioni di probabilità diverse che danno luogo allo stesso valore atteso, e ne vedremo ora diversi esempi. In particolare due distribuzioni binomiali $Bin(n', \theta')$ e $Bin(n'', \theta'')$ danno luogo allo stesso valore atteso qualora risulti $n' \cdot \theta' = n'' \cdot \theta''$. A questo proposito riprendiamo ora il caso visto nell'Esempio 3.4 della Lezione 3.

Esempio 3.4 rivisitato ("Paradosso del Cavalier De Méré"). *Il numero di volte in cui si ottiene il risultato "asso" in quattro lanci di un dado è una variabile aleatoria con distribuzione $Bin(4, \frac{1}{6})$; dunque il suo valore atteso è dato da $4 \cdot \frac{1}{6} = \frac{2}{3}$. Tale valore coincide anche con il valore atteso del numero di volte in cui si presenta il doppio asso in ventiquattro lanci di una coppia di dadi, che è dato da $24 \cdot \frac{1}{36}$.*

Possiamo concludere quindi che in entrambi i tipi di gioco d'azzardo si ha uguale valore atteso del numero dei successi. Tuttavia, come avevamo visto nell'Esempio 3.4, sono diverse le probabilità di ottenere almeno un successo nelle due diverse scommesse, e questo spiega il motivo per cui questo esempio viene ricordato come "paradosso".

***Osservazione* 9.1.** Consideriamo ora, in generale, n eventi, A_1, A_2, ...,A_n, e la variabile S_n che conta il numero dei successi su tali eventi, ovvero

$$S_n = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n} = X_1 + X_2 + \cdots + X_n,$$

con $X_i = \mathbf{1}_{A_i}$, $i = 1, 2, \ldots, n$. Poiché $\mathbb{E}(X_i) = \mathbb{E}(\mathbf{1}_{A_i}) = \mathbb{P}(A_i)$, per la proprietà di linearità del valore atteso, risulta

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n).$$

Ciò generalizza quanto visto nell'Esempio 9.3, in cui ciascuno di tali eventi abbia probabilità θ, e il valore atteso $\mathbb{E}(S_n)$ è uguale al prodotto $n \cdot \theta$.

È opportuno sottolineare, che, per giungere a tale conclusione ***non abbiamo fatto alcuna ipotesi circa l'indipendenza stocastica, o meno, fra tali eventi***. Naturalmente il tipo di correlazione fra tali eventi ha influenza sulla distribuzione di probabilità di S_n, ma non sul suo valore atteso che, con la sola ipotesi che gli eventi abbiano tutti probabilità θ, resta in ogni caso uguale a $n \cdot \theta$:

$$\mathbb{E}(S_n) = n\theta.$$

La variabile aleatoria

$$Y_n := \frac{S_n}{n}$$

può essere interpretata come la ***frequenza relativa* dei successi** sugli n eventi. Ancora per la proprietà di linearità, avremo, in ogni caso,

$$\mathbb{E}\left(Y_n\right) = \frac{1}{n}\,\mathbb{E}\left(S_n\right) = \theta.$$

Esempio 9.4 (Valore atteso di una v.a. ipergeometrica). *Consideriamo in particolare* ***una distribuzione ipergeometrica di parametri*** M, m_1, n. *Come sappiamo questa è la distribuzione di probabilità di una variabile aleatoria S che conta il numero di successi su n eventi (non indipendenti): ciascuno di probabilità $\frac{m_1}{M}$; e quindi si ha che* ***il suo valore atteso è dato da*** $n \cdot \frac{m_1}{M}$. *Si veda la nota seguente e la nota a pagina 126, per il calcolo del valore atteso usando la densità discreta.*

Possiamo scrivere

$$S = \sum_{i=1}^{n} \mathbf{1}_{A_i},$$

dove A_i={all'i-sima estrazione esce un elemento di tipo A}, per $i = 1, 2, ..., n$, le n estrazioni avvengono senza reinserimento da una popolazione/urna con M individui/palline di cui m_1 di tipo A e i rimanenti $m_2 = M - m_1$ di tipo B. Sappiamo che $\mathbb{P}(A_i) = \mathbb{P}(A_1) = \frac{m_1}{M}$, anche se si tratta di estrazioni con reinserimento, e quindi

$$\boxed{\mathbb{E}(S) = \mathbb{E}\Big(\sum_{i=1}^{n} \mathbf{1}_{A_i}\Big) = \sum_{i=1}^{n} \mathbb{E}\big(\mathbf{1}_{A_i}\big) = n\,\mathbb{P}(A_1) = n\,\frac{m_1}{M}.}$$

Estendiamo ora la discussione su un punto, già introdotto nella precedente ***Osservazione* 9.1**, che risulta di notevole importanza nella teoria della probabilità.

***Proposizione* 9.9 (Valore atteso di una media aritmetica).** *Consideriamo n variabili aleatorie $X_1, ..., X_n$, definite su uno stesso spazio $(\Omega, \mathcal{P}(\Omega))$ e che abbiano tutte lo stesso valore atteso; cioè, usando il simbolo μ per indicare brevemente $\mathbb{E}\left(X_j\right)$,*

$$\mathbb{E}\left(X_1\right) = ... = \mathbb{E}\left(X_n\right) = \mu.$$

Consideriamo ora, sullo stesso spazio $(\Omega, \mathcal{P}(\Omega))$, la variabile aleatoria Y_n definita come **media aritmetica** *delle $X_1, ..., X_n$, ossia*

$$Y_n\left(\omega_i\right) = \frac{\sum_{j=1}^{n} X_j\left(\omega_i\right)}{n}, \qquad i = 1, ..., N;$$

si ha

$$\mathbb{E}\left(Y_n\right) = \mu,$$

Dimostrazione. Basta osservare che, per la linearità del valore atteso

$$\mathbb{E}\left(Y_n\right) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_j) = \frac{n\mu}{n} = \mu.$$

□

Osservazione* 9.2 (Sulla distribuzione di una media aritmetica).** La ***Proposizione* 9.9** è quindi una generalizzazione del caso considerato nella precedente ***Osservazione 9.1. A parità di valore atteso possono però sussistere situazioni assai diverse per quanto riguarda la distribuzione di probabilità di Y_n, a seconda delle proprietà di ***correlazione*** fra le variabili $X_1, ..., X_n$. Questo è quanto vedremo meglio nella prossima lezione, introducendo i concetti di ***varianza*** di una variabile aleatoria e di ***covarianza*** fra due variabili aleatorie.

Un esempio, in un certo senso *estremo*, di comportamento di una media aritmetica è il seguente; questo mette anche in luce alcuni aspetti basilari del concetto di valore atteso.

Esempio 9.5. *Consideriamo una lotteria in cui vengono venduti n biglietti, numerati progressivamente.*

Supponiamo che tale lotteria distribuisca un totale di r ($r < n$) premi, di cui, ad esempio,

r_1 *"primi premi" di entità c_1,*

r_2 *"secondi premi" di entità c_2 e*

r_3 *"terzi premi" di entità c_3*

(dunque $r = r_1 + r_2 + r_3$ e $c_1 > c_2 > c_3 > 0$).

Indichiamo con $X_1, ..., X_n$ le vincite rispettivamente associate al biglietto 1, al biglietto 2, ..., al biglietto n.

Ovviamente $X_1, ..., X_n$ sono delle variabili aleatorie i cui valori possibili costituiscono l'insieme $\{0, c_1, c_2, c_3\}$. $X_1, ..., X_n$ hanno tutte, la stessa distribuzione di probabilità, data da[27]

$$\mathbb{P}(\{X_j = c_1\}) = \frac{r_1}{n}, \quad \mathbb{P}(\{X_j = c_2\}) = \frac{r_2}{n}, \quad \mathbb{P}(\{X_j = c_3\}) = \frac{r_3}{n},$$

mentre

$$\mathbb{P}(\{X_j = 0\}) = \frac{n-r}{n}\big(= 1 - \mathbb{P}(\{X_j = c_1\}) - \mathbb{P}(\{X_j = c_2\}) - \mathbb{P}(\{X_j = c_3\})\big);$$

e dunque risulta

$$\mathbb{E}\left(X_j\right) = \frac{r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3}{n}.$$

Si può pervenire anche più semplicemente a quest'ultima conclusione, sfruttando di nuovo la proprietà di linearità del valore atteso e ragionando come segue: per motivi di simmetria, è ovvio che $X_1, ..., X_n$ abbiano tutte uno stesso valore atteso $\mathbb{E}\left(X_j\right) = \mu$. Andiamo a considerare la variabile aleatoria

$$S_n = \sum_{j=1}^{n} X_j,$$

---

[27] Infatti si può pensare che vengano effettivamente messe in un'urna n palline con i numeri di tutti i possibili biglietti $1,\ 2\ \ldots, n$ e che vengano effettuate $r_1 + r_2 + r_3$ estrazioni senza reinserimento. L'evento *"vincita della cifra c_1"* viene allora espresso come l'evento *"viene estratto il numero j, nelle prime r_1 estrazioni"*. Analogamente l'evento *"vincita della cifra c_2"* viene espresso come l'evento *"viene estratto il numero j, in una delle estrazioni tra la $(r_1+1)-sima$ e la $(r_1+r_2)-sima$ estrazione"*. Infine l'evento *"vincita della cifra c_3"* viene espresso come l'evento *"viene estratto il numero j, in una delle estrazioni tra la $(r_1 + r_2 + 1) - sima$ e l'ultima estrazione, ovvero la $(r_1 + r_2 + r_3) - sima$"*. Basta poi notare che la probabilità che il numero j sia estratto alla $k - sima$ estrazione vale $\frac{1}{n}$, qualunque sia k per ottenere la distribuzione di X_j.

È interessante anche notare che si ottiene la stessa distribuzione anche nel caso in cui invece la lotteria sia del tipo ***gratta e vinci***. Ovvero ci sono n biglietti e si prende un biglietto a caso tra n in cui r_i danno diritto al premio c_i, per $i = 1, 2, 3$.

osserviamo che S_n ha il significato di ammontare complessivo dei premi distribuiti dalla lotteria, e quindi risulta essere una variabile aleatoria degenere, di valore $r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3$, ovvero

$$S_n(\omega) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3, \qquad \textit{qualunque sia } \omega \in \Omega$$

quindi

$$\mathbb{E}\left(S_n\right) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3.$$

D'altra parte, per la proprietà di linearità si deve avere

$$\mathbb{E}\left(S_n\right) = \mathbb{E}\left(\sum_{j=1}^{n} X_j\right) = n \cdot \mu$$

e dunque deve essere

$$\mu = \frac{\mathbb{E}\left(S_n\right)}{n} = \frac{r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3}{n}. \tag{63}$$

Ovviamente anche il valore atteso della variabile aleatoria $Y_n := \frac{S_n}{n}$ (media aritmetica di $X_1, ..., X_n$) è uguale alla quantità μ, come sappiamo deve essere; ma in questo speciale caso Y_n è una variabile aleatoria con distribuzione degenere, tutta concentrata sul valore μ. Cioè, dal confronto fra la distribuzione di X_j e quella di Y_n, vediamo che entrambe hanno valore atteso μ, ma la X_j non ha una distribuzione degenere, come invece accade per Y_n.

Esempio* 9.6. *La rilevanza della proprietà di linearità del valore atteso può essere illustrata dal seguente esempio. Supponiamo di possedere un biglietto di una lotteria A ed un biglietto della lotteria B:*

A distribuisce r'_1 premi di entità c'_1 e r'_2 premi di entità c'_2, su un totale di n' biglietti;

e

B distribuisce r''_1 premi di entità c''_1 e r''_2 premi di entità c''_2, su un totale di n'' biglietti.

Indicando con X la variabile aleatoria che indica la vincita complessivamente derivante dai due biglietti, calcolare $\mathbb{E}(X)$.

Soluzione: Indichiamo rispettivamente con X' e X'' le vincite relative ai due singoli biglietti, cosicché $X = X' + X''$. Con questa posizione X' è una variabile aleatoria che può assumere i valori $\{0, c'_1, c'_2\}$ e X'' è una variabile aleatoria che può assumere i valori $\{0, c''_1, c''_2\}$ e risulta, procedendo come nel precedente Esempio 9.5,

$$\mathbb{E}(X') = \frac{c'_1 r'_1 + c'_2 r'_2}{n'}, \qquad \mathbb{E}(X'') = \frac{c''_1 r''_1 + c''_2 r''_2}{n''}.$$

È importante sottolineare che, sempre grazie alla proprietà di linearità del valore atteso, ***per calcolare il valore atteso*** $\mathbb{E}(X)$ ***non c'è bisogno di calcolare interamente la distribuzione di probabilità di*** X; possiamo infatti ottenere immediatamente, in virtù della proprietà di linearità,

$$\mathbb{E}(X) = \mathbb{E}(X') + \mathbb{E}(X'') = \frac{c'_1 r'_1 + c'_2 r'_2}{n'} + \frac{c''_1 r''_1 + c''_2 r''_2}{n''}.$$

***Osservazione* 9.3.** Riprendiamo il caso di una singola lotteria, considerata nel precedente Esempio 9.5, ed indichiamo con c il costo di un singolo biglietto.

L'acquirente del biglietto j paga dunque il prezzo certo c ed ottiene in cambio il guadagno aleatorio X_j, il cui valore atteso è $\mathbb{E}\left(X_j\right) = \mu$, dove μ è ottenuto come in (63) dell'Esempio 9.5. Si ha invece che (nell'ipotesi che venda tutti gli n biglietti) l'organizzatore della lotteria ottiene un ricavo R pari a

$$R = n \cdot (c - \mu);$$

tale ricavo (positivo, negativo o nullo a seconda che sia μ minore, maggiore o uguale a c) è certo, nel senso che non dipende dal risultato aleatorio della lotteria (cioè da quali saranno i biglietti estratti). Su tale base, si può interpretare il valore atteso $\mu = \mathbb{E}(X_j)$ come il prezzo "equo" per l'acquisto di un singolo biglietto.

Tale interpretazione della nozione di valore atteso è fondamentale nel contesto della finanza matematica, in relazione al concetto di "non arbitraggio"[28]

Il ruolo della nozione di valore atteso nella comprensione del concetto di gioco equo è anche illustrata nel seguente esempio:

Esempio* 9.7. *Un giocatore, in possesso di un capitale iniziale di 31 Euro, gioca* **al raddoppio** *in una serie di puntate in ciascuna delle quali può vincere o perdere la cifra puntata: inizialmente punta 1 Euro, se vince ottiene 1 Euro e si ferma; se perde, raddoppia la puntata e continua così di seguito finché non vince la prima volta o finché non esaurisce il suo capitale iniziale.*

Supponiamo che, in ciascuna puntata, il giocatore vince o perde con probabilità rispettivamente date da θ o $1-\theta$.

Indicando con X la variabile aleatoria che rappresenta il capitale del giocatore al termine del gioco, vogliamo calcolare $\mathbb{E}(X)$.

Soluzione: Osserviamo innanzitutto che X è una variabile aleatoria che prende solo i due valori dell'insieme $\{0, 32\}$. Infatti il giocatore continua a giocare al raddoppio fino a che non raggiunge il capitale di 32 Euro, a meno che non perda per 5 volte consecutive, nel qual caso esaurisce appunto tutto il suo capitale iniziale di 31 = 1 + 2 + 4 + 8 + 16 Euro. Ne segue quindi

$$\mathbb{P}(\{X=0\}) = (1-\theta)^5\,, \qquad \mathbb{P}(\{X=32\}) = 1-(1-\theta)^5\,,$$

da cui

$$\mathbb{E}(X) = 0\cdot(1-\theta)^5 + 32\left(1-(1-\theta)^5\right) = 32 - 32\cdot(1-\theta)^5\,.$$

Vediamo quindi che, così facendo, il giocatore decide di scambiare il suo capitale certo di 31 Euro con un capitale aleatorio X di valore atteso $\mathbb{E}(X)$.

Osserviamo a tale proposito che risulta $\mathbb{E}(X) < 31, \mathbb{E}(X) = 31, \mathbb{E}(X) > 31$, a seconda che sia $\theta < \frac{1}{2}$, $\theta = \frac{1}{2}$, oppure $\theta > \frac{1}{2}$, cioè a seconda che il gioco sia sfavorevole, equo, oppure favorevole per il giocatore stesso.

***Osservazione* 9.4.** Supponiamo di distribuire su una retta delle masse m_1, m_2, ..., m_n, ponendole rispettivamente sui punti di ascissa $x_1, x_2, ..., x_n$. Osserviamo allora che, posto $p_i = \frac{m_i}{\sum_{k=1}^n m_k}$ la quantità $\sum_{j=1}^n p_j \cdot x_j$ equivale al baricentro di tale distribuzione di masse.

Così come il baricentro costituisce un valore che, a certi fini, risulta "riassuntivo" di tutta la distribuzione di masse, così il valore atteso di una variabile aleatoria costituisce un valore che, entro certi fini, riassume la conoscenza completa della distribuzione di probabilità.

In generale se le masse sono m_1, m_2, ..., m_n e se sono poste rispettivamente sui punti di ascissa $x_1, x_2, ..., x_n$, allora il baricentro è

$$\hat{x} = \frac{x_1\cdot m_1 + x_2\cdot m_2 + \cdots + x_n\cdot m_n}{m_1+m_2+\cdots+m_n}.$$

Posto $p_i = \frac{m_i}{m_1+m_2+\cdots+m_n}$, per $i = 1, \cdots, n$, possiamo intrepretare p_i come la densità discreta di una variabile aleatoria X a valori in $\{x_1, x_2, ..., x_n\}$: $p_i = p_X(x_i)$, di modo che la precedente espressione diviene

$$\hat{x} = x_1\cdot p_1 + x_2\cdot p_2 + \cdots + x_n\cdot p_n = \mathbb{E}(X).$$

[28]Si dice che in una operazione finanziaria c'è opportunità di arbitraggio se l'operazione finanziaria di sicuro non comporta perdite, e comporta un guadagno strettamente positivo con probabilità strettamente positiva.

Valore atteso di $X \sim Bin(n, \theta)$, con la densità discreta

Come ricordato, per ottenere che $\mathbb{E}(X) = n\theta$, si può procedere anche usando la densità discreta, ma a costo di un po' di calcoli, che invece la proprietà di linearità del valore atteso ci permette di risparmiare:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k p_X(k) = \sum_{k=0}^{n} k\, \mathbb{P}(X = k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \theta^k (1-\theta)^{n-k}, \quad \text{e quindi}
\end{aligned}$$

ponendo $h = k - 1$, e tenendo conto che $1 \le k \le n \Leftrightarrow 0 \le h \le n-1$, che $n - k = n - 1 - (k-1) = n - 1 - h$, $n! = n \cdot (n-1)!$, $k = h + 1$, si ha

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{h=0}^{n-1} n \frac{(n-1)!}{h!(n-1-h)!} \theta^{h+1} (1-\theta)^{n-1-h} \\
&= n\theta \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} \theta^{h} (1-\theta)^{n-1-h} = n\theta \sum_{h=0}^{n-1} \binom{n-1}{h} \theta^{h} (1-\theta)^{n-1-h},
\end{aligned}$$

infine, tenendo conto dello sviluppo della potenza del binomio, otteniamo

$$\mathbb{E}(X) = n\theta \left(\theta + (1-\theta)\right)^{n-1} = n\theta \cdot 1.$$

Valore atteso di $X \sim Hyp(M, m_1; n)$, con la densità discreta

Poniamo, come al solito $m_2 = M - m_1$,

$$\mathbb{E}(X) = \sum_{k=0}^{n} k p_X(k) = \sum_{k=1}^{n} k\, \mathbb{P}(X = k) = \sum_{k \ge 1}^{*} k \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{m_1+m_2}{n}},$$

dove nella somma si richiede che $k \le m_1$, $n - k \le m_2$ ed $1 < n \le M = m_1 + m_2$. Ricordando che per i coefficienti binomiali vale la seguente indentità

$$h\binom{r}{h} = r\binom{r-1}{h-1}, \quad \text{e quindi} \quad \binom{r}{h} = \frac{r}{h}\binom{r-1}{h-1}$$

si ottiene allora

$$\mathbb{E}(X) = \sum_{k \ge 1}^{*} \frac{m_1 \binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\frac{m_1+m_2}{n}\binom{m_1+m_2-1}{n-1}} = n \frac{m_1}{m_1+m_2} \sum_{k \ge 1}^{*} \overbrace{\frac{\binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\binom{m_1+m_2-1}{n-1}}}^{=1} = n \frac{m_1}{M},$$

in quanto, ponendo $h = k - 1$, da cui $0 \le h \le m_1 - 1$ e $0 \le n - 1 - h = n - k \le m_2$,

$$\sum_{k \ge 1}^{*} \frac{\binom{m_1-1}{k-1}\binom{m_2}{n-k}}{\binom{m_1+m_2-1}{n-1}} = \sum_{h \ge 0}^{**} \frac{\binom{m_1-1}{h}\binom{m_2}{n-1-h}}{\binom{m_1+m_2-1}{n-1}} = 1$$

9.3 Valori attesi di trasformazioni di variabili aleatorie

Una proprietà di fondamentale importanza è data nella seguente proposizione:

***Proposizione* 9.10.** *Siano X, Y due variabili aleatorie indipendenti definite su uno stesso spazio di probabilità. Allora*

$$\mathbb{E}\,(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Dimostrazione. Siano rispettivamente

$$X(\Omega) = \{x_1, ..., x_n\} \quad \text{e} \quad Y(\Omega) = \{y_1, ..., y_m\}.$$

Possiamo allora considerare la partizione di Ω costituita dagli eventi del tipo

$$H^{X,Y}_{h,k} = \{X = x_h\} \cap \{Y = y_k\} := \{\omega_i \in \Omega : \; X(\omega_i) = x_h, Y(\omega_i) = y_k\},$$

per $h = 1, 2, \ldots, n$, e $k = 1, 2, \ldots, m$.
I modo Possiamo scrivere

$$\begin{aligned}
\mathbb{E}\,(X \cdot Y) &= \sum_{i=1}^{N} p(\omega_i)\,[X(\omega_i) \cdot Y(\omega_i)] = \sum_{h=1}^{n}\sum_{k=1}^{m} \left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i)\,[X(\omega_i) \cdot Y(\omega_i)] \right] \\
&= \sum_{h=1}^{n}\sum_{k=1}^{m} \left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i) x_h \cdot y_k \right] = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i) \right] \\
&= \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k\, \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right).
\end{aligned}$$

II modo Si noti che a questo risultato si poteva arrivare anche considerando che

$$X \cdot Y = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \mathbf{1}_{H^{X,Y}_{h,k}}$$

da cui immediatamente, per la proprietà di linearità del valore atteso, si ha

$$\mathbb{E}(X \cdot Y) = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \mathbb{E}\left(\mathbf{1}_{H^{X,Y}_{h,k}}\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k\, \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right). \tag{64}$$

Si noti inoltre che fino a questo punto non si è usata l'ipotesi di indipendenza, mentre ora, tenendo conto del fatto che, per l'indipendenza $\mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right) = \mathbb{P}\left(\{X = x_h\}\right) \cdot \mathbb{P}\left(\{Y = y_k\}\right)$, si ottiene[29]

$$\begin{aligned}
\mathbb{E}\,(X \cdot Y) &= \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \cdot \mathbb{P}\,(X = x_h) \cdot \mathbb{P}(Y = y_k) \\
&= \sum_{h=1}^{n} x_h \cdot \mathbb{P}\,(X = x_h) \cdot \sum_{k=1}^{m} y_k \cdot \mathbb{P}(Y = y_k) = \mathbb{E}(X) \cdot \mathbb{E}(Y),
\end{aligned}$$

e la dimostrazione è terminata. □

[29] Se i passaggi che seguono risultassero difficili, si noti che

$$\begin{aligned}
&\sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \cdot \mathbb{P}\,(X = x_h) \cdot \mathbb{P}(Y = y_k) = \sum_{h=1}^{n} \left(\sum_{k=1}^{m} x_h \cdot y_k \cdot \mathbb{P}\,(X = x_h) \cdot \mathbb{P}(Y = y_k) \right) \\
&= \sum_{h=1}^{n} x_h \cdot \mathbb{P}\,(X = x_h) \left(\sum_{k=1}^{m} y_k \cdot \mathbb{P}(Y = y_k) \right) = \sum_{h=1}^{n} x_h \cdot \mathbb{P}\,(X = x_h)\, \mathbb{E}(Y) = \mathbb{E}(Y) \sum_{h=1}^{n} x_h \cdot \mathbb{P}\,(X = x_h) = \mathbb{E}(Y)\mathbb{E}(X)
\end{aligned}$$

In questa ultima parte estendiamo i risultati ottenuti in ***Proposizione*** **9.7** e ***Proposizione*** **9.10**, al caso di trasformazioni di coppie di variabili aleatorie.

***Proposizione* 9.11.** *Sia X una variabile aleatoria definita su uno spazio di probabilità finito, e sia data la funzione $h: \mathbb{R} \to \mathbb{R};\ x \mapsto h(x)$. Allora*

$$\mathbb{E}(h(X)) = \sum_{j=1}^{n} h(x_j) \cdot \mathbb{P}\left(\{X = x_j\}\right) = \sum_{j=1}^{n} h(x_j) \cdot p_X(x_j). \tag{65}$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilità finito, e sia data la funzione $g: \mathbb{R}^2 \to \mathbb{R}$; $(x, y) \mapsto g(x, y)$. Allora

$$\mathbb{E}\left(g(X, Y)\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k)\, \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k) \cdot p_{X,Y}(x_h, y_k). \tag{66}$$

Osservazione* 9.5.** Prima di dare la dimostrazione di questo risultato osserviamo che la precedente ***Proposizione **9.11** assicura che **per calcolare il valore atteso della variabile aleatoria di una trasformazione di variabili aleatorie (ovvero di $Z = h(X)$ o $W = g(X, Y)$) non è necessario calcolare la sua distribuzione di probabilità**, ma basta conoscere la densità discreta di X o la densità discreta congiunta di X, Y e applicare la (65) o la (66) rispettivamente.

Dimostrazione della **Proposizione** ***9.11***. La dimostrazione di (65) non è altro che l'immediata estensione della dimostrazione della ***Proposizione*** **9.7**: infatti, con le stesse notazioni usate nel (*II modo*) di tale dimostrazione, basta notare[30] che, posto $X_j = \mathbf{1}_{H_j^X}$, con $H_j^X = \{X = x_j\}$,

$$h(X) = \sum_{j=1}^{n} h(x_j) X_j = \sum_{j=1}^{n} h(x_j) \mathbf{1}_{H_j^X}$$

e quindi per linearità

$$\mathbb{E}(h(X)) = \sum_{j=1}^{n} h(x_j)\mathbb{E}(X_j) = \sum_{j=1}^{n} h(x_j)\, \mathbb{P}(H_j^X) = \sum_{j=1}^{n} h(x_j)\, \mathbb{P}(\{X = x_j\}).$$

La dimostrazione di (66) è simile:
con le stesse notazioni usate nella dimostrazione della ***Proposizione*** **9.10**, ed in particolare ponendo $H_{h,k}^{X,Y} = \{X = x_h\} \cap \{Y = y_k\}$,

$$g(X, Y) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k) \mathbf{1}_{H_{h,k}^{X,Y}}$$

da cui immediatamente, per la proprietà di linearità del valore atteso si ha

$$\mathbb{E}\big(g(X, Y)\big) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k)\mathbb{E}\left(\mathbf{1}_{H_{h,k}^{X,Y}}\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k)\, \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right). \tag{67}$$

□

Infine va notato che nel caso di indipendenza tra X e Y si ha anche il seguente risultato:

[30] Ovviamente se $\omega \in H_j^X = \{X = x_j\}$, ovvero è tale che $X(\omega) = x_j$, allora $\boldsymbol{Z(\omega) = h(X(\omega)) = h(x_j)}$.

***Proposizione* 9.12.** *Siano X, Y due variabili aleatorie **indipendenti** definite su uno stesso spazio di probabilità, e siano $h_i: \mathbb{R} \to \mathbb{R};\ x \mapsto h_i(x)$, per $i = 1, 2$ due funzioni reali, allora*

$$\mathbb{E}\left(h_1(X) \cdot h_2(Y)\right) = \mathbb{E}\left(h_1(X)\right) \cdot \mathbb{E}\left(h_2(Y)\right).$$

Dimostrazione. Basta prendere $g(x, y) = h_1(x) \cdot h_2(y)$ in (67) ottenendo così

$$\mathbb{E}\big(h_1(X) \cdot h_2(Y)\big) = \sum_{h=1}^{n} \sum_{k=1}^{m} h_1(x_h) \cdot h_2(y_k) \cdot \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right)$$

e sfruttando l'indipendenza tra X ed Y

$$= \sum_{h=1}^{n} \sum_{k=1}^{m} h_1(x_h) \cdot h_2(y_k) \cdot \mathbb{P}\left(\{X = x_h\}\right) \cdot \mathbb{P}\left(\{Y = y_k\}\right).$$

A questo punto la dimostrazione è identica alla dimostrazione della ***Proposizione* 9.10**, pur di sostituire $x_h \cdot y_k$ con $h_1(x_h) \cdot h_2(y_k)$. $\square$

Come già osservato, per calcolare il valore atteso della variabile aleatoria di una trasformazione di variabili aleatorie (ovvero di $W = h(X)$ o $Z = g(X, Y)$) non è necessario calcolare la sua distribuzione di probabilità. Tuttavia questo problema è un problema interessante, e abbiamo già calcolato alcune distribuzioni di trasformazioni di variabili aleatorie: si vedano i precedenti esempi ed esercizi che riguardano la somma di variabili aleatorie, il massimo o il minimo, etc., in particolare si ricordino Esempio 7.5, Esempio 7.6, Esercizio proposto 7.1, Esercizio proposto 8.1, Esercizio 7.1 (punti (b) e (c)), Esercizio 8.4 Esercizio 8.5

La prossima proposizione riassume il metodo generale per ottenere la distribuzione di trasformazioni di variabili aleatorie. La prima parte riguarda la trasformazione di una sola variabile aleatoria e riassume quanto già dimostrato nella Sezione 7.4; la seconda parte riguarda invece le trasformazioni di una coppia di variabili aleatorie e si dimostra in modo del tutto simile alla prima parte.

***Proposizione* 9.13.** *Sia X una variabile aleatoria definita su uno spazio di probabilità finito, e sia data la funzione $h: \mathbb{R} \to \mathbb{R};\ x \mapsto h(x)$. Allora, posto $Z = h(X)$, e $Z(\Omega) = \{z_1, z_2, \ldots, z_\ell\}$ si ha*

$$\mathbb{P}(Z = z_k) = \mathbb{P}(h(X) = z_k) = \sum_{j:\ h(x_j) = z_k} \mathbb{P}\big(\{X = x_j\}\big), \qquad k = 1, 2, \ldots, \ell \tag{68}$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilità finito, e sia data la funzione $g: \mathbb{R}^2 \to \mathbb{R};\ (x, y) \mapsto g(x, y)$. Allora, posto $W = g(X, Y)$ e $W(\Omega) = \{w_1, w_2, \ldots, w_r\}$ si ha

$$\mathbb{P}\left(g(X, Y) = w_j\right) = \sum_{h,k:\ g(x_h, y_k) = w_j}^{1 \le h \le n,} \ \sum^{1 \le k \le m} \mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right), \qquad j = 1, 2, \ldots, r. \tag{69}$$

Esempio 9.8. *Siano X ed Y le variabili aleatorie definite nell'Esempio 8.1 e sia $g(x, y) = 4\,x^2\,y$*

$Y \diagdown X$	-1		0		1	
1/4	0.10	$g(-1, 1/4) = 1$	0.20	$g(0, 1/4) = 0$	0	$g(1, 1/4) = 1$
1/2	0	$g(-1, 1/2) = 2$	0.12	$g(0, 1/2) = 0$	0.05	$g(1, 1/2) = 2$
3/4	0.05	$g(-1, 3/4) = 3$	0.10	$g(0, 3/4) = 0$	0.04	$g(1, 3/4) = 3$
1	0.10	$g(-1, 1) = 4$	0.04	$g(0, 1) = 0$	0.20	$g(1, 1) = 4$

dove, nei riquadri a fianco ai valori della densità discreta congiunta di X, Y, ci sono i valori che può assumere la variabili aleatoria $Z = g(X,Y) = 4\,X^2\,Y$. Per calcolare il valore atteso di Z, non è necessario calcolare la densità discreta di Z infatti

$$\begin{aligned}
\mathbb{E}(Z) &= \mathbb{E}\big(g(X.Y)\big) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h, y_k)\,\mathbb{P}\left(\{X = x_h\} \cap \{Y = y_k\}\right) \\
&= g(-1, 1/4) \cdot \mathbb{P}(X = -1, Y = 1/4) + g(0, 1/4) \cdot \mathbb{P}(X = 0, Y = 1/4) + g(1, 1/4) \cdot \mathbb{P}(X = 1, Y = 1/4) \\
&+ g(-1, 1/2) \cdot \mathbb{P}(X = -1, Y = 1/2) + g(0, 1/2) \cdot \mathbb{P}(X = 0, Y = 1/2) + g(1, 1/2) \cdot \mathbb{P}(X = 1, Y = 1/2) \\
&+ g(-1, 3/4) \cdot \mathbb{P}(X = -1, Y = 3/4) + g(0, 3/4) \cdot \mathbb{P}(X = 0, Y = 3/4) + g(1, 3/4) \cdot \mathbb{P}(X = 1, Y = 3/4) \\
&+ g(-1, 1) \cdot \mathbb{P}(X = -1, Y = 1) + g(0, 1) \cdot \mathbb{P}(X = 0, Y = 1) + g(1, 1) \cdot \mathbb{P}(X = 1, Y = 1) \\
&= 1 \cdot 0.10 + 0 \cdot 0.20 + 1 \cdot 0 \quad + 2 \cdot 0 \quad + 0 \cdot 0.12 + 2 \cdot 0.05 \\
&+ 3 \cdot 0.05 + 0 \cdot 0.10 + 3 \cdot 0.04 + 4 \cdot 0.10 + 0 \cdot 0.04 + 4 \cdot 0.20 \\
&= 1 \cdot 0.10 + 2 \cdot 0.05 + 3 \cdot 0.05 + 3 \cdot 0.04 + 4 \cdot 0.10 + 4 \cdot 0.20 = 1.67
\end{aligned}$$

Se invece è necessario trovare la distribuzione di Z, prima di tutto è immediato capire che $Z(\Omega) = \{0, 1, 2, 3, 4\}$. Inoltre, sommando sui valori di $p_{X,Y}(x_h, y_k)$ per i quali $g(x_h, y_k) = i$, per $i = 0, 1, 2, 3, 4$. Si vede immediatamente che

$$\begin{aligned}
p_Z(1) &= \mathbb{P}(X = -1, Y = 1/4) + \mathbb{P}(X = 1, Y = 1/4) = 0.10, \\
p_Z(2) &= \mathbb{P}(X = -1, Y = 1/2) + \mathbb{P}(X = 1, Y = 1/2) = 0.05, \\
p_Z(3) &= \mathbb{P}(X = -1, Y = 3/4) + \mathbb{P}(X = 1, Y = 3/4) = 0.05 + 0.04 = 0.09, \\
p_Z(4) &= \mathbb{P}(X = -1, Y = 1) + \mathbb{P}(X = 1, Y = 1) = 0.10 + 0.20 = 0.30.
\end{aligned}$$

Infine il valore di $p_Z(0)$ si può ottenere

$$\begin{aligned}
\textit{sia come}\quad p_Z(0) &= \mathbb{P}(X = 0, Y = 1/4) + \mathbb{P}(X = 0, Y = 1/2) + \mathbb{P}(X = 0, Y = 3/4) + \mathbb{P}(X = 0, Y = 1) \\
&= 0.20 + 0.12 + 0.10 + 0.04 = 0.46, \\
\textit{sia come}\quad p_Z(0) &= 1 - \big(p_Z(1) + p_Z(2) + p_Z(3) + p_Z(4)\big) = 1 - (0.10 + 0.05 + 0.09 + 0.30) = 1 - 0.54 = 0.46.
\end{aligned}$$

Si suggerisce infine di ritrovare il valore atteso di Z usando invece la formula $\mathbb{E}(Z) = \sum_{i=0}^{4} i \cdot p_Z(i)$.

9.4 Valore atteso condizionato e formula del valore atteso totale

Supponiamo di essere venuti a conoscenza del fatto che si è verificato un evento A, con $\mathbb{P}(A) > 0$, allora, per valutare la probabilità di un evento E va utilizzata la probabilità condizionata ad A, cioè $\mathbb{P}(E|A) = \mathbb{P}(E \cap A)/\mathbb{P}(A)$, invece della probabilità $\mathbb{P}(E)$ che avremmo usato, se non avessimo avuto l'informazione che si è verificato l'evento A: in modo del tutto analogo, per valutare il valore atteso di una variabile aleatoria, va utilizzata la probabilità condizionata ad A, e ciò suggerisce la seguente definizione.

Definizione 9.2 (Valore atteso condizionato ad un evento di una variabile aleatoria discreta)**.** *Sia A un evento con $\mathbb{P}(A) > 0$ e sia X una variabile aleatoria a valori in $X(\Omega) = \{x_1, \ldots, x_n\}$. Allora si definisce il* **valore atteso di X, condizionato all'evento A** *come*

$$\mathbb{E}(X|A) = \sum_{i=1}^{N} X(\omega_i)\ \mathbb{P}(\{\omega_i\}|A).$$

La ***Proposizione*** 9.7, applicata alla probabilità condizionata, invece che a $\mathbb{P}$, implica che

$$\mathbb{E}(X|A) = \sum_{k=1}^{n} x_k\ \mathbb{P}(X = x_k|A). \tag{70}$$

Dalla precedente definizione e dalla formula delle probabilità totali si ottiene immediatamente il seguente importante risultato, che mette in relazione il valore atteso $\mathbb{E}(X)$ con i valori attesi condizionati $\mathbb{E}(X|H_j)$, $j = 1, ..., m$, quando gli eventi H_j formano una partizione.

Proposizione **9.14** (Formula del valore atteso totale)**.** *Se $\{H_1, \ldots, H_m\}$ formano una partizione e se $\mathbb{P}(H_i) > 0$ per ogni $i = 1, \ldots, m$, allora vale la seguente formula*

$$\mathbb{E}(X) = \sum_{j=1}^{m} \mathbb{E}(X|H_j)\ \mathbb{P}(H_j), \tag{71}$$

detta ***formula del valore atteso totale****.*

Dimostrazione. Sappiamo che, per ogni $j = 1, \ldots, m$

$$\mathbb{E}(X|H_j) = \sum_{k=1}^{n} x_k\ \mathbb{P}(X = x_k|H_j)$$

per cui

$$\sum_{j=1}^{m} \mathbb{E}(X|H_j)\ \mathbb{P}(H_j) = \sum_{j=1}^{m}\sum_{k=1}^{n} x_k\ \mathbb{P}(X = x_k|H_j)\ \mathbb{P}(H_j).$$

Essendo una somma finita, possiamo scambiare l'ordine di sommatoria e quindi ottenere la tesi osservando che

$$\begin{aligned}\sum_{j=1}^{m} \mathbb{E}(X|H_j)\ \mathbb{P}(H_j) &= \sum_{k=1}^{n}\sum_{j=1}^{m} x_k\ \mathbb{P}(X = x_k|H_j)\ \mathbb{P}(H_j) = \sum_{k=1}^{n} x_k \sum_{j=1}^{m} \mathbb{P}(X = x_k|H_j)\ \mathbb{P}(H_j)\\ &= \sum_{k=1}^{n} x_k\, \mathbb{P}(X = x_k),\end{aligned}$$

dove l'ultima uguaglianza deriva della formula delle probabilità totali applicata all'evento $\{X = x_k\}$ e alla partizione $\{H_1, \ldots, H_m\}$. □

Un caso particolarmente interessante è il caso in cui la partizione è data dalla partizione generata da una variabile aleatoria Y a valori in $Y(\Omega) = \{y_1, \dots, y_m\}$

$$H_j = H_j^Y := \{Y = y_j\}, \qquad j = 1, \dots m.$$

È interessante allora applicare il risultato precedente al caso precedente, specialmente se la distribuzione (discreta) congiunta di X e di Y è data attraverso la distribuzione marginale di Y, o attraverso la sua densità discreta

$$p_Y(y_j) = \mathbb{P}(Y = y_j), \qquad j = 1, \dots, m.$$

e la distribuzione condizionata di X data Y, ovvero attraverso

$$p_{X|Y}(x_k|y_j) = \mathbb{P}(X = x_k|\{Y = y_j\}), \quad k = 1, \dots n$$

per ogni $j = 1, \dots, m$.

In questo caso la formula (71) diviene

$$\begin{aligned} \mathbb{E}(X) &= \sum_{j=1}^{m} \mathbb{E}(X|H_j^Y)\ \mathbb{P}(H_j^Y) = \sum_{j=1}^{m} \mathbb{E}(X|\{Y = y_j\})\ \mathbb{P}(Y = y_j) \\ &= \sum_{j=1}^{m} \Big(\sum_{k=1}^{n} x_k\, \mathbb{P}(X = x_k|\{Y = y_j\}) \Big)\, \mathbb{P}(Y = y_j) = \sum_{j=1}^{m} \Big(\sum_{k=1}^{n} x_k p_{X|Y}(x_k|y_j) \Big) p_Y(y_j) \end{aligned} \tag{72}$$

Vediamo come si può applicare questa formula.

Esempio 9.9. *Sia Y una variabile aleatoria uniforme nell'insieme $\{1, \dots, 10\}$, e supponiamo di sapere che, condizionatamente a $\{Y = j\}$, X ha distribuzione binomiale di parametri $n = j$ e $p = \frac{1}{4}$, cioè, per $j = 1, \dots, 10$,*

$$\mathbb{P}(X = k|Y = j) = \binom{j}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{j-k}, \qquad k = 0, 1, \dots, j.$$

Ad esempio, questo è il caso, se prima viene estratto un numero Y a caso tra 1 e 10, e poi si lanciano due monete ben equilibrate, per Y volte, e si conta il numero X delle volte che esce doppia testa, di modo che, condizionatamente a $\{Y = j\}$, la variabile aleatoria X ha distribuzione $Bin(j, \frac{1}{4})$. Chiaramente $\mathbb{E}(X|\{Y = j\}) = j\,\frac{1}{4}$, e quindi si ottiene

$$\mathbb{E}(X) = \sum_{j=1}^{10} \mathbb{E}(X|Y = j)\ \mathbb{P}(Y = j) = \sum_{j=1}^{10} j\, \frac{1}{4}\, \frac{1}{10} = \frac{1}{4}\, \frac{10 \cdot 11}{2}\, \frac{1}{10} = \frac{11}{8}.$$

Più in generale se Y una variabile aleatoria a valori nell'insieme $\{1, \dots, m\}$ e, condizionatamente a $\{Y = j\}$, X ha distribuzione $Bin(j, p)$, si ha

$$\mathbb{E}(X) = \sum_{j=1}^{m} \mathbb{E}(X|Y = j)\ \mathbb{P}(Y = j) = \sum_{j=1}^{m} j\, p\ \ \mathbb{P}(Y = j) = \mathbb{E}(Y\, p) = p\, \mathbb{E}(Y).$$

L'interesse della formula del valore atteso totale (71) (e similmente della formula (72)) sta nel fatto che permette di calcolare il valore atteso della variabile aleatoria X, senza bisogno di calcolarne la distribuzione, qualora siano noti i valori attesi $\mathbb{E}(X|H_j)$. In particolare, in situazioni simili a quella dell'Esempio 9.9, appena esaminato, basta conoscere il valore atteso della variabile aleatoria Y. Nella Lezione 11 sul campionamento, vedremo alcune applicazioni della formula del valore atteso totale.

9.5 Esercizi di verifica

Esercizio 9.1. Sia X una variabile aleatoria tale che

$$\mathbb{P}(X=-1)=q, \qquad \mathbb{P}(X=1)=p$$

con $0<q=1-p<1$. Calcolare $\mathbb{E}\left(X\right)$.

Esercizio 9.2. Sia X una variabile aleatoria a valori nell'insieme $\{1,2,...,n\}$ con distribuzione di probabilità data dalla posizione

$$\mathbb{P}(\{X=k\}) \propto k, \qquad k=1,...,n.$$

Calcolare[31] $\mathbb{P}(\{X=k\})$, per $k=1,...,n$, e $\mathbb{E}(X)$.

Esercizio 9.3. Calcolare il valore atteso di una variabile aleatoria X a valori nell'insieme

$$\{0,1,2,...,n\},$$

sapendo che risulta

$$\mathbb{P}(\{X=k\}) \propto \frac{\rho^k}{k!(n-k)!}, \qquad k=0,1,...,n.$$

con ρ costante positiva assegnata.

Esercizio 9.4. Siano $g_1,...,g_n$ dei numeri positivi assegnati e $x_1<...<x_n$ assegnati numeri reali.
Sia X una variabile aleatoria a valori nell'insieme $\{x_1,...,x_n\}$ e tale che

$$\mathbb{P}(\{X=x_k\}) \propto g_k, \qquad k=1,...,n.$$

Ottenere la formula per calcolare il valore atteso di X.

Esercizio 9.5. Una persona possiede un biglietto di una lotteria in cui si distribuisce un premio da 5000 euro e tre premi da 2000 Euro ciascuno; possiede inoltre due biglietti di una lotteria che distribuisce un solo premio di 10000 Euro. Per entrambe le lotterie sono stati emessi 1000 biglietti. Indichiamo con X la variabile aleatoria che rappresenta l'importo complessivo della vincita di questa persona.
(a) Calcolare la probabilità di vincere almeno un premio, cioè $\mathbb{P}(X>0)$.
(b) Calcolare il valore atteso di X.

Esercizio 9.6. Tizio punta sul lancio di due dadi; relativamente al primo dado vince 1 Euro se si presenta il punto "3" o il punto "4" e vince 2 Euro se si presenta il punto "5" o il punto "6"; relativamente al secondo dado vince 1 Euro se si presenta il punto "3" e vince 2 Euro se si presenta un qualunque punto pari. In tutti gli altri casi non riceve alcuna vincita.
(a) Determinare il valore atteso della vincita complessiva.
(b) Calcolare la probabilità che Tizio vinca su entrambi i dadi.

Esercizio 9.7. Renato e Stefano lanciano una moneta perfetta n volte ciascuno; indichiamo con X_r ed X_s il numero di risultati testa da loro rispettivamente ottenuti.
Calcolare

$$\mathbb{E}\left(X_r^2 - X_r \cdot X_s\right).$$

[31] Si ricordi che

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \qquad \sum_{k=1}^{n} k^2 = \frac{1}{3}n(n+\frac{1}{2})(n+1)$$

Esercizio 9.8. Renato e Stefano lanciano una moneta perfetta n volte ciascuno ed il vincitore è quello fra i due che realizza il maggior numero di risultati testa.
Indichiamo con X il punteggio del vincitore e con Y il punteggio del perdente (se Renato e Stefano pareggiano, cioè se, utilizzando le notazioni dell'esercizio precedente, $X_r = X_s$ allora $X = Y = X_r = X_s$).
Trovare $\mathbb{E}(X - Y)$.

Esercizio 9.9. Verificare che risulta, per $n = 1, 2, 3, ...$,

$$2^n - 1 = \sum_{k=0}^{n-1} 2^k$$

e riformulare più in generale il testo e la soluzione dell'Esempio 9.7 (sostituendo un valore n generico al posto di $n = 5$).

Soluzione dell'Esercizio 9.8 e di una sua variante

In generale, per ogni a e b numeri reali, $\max(a,b) - \min(a,b) = |a-b|$. Per le due variabili aleatorie X_r e X_s, rispettivamente il punteggio di Renato e di Stefano, si ha che il punteggio del vincitore è $Y = \max(X_r, X_s)$ e quello del perdente è $X = \min(X_r, X_s)$, e quindi $Y - X = \max(X_r, X_s) - \min(X_r, X_s) = |X_r - X_s|$. Poiché X_r e X_s sono variabili aleatorie indipendenti, entrambe con distribuzione $Bin(n, 1/2)$, le variabili aleatorie X_r e $n - X_s$, hanno la stessa distribuzione congiunta di X_r e X_s, e quindi

$$\mathbb{E}\big[|X_r - X_s|\big] = \mathbb{E}\big[|X_r - (n - X_s)|\big] = \mathbb{E}\big[|X_r + X_s - n|\big].$$

Come sappiamo (si veda la sezione 8.3.1) la somma di due variabili aleatorie indipendenti con distribuzione binomiale $Bin(n,p)$ e $Bin(m,p)$, rispettivamente, ha ancora distribuzione binomiale di parametri $n+m$ e p. Quindi $Z = X_r + X_s$ ha distribuzione $Bin(2n, p)$ e si tratta di calcolare

$$\begin{aligned}
\mathbb{E}\big[|Z-n|\big] &= \sum_{k=0}^{2n} |k-n| \, \mathbb{P}(Z=k) = \sum_{k=0}^{2n} |k-n| \binom{2n}{k} \frac{1}{2^{2n}} \\
&= \sum_{k=0}^{n} (n-k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{k=n}^{2n} (k-n) \binom{2n}{k} \frac{1}{2^{2n}},
\end{aligned}$$

(si noti che il caso $k = n$ è ripetuto due volte, ma dà un contributo nullo alla somma)

$$= \sum_{k=0}^{n} (n-k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{k=n}^{2n} (k-n) \binom{2n}{2n-k} \frac{1}{2^{2n}}$$

ponendo $h = 2n - k$ di modo che $k - n = n - h$

$$\begin{aligned}
&= \sum_{k=0}^{n} (n-k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{h=0}^{n} (n-h) \binom{2n}{h} \frac{1}{2^{2n}} \\
&= 2 \sum_{k=0}^{n} (n-k) \binom{2n}{k} \frac{1}{2^{2n}} = 2 \frac{1}{2^{2n}} \left[n \sum_{k=0}^{n} \binom{2n}{k} - \sum_{k=0}^{n} k \binom{2n}{k} \right] \\
&= \frac{n}{2^{2n-1}} \left[n \sum_{k=0}^{n} \binom{2n}{k} - \sum_{k=1}^{n} 2n \binom{2n-1}{k-1} \right]
\end{aligned}$$

Bisogna ora ricordare che

$$\binom{2n}{k} = \binom{2n-1}{k} + \binom{2n-1}{k-1}, \qquad k \geq 1,$$

da cui

$$\begin{aligned}
\mathbb{E}\big[|Z-n|\big] &= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^{n} \binom{2n-1}{k} + \sum_{k=1}^{n} \binom{2n-1}{k-1} - 2 \sum_{k=1}^{n} \binom{2n-1}{k-1} \right] \\
&= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^{n} \binom{2n-1}{k} - \sum_{k=1}^{n} \binom{2n-1}{k-1} \right] \\
&= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^{n} \binom{2n-1}{k} - \sum_{h=0}^{n-1} \binom{2n-1}{h} \right] \\
&= \frac{n}{2^{2n-1}} \binom{2n-1}{n}.
\end{aligned}$$

10 Varianza, Covarianza e comportamento delle medie aritmetiche di variabili aleatorie

10.1 Varianza e Covarianza

Supponiamo di considerare due variabili aleatorie con lo stesso valore atteso, ma con distribuzione diversa. A seguito di quanto visto nella precedente Lezione 9 (ad esempio in ***Osservazione*** **9.2**), è interessante mettere in evidenza – in modo sintetico – le differenze tra le due distribuzioni. Più precisamente l'idea è di trovare altri parametri, oltre al valore atteso, per cui si possano differenziare le distribuzioni stesse. Una nozione utile a tale riguardo è quella di ***varianza***.

In questa lezione introdurremo dunque il concetto di varianza di una variabile aleatoria e ne vedremo alcune proprietà fondamentali; verrà introdotto, a tale proposito, anche il concetto di covarianza di una coppia di variabili aleatorie.

Come si vedrà, tali due nozioni forniscono, fra l'altro, utili informazioni relativamente al comportamento probabilistico di medie aritmetiche di una collezione di variabili aleatorie.

Cominciamo con la seguente

Definizione 10.1 (Varianza). *Sia $X : \Omega := \{\omega_1, ..., \omega_N\} \to X(\Omega) := \{x_1, ..., x_n\}$ una variabile aleatoria ed indichiamo brevemente con μ il valore atteso di X. La* **varianza** *di X viene definita come il valore atteso di $(X-\mu)^2$ e si indica brevemente con il simbolo $Var\,(X)$, cioè*

$$Var\,(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) := \sum_{\omega_i \in \Omega} p(\omega_i)\,(X\,(\omega_i) - \mu)^2 .$$

Osservazione **10.1.** Come il valore atteso $\mathbb{E}(X)$, anche la varianza $Var\,(X)$ dipende soltanto dalla distribuzione di probabilità di X. Si veda a questo proposito anche la ***Proposizione*** **9.7** e la formula (65), con $h(x) = x^2$ (oppure $h(x) = (x-\mu)^2$, con $\mu = \mathbb{E}[X]$), per cui vale

$$Var(X) = \sum_{k=1}^{n} (x_k - \mu)^2\, p_X(x_k)$$

Le proprietà fondamentali di tale definizione sono elencate qui di seguito.

Proposizione **10.1.** *La varianza di una variabile aleatoria X è sempre non negativa, ovvero $Var\,(X) \geq 0$. Inoltre, se $p(\omega_i) > 0$ per ogni $\omega_i \in \Omega$, si ha $Var\,(X) = 0$ se e solo se X è una variabile aleatoria degenere*[32]*.*

[32] Nel caso in cui Ω è un insieme finito, l'ipotesi che $p(\omega_i) > 0$ appare del tutto naturale. Tuttavia vale la pena di esaminare il caso generale in cui possa accadere che $p(\omega_i) = 0$ per qualche $\omega_i \in \Omega$. In tale caso non è più vero che $Var\,(X) = 0$ sia equivalente all'esistenza di un valore $\hat{x}$, tale che $X(\omega_i) = \hat{x}$ per ogni $\omega_i \in \Omega$. Si può invece affermare che $Var\,(X) = 0$ se e solo se esiste un $\hat{x} \in X(\Omega)$ tale che $\mathbb{P}(X = \hat{x}) = 1$. In entrambi i casi, $\hat{x}$ coincide con il valore atteso $\mu := \mathbb{E}(X)$.

Infatti, posto $\mathbb{E}(X) = \mu$, si ha

$$Var\,(X) = \sum_{i=1}^{N} p(\omega_i)\big(X(\omega_i) - \mu\big)^2 = 0$$

se e solo se

$$p(\omega_i)\big(X(\omega_i) - \mu\big)^2 = 0, \qquad \text{per ogni } i = 1, \ldots, N$$

in quanto una somma di termini non negativi è nulla se e solo se tutti gli addendi sono nulli. Di conseguenza almeno uno tra $p(\omega_i)$ e $X(\omega_i) - \mu$ deve essere nullo, cioè

$$\begin{cases} \text{o } X(\omega_i) - \mu = 0 & \text{e allora può essere sia } p(\omega_i) > 0, \text{ sia } p(\omega_i) = 0, \\ \text{o } X(\omega_i) - \mu \neq 0 & \text{e allora necessariamente } p(\omega_i) = 0. \end{cases}$$

Basta poi osservare che $\mathbb{P}(\{\omega : \; X(\omega) \neq \mu\}) = \sum_{i:\; X(\omega_i) - \mu \neq 0} p(\omega_i) = 0$, e che ciò equivale a $\mathbb{P}(\{X(\omega_i) = \mu\}) = 1$.

Un modo equivalente di esprimere la precedente proprietà è il seguente: $Var(X) = 0$ se e solo se X coincide con la variabile aleatoria degenere identicamente uguale a $\mu := \mathbb{E}(X)$ con probabilità 1.

La dimostrazione è immediata e viene lasciata come esercizio.

***Proposizione* 10.2.** *Per ogni variable aleatoria X è possibile calcolare la sua varianza anche attraverso la seguente formula:*

$$Var\,(X) = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2\,. \tag{73}$$

Dimostrazione. Osservando che $(X-\mu)^2 = X^2 - 2\mu X + \mu^2$, e applicando la proprietà di linearità del valore atteso, si ha immediatamente

$$Var\,(X) = \mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mu^2.$$

□

Esempio 10.1. *Sia X una variabile aleatoria binaria, ad esempio l'indicatore di un evento E, con $\mathbb{P}(E) = p$.*
Allora si ha, ricordando anche la **Proposizione** *9.2 della precedente Lezione 9*

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X) - \mathbb{E}(X)^2 = \mathbb{E}(X)\,(1-\mathbb{E}(X)) = p\,(1-p)\,.$$

Si tenga anche conto che se $X = \mathbf{1}_E$, allora $X^2 = X$, in quanto ovviamente $0^2 = 0$ e $1^2 = 1$, e che quindi

$$\mathbb{E}(X^2) = \mathbb{E}(X) = \mathbb{E}(\mathbf{1}_E) = 0\cdot\mathbb{P}(X=0) + 1\cdot\mathbb{P}(X=1) = p.$$

Anche per le seguenti due proprietà le dimostrazioni sono immediate, tenendo conto ad esempio di (73), e vengono lasciate per esercizio.

***Proposizione* 10.3.** *Sia $Y = X + b$, essendo $b \in \mathbb{R}$. Allora $Var\,(Y) = Var\,(X+b) = Var\,(X)\,.$*

***Proposizione* 10.4.** *Sia $Y = a\cdot X$, essendo $a \in \mathbb{R}$. Allora $Var\,(Y) = Var\,(aX) = a^2 \cdot Var\,(X)\,.$*

Dimostrazione di $Var(aX+b) = a^2\,Var(X)$

Iniziamo osservando che

$$\mathbb{E}(aX+b) = a\,\mathbb{E}(X) + b$$

quindi

$$\big(aX+b-\mathbb{E}(aX+b)\big)^2 = \big(aX+\cancel{b}-a\,\mathbb{E}(X)-\cancel{b}\big)^2 = \big[a\big(X-\mathbb{E}(X)\big)\big]^2 = a^2\,\big(X-\mathbb{E}(X)\big)^2$$

da cui

$$\begin{aligned} Var(aX+b) &= \mathbb{E}\big[\big(aX+b-\mathbb{E}(aX+b)\big)^2\big] = \mathbb{E}\big[a^2\,\big(X-\mathbb{E}(X)\big)^2\big] \\ &= a^2\,\mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big] = a^2\,Var(X) \end{aligned}$$

Esempio 10.2. *Consideriamo una variabile aleatoria X tale*

$$\mathbb{P}(X=-1) = q, \qquad \mathbb{P}(X=1) = p$$

con $0 < q = 1-p < 1$. Allora si ha $\mathbb{E}\,(X) = 2p-1$ e, osservando che X^2 è una variabile aleatoria con distribuzione degenere su 1,

$$Var\,(X) = \mathbb{E}\left[X^2\right] - (2p-1)^2 = 4p\,(1-p)\,.$$

Consideriamo ora, per un dato $a > 0$, la variabile $W = a \cdot X$:

$$\mathbb{P}(W = -a) = q, \qquad \mathbb{P}(W = a) = p;$$

allora W^2 è una variabile aleatoria con distribuzione degenere su a^2 e si ha

$$\mathbb{E}\,(W) = a\mathbb{E}\,(X) = a\,(2p-1)\,;$$

$$Var\,(W) = a^2 - a^2\,(2p-1)^2 = a^2 \cdot 4p\,(1-p) = a^2 \cdot Var\,(X)\,,$$

*come del resto era ovvio, tenendo conto della **Proposizione** 10.4*

***Proposizione* 10.5.** *Siano X, Y due variabili aleatorie definite su $(\Omega, \mathcal{P}(\Omega))$. Allora*

$$Var\,(X+Y) = Var\,(X) + Var\,(Y) + 2\,\mathbb{E}\big[\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\big] \tag{74}$$

Dimostrazione. Anche in questo caso la dimostrazione è immediata; ricordando la definizione di varianza

$$Var\,(X+Y) = \mathbb{E}\left[\big(X+Y-\mathbb{E}(X+Y)\big)^2\right]$$

e osservando che

$$\begin{aligned}\big(X+Y-\mathbb{E}(X+Y)\big)^2 &= \big(X-\mathbb{E}(X)+Y-\mathbb{E}(Y)\big)^2 \\ &= \big(X-\mathbb{E}(X)\big)^2 + \big(Y-\mathbb{E}(Y)\big)^2 + 2\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\end{aligned}$$

si ha immediatamente, per la linearità del valore atteso,

$$Var\,(X+Y) = \mathbb{E}\big[\big(X-\mathbb{E}(X)\big)^2\big] + \mathbb{E}\big[\big(Y-\mathbb{E}(Y)\big)^2\big] + 2\,\mathbb{E}\big[\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\big]$$

□

Definizione 10.2 (Covarianza). *Siano date, su uno stesso spazio di probabilità, due variabile aleatorie X ed Y e, per comodità, indichiamo brevemente con μ_X e μ_Y i loro rispettivi valori attesi. Si definisce* **covarianza** *fra X, Y la quantità*

$$Cov\,(X,Y) := \mathbb{E}\,[(X-\mu_X)\cdot(Y-\mu_Y)]\,\Big(=\mathbb{E}\,\big[\big(X-\mathbb{E}(X)\big)\cdot\big(Y-\mathbb{E}(Y)\big)\big]\Big). \tag{75}$$

***Osservazione* 10.2.** Si osservi che $Cov(X,Y) = Cov(Y,X)$ e che,

$$\boxed{\text{se } Y = X, \text{ allora } Cov(X,Y) = Cov(X,X) = Var(X).}$$

Inoltre, svolgendo il prodotto $\big(X-\mu_X\big)\cdot\big(Y-\mu_Y\big)$ nel secondo membro della (75) ed applicando la proprietà di linearità del valore atteso, si ottiene immediatamente[33]

$$Cov\,(X,Y) = \mathbb{E}\,(X\cdot Y) - \mu_X\cdot\mu_Y\Big(=\mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)\Big). \tag{76}$$

Infine, tenendo presente la (75), possiamo scrivere la (74),

$$\boxed{Var(X+Y) = Var\,(X) + Var\,(Y) + 2\,Cov\,(X,Y)\,.} \tag{77}$$

[33]Infatti

$$(X-\mu_X)\cdot(Y-\mu_Y) = X\cdot Y - X\cdot\mu_Y - \mu_X\cdot Y + \mu_X\cdot\mu_Y$$

da cui, prendendo il valore atteso,

$$\begin{aligned}\mathbb{E}\Big((X-\mu_X)\cdot(Y-\mu_Y)\Big) &= \mathbb{E}(X\cdot Y) - \mathbb{E}(X\cdot\mu_Y) - \mathbb{E}(\mu_X\cdot Y) + \mathbb{E}(\mu_X\cdot\mu_Y) \\ &= \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mu_Y - \mu_X\cdot\mathbb{E}(Y) + \mu_X\cdot\mu_Y \\ &= \mathbb{E}(XY) - \mu_X\cdot\mu_Y - \mu_X\cdot\mu_Y + \mu_X\cdot\mu_Y = \mathbb{E}(XY) - \mu_X\cdot\mu_Y\,.\end{aligned}$$

Esercizio proposto 10.1. *Verificare che, date tre variabili aleatorie X, Y_1, Y_2 e due costanti α, β risulta*

$$Cov\left(X, \alpha Y_1 + \beta Y_2\right) = \alpha\, Cov\left(X, Y_1\right) + \beta\, Cov\left(X, Y_2\right). \tag{78}$$

Per la simmetria della covarianza, la proprietà (78) implica ovviamente anche che, date tre variabili aleatorie X_1, X_2, Y, e due costanti a e b

$$Cov\left(aX_1 + bX_2, Y\right) = aCov\left(X_1, Y\right) + bCov\left(X_2, Y\right).$$

La precedente proprietà insieme alla (78) permette di ottenere che

$$\begin{aligned} &Cov\left(aX_1 + bX_2, \alpha Y_1 + \beta Y_2\right) \\ &= a \cdot \alpha Cov\left(X_1, Y_1\right) + a \cdot \beta Cov\left(X_1, Y_2\right) + b \cdot \alpha Cov\left(X_2, Y_1\right) + b \cdot \beta Cov\left(X_2, Y_2\right), \end{aligned}$$

ossi la proprietà di bilinearità della covarianza.

Come corollario della ***Proposizione* 9.10** della precedente Lezione 9, ed in virtù della (77), otteniamo immediatamente

***Proposizione* 10.6.** *Siano date X ed Y due variabili aleatorie definite su uno stesso spazio di probabilità. Se X ed Y sono stocasticamente indipendenti, allora*

$$Cov(X, Y) = 0 \quad e \quad Var(X + Y) = Var\left(X\right) + Var\left(Y\right).$$

Dimostrazione. Grazie alla (77), basta dimostrare che $Cov(X, Y) = 0$. A questo scopo basta ricordare la (76) e che, per la ***Proposizione* 9.10** se X ed Y sono indipendenti allora $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$. □

***Osservazione* 10.3.** Siano X ed Y due variabili aleatorie binarie, ossia a valori $in\{0, 1\}$, definite su uno stesso spazio di probabilità. Osserviamo che anche il loro prodotto $X \cdot Y$ è una variabile binaria, con

$$\mathbb{E}\left(X \cdot Y\right) = \mathbb{P}\left(\{X \cdot Y = 1\}\right) = \mathbb{P}\left(\{X = 1\} \cap \{Y = 1\}\right)$$

e dunque

$$Cov\left(X, Y\right) = \mathbb{P}\left(\{X = 1\} \cap \{Y = 1\}\right) - \mathbb{P}\left(\{X = 1\}\right) \cdot \mathbb{P}\left(\{Y = 1\}\right).$$

A questo proposito è utile osservare che,

se $X = \mathbf{1}_A$ ed $Y = \mathbf{1}_B$ allora $X \cdot Y = \mathbf{1}_A \cdot \mathbf{1}_B = \mathbf{1}_{A\cap B}$,

come si vede subito, considerando che
$\mathbf{1}_A \cdot \mathbf{1}_B(\omega) = 1$ se e solo se $\omega \in A$ e simultaneamente $\omega \in B$, ovvero $\omega \in A \cap B$
e anche
$\mathbf{1}_{A\cap B}(\omega) = 1$ se e solo se $\omega \in A \cap B$.

Notiamo inoltre che la condizione $Cov(X, Y) = 0$, ovvero

$$\mathbb{E}\left(X \cdot Y\right) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{79}$$

implica in tale caso che X ed Y sono stocasticamente indipendenti in quanto

$$\begin{aligned} \mathbb{P}\left(\{X = 1\} \cap \{Y = 1\})\right\} &= \mathbb{E}\left(X \cdot Y\right) \\ &= \mathbb{E}(X) \cdot \mathbb{E}(Y) = \mathbb{P}\left(\{X = 1\}\right) \cdot \mathbb{P}\left(\{Y = 1\}\right). \end{aligned}$$

Si osservi che se, come prima, $X = \mathbf{1}_A$ ed $Y = \mathbf{1}_B$ allora la distribuzione congiunta di X e di Y è individuata da

$$\mathbb{P}(\{X=0\}\cap\{Y=0\}) = \mathbb{P}\left(\overline{A}\cap\overline{B}\right), \quad \mathbb{P}(\{X=0\}\cap\{Y=1\}) = \mathbb{P}\left(\overline{A}\cap B\right),$$
$$\mathbb{P}(\{X=1\}\cap\{Y=0\}) = \mathbb{P}\left(A\cap\overline{B}\right), \quad \mathbb{P}(\{X=1\}\cap\{Y=1\}) = \mathbb{P}\left(A\cap B\right),$$

e che quindi l'indipendenza degli eventi $A = \{X=1\}$ e $B = \{Y=1\}$, è necessaria e sufficiente per ottenere l'indipendenza delle variabili aleatorie X ed Y.

***Osservazione* 10.4.** In generale, cioè per coppie di variabili non entrambe binarie, ***la condizione* (79) *(ovvero* $Cov(X,Y) = 0$*)* è soltanto necessaria, ma non sufficiente per l'indipendenza stocastica**, come mostra infatti il seguente controesempio.
Controesempio: Mostreremo due variabili aleatorie X e Y, con $Cov(X,Y) = 0$, ma che non sono indipendenti.
Sia X una variabile aleatoria tale che

$$\mathbb{P}(X=-1) = \mathbb{P}(X=0) = \mathbb{P}(X=+1) = \frac{1}{3}$$

e poniamo $Y := X^2$, cosicché $X \cdot Y = X^3$. Ovviamente X, Y non sono stocasticamente indipendenti.

Il fatto che X ed $Y = X^2$ non siano indipendenti si vede immediatamente: ad esempio
$\mathbb{P}(Y=0) = \mathbb{P}(X^2=0) = \mathbb{P}(X=0) = 1/3 > 0$, e quindi $\mathbb{P}(Y=0)\cdot\mathbb{P}(X=1) = (1/3)\cdot(1/3) > 0$, mentre $\mathbb{P}(\{Y=0\}\cap\{X=1\}) = \mathbb{P}(\emptyset) = 0$.

Per vedere che $Cov(X,Y) = 0$ notiamo che, in questo caso, $XY = X^3$ ha la stessa distribuzione di probabilità di X, in quanto addirittura $X^3 = X$ (infatti $x^3 = x$ per $x = -1, 0, +1$), inoltre $\mathbb{E}(X) = \mathbb{E}\left(X^3\right) = 0$, e quindi risulta

$$\mathbb{E}(X\cdot Y) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X)\cdot\mathbb{E}(Y) \quad \Leftrightarrow \quad Cov(X,Y) = 0.$$

La precedente osservazione ci porta naturalmente a dare la seguente definizione

Definizione 10.3 (Variabili aleatorie non correlate, correlate positivamente e correlate negativamente)**.** *Diremo che due variabili aleatorie X ed Y, definite su uno stesso spazio di probabilità, sono* **non correlate** *(o anche* **scorrelate***) se risulta verificata la condizione* (79)*, ossia*

$$\mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y) \quad \Leftrightarrow \quad Cov(X,Y) = 0.$$

Se invece $\mathbb{E}(XY) > \mathbb{E}(X)\mathbb{E}(Y)$, ossia se $Cov(X,Y) > 0$, si dice che X e Y sono ***correlate positivamente****, mentre se $\mathbb{E}(XY) < \mathbb{E}(X)\mathbb{E}(Y)$, ossia se $Cov(X,Y) < 0$, si dice che X e Y sono* **correlate negativamente***.*

Esercizio proposto 10.2. *Generalizzare il controesempio dell'***Osservazione 10.4***, mostrando che le variabili aleatorie X ed $Y = X^2$, sono scorrelate, ma non indipendenti, ogni volta che la variabile aleatoria X è simmetrica rispetto all'origine, ossia con $X(\Omega) = \{\pm x_1, \pm x_2, \ldots, \pm x_r\} \cup \{0\}$, per un $r \geq 1$, (oppure $X(\Omega) = \{\pm x_1, \pm x_2, \ldots, \pm x_r\}$, per un $r \geq 2$), e con*

$$\mathbb{P}(X = x_i) = \mathbb{P}(X = -x_i)$$

per ogni $i = 1, \ldots, r$.
Perché il caso $X(\Omega) = \{-x_1, x_1\}$ non fornisce un controesempio?

Prima di proseguire è utile generalizzare la relazione (77) che permette di calcolare la varianza della somma di due variabili aleatorie al caso della somma di un numero finito $X_1, X_2, \ldots, X_n$ di variabili aleatorie:

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{h=1}^{n}\sum_{k=1}^{n} Cov(X_h, X_k) \tag{80}$$

$$= \sum_{k=1}^{n} Var(X_k) + \sum\sum_{h\neq k}^{1\leq h,k\leq n} Cov(X_h, X_k) \tag{81}$$

$$= \sum_{k=1}^{n} Var(X_k) + 2\sum_{h=1}^{n-1}\sum_{k=h+1}^{n} Cov(X_h, X_k). \tag{82}$$

Basterà mostrare (80), in quanto le (81) e (82), sono solo forme differenti della prima espressione, in cui si usa il fatto che $Cov(X_k, X_k) = Var(X_k)$ e che $Cov(X_k, X_h) = Cov(X_h, X_k)$. Tuttavia rimandiamo la dimostrazione della (80), nell'Appendice alla fine di questa Lezione, e preferiamo illustrare prima le sue applicazioni.

Esempio* 10.3 (Varianza di una variabile binomiale). *Sia X una variabile aleatoria con distribuzione binomiale $Bin(n, \theta)$. Quanto vale $Var(X)$?*

Soluzione: Conviene ragionare lungo la linea già svolta nella soluzione dell'Esempio 9.3 della precedente lezione, cioè *riguardiamo* X come la somma di n variabili aleatorie binarie indipendenti $X_1, \ldots, X_n$: per ogni $i = 1, \ldots, n$, possiamo prendere $X_i = \mathbf{1}_{A_i}$, dove $A_1, A_2, \ldots, A_n$ formano uno schema di Bernoulli, ossia sono eventi indipendenti e tutti di probabilità θ. Ciascuna delle $X_i = \mathbf{1}_{A_i}$ ha valore atteso θ e di varianza $\theta \cdot (1-\theta)$; dunque $Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = n\theta \cdot (1-\theta)$, per la (81), in quanto $Cov(X_h, X_k) = 0$, per $h \neq k$.

Riguardare X come la somma di n variabili aleatorie $X_1, \ldots, X_n$ binarie indipendenti (e quindi non correlate: $Cov(X_i, X_j) = 0$ per $i \neq j$), ciascuna di valore atteso θ e di varianza $\theta \cdot (1-\theta)$, equivale a considerare che X ha la stessa distribuzione di $S := \sum_{i=1}^{n} X_i$, e quindi

$$Var(X) = Var(S) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum\sum_{h\neq k}^{1\leq h,k\leq n} \underbrace{Cov(X_h, X_k)}_{=0}$$

$$= \sum_{i=1}^{n} Var(X_i) = n\theta(1-\theta).$$

Esempio 10.4 (Varianza di una variabile ipergeometrica). *Vogliamo ora calcolare la varianza di una variabile aleatoria X con distribuzione ipergeometrica di parametri M, m_1, n. Come nel precedente esempio, possiamo riguardare X come la somma di n variabili aleatorie binarie $X_1, \ldots, X_n$: per ogni $i = 1, \ldots, n$, possiamo prendere $X_i = \mathbf{1}_{A_i}$, dove $A_i =$*{all'i-sima estrazione esce un elemento di tipo A}*, quando si fanno n estrazioni senza reinserimento da una popolazione/urna che contiene M individui/palline, di cui m_1 di tipo A e $m_2 = M - m_1$ di tipo B. Ciascuna delle $X_i = \mathbf{1}_{A_i}$ ha valore atteso $\frac{m_1}{M}$ e dunque di varianza $\frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right)$.*

Nel presente caso però $X_1, \ldots, X_n$ non sono indipendenti; è chiaro che esse, prese a due a due, esse hanno una stessa covarianza, cioè $Cov(X_h, X_k) = Cov(X_1, X_2)$ e quindi possiamo scrivere

$$Var(X) = \sum_{k=1}^{n} Var(X_k) + \sum\sum_{h\neq k}^{1\leq h,k\leq n} Cov(X_h, X_k) = n \cdot \frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right) + n \cdot (n-1) \cdot Cov(X_1, X_2).$$

Pensando al caso delle estrazioni senza reinserimento, per ogni $h \in \{1, 2, ..., n\}$ si ha $\mathbb{P}(A_h) = \mathbb{P}(A_1) = \frac{m_1}{M}$ e , per ogni $h \neq k$ in $\{1, 2, ..., n\}$,

$$\mathbb{E}(X_h \cdot X_k) = \mathbb{E}\big(\mathbf{1}_{A_h} \cdot \mathbf{1}_{A_k}\big) = \mathbb{E}\big(\mathbf{1}_{A_h \cap A_k}\big) = \mathbb{P}(A_h \cap A_k) = \mathbb{P}(A_1 \cap A_2)$$

e inoltre $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1) = \frac{m_1}{M}\,\frac{m_1-1}{M-1}$

Per vedere che $\mathbb{P}(A_h) = \mathbb{P}(A_1)$ e che $\mathbb{P}(A_h \cap A_k) = \mathbb{P}(A_1 \cap A_2)$ iniziamo ricordando che gli m_1 elementi di tipo A si possono numerare da 1 a m_1, e gli $m_2 = M - m_1$ di tipo B si possono numerare da $m_1 + 1$ a $m_1 + m_2 = M$, di modo che Ω è l'insieme delle disposizioni $(j_1, j_2, ..., j_n)$ di M elementi di classe n, e $A_i = \big\{(j_1, j_2, ..., j_n) \in \Omega :\ j_i \in \{1, ..., m_1\}\big\}$; basta poi osservare che $|A_h| = |A_1|$ e $|A_h \cap A_k| = |A_1 \cap A_2|$ in quanto esiste una corrispondenza biunivoca tra A_h e A_1, data dallo scambio di j_1 con j_h, e similmente ne esiste una anche tra $A_h \cap A_k$ e $A_1 \cap A_2$, data dagli scambi di j_1 con j_h, e j_2 con j_k.

Ora, essendo X_1, X_2 variabili aleatorie binarie, si ha

$$\begin{aligned}
\boxed{Cov\,(X_1, X_2)} &= \mathbb{P}\left(\{X_1 = 1\} \cap \{X_2 = 1\}\right) - \mathbb{P}\left(\{X_1 = 1\}\right) \cdot \mathbb{P}\left(\{X_2 = 1\}\right) \\
&= \mathbb{P}(\{X_1 = 1\}) \cdot \mathbb{P}(\{X_2 = 1\}|\{X_1 = 1\}) - \left(\frac{m_1}{M}\right)^2 \\
&= \frac{m_1}{M} \cdot \frac{m_1 - 1}{M - 1} - \left(\frac{m_1}{M}\right)^2 = \frac{m_1}{M}\left(\frac{m_1 - 1}{M - 1} - \frac{m_1}{M}\right) = \frac{m_1}{M} \cdot \frac{\cancel{m_1 M} - M - \cancel{m_1 M} + m_1}{M(M-1)} \\
&= -\frac{m_1}{M} \cdot \frac{M - m_1}{M(M-1)} \boxed{= -\frac{m_1}{M}\left(1 - \frac{m_1}{M}\right)\frac{1}{M-1}}
\end{aligned}$$

Possiamo concludere scrivendo

$$\begin{aligned}
\boxed{Var(X)} &= \sum_{k=1}^{n} Var(X_k) + n(n-1)\,Cov(X_1, X_2) \\
&= n\,\frac{m_1}{M}\left(1 - \frac{m_1}{M}\right) + n(n-1)\left(-\frac{m_1}{M}\right) \cdot \left(1 - \frac{m_1}{M}\right) \cdot \frac{1}{M-1} \\
&= n \cdot \frac{m_1}{M} \cdot \frac{M - m_1}{M}\left(1 - \frac{n-1}{M-1}\right) \boxed{= np(1-p)\left(1 - \frac{n-1}{M-1}\right)},
\end{aligned}$$

dove si è posto $p = \frac{m_1}{M}$, la percentuale di palline bianche presenti nell'urna.

Osserviamo che per $n = 1$ si ha una sola estrazione e la varianza viene, come ci si aspetta $p(1-p)$ mentre se $n = M$, ossia se estraiamo tutte le palline, allora certamente $X = m_1$, ossia X è una variabile aleatoria degenere, e quindi $Var(X) = 0$, ed in effetti se $n = M$ allora $np(1-p)\left(1 - \frac{n-1}{M-1}\right) = np(1-p)\left(1 - \frac{M-1}{M-1}\right) = 0$.

*Osserviamo inoltre che se il numero totale M delle palline presenti nell'urna è molto grande rispetto al numero n delle estrazioni, allora $np(1-p)\left(1 - \frac{n-1}{M-1}\right)$, cioè la varianza del numero di palline estratte (**senza reinserimento**) dall'urna è molto vicina alla varianza $np(1-p)$, con $p = \frac{m_1}{M}$ del numero delle palline estratte nel caso in cui le estrazioni siano **con reinserimento**, in accordo con quanto osservato nell'**Osservazione** 6.3.*

Esempio 10.5. *Consideriamo le vincite X_h, per $h = 1, 2, ..., n$, associate agli n diversi biglietti nella lotteria considerata nell'Esempio 9.5 della precedente Lezione 9. Abbiamo visto che che $\mathbb{E}(X_h) = \frac{c_1 r_1 + c_2 r_2 + c_3 r_3}{n}$ e possiamo facilmente calcolare $Var(X_h) = \mathbb{E}(X_h^2) - \big(\mathbb{E}(X_h)\big)^2$, tenendo presente che $\mathbb{E}(X^2) = \frac{c_1^2 r_1 + c_2^2 r_2 + c_3^2 r_3}{n}$, in quanto $\mathbb{P}(X_h = c_i) = \frac{r_i}{n}$, $i = 1, 2, 3$, e $\mathbb{P}(X_h = 0) = 1 - \frac{r}{n}$, dove $r = r_1 + r_2 + r_3$,*

$$Var(X_h) = \frac{c_1^2 r_1 + c_2^2 r_2 + c_3^2 r_3}{n} - \left(\frac{c_1 r_1 + c_2 r_2 + c_3 r_3}{n}\right)^2.$$

Volendo calcolare $Cov\,(X_h, X_k)$, per $h \neq k$ potremmo procedere semplicemente come segue. Innanzitutto, come nel precedente esempio, possiamo osservare che $Cov\,(X_h, X_k)$ non dipende dalla coppia di indici h, k (purché $h \neq k$) e da ciò segue

$$Var\,(S_n) = n \cdot Var\,(X_1) + n(n-1)Cov\,(X_1, X_2)\,;$$

si ha d'altra parte che la distribuzione di probabilità di $S_n = X_1 + X_2 + \cdots + X_n$ è degenere; ne segue

$$Var\,(S_n) = 0.$$

Possiamo dunque concludere

$$Cov\,(X_1, X_2) = -\frac{Var\,(X_1)}{n-1}.$$

10.2 Media aritmetica, disuguaglianza di Chebyshev e primi passi verso la Legge dei Grandi Numeri

Consideriamo ora n variabili aleatorie $X_1, ..., X_n$ (non necessariamente binarie) e, per semplicità, le assumiamo tali che

$$\mathbb{E}(X_1) = ... = \mathbb{E}(X_n) = \mu, \tag{83}$$
$$Var(X_1) = ... = Var(X_n) = \sigma^2 \tag{84}$$
$$Cov(X_h, X_k) = \varphi, \qquad 1 \leq h \neq k \leq n. \tag{85}$$

Posto Y_n la loro ***media aritmetica***, ossia

$$Y_n := \frac{1}{n} \sum_{h=1}^{n} X_h, \tag{86}$$

che cosa possiamo dire circa la loro ***media aritmetica***?

A tale proposito è immediato verificare la seguente proposizione (basta tener presente sia la linearità del valore atteso che le regole viste qui sopra circa il calcolo della varianza).

***Proposizione* 10.7.** *Siano date n variabili aleatorie $X_1, ..., X_n$, che verificano (83), (84) e (85), e sia Y_n è la media aritmetica di $X_1, ..., X_n$, definita come sopra in (86). Si ha*

$$\mathbb{E}(Y_n) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \mu$$

e

$$Var(Y_n) = Var\left(\frac{1}{n} \sum_{k=1}^{n} X_k\right) = \frac{1}{n}\left[\sigma^2 + (n-1)\varphi\right].$$

Dimostrazione. La dimostrazione è immediata

$$\mathbb{E}(Y_n) = \mathbb{E}\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) = \frac{1}{n}\,\mathbb{E}\Big(\sum_{k=1}^{n} X_k\Big) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k) = \frac{1}{n}\,n\mu = \mu,$$

$$Var(Y_n) = Var\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) = \Big(\frac{1}{n}\Big)^2 \; Var\Big(\sum_{k=1}^{n} X_k\Big) = \frac{1}{n^2}\left[\sum_{k=1}^{n}\overbrace{Var(X_k)}^{=\sigma^2} + \sum_{\substack{1\le h,k\le n\\ h\neq k}}\sum \overbrace{Cov(X_h,X_k)}^{=\varphi}\right]$$

$$= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)\varphi\right] = \frac{1}{n}\left[\sigma^2 + (n-1)\varphi\right].$$

□

***Osservazione* 10.5.** La varianza $Var(X)$ di una variabile aleatoria X è un indice del grado di dispersione della distribuzione di probabilità rispetto al valore atteso; è chiaro dalla definizione di varianza che, a parità di valore atteso μ, scarti grandi (in modulo) di X rispetto a μ sono tanto più probabili quanto più risulta grande $Var(X)$.

A tale proposito è utile osservare quanto segue: così come $\mathbb{E}(X)$ corrisponde al concetto di baricentro quando le probabilità $p_j = \mathbb{P}(X = x_j)$ vengono interpretate come delle masse concentrate sui diversi punti $x_1, ..., x_n$, analogamente $Var(X)$, nella stessa interpretazione, corrisponde al concetto di ***momento di inerzia*** della distribuzione stessa.

In collegamento con quanto appena detto, possiamo affermare che la conoscenza del valore atteso e della varianza di una variabile aleatoria fornisce un'idea indicativa, seppur riassuntiva, della sua distribuzione di probabilità e se ne possono ricavare alcune utili disuguaglianze; in particolare vale la seguente fondamentale disuguaglianza:

***Proposizione* 10.8 (Disuguaglianza di Chebyshev).** *Sia X una variabile aleatoria, con valore atteso uguale a μ e varianza uguale a σ^2. Allora*

$$\mathbb{P}(\{|X-\mu| > \varepsilon\}) \le \frac{\sigma^2}{\varepsilon^2}, \qquad \forall\, \varepsilon > 0,$$

o equivalentemente

$$\mathbb{P}\big(\{|X-\mathbb{E}(X)| > \varepsilon\}\big) \le \frac{Var(X)}{\varepsilon^2}, \qquad \forall\, \varepsilon > 0.$$

Si osservi che ovviamente nella diseguaglianza di Chebyshev si prende $\varepsilon > 0$ in quanto, per $\varepsilon < 0$ si ha $\mathbb{P}(\{|X-\mu| > \varepsilon\}) = \mathbb{P}(\Omega) = 1$, mentre per $\varepsilon = 0$ non avrebbe senso il secondo membro della diseguaglianza.
Infine va osservato che la diseguaglianza di Chebyshev è interessante solo se $\frac{\sigma^2}{\varepsilon^2} < 1$, in quanto in caso contrario si ottiene solo una banalità, ovvero che $\mathbb{P}(\{|X-\mu| > \varepsilon\})$ è minore o uguale di un numero maggiore o uguale di 1, ma ciò è ovvio in quanto la probabilità di ogni evento è un numero minore o uguale ad 1.

.

Dimostrazione. La dimostrazione segue immediatamente dalla definizione stessa di varianza, infatti possiamo scrivere

$$\sigma^2 = \sum_{\omega_i\in\Omega} p(\omega_i)\cdot(X(\omega_i)-\mu)^2 = \sum_{\omega_i:|X(\omega_i)-\mu|>\varepsilon} p(\omega_i)\cdot(X(\omega_i)-\mu)^2 + \sum_{\omega_i:|X(\omega_i)-\mu|\le\varepsilon} \overbrace{p(\omega_i)\cdot(X(\omega_i)-\mu)^2}^{\ge 0}$$

$$\ge \sum_{\omega_i:|X(\omega_i)-\mu|>\varepsilon} p(\omega_i)\cdot(X(\omega_i)-\mu)^2 \ge \sum_{\omega_i:|X(\omega_i)-\mu|>\varepsilon} p(\omega_i)\cdot\varepsilon^2 = \varepsilon^2\cdot \sum_{\omega_i:|X(\omega_i)-\mu|>\varepsilon} p(\omega_i)$$

$$= \varepsilon^2\cdot\mathbb{P}(\{|X-\mu| > \varepsilon\}).$$

La dimostrazione è quindi terminata. $\square$

Si osservi che la dimostrazione della Disuguaglianza di Chebyshev rimane pressoché invariata se invece di considerare l'evento l'evento $\{|X-\mu| \geq \varepsilon\}$, per cui vale anche

$$\mathbb{P}(\{|X-\mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

e che, essendo $\{|X-\mu| > \varepsilon\} \subseteq \{|X-\mu| \geq \varepsilon\}$, la precedente disuguaglianza è anche più forte della disuguaglianza appena dimostrata.

***Proposizione* 10.9.** *Siano date n variabili aleatorie $X_1, ..., X_n$ ed indichiamo con $Y_n = (\sum_{k=1}^{n} X_k)/n$ la loro media aritmetica. Se $X_1, ..., X_n$ verificano (83), (84) e (85), con $\varphi = 0$ ovvero hanno lo stesso valore atteso, la stessa varianza e non sono correlate, cioè sono tali che*

$$\begin{aligned} \mathbb{E}(X_1) &= ... = \mathbb{E}(X_n) = \mu, \\ Var(X_1) &= ... = Var(X_n) = \sigma^2 \\ Cov(X_h, X_k) &= 0, \qquad 1 \leq h \neq k \leq n, \end{aligned}$$

allora

$$\mathbb{P}(\{|Y_n - \mu| > \varepsilon\}) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| > \varepsilon\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}, \qquad \textit{qualunque sia } \varepsilon > 0.$$

Dimostrazione. La dimostrazione segue immediatamente ricordando le precedenti ***Proposizioni* 10.7 e 10.8**. Infatti si ha prima di tutto per la diseguaglianza di Chebyshev (***Proposizione* 10.8**) si ha

$$\mathbb{P}(|Y_n - \mathbb{E}(Y_n)| > \varepsilon) \leq \frac{Var(Y_n)}{\varepsilon^2}.$$

Inoltre per la ***Proposizione* 10.7** si ha che $\mathbb{E}(Y_n) = \mu$ e che $Var(Y_n) = \frac{1}{n}\sigma^2$. $\square$

***Osservazione* 10.6.** La tesi della ***Proposizione* 10.9** si può anche riscrivere come

$$\mathbb{P}(|Y_n - \mu| \leq \varepsilon) \geq 1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2},$$

o equivalentemente

$$\mathbb{P}(\mu - \varepsilon \leq Y_n \leq \mu + \varepsilon) \geq 1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}.$$

In altre parole si può dire che l'evento *"la media aritmetica Y_n di X_1, X_2,..., X_n, differisce dal valore atteso μ (comune a tutte le v.a. X_i) meno di ε"* ha probabilità maggiore o uguale a $1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}$. Se n è *"molto grande"*, in modo che $1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}$ sia *"vicino"* ad 1, la tesi si può parafrasare anche dicendo che, se n è *"molto grande"*, media aritmetica e valore atteso differiscono tra loro meno di ε, con probabilità *"vicina"* ad 1.
In questo senso possiamo affermare che la ***Proposizione* 10.9** è un **primo passo verso la Legge dei Grandi Numeri**.

Dal punto di vista applicativo, è interessante anche il fatto che, sempre grazie a tale risultato, siamo in grado di rispondere ***alla seguente domanda***:
Nelle condizioni della* Proposizione *10.9, quante prove si devono effettuare, ovvero quanto si deve prendere grande n, affinché, con probabilità almeno $1-\delta$, la media aritmetica differisca dal valore atteso μ meno di ε?

La risposta alla precedente domanda è molto semplice: ***è sufficiente prendere*** $\boxed{n \geq \frac{\sigma^2}{\varepsilon^2\,\delta},}$

infatti in tale caso $\frac{\sigma^2}{\varepsilon^2\,n} \leq \delta$ è equivalente a $1 - \frac{\sigma^2}{\varepsilon^2\,n} \geq 1 - \delta$, e quindi

$$\mathbb{P}(|Y_n - \mu| > \varepsilon) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} \leq \delta, \quad \Leftrightarrow \quad \mathbb{P}(\mu - \varepsilon \leq Y_n \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2\,n} \geq 1 - \delta.$$

Nel caso in cui non fosse noto esplicitamente il valore σ^2 della varianza, ma fosse nota solo una sua maggiorazione, cioè se si conoscesse esplicitamente un valore $\overline{\sigma}$ tale che $\sigma^2 \leq \overline{\sigma}^2$, allora, sempre allo scopo di ottenere che la media aritmetica differisca dal valore atteso μ meno di ε con probabilità maggiore o uguale a $1 - \delta$, basterebbe prendere

$$n \geq \frac{\overline{\sigma}^2}{\varepsilon^2\,\delta} \geq \frac{\sigma^2}{\varepsilon^2\,\delta}.$$

Esempio* 10.6. *Una coppia di dadi perfetti a sei facce viene lanciata n volte ed indichiamo con S_n il numero dei lanci in cui il maggiore fra i due punteggi risulta maggiore o uguale a 5.*

Calcolare il minimo valore di n per il quale, in base alla disuguaglianza di Chebyshev, si possa scrivere

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{5}{9}\right| > \frac{1}{30}\right) \leq \frac{1}{10}$$

Soluzione: Per iniziare osserviamo che S_n è la somma di n variabili aleatorie binarie X_i indipendenti ed ugualmente distribuite. Posto Z il valore del primo dado e W il valore del secondo dado (entrambi al primo lancio) si ha

$$\begin{aligned}\mathbb{P}(X_1 = 1) &= \mathbb{P}(\max(Z,W) \geq 5) = 1 - \mathbb{P}(\max(Z,W) < 5) = 1 - \mathbb{P}(Z < 5,\ W < 5)\\ &= 1 - \mathbb{P}(Z < 5)\,\mathbb{P}(W < 5) = 1 - \mathbb{P}(Z \leq 4)\,\mathbb{P}(W \leq 4) = 1 - \left(\frac{4}{6}\right)^2 = \frac{20}{36} = \frac{5}{9} = \theta.\end{aligned}$$

Quindi in questo caso $\mu = \theta = \frac{5}{9}$, mentre $\sigma^2 = \theta(1-\theta) = \frac{5}{9}\frac{4}{9}$, e infine $\varepsilon = \frac{1}{30}$, di conseguenza

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{5}{9}\right| > \frac{1}{30}\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} = \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} = \frac{1}{n}\frac{\frac{5}{9}\frac{4}{9}}{\left(\frac{1}{30}\right)^2} = \frac{1}{n}\frac{5\cdot 4\cdot 3^2\cdot 10^2}{9\cdot 9} = \frac{1}{n}\frac{2000}{9} \leq \frac{1}{10}$$

$$\Updownarrow$$

$$n \geq \frac{20000}{9} \simeq 2222,22 \qquad \Leftrightarrow \qquad n \geq 2223$$

Esempio* 10.7 (intervalli di fiducia o di confidenza)**.** *Si consideri una moneta truccata con parametro p incognito, dove p rappresenta la probabilità che esca testa. Al fine di determinare p si lancia una la moneta n volte e si stima p con la frequenza di testa $Y_n = S_n/n$, dove S_n è il numero delle volte in cui esce testa, tra gli n lanci, ossia $S_n = \mathbf{1}_{E_1} + \mathbf{1}_{E_2} + \cdots + \mathbf{1}_{E_n}$, dove $E_i =$*{esce testa all'i-simo lancio}*, sono eventi indipendenti e inoltre $\mathbb{E}(\mathbf{1}_{E_i}) = \mathbb{P}(E_i) = p$ e $Var(\mathbf{1}_{E_i}) = \mathbb{P}(E_i)(1 - \mathbb{P}(E_i)) = p(1-p)$, per cui $\mathbb{Y}_\ltimes = p$ e $Var(Y_n) = p(1-p)/n$. Fissato $\varepsilon > 0$ vogliamo determinare come possiamo prendere n in modo che la probabilità che $|Y_n - p| = |S_n/n - p| < \varepsilon$ sia almeno il 95% (usando la disuguaglianza di Chebyshev).*

*Prima di capire come si può trovare n, vale la pena osservare che, una volta risolto tale problema potremo affermare non solo che $\mathbb{P}(p - \varepsilon \leq Y_n \leq p + \varepsilon) \geq 0.95$ (NOTA BENE si tratta della probabilità a priopri, ossia prima di lanciare la moneta) ma potremo anche riparafrasare tale risultato dicendo che il valore incognito p apparterrà all'**intervallo aleatorio** $[Y_n - \varepsilon, Y_n + \varepsilon]$ con probabilità maggiore o uguale a 0.95, ossia*

$$\mathbb{P}(Y_n - \varepsilon \leq p \leq Y_n + \varepsilon) \geq 0.95$$

*L'intervallo aleatorio $[Y_n - \varepsilon, Y_n + \varepsilon]$ viene detto allora **intervallo di fiducia** o anche **intervallo di confidenza** (al 95%) per il valore incognito p.*

Per risolvere il problema consideriamo che $95\% = 0.95 = 1 - \delta$, per $\delta = 0.05$, e quindi, da quanto visto in precedenza, sappiamo che basterebbe prendere, per tale valore di $\delta = 0.05$, e $\sigma^2 = p(1-p)$,

$$n \geq \frac{\sigma^2}{\delta \varepsilon^2} = \frac{p(1-p)}{0.05\varepsilon^2}.$$

Tuttavia sorge una difficoltà: non conosciamo $p = \mathbb{P}(E_i)$, e quindi neanche il valore della varianza; tuttavia sappiamo che per $p \in [0,1]$ e quindi[34] il valore $\sigma^2 = p(1-p) \leq 1/4$. Questa osservazione ci permette di affermare che basta prendere

$$n \geq \frac{5}{\varepsilon^2} = \frac{1}{4 \cdot 0.05\varepsilon^2} \geq \frac{p(1-p)}{0.05\varepsilon^2}.$$

Ad esempio per $\epsilon = 0.1 = \frac{1}{10}$, otteniamo che basta prendere $n \geq \frac{5}{\frac{1}{10^2}} = 500$.

Definizione 10.4 (Variabili aleatorie standard). *Una variabile aleatoria Z si dice* **standard** *quando*

$$\mathbb{E}(Z) = 0, \qquad Var(Z) = 1.$$

Come applicazione della disuguaglianza di Chebyshev, si ha che, se Z è una variabile aleatoria standard allora[35] , $\forall \varepsilon > 0$,

$$\mathbb{P}(\{|Z| > \varepsilon\}) \leq \frac{1}{\varepsilon^2}. \tag{87}$$

Data una variabile aleatoria X, con valore atteso μ e varianza $\sigma^2 > 0$, possiamo costruire una variabile aleatoria standard e che sia funzione[36] di X, ponendo

$$X^* = \frac{X-\mu}{\sigma} \left(= \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} \right),$$

con $\sigma = \sqrt{\sigma^2} > 0$. La variabile X^* viene detta ***standardizzata*** di X; inoltre il valore $\sigma = \sqrt{\sigma^2}$ prende il nome di ***scarto standard*** o anche ***scarto quadratico medio***. In altre parole si può dire che una variabile aleatoria X, con valore atteso μ e varianza σ^2 si può sempre scrivere nella forma

$$X = \sigma Z + \mu,$$

essendo Z una variabile aleatoria standard: infatti se $\sigma^2 > 0$ possiamo prendere $Z = X^*$, mentre se $\sigma^2 = 0$ allora X è una variabile aleatoria degenere e quindi coincide con il suo valore atteso μ. Inoltre la (87) diviene

$$\mathbb{P}\left(|X^*| > \varepsilon\right) = \mathbb{P}\left(\left| \frac{X-\mu}{\sigma} \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2}.$$

Consideriamo ora n prove bernoulliane $E_1, ..., E_n$ di probabilità θ.

[34] La funzione $x \in \mathbb{R} \mapsto f(x) = x(1-x)$ è tale che $f'(x) = 1 - 2x$ e $f''(x) = -1 < 0$, con $f'(x) = 0$ se e solo se $x = 1/2$ e inoltre $f(\frac{1}{2}) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}$, e cioè il punto di massimo è $x_{\max} = \frac{1}{2}$ e il valore massimo è $f(\frac{1}{2}) = \frac{1}{4}$.

[35] Ovviamente la (87) ha interesse solo se $\frac{1}{\varepsilon^2} < 1$, cioè solo se $\varepsilon > 1$.

[36] Si tratta di una funzione affine di X, ovvero $X^* = aX + b$, con $(a,b) = \left(\frac{1}{\sigma}, -\frac{\mu}{\sigma}\right)$. Si noti che $\left(\frac{1}{\sigma}, -\frac{\mu}{\sigma}\right)$ è l'unica coppia di valori (a,b) per cui $aX + b$ è una variabile aleatoria standard.

Indichiamo con $X_1, ..., X_n$ gli indicatori di $E_1, ..., E_n$ e con $S_n = \sum_{j=1}^n X_j = \sum_{j=1}^n \mathbf{1}_{E_j}$ il numero di successi sulle n prove, e quindi la loro media aritmetica

$$Y_n = \frac{S_n}{n},$$

è la variabile aleatoria con il significato di ***frequenza relativa dei successi sulle*** n ***prove***.

Relativamente alla loro media aritmetica $Y_n = \frac{S_n}{n}$, abbiamo il seguente risultato.

***Proposizione* 10.10.** *Per ogni* $\varepsilon > 0$

$$\mathbb{P}\left(\left|\sqrt{n}\frac{Y_n - \theta}{\sqrt{\theta\,(1-\theta)}}\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}.$$

Dimostrazione. Basta ricordare che $\mathbb{E}(S_n) = n\theta$ e (Esempio 10.3) che $Var(S_n) = n \cdot \theta\,(1-\theta)$ e dunque

$$\mathbb{E}(Y_n) = \theta, \qquad Var(Y_n) = \frac{\theta\,(1-\theta)}{n}$$

e quindi applicare la (87) alla variabile aleatoria standardizzata di Y_n

$$Y_n^* = \frac{Y_n - \mathbb{E}(Y_n)}{Var(Y_n)} = \sqrt{n}\frac{Y_n - \theta}{\sqrt{\theta\,(1-\theta)}}.$$

□

Chiudiamo questo paragrafo notando che l'interesse della precedente ***Proposizione* 10.10** risiede nel fatto che, per la variabile standardizzata $\sqrt{n}\frac{Y_n-\theta}{\sqrt{\theta(1-\theta)}}$ della media aritmetica, le probabilità di differire dal valore atteso (cioè zero) più di ε non si può rendere piccola (nemmeno prendendo n grande), come invece accade per la variabile Y_n. Questa proprietà è connessa con il Teorema Centrale del Limite, che verrà discusso più avanti nella Lezione 16.2.

10.3 Diseguaglianza di Cauchy e coefficiente di correlazione

Siano date due variabili aleatorie X ed Y, vale allora la seguente diseguaglianza

$$|Cov(X, Y)| \le \sqrt{Var(X)} \cdot \sqrt{Var(Y)}. \tag{88}$$

Tale diseguaglianza è nota come ***diseguaglianza di Cauchy*** e generalizza la diseguaglianza di Cauchy-Schwartz per vettori, ovvero, se $\boldsymbol{u}, \boldsymbol{v}$ sono vettori di $\mathbb{R}^n$, allora

$$| < \boldsymbol{u}, \boldsymbol{v} > | \le \|\boldsymbol{u}\|\|\boldsymbol{v}\|\Big(\sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} \cdot \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}\Big),$$

dove con $< \boldsymbol{u}, \boldsymbol{v} >= \sum_{i=1}^n u_i \cdot v_i$ si indica il prodotto scalare tra i vettori $\boldsymbol{u}$ e $\boldsymbol{v}$, e con $\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$ si indica il modulo del vettore $\boldsymbol{u}$.

Infatti si consideri il caso particolare in cui $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ ed $Y(\Omega) = \{y_1, y_2, \ldots, y_n\}$, e si abbia

$$\mathbb{P}(X = x_i, Y = y_i) = \frac{1}{n}, \quad i = 1, 2, \ldots, n$$
$$\mathbb{P}(X = x_i, Y = y_j) = 0, \qquad i \ne j, i, j = 1, 2, \ldots, n.$$

Allora, posto $\mu = \mathbb{E}(X)$ e $\nu = \mathbb{E}(Y)$, la diseguaglianza (88) diviene

$$\left|\sum_{i=1}^{n}(x_i-\mu)(y_i-\nu)\frac{1}{n}\right| \le \sqrt{\sum_{i=1}^{n}(x_i-\mu)^2\frac{1}{n}}\sqrt{\sum_{i=1}^{n}(y_i-\nu)^2\frac{1}{n}}$$

che è esattamente la diseguaglianza di Cauchy-Schwartz per i vettori $\boldsymbol{u}$ e $\boldsymbol{v}$ con $u_i = x_i - \mu$ e $v_i = y_i - \nu$, per $i = 1, 2, \ldots, n$, a parte per il fattore $\frac{1}{n}$.

La dimostrazione della diseguaglianza (88) di Cauchy è basata sull'osservazione che la funzione

$$\varphi(x) := \mathbb{E}\big[\big((X-\mu) - x(Y-\nu)\big)^2\big]$$

gode di due proprietà:
$\varphi(x) \ge 0$ per ogni x, in quanto valore atteso di una variabile aleatoria non negativa,
$\varphi(x) = Var(X) - 2x\,Cov(X,Y) + x^2\,Var(Y)$, come si vede subito per la linearità del valore atteso e considerando che

$$\big((X-\mu) - x(Y-\nu)\big)^2 = (X-\mu)^2 - 2x(X-\mu)(Y-\nu) + x^2(Y-\nu)^2.$$

Di conseguenza $\varphi(x) = a\,x^2 + b\,x + c$, con $a = Var(Y)$, $b = -2Cov(X,Y)$ e $c = Var(X)$, ed il discriminante $b^2 - 4ac = 4\big(Cov(X,Y)\big)^2 - 4Var(X)Var(Y)$ è minore o uguale a zero, ovvero

$$\big(Cov(X,Y)\big)^2 \le Var(X)Var(Y).$$

Per ottenere la (88) basta estrarre la radice quadrata.

Come conseguenza il rapporto

$$\rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

che è detto ***coefficiente di correlazione*** tra X ed Y è sempre minore o uguale ad 1 in valore assoluto.

Seguitando l'analogia con il caso vettoriale, in qualche senso il coefficiente di correlazione $\rho_{X,Y}$ generalizza il coseno dell'angolo tra due vettori, infatti è noto che, se $\widehat{\boldsymbol{uv}}$ indica l'angolo compreso fra due vettori,

$$\cos(\widehat{\boldsymbol{uv}}) = \frac{<\boldsymbol{u},\boldsymbol{v}>}{\|\boldsymbol{u}\|\cdot\|\boldsymbol{v}\|}.$$

Per continuare l'analogia notiamo che se $Y = \alpha\,X$, allora $\rho_{X,Y} = \pm 1$ a seconda del segno di α, come del resto accade che $\cos(\widehat{\boldsymbol{uv}}) = \pm 1$, nel caso in cui $\boldsymbol{v} = \alpha\boldsymbol{u}$.

10.4 Appendice: Covarianza della somma di n variabili aleatorie

Se $X_1, X_2, \ldots, X_n$ sono n variabili aleatorie, definite sullo stesso spazio di probabilità, allora

$$\begin{aligned}
Var\left(\sum_{k=1}^{n} X_k\right) &= \sum_{h=1}^{n}\sum_{k=1}^{n} Cov(X_h, X_k)\\
&= \sum_{k=1}^{n} Var(X_k) + \sum_{h\neq k}^{1\le h,k\le n}\!\!\sum Cov(X_h, X_k)\\
&= \sum_{k=1}^{n} Var(X_k) + 2\sum_{h=1}^{n-1}\sum_{k=h+1}^{n} Cov(X_h, X_k).
\end{aligned}$$

Come già osservato basta mostrare la prima uguaglianza, in quanto le altre due sono solo forme differenti della prima espressione.

Per iniziare si ponga per semplicità $\mathbb{E}(X_i) = \mu_i$, così $\mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mu_i$. Si tratta quindi di calcolare il valore atteso di

$$\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)^2 = \left(\sum_{i=1}^n (X_i - \mu_i)\right)^2 = \left(\sum_{h=1}^n (X_h - \mu_h)\right) \cdot \left(\sum_{k=1}^n (X_k - \mu_k)\right)$$
$$= \sum_{h=1}^n \sum_{k=1}^n (X_h - \mu_h)(X_k - \mu_k).$$

dove l'ultima uguaglianza dipende dalla propiretà distributiva della somma rispetto al prodotto: si veda la nota a pagine 35, con $R = S = n$ e $b_k = a_k = (X_k - \mu_k)$.

A questo punto basta passare al valore atteso e sfruttarne la proprietà di linearità:

$$Var\left(\sum_{i=1}^n X_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)^2\right] = \mathbb{E}\left[\sum_{h=1}^n \sum_{k=1}^n (X_h - \mu_h)(X_k - \mu_k)\right]$$
$$= \sum_{h=1}^n \sum_{k=1}^n \mathbb{E}\left[(X_h - \mu_h)(X_k - \mu_k)\right] = \sum_{h=1}^n \sum_{k=1}^n Cov(X_h, X_k).$$

10.5 Esercizi di verifica

Esercizio 10.1. Sia X il punteggio ottenuto nel lancio di un dado a sei facce. Calcolare $Var\,(X)$.

Esercizio 10.2. Sia X la vincita associata ad un biglietto di una lotteria che, su un totale di 10000 biglietti distribuisce 10 premi da 200 Euro e 20 premi da 100 Euro. Calcolare $Var\,(X)$.

Esercizio 10.3. Siano Y_1 ed Y_2 i primi due numeri estratti su una ruota del lotto e poniamo

$$E_1 := \{Y_1 > 45\}, \quad E_2 := \{Y_2 < 45\}.$$

Calcolare la covarianza fra gli indicatori $X_1 = \mathbf{1}_{E_1}$ ed $X_2 = \mathbf{1}_{E_2}$ degli eventi E_1, E_2.

Esercizio 10.4. Sia S_{100} il numero di elettori per lo schieramento A in un campione casuale (senza reinserimento) di 100 elettori estratti da una popolazione di 1000 elettori di cui m votano per A e $(1000-m)$ votano per B. Che cosa si ottiene applicando la diseguaglianza di Chebyshev alla variabile aleatoria S_{100}? Considerare il caso $m = 500$, con $\varepsilon = 20$, $\varepsilon = 10$ e $\varepsilon = 1$.

Esercizio 10.5. Il Dipartimento di Matematica acquista 20 copie di un software; ciascuna copia ha probabilità $\frac{1}{100}$ di dare degli errori di funzionamento, indipendentemente dal comportamento delle altre. Indichiamo con S, la variabile aleatoria che conta il numero di copie che danno errori. Scrivere la disuguaglianza che si ottiene applicando ad S la *diseguaglianza di Chebyshev*, e considerare il caso $\varepsilon = 3$, $\varepsilon = 2$ e $\varepsilon = 1$.

Esercizio 10.6. Siano X, Y due variabili aleatorie standardizzate e consideriamo,$\forall t \in \mathbb{R}$, la variabile aleatoria

$$T_t := (X - tY)^2\,.$$

(a) Calcolare $\mathbb{E}\,(T_t)$
(b) Tenendo conto che deve risultare $\mathbb{E}\,(T_t) \geq 0, \forall t \in \mathbb{R}$, dimostrare che risulta

$$|Cov(X, Y)| \leq 1.$$

Esercizio 10.7. Sia X una variabile aleatoria con valore atteso μ e varianza σ^2. Definiamo, per $t \in \mathbb{R}$, la funzione

$$f(t) := \mathbb{E}\left((X-t)^2\right).$$

(a) Calcolare esplicitamente $f(t)$.
(b) Mostrare che μ è il punto di minimo di f e che

$$\sigma^2 = \min_{t\in\mathbb{R}} f(t).$$

In altre parole mostrare che, per ogni $t \in \mathbb{R}$,

$$Var(X) = \mathbb{E}\left((X-\mu)^2\right) \leq \mathbb{E}\left((X-t)^2\right).$$

11 Campionamento da popolazioni con composizione incognita; indipendenza condizionata

In questa lezione analizzeremo una generalizzazione del classico modello di un numero n fissato di estrazioni (con o senza reinserimento) da una stessa urna/popolazione, che contiene un numero noto M di palline/individui di due tipi A e B, ma la cui composizione è incognita, nel senso che non è noto il numero di palline/individui di tipo A che l'urna contiene. In altre parole il numero di palline/individui di tipo A viene considerato come una variabile aleatoria che denoteremo con $R = R^A$. Denoteremo invece con $S = S_n^A$ il numero delle palline/individui di tipo A estratti nelle n estrazioni.

Assumeremo note la distribuzione di R, ossia $\mathbb{P}(R = r)$, $r = 0, 1, ..., M$, e la distribuzione condizionata di S dato $R = r$, ossia $\mathbb{P}(S = s|R = r)$. Chiaramente, avendo considerato solo estrazioni con o senza reinserimento, la distribuzione condizionata di S dato $R = r$ potrà essere solo di due tipi binomiale o ipergeometrica.

Lo scopo principale di questa lezione è quindi quello di studiare la distribuzione congiunta di R ed S, e la distribuzione condizionata di R dato $S = s$ (sulla motivazione vedere la successiva ***Osservazione* 11.1**.). Tuttavia sono interessanti anche il valore atteso di R dato $S = s$ e la distribuzione marginale di S.

Si tratta quindi di sviluppare un'analisi di casi piuttosto particolari. Tale analisi può risultare interessante per varie applicazioni e per le connessioni con problematiche di tipo statistico; essa permetterà inoltre di illustrare ulteriormente varie nozioni viste nelle precedenti lezioni.

Alla fine di questa lezione considereremo anche il concetto di indipendenza condizionata, che è legata solo al caso di estrazioni con reinserimento, ma ci limiteremo al caso di due eventi e di due variabili aleatorie.

La generalizzazione consiste nel considerare una coppia di variabili aleatorie R, S (a valori interi non negativi), nei casi in cui la distribuzione di R sia nota e la distribuzione condizionata di S dato R sia binomiale oppure ipergeometrica.

11.1 Descrizione formale del modello

Sia data una coppia di variabili aleatorie R, S (a valori interi non negativi), di cui sia nota la distribuzione di R e la distribuzione condizionata di S dato R.

Pur se con qualche modifica nella notazione, iniziamo innanzitutto richiamando e sviluppando alcuni aspetti ed alcuni esempi, ai quali si è già accennato nelle lezioni precedenti.

Consideriamo una ***popolazione*** costituita da un totale di M ***elementi***, di cui alcuni di ***tipo*** A ed altri di ***tipo*** B.

Qui analizziamo il caso in cui ***il numero complessivo di elementi di tipo A*** (e quindi anche di tipo B) sia ***non noto*** e viene visto come una variabile aleatoria, ***che indicheremo con il simbolo*** R; ovviamente R è dunque una variabile aleatoria a valori in $\{0, 1, ..., M\}$.

Lo stato di informazione su R viene descritto attraverso la distribuzione di probabilità:

$$p_R(r) := \mathbb{P}(\{R = r\}), \qquad r = 0, 1, ..., M. \tag{89}$$

Eseguiamo ora n estrazioni dalla popolazione, registrando il tipo di elemento (A o B) che, man mano, viene estratto ed ***indichiamo con S il numero di elementi di tipo A estratti,*** o, meglio, risultanti nel ***campione*** estratto.

Ovviamente S è, in generale, una variabile aleatoria; la distribuzione di probabilità di S, condizionata al valore assunto da R, è data dalle probabilità condizionate

$$p_S(s|\{R = r\}) := \mathbb{P}(\{S = s\}|\{R = r\}), \qquad s = 0, 1, ..., n. \tag{90}$$

È chiaro che tale distribuzione condizionata viene determinata in base alle modalità con cui vengono effettuate le n estrazioni.

Una volta che siano state assegnate la distribuzione marginale della variabile R e le distribuzioni condizionate di S data R, ne risulta univocamente determinata la distribuzione di probabilità congiunta della coppia (R, S), attraverso la formula[37]

$$p_{R,S}(r,s) := \mathbb{P}(\{R=r, S=s\}) = p_R(r) \cdot p_S(s|R=r), \qquad 0 \leq s \leq n, \quad 0 \leq r \leq M. \tag{91}$$

A partire da tale distribuzione congiunta possiamo ottenere la distribuzione marginale di S attraverso la formula delle probabilità totali[38]

$$p_S(s) = \sum_{r=0}^{M} p_{R,S}(r,s) = \sum_{r=0}^{M} p_R(r) \cdot p_S(s|R=r), \qquad s = 0, 1, ..., n. \tag{92}$$

Attraverso l'uso della Formula di Bayes[39] possiamo ora anche ottenere la distribuzione condizionata di R data l'osservazione di un valore s per S:
possiamo scrivere, ponendo $p_R(r|S=s) = \mathbb{P}(\{R=r\}|\{S=s\})$,

$$p_R(r|S=s) = \frac{p_R(r) \cdot p_S(s|R=r)}{p_S(s)}, \qquad r = 0, 1, ..., M. \tag{93}$$

***Osservazione* 11.1 (di carattere euristico).** Il problema risolto dalla formula (93), cioè quello di calcolare la distribuzione di probabilità condizionata della variabile R (numero complessivo di elementi di tipo A fra gli M elementi della popolazione) data l'osservazione di un valore s per la variabile S (numero di elementi di tipo A fra tutti gli n elementi esaminati) si riallaccia chiaramente ad una problematica di tipo statistico.

Tale problema è legato infatti all'esigenza di ricavare, in merito al numero di elementi di un tipo fissato presenti all'interno di una popolazione, dell'informazione rilevante senza scrutinare tutta la popolazione, ma bensì scrutinandone soltanto una parte. Problemi di tale genere si presentano frequentemente in molti campi applicativi, ad esempio nel controllo di qualità che si deve effettuare su pezzi di una produzione industriale o nelle proiezioni di un risultato elettorale, etc....

Vi sono, nella pratica, vari metodi per formalizzare ed affrontare tali problemi. Il metodo qui esaminato (che in un certo senso è quello più puramente probabilistico) si può riassumere come segue:

$*$ assegnando una distribuzione di probabilità (marginale) ad R, si esprime ***lo stato di informazione***, circa il valore che può assumere tale variabile, di cui si dispone ***prima di fare il "campionamento"*** (cioè prima delle estrazioni degli n elementi da esaminare)

$*$ quindi si assegnano le distribuzioni condizionate di S date le possibili ipotesi sul valore assunto da R; tali distribuzioni condizionate riflettono ***le modalità con cui vengono effettuate le estrazioni*** degli n elementi

$*$ in base ai due ingredienti fin qui descritti si ottiene, applicando la Formula di Bayes, la distribuzione condizionata di R dato il valore s osservato per la variabile S; tale distribuzione condizionata si interpreta come quella distribuzione di probabilità che rappresenta ***lo stato di informazione su*** R***, a cui si perviene dopo aver osservato l'evento*** $\{S=s\}$. Si suggerisce a tale proposito di tornare all'***Osservazione* 4.3** della Lezione 4.

Le considerazioni svolte nella precedente ***Osservazione 11.1*** mettono anche in luce l'interesse di calcolare

$$\mathbb{E}(R|S=s) := \sum_{i=1}^{N} R(\omega_i)\, \mathbb{P}(\{\omega_i\}|\{S=s\}),$$

[37]Si tratta della solita formula delle probabilità composta:

$$\mathbb{P}(\{R=r, S=s\}) = \mathbb{P}(\{R=r\} \cap \{S=s\}) = \mathbb{P}(\{R=r\})\, \mathbb{P}(\{S=s\}|\{R=r\}).$$

[38]La formula delle probabilità totali viene applicata all'evento $\{S=s\}$ e alla partizione $H_r = \{R=r\}$

[39]La fomrula di Bayes viene applicata alll'evento $\{S=s\}$, e alla partizione $H_r = \{R=r\}$

cioè il ***valore atteso condizionato*** di R data l'osservazione di un valore s per S.

Come abbiamo già visto nel paragrafo 9.4, ed in particolare in (70),

$$\mathbb{E}(R|S=s) = \sum_{r=0}^{M} r \cdot \mathbb{P}(\{R=r\}|\{S=s\}) = \sum_{r=0}^{M} r \cdot p_R(r|\{S=s\}).$$

Tale quantità non è altro che il valore atteso calcolato rispetto alla distribuzione di R condizionata ad $\{S=s\}$ ed espressa nella (93) e cioè

$$\mathbb{E}(R|S=s) = \frac{\sum_{r=0}^{M} r \cdot p_R(r) \cdot p_S(s|R=r)}{p_S(s)}. \tag{94}$$

Consideriamo ora in dettaglio due particolari modalità di estrazioni dalla popolazione:
(i) estrazioni casuali con reinserimento
e
(ii) estrazioni casuali senza reinserimento.

Questi casi danno luogo a due specifici modelli per quanto riguarda le distribuzioni condizionate per la variabile S dati i valori della variabile R.

Come sappiamo, nel caso ***(i)*** delle ***estrazioni casuali con reinserimento*** risulta (si veda a questo proposito l'Esempio 7.3 nella Lezione 7)

$$p_S(s|R=r) = \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}, \qquad s = 0, 1, ..., n. \tag{95}$$

cioè la distribuzione condizionata di S dato $(R=r)$ è una distribuzione binomiale di parametri n ed $\frac{r}{M}$.

Nel caso ***(ii)*** delle ***estrazioni casuali senza reinserimento***, per cui necessariamente $n \leq M$, risulta invece (si veda ora l'Esempio 7.4 nella Lezione 7)

$$p_S(s|R=r) = \frac{\binom{r}{s}\binom{M-r}{n-s}}{\binom{M}{n}}, \qquad \max(0, r+n-M) \leq s \leq \min(r, n), \tag{96}$$

cioè la distribuzione condizionata è una distribuzione ipergeometrica $Hyp(M, r; n)$.

Imponendo rispettivamente tali due condizioni, le precedenti formule (92), (93) e (94) diventano dunque:
(i) estrazioni casuali con reinserimento

$$p_S(s) = \frac{\binom{n}{s}}{M^n} \sum_{r=0}^{M} p_R(r) \cdot r^s \cdot (M-r)^{n-s}, \qquad s = 0, 1, ..., n \tag{97}$$

$$p_R(r|S=s) = \frac{p_R(r) \cdot \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}, \qquad r = 0, 1, ..., M \tag{98}$$

e di conseguenza, sempre per ogni $s = 0, 1, ..., n$,

$$\mathbb{E}(R|S=s) = \frac{\sum_{r=0}^{M} r \cdot p_R(r) \cdot \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}. \tag{99}$$

(ii) estrazioni casuali senza reinserimento

$$p_S(s) = \sum_{r=s}^{M-n+s} p_R(r) \cdot \frac{\binom{r}{s}\binom{M-r}{n-s}}{\binom{M}{n}}, \qquad s = 0, 1, ..., n \tag{100}$$

$$p_R(r|S=s) = \frac{p_R(r) \cdot \binom{r}{s}\binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}}, \qquad r = s, s+1, ..., M-n+s, \tag{101}$$

e di conseguenza[40], sempre per ogni $s = 0, 1, ..., n$,

$$\mathbb{E}(R|S=s) = \frac{\sum_{r=s}^{M-n+s} r \cdot p_R(r) \cdot \binom{r}{s}\binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}} \qquad . \tag{102}$$

A questo punto vediamo che $p_S(s)$, $p_R(r|S=s)$ e $\mathbb{E}(R|S=s)$ sono completamente determinati, in modi diversi a seconda che ci si trovi nel caso ***(i)*** o nel caso ***(ii)***, una volta specificata la distribuzione marginale di R.

È interessante ora analizzare in particolare il caso in cui R segue una distribuzione binomiale.

11.2 Esempi elementari

Esempio* 11.1. *Un'urna contiene 5 oggetti, alcuni di colore Arancio ed altri di colore Blu.*

Il numero di oggetti di colore Arancio non è a noi noto, bensì è una variabile aleatoria R, con distribuzione data da

$$p_R(0) = \mathbb{P}(R=0) = \frac{1}{20}, \quad p_R(1) = \mathbb{P}(R=1) = \frac{2}{10}, \quad p_R(2) = \mathbb{P}(R=2) = \frac{1}{20},$$
$$p_R(3) = \mathbb{P}(R=3) = 0, \quad p_R(4) = \mathbb{P}(R=4) = \frac{5}{10}, \quad p_R(5) = \mathbb{P}(R=5) = \frac{2}{10}.$$

Eseguiamo 3 estrazioni casuali senza reinserimento da quest'urna ed indichiamo con S il numero di oggetti di colore Arancio ottenuti in tali estrazioni.
Qual è la distribuzione di probabilità di S?

Soluzione: Ovviamente risulta $\mathbb{P}(\{S=h\}|\{R=r\}) = \frac{\binom{r}{h}\binom{5-r}{3-h}}{\binom{5}{3}}$ per $0 \leq h \leq r$ e $0 \leq 3-h \leq 5-r$, da cui:

$$\begin{aligned}\mathbb{P}(\{S=0\}) &= \sum_{r=0}^{5} \mathbb{P}(\{R=r\})\ \mathbb{P}(\{S=0\}|\{R=r\}) \\ &= \frac{1}{20} \cdot 1 + \frac{2}{10} \cdot \frac{2}{5} + \frac{1}{20} \cdot \frac{1}{10} = \frac{27}{200};\end{aligned}$$

[40]Si noti che la condizione che $r = s, s+1, ..., M-n+s$ in (101), deriva immediatamente dalle condizioni su s in (96), che a loro volta derivano da

$$\begin{cases} 0 \leq s \leq r \\ 0 \leq n-s \leq M-r. \end{cases}$$

Queste ultime, viste come condizioni su r, per s fissato, divengono

$$\begin{cases} r \geq s \\ r \leq M-(n-s) \end{cases}$$

Del resto, nel caso ***estrazioni casuali senza reinserimento***, sapere che $\{S=s\}$ equivale ad aver osservato s elementi di tipo A ed $n-s$ elementi di tipo B, e quindi equivale a sapere che il numero totale di elementi di tipo A della popolazione è almeno s, e che il numero totale di elementi di tipo B è almeno $n-s$, e quest'ultima condizione implica che gli elementi di tipo A non possono essere più di $M-(n-s) = M-n+s$.

analogamente

$$\begin{aligned}\mathbb{P}(\{S=1\}) &= \frac{2}{10}\cdot\frac{6}{10}+\frac{1}{20}\cdot\frac{6}{10}=\frac{30}{200};\\ \mathbb{P}(\{S=2\}) &= \frac{1}{20}\cdot\frac{3}{10}+\frac{5}{10}\cdot\frac{6}{10}=\frac{63}{200};\\ \mathbb{P}(\{S=3\}) &= \frac{5}{10}\cdot\frac{4}{10}+\frac{2}{10}\cdot 1=\frac{80}{200}.\end{aligned}$$

Esempio* 11.2. *Nelle condizioni del precedente Esempio 11.1, qual è la distribuzione condizionata di R data l'osservazione $\{S=2\}$? Inoltre, posto $T := R-S$ il numero di palline color Arancio rimaste nell'urna, le variabili aleatorie S e T sono indipendenti?*

Soluzione: Se nelle tre estrazioni senza reinserimento, sono state estratte 2 palline di color Arancio (e quindi 1 di colore Blu), risulta ovviamente che ci devono essere almeno 2 palline color Arancio e almeno 1 di colore Blu e quindi

$$\mathbb{P}(\{R=0\}|\{S=2\})=\mathbb{P}(\{R=1\}|\{S=2\})=\mathbb{P}(\{R=5\}|\{S=2\})=0;$$

inoltre, essendo $\mathbb{P}(\{R=3\})=0$ si ha[41]

$$\mathbb{P}(\{R=3\}|\{S=2\})=0;$$

infine

$$\begin{aligned}\mathbb{P}(\{R=2\}|\{S=2\}) &= \frac{\frac{1}{20}\cdot\frac{3}{10}}{\frac{63}{200}}=\frac{3}{63},\\ \mathbb{P}(\{R=4\}|\{S=2\}) &= \frac{\frac{5}{10}\cdot\frac{6}{10}}{\frac{63}{200}}=\frac{60}{63}.\end{aligned}$$

È facile verificare che, in questo caso, le due variabili S e $T := R-S$ non possono essere stocasticamente indipendenti: ad esempio

$$\begin{aligned}\mathbb{P}(\{R-S=0\}) &= \sum_{r=0}^{5}\mathbb{P}(\{R=r\})\;\mathbb{P}(\{S=r\}|\{R=r\})=\sum_{r=0}^{3}\mathbb{P}(\{R=r\})\;\frac{\binom{r}{r}\binom{5-r}{3-r}}{\binom{5}{3}}\\ &= \frac{1}{20}+\frac{6}{10}\,\frac{2}{10}+\frac{3}{10}\,\frac{1}{20}=\frac{10+24+3}{200}=\frac{37}{200}\end{aligned}$$

mentre

$$\mathbb{P}(\{R-S=0\}|\{S=2\})=\mathbb{P}(\{R=2\}|\{S=2\})=\frac{3}{63}$$

Esempio* 11.3. *Continuando sempre a considerare l'urna dell'Esempio 11.1, esaminiamo ora il caso in cui* **le tre estrazioni siano con reinserimento***. Qual è la distribuzione di probabilità di S?*

[41]Si osservi che $\mathbb{P}(\{R=3\}|\{S=2\})=0$, in quanto per ipotesi $\mathbb{P}(\{R=3\})=0$ e quindi

$$0\le\mathbb{P}(\{R=3\}\cap\{S=2\})\le\mathbb{P}(\{R=3\})=0.$$

Di conseguenza

$$\mathbb{P}(\{R=3\}|\{S=2\})=\frac{\mathbb{P}(\{R=3\}\cap\{S=2\})}{\mathbb{P}(\{S=2\})}=0.$$

Soluzione: Si ha

$$\mathbb{P}(\{S=k\}) = \sum_{r=0}^{5} \mathbb{P}(\{R=r\}) \binom{3}{k} \left(\frac{r}{5}\right)^k \left(\frac{5-r}{5}\right)^{3-k}, \qquad k=0,1,2,3.$$

In particolare[42]

$$\mathbb{P}(\{S=2\}) = \frac{1}{125}\left(0+\frac{2}{10}\cdot 12+\frac{1}{20}\cdot 36+\frac{5}{10}\cdot 48\right)$$

Esempio* 11.4. *Nelle stesse condizioni del precedente Esempio 11.3 qual è in questo caso la distribuzione condizionata di R data l'osservazione $\{S=2\}$?*

Soluzione: Si debbono ovviamente ancora escludere i casi $\{R=0\}$, $\{R=5\}$, nel senso che le rispettive probabilità condizionate sono nulle, e si ha[43]

$$\mathbb{P}(\{R=1\}|\{S=2\}) = \frac{\mathbb{P}(\{R=1\})\,\mathbb{P}(\{S=2\}|\{R=1\})}{\mathbb{P}(\{S=2\})} = \frac{4}{47}$$

analogamente

$$\mathbb{P}(\{R=2\}|\{S=2\}) = \frac{3}{47}, \qquad \mathbb{P}(\{R=4\}|\{S=2\}) = \frac{40}{47}$$

Di nuovo $\mathbb{P}(\{R=3\}|\{S=2\})=0$, in quanto per ipotesi $\mathbb{P}(\{R=3\})=0$.

11.3 Caso di estrazioni casuali senza reinserimento e con distribuzione binomiale per R.

Ci sono alcuni aspetti da notare nel caso particolare in cui R segua una distribuzione binomiale $Bin(M,\theta)$ e ci si ponga nel caso ***(ii)*** di estrazioni senza reinserimento.

Come vedremo in tale caso anche il numero aleatorio S ha una distribuzione binomiale, e precisamente $Bin(n,\theta)$. Qui sotto ne diamo la dimostrazione *analitica*, ma consigliamo di leggere attentamente l'***Osservazione* ??**, per una dimostrazione euristica/probabilistica.

[42] Si consideri che per $k=2$, si ha $3-k=1$ e quindi, se sono state estratte (con reinserimento) 2 palline di color Arancio è stata estratta anche 1 pallina di colore Blu, per cui il numero di palline di color Arancio è almeno 1, e lo stesso vale per il numero delle palline di colore Blu, o equivalentemente, dal punto di vista puramente algebrico, per $r=0$ si ha $\left(\frac{0}{5}\right)^2=0$, che, per $r=5$ si ha $\left(\frac{5-5}{5}\right)^1=0$. Di conseguenza, ricordando anche che $\mathbb{P}(\{R=3\})=0$,

$$\begin{aligned}\mathbb{P}(\{S=2\}) &= \mathbb{P}(\{R=1\})\binom{3}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^1 + \mathbb{P}(\{R=2\})\binom{3}{2}\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^1 + \mathbb{P}(\{R=4\})\binom{3}{2}\left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right)^1\\ &= \frac{3}{125}\left(\frac{2}{10}\cdot 4+\frac{1}{20}\cdot 2^2\cdot 3+\frac{5}{10}\cdot 4^2\right) = \frac{1}{125}\left(\frac{2}{10}\cdot 12+\frac{1}{20}\cdot 36+\frac{5}{10}\cdot 48\right)\\ &= \frac{3}{125}\left(\frac{1}{5}\cdot 4+\frac{1}{5}\cdot 3+\frac{5}{5}\cdot 8\right) = \frac{3}{625}(4+3+40) = \frac{3}{625}\cdot 47.\end{aligned}$$

[43] Infatti

$$\begin{aligned}\mathbb{P}(\{R=1\}|\{S=2\}) &= \frac{\mathbb{P}(\{R=1\})\,\mathbb{P}(\{S=2\}|\{R=1\})}{\mathbb{P}(\{S=2\})} = \frac{\frac{3}{625}\cdot 4}{\frac{3}{625}\cdot(4+3+40)} = \frac{4}{47};\\ \mathbb{P}(\{R=2\}|\{S=2\}) &= \frac{\mathbb{P}(\{R=2\})\,\mathbb{P}(\{S=2\}|\{R=1\})}{\mathbb{P}(\{S=2\})} = \frac{\frac{3}{625}\cdot 3}{\frac{3}{625}\cdot(4+3+40)} = \frac{3}{47};\\ \mathbb{P}(\{R=4\}|\{S=2\}) &= \frac{\mathbb{P}(\{R=4\})\,\mathbb{P}(\{S=2\}|\{R=1\})}{\mathbb{P}(\{S=2\})} = \frac{\frac{3}{625}\cdot 40}{\frac{3}{625}\cdot(4+3+40)} = \frac{40}{47};\end{aligned}$$

Ricordiamo che R può prendere valori nell'insieme $\{0, 1, ..., M\}$, ed R segue una distribuzione binomiale, allora è naturale pensare che $R \sim Bin(M, \theta)$ per un qualche valore $0 < \theta < 1$, ossia :

$$p_R(r) = \binom{M}{r} \theta^r (1-\theta)^{M-r}, \qquad r = 0, 1, ..., M.$$

Sostituendo tale espressione nelle (100) e (101) otteniamo quanto segue

$$\begin{aligned}
p_S(s) &= \sum_{r=s}^{M-n+s} \binom{M}{r} \theta^r (1-\theta)^{M-r} \cdot \frac{\binom{r}{s}\binom{M-r}{n-s}}{\binom{M}{n}} \\
&= \sum_{k=0}^{M-n} \binom{M}{s+k} \theta^{s+k} (1-\theta)^{M-(s+k)} \cdot \frac{\binom{s+k}{s}\binom{M-(s+k)}{n-s}}{\binom{M}{n}} \\
&= \theta^s (1-\theta)^{n-s} \sum_{k=0}^{M-n} \binom{M}{s+k} \theta^k (1-\theta)^{M-n-k} \cdot \frac{(k+s)!\, n!\, (M-n)!\, (M-k-s)!}{s!\, k!\, M!\, (n-s)!\, (M-n-k)!} \\
&= \binom{n}{s} \theta^s (1-\theta)^{n-s} \sum_{k=0}^{M-n} \cancel{\binom{M}{s+k}} \binom{M-n}{k} \theta^k (1-\theta)^{M-n-k} \cdot \cancel{\frac{(k+s)!\,(M-k-s)!}{M!}} \\
&= \binom{n}{s} \theta^s (1-\theta)^{n-s} \sum_{k=0}^{M-n} \binom{M-n}{k} \theta^k (1-\theta)^{M-n-k} = \boxed{\binom{n}{s} \theta^s (1-\theta)^{n-s}}
\end{aligned}$$

Abbiamo quindi mostrato che ***anche la distribuzione marginale di S è binomiale***, più precisamente abbiamo che $S \sim Bin(n, \theta)$.

Per quanto riguarda la distribuzione condizionata di R data S, abbiamo

$$\begin{aligned}
p_R(r|S=s) &= \frac{\binom{M}{r} \theta^r (1-\theta)^{M-r} \cdot \binom{r}{s} \binom{M-r}{n-s}}{\binom{n}{s} \theta^s (1-\theta)^{n-s} \cdot \binom{M}{n}} \qquad r = s, s+1, ..., M-(n-s) \\
&= \theta^{r-s} (1-\theta)^{M-n-(r-s)} \frac{M!\, r!\, (M-r)!\, s!\, (n-s)!\, n!\, (M-n)!}{r!\, (M-r)!\, s!\, (r-s)!\, (n-s)!\, [(M-r)-(n-s)]!\, n!\, M!} \\
&= \theta^{r-s} (1-\theta)^{M-n-(r-s)} \frac{(M-n)!}{(r-s)!\, [(M-r)-(n-s)]!}
\end{aligned}$$

Possiamo concludere dunque scrivendo ***la distribuzione condizionata di R dato $\{S = s\}$ è individuata da***

$$p_R(r|S=s) = \binom{M-n}{r-s} \cdot \theta^{r-s} (1-\theta)^{M-n-(r-s)}, \qquad r = s, s+1, ..., M-(n-s). \tag{103}$$

Per quanto riguarda invece $\mathbb{E}(R|S = s)$ osserviamo che ovviamente $R = S + R - S$ e quindi

$$\mathbb{E}(R|S=s) = \mathbb{E}(S+R-S|S=s) = s + \mathbb{E}(R-S|S=s)$$

Consideriamo quindi ora la variabile aleatoria

$$T := R - S,$$

che rappresenta ***il numero di elementi di tipo A fra gli $M - n$ non estratti***.

Ragionando analogamente a quanto si è fatto per S possiamo dedurre che anche T deve avere una distribuzione binomiale. Più precisamente si ottiene che T ha distribuzione $Bin(M - n, \theta)$ ed è indipendente da S. Questo fatto implica che $\mathbb{E}(T|S = s) = \mathbb{E}(T)$ e ci permette di ottenere il valore atteso condizionato di R dato $S = s$:

$$\mathbb{E}(R|S=s) = s + \mathbb{E}(R-S|S=s) = s + \mathbb{E}(T|S=s) = s + \mathbb{E}(T) = s + (M-n)\theta.$$

Iniziamo con la dimostrazione *analitica* che la distribuzione condizionata di $T = R - S$ dato $S = s$ è $Bin(M - n, \theta)$, ma consigliamo di leggere attentamente la dimostrazione euristica/probabilistica della successiva ***Osservazione*** 11.2.
Dalla (103) possiamo poi dedurre[44] ***la distribuzione di probabilità condizionata di*** T***, dato*** $\{S = s\}$,

$$p_T\,(t|S = s) = \mathbb{P}(\{R - S = t\}|\{S = s\}) = \mathbb{P}(\{R = s + t\}|\{S = s\})$$
$$= \binom{M-n}{t} \cdot \theta^t\,(1-\theta)^{M-n-t}, \qquad t = 0, 1, \ldots, M - n.$$

Abbiamo cioè dimostrato che la distribuzione di probabilità condizionata di T, dato $\{S = s\}$, qualunque sia il valore s, è uguale alla distribuzione di probabilità marginale di T e dunque[45] S e T sono variabili aleatorie stocasticamente indipendenti!

Osservazione **11.2 (di carattere euristico).** È utile soffermarsi ad illustrare lo specifico significato intuitivo che è possibile rintracciare, riguardando a posteriori quanto abbiamo qui ottenuto. ***Si tratta di vedere come si sarebbe potuto arrivare alle stesse conclusioni anche sulla base di un ragionamento intuitivo***.
L'assegnazione della distribuzione binomiale $Bin(M, \theta)$ alla variabile R traduce la seguente condizione: ogni elemento nella popolazione ha probabilità θ di essere di tipo A e probabilità $(1 - \theta)$ di essere di tipo B; inoltre ogni elemento si comporta in modo indipendente dagli altri.

Estraiamo ora a caso n elementi dalla popolazione; la circostanza che l'estrazione sia casuale permette di asserire che anche ciascuno degli elementi estratti ha probabilità θ di essere di tipo A e che si comporta in modo indipendente dagli altri.

Questa osservazione ci permette subito di concludere (senza fare troppi calcoli) che la distribuzione di S ***deve essere*** $Bin(n, \theta)$.

Guardiamo ora alla distribuzione condizionata di R data S. Decomponiamo $R = S + T$, come la somma delle due variabili aleatorie S (numero di elementi di tipo A fra gli n estratti) e $T := R - S$ (numero di elementi di tipo A fra gli $M - n$ non estratti). Per quanto sopra osservato circa il significato della posizione $R \sim Bin(M, \theta)$, possiamo vedere intuitivamente che anche T deve avere una distribuzione binomiale $Bin(M - n, \theta)$ e che T, S debbono essere stocasticamente indipendenti; dunque possiamo scrivere

$$p_R(r|S = s) = \mathbb{P}\,(R = r|S = s) = \mathbb{P}(R - S = r - s|S = s)$$
$$= \mathbb{P}(T = r - s|S = s) = \mathbb{P}(T = r - s)$$
$$= \binom{M-n}{r-s} \cdot \theta^{r-s}\,(1-\theta)^{M-n-(r-s)} \qquad r = s, s + 1, \ldots, M - (n - s)$$

e cioè ritrovare la (103).

Osservazione **11.3 (ancora di carattere euristico).** Abbiamo dunque notato che, se R segue una distribuzione binomiale, allora T ed S sono stocasticamente indipendenti. Bisogna tener bene presente che ***in generale ciò non accade, se si attribuisce ad*** R ***un tipo di distribuzione di probabilità diverso dalla binomiale***.

[44] Il fatto che $\mathbb{P}(\{R - S = t\}|\{S = s\}) = \mathbb{P}(\{R = s + t\}|\{S = s\})$ è intuitivo, ma può essere dedotto facilmente ragionando come segue:

$$\mathbb{P}(\{R - S = t\}|\{S = s\}) = \frac{\mathbb{P}(\{R - S = t\} \cap \{S = s\})}{\mathbb{P}(\{S = s\})},$$

inoltre $\{R - S = t\} \cap \{S = s\} = \{R - s = t\} \cap \{S = s\}$, e quindi

$$\mathbb{P}(\{R - S = t\}|\{S = s\}) = \frac{\mathbb{P}(\{R - s = t\} \cap \{S = s\})}{\mathbb{P}(\{S = s\})} = \mathbb{P}(\{R = s + t\}|\{S = s\})$$

[45] Il fatto che la distribuzione di probabilità condizionata di T, dato $\{S = s\}$, sia la stessa qualunque sia il valore s, implica l'indipendenza delle variabili aleatorie T ed S, si veda a questo proposito l'Esercizio proposto 8.3, ed in particolare l'equivalenza tra le condizioni *(i)* e *(iii)*, ivi indicate.

L'indipendenza stocastica fra S e T esprime il fatto che l'osservazione di S non apporta dell'informazione circa il valore di T. Possiamo dunque concludere che ***l'assunzione che la distribuzione marginale di R sia binomiale non è molto realistica nel problema del campionamento***, fin qui illustrato. In altre parole potremo dire che, se si assume che R sia binomiale, allora non è molto utile eseguire un campionamento allo scopo di trarre dell'informazione rilevante, circa il comportamento degli elementi non scrutinati, sulla base del comportamento degli elementi già scrutinati.

Ciò può risultare abbastanza evidente, ad esempio, nel problema del sondaggio elettorale: supponiamo di assumere che ogni elettore (in un gruppo di M elettori) abbia una probabilità fissata, p, di votare per lo schieramento A, indipendentemente dal comportamento degli altri elettori (in tal caso il numero complessivo R di elettori per A ha una distribuzione $Bin(M, p)$). È chiaro intuitivamente che in un tal caso è poco utile eseguire un sondaggio, in quanto la risposta di un elettore non fornisce indicazioni circa il comportamento degli altri.

La situazione più frequente, comunque, è quella in cui non sussiste indipendenza stocastica fra i vari elettori.

A questo punto intervengono aspetti piuttosto delicati circa la condizione di indipendenza stocastica nel caso di estrazioni da una popolazione con caratteristiche non note. Anche se qui non è ne' possibile ne' particolarmente opportuno chiarire completamente tali aspetti, essenzialmente connessi a problematiche di tipo statistico, sarà però utile avviare in proposito un discorso circa il ruolo della nozione di ***indipendenza condizionata***. Ciò sarà l'argomento del prossimo paragrafo.

11.4 Indipendenza condizionata

Cominciamo questo paragrafo insistendo su nozioni che dovrebbero essere ormai chiare, per passare subito dopo a sottolineare aspetti critici, relativi al caso di estrazioni casuali da una popolazione con composizione aleatoria.

Consideriamo allora di nuovo il caso di n estrazioni casuali da una popolazione che contiene oggetti di due tipi, ad esempio A e B.

Poniamo

$$E_j := \{\textit{oggetto di tipo } \mathrm{A} \textit{ alla j-esima estrazione}\}, \qquad j = 1, 2, ..., n;$$

poniamo anche

$$S := \sum_{j=1}^{n} \mathbf{1}_{E_j}$$

Supponiamo di sapere che la popolazione contiene r elementi di tipo A e $(M - r)$ elementi di tipo B. In tale caso, ***sotto la condizione che le estrazioni siano senza reinserimento***, c'è ***dipendenza stocastica*** fra $E_1, ..., E_n$, ed inoltre S ha ***distribuzione condizionata ipergeometrica*** $Hyp(M, r; n)$.

Sotto la condizione che le estrazioni siano con reinserimento, $E_1, ..., E_n$ sono ***eventi stocasticamente indipendenti***, di probabilità $\frac{r}{M}$, e S ha invece ***distribuzione condizionata binomiale*** $Bin(n, \frac{r}{M})$.

Ora dobbiamo sottolineare quanto segue.

A proposito di questo ultimo caso (di estrazioni con reinserimento), si deve fare attenzione al fatto che l'indipendenza stocastica fra $E_1, ..., E_n$ sussiste in virtù della concomitanza fra due diverse circostanze:

(a) le ***estrazioni*** (casuali) sono con reinserimento

(b) ***conosciamo la composizione della popolazione*** da cui si effettua il campionamento (cioè sono note le proporzioni $\frac{r}{M}$ e $\frac{M-r}{M}$ degli elementi di tipo rispettivamente A e B).

Vedremo infatti qui di seguito, iniziando con un semplice esempio, che, fermo restando la condizione **(a)** di estrazioni con reinserimento, ***non vi può essere in generale indipendenza stocastica fra $E_1, ..., E_n$ se viene a mancare la condizione (b)***.

Notiamo d'altra parte che, come abbiamo visto prima, la motivazione per effettuare un campionamento è data proprio dall'esigenza di ricavare informazioni circa la composizione di una popolazione; è dunque realistico pensare che, se si effettua il campionamento, il numero di elementi del tipo A sia una variabile aleatoria R, piuttosto che un valore noto r; ed in tale caso, ripetiamo, pur se le estrazioni casuali sono effettuate con reinserimento, non vi è in generale indipendenza stocastica fra $E_1, ..., E_n$.

Prima di proseguire vediamo infatti il seguente esempio illustrativo a cui avevamo accennato poco fa.

Esempio* 11.5. *Due urne, U_1 e U_2, contengono 10 palline ciascuna: U_1 contiene 3 palline verdi e 7 blu, mentre U_2 ne contiene 3 blu e 7 verdi. Ci viene data a caso una delle due urne (non sappiamo quale e non abbiamo la possibilità di esaminare l'urna) e da tale urna eseguiamo due successive estrazioni (ciò significa che stiamo facendo delle estrazioni da una popolazione in cui il numero delle palline verdi è una variabile aleatoria R che può assumere, con uguale probabilità, il valore 3 oppure il valore 7). Poniamo*

$$E_j := \{\textit{pallina verde alla j-esima estrazione}\}, \qquad j = 1, 2$$

Calcolare $\mathbb{P}(E_1)$, $\mathbb{P}(E_2)$ e $\mathbb{P}(E_2|E_1)$, confrontandoli tra loro.

Soluzione: Allo scopo di calcolare $\mathbb{P}(E_1)$, $\mathbb{P}(E_2)$ e $\mathbb{P}(E_2|E_1)$ osserviamo che vi sono due ipotesi alternative:

$$H_1 := \{\text{abbiamo eseguito le estrazioni da } U_1\}$$
$$H_2 := \{\text{abbiamo eseguito le estrazioni da } U_2\};$$

certamente una di queste due ipotesi è vera, ma non sappiamo quale ed attribuiamo le probabilità (visto che l'urna è stata scelta "a caso")

$$\mathbb{P}(H_1) = \mathbb{P}(H_2) = \frac{1}{2}.$$

Abbiamo intanto, applicando la definizione di probabilità condizionata e poi la formula delle probabilità totali:

$$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)} = \frac{\mathbb{P}(H_1)\,\mathbb{P}(E_1 \cap E_2|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(E_1 \cap E_2|H_2)}{\mathbb{P}(H_1)\,\mathbb{P}(E_1|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(E_1|H_2)}$$

e quindi **in questo caso specifico** in cui $\mathbb{P}(H_i) = \frac{1}{2}$,

$$= \frac{\frac{1}{2}\,\mathbb{P}(E_1 \cap E_2|H_1) + \frac{1}{2}\,\mathbb{P}(E_1 \cap E_2|H_2)}{\frac{1}{2}\,\mathbb{P}(E_1|H_1) + \frac{1}{2}\,\mathbb{P}(E_1|H_2)} = \frac{\mathbb{P}(E_1 \cap E_2|H_1) + \mathbb{P}(E_1 \cap E_2|H_2)}{\mathbb{P}(E_1|H_1) + \mathbb{P}(E_1|H_2)}.$$

Come già accennato vogliamo analizzare specificamente il caso in cui le estrazioni siano casuali e con reinserimento. In tal caso si ha che, sotto l'ipotesi di eseguire le estrazioni da U_1, ossia condizionatamente ad H_1, la probabilità di ottenere una pallina verde in una singola estrazione è uguale a $\frac{3}{10}$ e, visto che le estrazioni avvengono con reinserimento da un'urna contenente 3 palline verdi e 7 blu, la probabilità di ottenere due volte pallina verde in due successive estrazione è uguale a $\left(\frac{3}{10}\right)^2$; possiamo scrivere in formule

$$\mathbb{P}(E_1|H_1) = \mathbb{P}(E_2|H_1) = \frac{3}{10}, \qquad \mathbb{P}(E_1 \cap E_2|H_1) = \mathbb{P}(E_1|H_1)\,\mathbb{P}(E_2|H_1) = \left(\frac{3}{10}\right)^2.$$

Analogamente, per quanto riguarda il condizionamento all'ipotesi H_2, possiamo scrivere

$$\mathbb{P}(E_1|H_2) = \mathbb{P}(E_2|H_2) = \frac{7}{10}, \qquad \mathbb{P}(E_1 \cap E_2|H_2) = \mathbb{P}(E_1|H_2)\,\mathbb{P}(E_2|H_2) = \left(\frac{7}{10}\right)^2.$$

Possiamo concludere quindi

$$\mathbb{P}(E_1) = \mathbb{P}(H_1)\,\mathbb{P}(E_1|H_1) + \mathbb{P}(H_2)\,\mathbb{P}(E_1|H_2) = \frac{1}{2}\left(\frac{3}{10} + \frac{7}{10}\right) = \frac{1}{2}$$

e ovviamente

$$\mathbb{P}\left(E_2\right)\Big(=\mathbb{P}(H_1)\,\mathbb{P}\left(E_2|H_1\right)+\mathbb{P}(H_2)\,\mathbb{P}\left(E_2|H_2\right)\Big)=\mathbb{P}\left(E_1\right)=\frac{1}{2},$$

mentre

$$\mathbb{P}(E_2|E_1)=\frac{\left(\frac{3}{10}\right)^2+\left(\frac{7}{10}\right)^2}{\frac{3}{10}+\frac{7}{10}}=\frac{58}{100}.$$

Dunque

$$\mathbb{P}(E_2|E_1)>\mathbb{P}\left(E_2\right)$$

da cui vediamo che $E_1,\ E_2$ ***non sono stocasticamente indipendenti***, bensì ***positivamente correlati***. □

Esercizio proposto 11.1. *Calcolare la probabilità condizionata $\mathbb{P}(E_1|E_2)$ nel caso di estrazioni casuali dall'urna considerata nel precedente Esempio 11.5.*

Esercizio proposto 11.2. *Calcolare la probabilità condizionata $\mathbb{P}(E_2|E_1)$ e $\mathbb{P}(E_1|E_2)$ nel caso di estrazioni casuali dall'urna considerata nel precedente Esempio 11.3, in cui R è aleatorio.*

Estrazioni con reinserimento da una popolazione con composizione incognita:

Generalizziamo ora quanto visto nell'Esempio 11.5 e nei precedenti Esercizi proposti 11.1 e 11.2.

Una popolazione contiene M oggetti, alcuni di tipo A ed altri di tipo B. Supponiamo che il numero complessivo di quelli di tipo A sia una variabile aleatoria R, con distribuzione di probabilità data da

$$p_R(0)=\mathbb{P}(\{R=0\}),\quad p_R(1)=\mathbb{P}(\{R=1\}),\ \dots\ ,\ p_R(M)=\mathbb{P}(\{R=M\})$$

con $p_R(r)\geq 0$ e $\sum_{r=0}^M p_R(r)=1$. Eseguiamo delle estrazioni casuali con reinserimento dalla popolazione e poniamo

$$E_j:=\{\textit{oggetto di tipo } \mathrm{A} \textit{ alla } j\textit{-esima estrazione}\},\qquad j=1,2,\dots$$

Vogliamo calcolare $\mathbb{P}(E_1)$, $\mathbb{P}(E_2)$, $\mathbb{P}(E_2|E_1)$ e $\mathbb{P}(E_1|E_2)$ e mostrare che $\mathbb{P}(E_1)=\mathbb{P}(E_2)$ e $\mathbb{P}(E_2|E_1)=\mathbb{P}(E_1|E_2)$ Inoltre E_1 ed E_2 sono correlati positivamente, tranne nel caso in cui R è una variabile degenere concentrata su un unico valore noto $\hat{r}$, ossia quando la composizione dell'urna è nota, e in tale caso i due eventi sono indipendenti.

Estendendo quanto svolto nel precedente Esempio 11.5, abbiamo

$$\mathbb{P}\left(E_1\right)=\mathbb{P}\left(E_2\right)=\sum_{r=0}^M\mathbb{P}\left(E_i|R=r\right)\cdot p_M(r)=\sum_{r=0}^M\frac{r}{M}\cdot p_M(r)=\frac{\mathbb{E}(R)}{M}$$

$$\begin{aligned}\mathbb{P}\left(E_1\cap E_2\right)&=\sum_{r=0}^M\mathbb{P}\left(E_1\cap E_2|R=r\right)\cdot p_R(r)=\sum_{r=0}^M\underbrace{\mathbb{P}\left(E_1|R=r\right)}_{=\frac{r}{M}}\cdot\underbrace{\mathbb{P}\left(E_2|R=r\right)}_{=\frac{r}{M}}\cdot p_R(r)\\&=\sum_{r=0}^M\left(\frac{r}{M}\right)^2\cdot p_R(r)=\frac{\mathbb{E}(R^2)}{M^2}\end{aligned}$$

da cui

$$\mathbb{P}(E_2|E_1)=\frac{\mathbb{P}\left(E_1\cap E_2\right)}{\mathbb{P}(E_1)}=\frac{\mathbb{P}\left(E_1\cap E_2\right)}{\mathbb{P}(E_2)}=\mathbb{P}(E_1|E_2)=\frac{1}{M}\,\frac{\mathbb{E}(R^2)}{\mathbb{E}(R)}.$$

È facile verificare che E_1 ed E_2 non possono essere correlati negativamente e sono stocasticamente indipendenti se e solo se R è una variabile degenere[46], ossia se e solo se esiste un $\hat{r} \in \{0, 1, \ldots, M\}$ tale che $R(\omega_i) = \hat{r}$ per ogni $\omega_i \in \Omega$, nel qual caso $Var(R) = 0$ e, quindi, ricordando la ***Proposizione* 10.1**, si ha $\mathbb{E}(R^2) = (\mathbb{E}(R))^2 = \hat{r}^2$.

Quanto svolto fin qui suggerisce le seguenti definizioni

Definizione 11.1 (indipendenza condizionata rispetto ad un evento e rispetto ad una partizione)**.** *Siano E_1, E_2, H tre eventi. Diremo che E_1 ed E_2* **sono condizionatamente indipendenti dato l'evento** *H se risulta*

$$\mathbb{P}(E_1 \cap E_2 | H) = \mathbb{P}(E_1|H) \cdot \mathbb{P}(E_2|H).$$

Sia ora $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ una partizione dell'evento certo. Diremo che E_1 ed E_2 **sono condizionatamente indipendenti data la partizione** *$\mathcal{H}$ se risulta*

$$\mathbb{P}(E_1 \cap E_2 | H_k) = \mathbb{P}(E_1|H_k) \cdot \mathbb{P}(E_2|H_k), \qquad \textit{per ogni } k = 1, 2, \ldots, m.$$

Definizione 11.2 (indipendenza condizionata rispetto ad una variabile aleatoria)**.** *Siano E_1, E_2 due eventi e sia Z una variabile aleatoria con $Z(\Omega) = \{z_1, z_2, ..., z_\ell\}$. Diremo che E_1 ed E_2 sono condizionatamente indipendenti data Z se risulta*

$$\mathbb{P}(E_1 \cap E_2 | \{Z = z\}) = \mathbb{P}(E_1|\{Z = z\}) \cdot \mathbb{P}(E_2|\{Z = z\}), \qquad \textit{per ogni } z \in Z(\Omega) = \{z_1, z_2, \ldots, z_\ell\}.$$

Si osservi che la precedente definizione equivale all'indipendenza condizionata di E_1 ed E_2 data la partizione $\mathcal{H} = \Big\{\{Z = z_k\},\ k = 1, 2, \ldots, \ell\Big\}$.

Definizione 11.3 (indipendenza condizionata di due v.a. rispetto ad una v.a.)**.** *Siano X, Y, Z tre variabili aleatorie. Diremo che X ed Y* **sono condizionatamente indipendenti data** *Z se risulta*

$$\mathbb{P}\left(\{X = x\} \cap \{Y = y\}\right) | \{Z = z\}\} = \mathbb{P}(X = x | Z = z) \cdot \mathbb{P}(Y = y | Z = z),$$

per ogni $x \in X(\Omega) = \{x_i,\ i = 1, 2, \ldots, n\}$, per ogni $y \in Y(\Omega) = \{y_j,\ j = 1, 2, \ldots, m\}$, per ogni $z \in Z(\Omega) = \{z_k,\ k = 1, 2, \ldots, \ell\}$.

Coppie di eventi condizionatamente indipendenti e coppie di variabili aleatorie condizionatamente indipendenti si incontrano comunemente in varie problematiche, in particolare nelle situazioni di tipo statistico. Ciò accade anche in situazioni al di fuori degli schemi di estrazioni casuali (con reinserimento) da popolazioni con composizione aleatoria. Guardiamo in proposito il seguente esempio:

Esempio* 11.6. *Vi sono a disposizione tre diversi canali di comunicazione, C_1, C_2 e C_3, per spedire dei messaggi. Ogni messaggio può essere spedito da ciascuno dei tre canali. La probabilità di trasmettere il messaggio correttamente tramite C_1 è uguale a $p^{(1)} = 0.9$. Le analoghe probabilità per C_2 e C_3 sono rispettivamente date da $p^{(2)} = 0.6$ e $p^{(3)} = 0.3$. Supponiamo ora che il canale venga scelto a caso da un meccanismo e* **non sia noto** *all'operatore. Questi spedisce il messaggio due volte consecutive (sempre sullo stesso canale, che gli è stato riservato per quel messaggio) al fine di aumentare l'affidabilità della trasmissione.*
(a) Trovare la probabilità che il messaggio sia trasmesso correttamente in almeno una delle due volte.
(b) Si supponga di sapere in seguito che il messaggio è stato trasmesso correttamente la seconda volta, ma non la prima volta. Condizionatamente a questa osservazione, come bisogna valutare le probabilità che sia stato utilizzato il canale C_1, C_2 e C_3, rispettivamente?

---

[46]Infatti

$$\mathbb{P}(E_2|E_1) \geq \mathbb{P}(E_2) \quad \Leftrightarrow \quad \frac{1}{M}\frac{\mathbb{E}(R^2)}{\mathbb{E}(R)} \geq \frac{1}{M}\mathbb{E}(R) \quad \Leftrightarrow \quad Var(R) = \mathbb{E}(R^2) - (\mathbb{E}(R))^2 \geq 0,$$

e inoltre $\mathbb{P}(E_2) = \mathbb{P}(E_2|E_1)$ se e solo se $Var(R) = 0$.

Soluzione: Poniamo

$$\begin{aligned} &E_i := \{\textit{messaggio trasmesso correttamente nell'i-esimo tentativo}\}, \qquad && i = 1,2 \\ &H_j := \{\textit{è stato assegnato il canale } C_j\}, && j = 1,2,3. \end{aligned}$$

(a) Dobbiamo calcolare $\mathbb{P}(E_1 \cup E_2)$. Applicando la formula delle probabilità totali possiamo scrivere

$$\mathbb{P}(E_1 \cup E_2) = \sum_{j=1}^{3} \mathbb{P}(E_1 \cup E_2 | H_j) \cdot \mathbb{P}(H_j)$$

Visto che il canale si assume scelto a caso possiamo porre

$$\mathbb{P}(H_j) = \frac{1}{3}, \qquad j = 1,2,3.$$

Ora possiamo osservare che gli eventi $E_1,\ E_2$ ***non sono indipendenti*** bensì ***sono condizionatamente indipendenti***, dati gli eventi della partizione $\{H_1,\, H_2,\, H_3\}$; quindi per la formula di inclusione ed esclusione, applicata alla probabilità $\mathbb{P}(\cdot|H_j)$,

$$\mathbb{P}(E_1 \cup E_2 | H_j) = \mathbb{P}(E_1|H_j) + \mathbb{P}(E_2|H_j) - \mathbb{P}(E_1 \cap E_2|H_j) =$$

$$= \mathbb{P}(E_1|H_j) + \mathbb{P}(E_2|H_j) - \mathbb{P}(E_1|H_j) \cdot \mathbb{P}(E_2|H_j) = p^{(j)} \left(2 - p^{(j)}\right).$$

Dunque

$$\mathbb{P}(E_1 \cup E_2) = \frac{1}{3} \left[0.9 \times 1.1 + 0.6 \times 1.4 + 0.3 \times 1.7\right]$$

(b) Dobbiamo calcolare $\mathbb{P}(H_j|\overline{E}_1 \cap E_2)$. Si ha, ancora in virtù della condizione di indipendenza condizionata

$$\begin{aligned} \mathbb{P}(\overline{E}_1 \cap E_2) &= \sum_{j=1}^{3} \mathbb{P}(\overline{E}_1 \cap E_2|H_j) \cdot \mathbb{P}(H_j) = \sum_{j=1}^{3} \mathbb{P}(\overline{E}_1|H_j) \cdot \mathbb{P}(E_2|H_j)\, \mathbb{P}(H_j) \\ &= (0.1 \times 0.91 + 0.4 \times 0.6 + 0.3 \times 0.7)\, \frac{1}{3} = \frac{1}{3}\, 0.54 \ = 0.18; \end{aligned}$$

utilizzando la Formula di Bayes abbiamo dunque

$$\mathbb{P}(H_j|\overline{E}_1 \cap E_2) = \frac{\mathbb{P}(H_j)\, \mathbb{P}(\overline{E}_1 \cap E_2|H_j)}{\mathbb{P}(\overline{E}_1 \cap E_2)} = \frac{\frac{1}{3}\, \mathbb{P}(\overline{E}_1|H_j) \cdot \mathbb{P}(E_2|H_j)}{\mathbb{P}(\overline{E}_1 \cap E_2)} = \frac{\left(1 - p^{(j)}\right) p^{(j)}}{0.54}.$$

e quindi

$$\mathbb{P}(H_1|\overline{E}_1 \cap E_2) = \frac{9}{54} = \frac{3}{18}, \qquad \mathbb{P}(H_2|\overline{E}_1 \cap E_2) = \frac{24}{54} = \frac{8}{18}, \qquad \mathbb{P}(H_3|\overline{E}_1 \cap E_2) = \frac{21}{54} = \frac{7}{18}.$$

Una proprietà fondamentale della nozione di indipendenza condizionata è mostrata dalla seguente ***Proposizione* 11.1**; la dimostrazione è lasciata al lettore per esercizio[47].

***Proposizione* 11.1.** *Siano E_1, E_2 due eventi condizionatamente indipendenti data una variabile aleatoria Z con $Z(\Omega) = \{z_1, z_2, ..., z_\ell\}$. Allora*

$$\mathbb{P}(E_2|E_1 \cap \{Z = z_j\}) = \mathbb{P}(E_2|\{Z = z_j\})$$

$$\mathbb{P}(E_2|E_1) = \sum_{j=1}^{\ell} \mathbb{P}(E_2|\{Z = z_j\})\, \mathbb{P}(\{Z = z_j\}|E_1).$$

11.5 Esercizi di verifica

Esercizio 11.1. Due urne, U_1 e U_2, contengono 10 palline ciascuna. L'urna U_1 contiene 10 palline blu, mentre l'urna U_2 ne contiene 5 blu e 5 verdi. Viene scelta a caso una delle due urne (non sappiamo quale) e da tale urna eseguiamo due successive estrazioni casuali con reinserimento, ottenendo ciascuna volta pallina blu. Condizionatamente a tale evento, qual è la probabilità che sia stata scelta l'urna U_1?

Esercizio 11.2. Una pianta produce R semi, dove R è una variabile aleatoria binomiale con parametri n e p. Supponiamo inoltre che ciascun seme, fra gli R prodotti, germogli con probabilità θ, indipendentemente dagli altri. Sia S il numero dei germogli risultanti.
(a) Calcolare $\mathbb{P}(S = j|R = i)$ e $\mathbb{P}(S = j, R = i)$.
(b) Calcolare $\mathbb{P}(S = j)$.
(c) Calcolare $\mathbb{P}(R = i|S = j)$.
(d) Calcolare $\mathbb{E}(R|S = j)$.

Esercizio 11.3. Il testo di esame scritto consiste di quattro esercizi. Ogni esercizio può contenere un errore con probabilità 0.1, indipendentemente dagli altri. Supponiamo che, dopo aver redatto il testo, ciascun esercizio venga ricontrollato e che la presenza di un eventuale errore sia rilevata con probabilità 0.8. Gli errori inizialmente presenti e poi rilevati vengono corretti.

---

[47] Come suggerimento per la dimostrazione diamo i seguenti elementi:

$$\mathbb{P}(E_2|E_1 \cap \{Z = z_j\}) = \frac{\mathbb{P}(E_2 \cap E_1 \cap \{Z = z_j\})}{\mathbb{P}(E_1 \cap \{Z = z_j\})} = \frac{\mathbb{P}(E_2 \cap E_1|\{Z = z_j\})\, \mathbb{P}(Z = z_j)}{\mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(Z = z_j)}$$

per l'ipotesi di indipendenza condizionata

$$= \frac{\mathbb{P}(E_2|\{Z = z_j\})\, \mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(Z = z_j)}{\mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(Z = z_j)}.$$

Inoltre

$$\begin{aligned}\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_1)} &= \frac{\sum_{j=1}^{n} \mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(E_2|\{Z = z_j\})\, \mathbb{P}(\{Z = z_j\})}{\sum_{i=1}^{n} \mathbb{P}(E_1|\{Z = z_i\})\, \mathbb{P}(\{Z = z_i\})} \\ &= \sum_{j=1}^{n} \mathbb{P}(E_2|\{Z = z_j\})\, \frac{\mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(\{Z = z_j\})}{\sum_{i=1}^{n} \mathbb{P}(E_1|\{Z = z_i\})\, \mathbb{P}(\{Z = z_i\})},\end{aligned}$$

e infine

$$\mathbb{P}(\{Z = z_j\}|E_1) = \frac{\mathbb{P}(E_1|\{Z = z_j\})\, \mathbb{P}(\{Z = z_j\})}{\sum_{i=1}^{n} \mathbb{P}(E_1|\{Z = z_i\})\, \mathbb{P}(\{Z = z_i\})}.$$

(a) Determinare la probabilità che, dopo il controllo, non vi siano più esercizi contenenti errori
(b) Condizionatamente al fatto che non vi siano più esercizi contenenti errori dopo il controllo, qual è la probabilità che non vi fossero errori neanche prima del controllo?

Esercizio 11.4. Indichiamo con R il numero di pezzi difettosi un lotto di 20 pezzi e facciamo la seguente valutazione di probabilità:

$$\mathbb{P}(R = 0) = 0.5, \quad \mathbb{P}(R = 10) = 0.4, \quad \mathbb{P}(R = 20) = 0.1.$$

Qual è la probabilità di avere entrambi i pezzi difettosi, scegliendo a caso (senza reinserimento) due pezzi dal lotto?

Esercizio 11.5. Si hanno m esemplari di un certo tipo di telecomando (TC) per televisore; ciascun TC ha bisogno di due batterie per il suo funzionamento. Si hanno a disposizione $2m$ batterie, di cui alcune possono essere scariche. Da tale gruppo di batterie vengono costituite in modo casuale m coppie, che vengono inserite negli m TC.
Indichiamo con R il numero complessivo delle batterie cariche; supponiamo che ciascuna batteria sia carica con probabilità $\frac{1}{2}$ ed indipendentemente da quello che accade alle altre.
(a) Calcolare la probabilità che un fissato TC abbia entrambe le batterie cariche.
(b) Sia S il numero complessivo dei TC con entrambe le batterie cariche. Calcolare $\mathbb{E}(S)$.

Esercizio 11.6. Abbiamo un dado di cui non siamo certi se è regolare oppure se è truccato. Truccato qui significa che esso fornisce il risultato “sei” con probabilità doppia rispetto a quella di tutti gli altri risultati. Attribuiamo probabilità 0.9 all'ipotesi che esso sia regolare e probabilità 0.1 all'ipotesi che invece sia truccato. Viene eseguito un lancio del dado e si ottiene il risultato “sei”.
(a) Condizionatamente a questo risultato, qual è la probabilità che il dado sia truccato?
(b) Condizionatamente a questo risultato, qual è la probabilità che si ottenga il risultato “sei” anche in un secondo lancio?

Esercizio 11.7. Una popolazione, composta da 100 elementi, contiene R elementi di tipo A e $100 - R$ elementi di tipo B, R essendo una variabile con una distribuzione binomiale $Bin\left(100, \frac{3}{4}\right)$.
Da tale popolazione si eseguono cinquanta estrazioni casuali senza reinserimento ed indichiamo con S il numero di elementi di tipo A fra quelli estratti.
(a) Trovare la distribuzione di probabilità di S
(b) Trovare la distribuzione di probabilità condizionata di R dato $\{S = 40\}$
(c) Calcolare il valore atteso condizionato di R dato $\{S = 40\}$
(d) Calcolare la varianza della distribuzione condizionata di R dato $\{S = 40\}$ e, pensando ad R come la somma di $S + (R - S)$ calcolare la minorazione fornita dalla diseguaglianza di Chebychev per $\mathbb{P}(\{45 \leq R \leq 65\}|\{S = 40\})$

Esercizio 11.8. Una popolazione, composta da 9 elementi, contiene R elementi di tipo A e $9 - R$ elementi di tipo B, R essendo una variabile con una distribuzione uniforme su $\{0, 1, 2, ..., 9\}$:

$$\mathbb{P}(\{R = r\}) = \frac{1}{10}, \quad r = 0, 1, ..., 9.$$

Da tale popolazione si eseguono quattro estrazioni casuali senza reinserimento ed indichiamo con S il numero di elementi di tipo A fra i quattro estratti.
(a) Trovare la distribuzione di probabilità di S
(b) Trovare la distribuzione di probabilità condizionata di R dato $\{S = 1\}$
(c) Calcolare il valore atteso condizionato di R dato $\{S = 1\}$.

12 Approfondimenti di Calcolo Combinatorio

In questa lezione introdurremo, inizialmente, ulteriori nozioni di calcolo combinatorio, quali *combinazioni con ripetizione* e *coefficienti multinomiali*, anche esse fondamentali nel Calcolo delle Probabilità.

12.1 Combinazioni con ripetizione e coefficienti multinomiali

Prima di affrontare la teoria, esaminiamo due esempi.

Esempio 12.1. *In un bar ci sono $n = 3$ tipi di paste, alla panna, al cioccolato e allo zabaione. Un cliente vuole acquistare $r = 4$ paste. Ci poniamo il problema di quali (e quanti) siano tutti i possibili vassoi diversi da 4 paste che un cliente può richiedere.*
Chiaramente per ciascun tipo di pasta, il cliente può richiederne da 0 fino ad un massimo di 4 (siamo anche implicitamente assumendo che per ogni tipo di pasta ci siano almeno 4 paste a disposizione nel bar). Inoltre l'ordine in cui le paste vengono ordinate non ha impostanza. Indicando con i simboli 1, 2 e 3 i tre tipi di pasta (ad esempio panna=1, cioccolato=2 e zabaione=3, rispettivamente), possiamo elencare tutti i possibili vassoi nel seguente modo:

$$\{1,1,1,1\},\ \{1,1,1,2\},\ \{1,1,1,3\},\ \{1,1,2,2\},\ \{1,1,2,3\},$$
$$\{1,1,3,3\},\ \{1,2,2,2\},\ \{1,2,2,3\},\ \{1,2,3,3\},\ \{1,3,3,3\},$$
$$\{2,2,2,2\},\ \{2,2,2,3\},\ \{2,2,3,3\},\ \{2,3,3,3\},\ \{3,3,3,3\},$$

*dove abbiamo usato le parentesi graffe, **non** per indicare un insieme, ma per indicare che l'ordine degli elementi non ha importanza, ossia, ad esempio, $\{1,1,1,2\}$ è uguale a $\{1,1,2,1\}$ e anche a $\{1,2,1,1\}$ o a $\{2,1,1,1\}$ e indicano il vassoio con tre paste alla panna (ossia del tipo 1) e una al cioccolato (ossia del tipo 2). Quello appena visto è un primo esempio di combinazioni con ripetizione di $n = 3$ elementi di classe $r = 4$; daremo la definizione precisa e generale più avanti (vedere la Definizione 12.1).*
Un modo alternativo, ma equivalente di elencare tutti i possibili vassoi è associare a ciascun vassoio una terna di numeri (k_1, k_2, k_3) dove k_i indica il numero delle paste di tipo i, per $i = 1,2,3$. Ovviamente $k_i \geq 0$ e $k_1+k_2+k_3 = 4$. Ad esempio i precedenti 15 possibili vassoi, possono essere rappresentati dalle terne

$$(4,0,0),\ (3,1,0),\ (3,0,1),\ (2,2,0),\ (2,1,1),$$
$$(2,0,2),\ (1,3,0),\ (1,2,1),\ (1,1,2),\ (1,0,3),$$
$$(0,4,0),\ (0,3,1),\ (0,2,2),\ (0,1,3),\ (0,0,4).$$

In questo elenco invece l'ordine è importante perché ad esempio $(2,1,1)$ rappresenta il vassoio con due paste alla panna, una al cioccolato e una allo zabaione, mentre $(1,1,2)$ rappresenta il vassoio con una pasta alla panna, una al cioccolato e due allo zabaione.

Esempio 12.2. *Quali (e quanti) sono i monomi di 4^o grado che si possono formare con le variabili a, b e c?*
Sono tutti i prodotti di quaterne non ordinate contenenti i simboli a, b, c, che possono essere ripetuti più volte. L'ordine non ha importanza, in quanto, conta solo il risultato: ad esempio $a \cdot a \cdot a \cdot b = a \cdot a \cdot b \cdot a = a \cdot b \cdot a \cdot a = b \cdot a \cdot a \cdot a = a^3 \cdot b^1 = a^3 \cdot b^1 \cdot c^0$.
Tutti i monomi sono:

$$a^4 = a^4b^0c^0,\ a^3b = a^3b^1c^0,\ a^3c = a^3b^0c^1,\ a^2b^2 = a^2b^2c^0,\ a^2b\,c = a^2b^1c^1,$$
$$a^2b^2 = a^2b^2c^0,\ a\,b^3 = a^1b^3c^0,\ a\,b^2c = a^1b^2c^1,\ a\,b\,c^2 = a^1b^1c^2,\ a\,c^3 = a^1b^0c^3,$$
$$b^4 = a^0b^4c^0,\ b^3c = a^0b^3c^1,\ b^2c^2 = a^0b^2c^2,\ b\,c^3 = a^0b^1c^3,\ c^4 = a^0b^0c^4.$$

È evidente che sostanzialmente si tratta dello stesso problema affrontato nell'esempio precedente, in particolare è evidente che vale la rappresentazione alternativa ivi discussa.

Passiamo ora a generalizzare i due esempi esaminati, tramite una definizione astratta.

Definizione 12.1 (Combinazioni con ripetizione). *Sia dato un insieme con n elementi. Una* ***combinazione con ripetizione di n elementi di classe r****, è un raggruppamento contenente r oggetti, ciascuno dei quali è un elemento dell'insieme di partenza, e che può essere ripetuto più volte; è fondamentale precisare che non ha importanza l'ordine in cui gli r oggetti sono disposti nel raggruppamento.*

Come abbiamo visto negli esempi precedenti, per le combinazioni con ripetizione è possibile prendere $r > n$, a differenza delle combinazioni usuali (senza ripetizione). Quest'ultime vengono chiamate anche *combinazioni semplici*, specialmente se in contrapposizione con le combinazione con ripetizione. Inoltre, generalizzando quanto visto negli esempi precedenti, è facile convincersi che una combinazione con ripetizione è individuata in maniera biunivoca da una n-pla ordinata di numeri $(k_1, k_2, ..., k_n)$, dove k_j rappresenta il numero di volte che l'elemento j compare nella combinazione stessa; in particolare $k_j = 0$ significa che l'elemento j non compare mai nella combinazione con ripetizione. È chiaro che $(k_1, k_2, ..., k_n)$ deve soddisfare la condizione che

$$k_j \geq 0, \text{ e interi, per ogni } j = 1, 2, ..., n, \quad \text{e} \quad \sum_{j=1}^{n} k_j = r. \tag{104}$$

Nel seguito indicheremo con il simbolo $A_{n,r}$ l'insieme di tutte le n-ple di numeri interi che soddisfano la precedente condizione (104):

$$A_{n,r} = \big\{\mathbf{k} := (k_1, ..., k_n) : \sum_{j=1}^{n} k_j = r,\ k_j \in \mathbb{N} \cup \{0\},\ j = 1, 2, \ldots, n\big\}. \tag{105}$$

Ad esempio, per $n = 3$ ed $r = 4$, ritroviamo, come negli Esempi 12.1 e 12.2,

$$\begin{aligned} A_{3,4} = \Big\{&(4,0,0),\ (0,4,0),\ (0,0,4),\\ &(3,1,0),\ (3,0,1),\ (1,3,0),\ (1,0,3),\ (0,3,1),\ (0,1,3),\\ &(2,1,1),\ (1,2,1),\ (1,1,2),\\ &(2,2,0),\ (2,0,2),\ (0,2,2)\Big\}\end{aligned}$$

Le n-ple di $A_{n,r}$ vengono anche dette ***vettori (o n-ple) di numeri di occupazione***. Il significato di questo nome verrà chiarito nella successivo paragrafo e nella successiva Lezione 13.

Ci poniamo ora la seguente domanda:

Quanti sono vettori di numeri di occupazione, ovvero, quanti sono gli elementi di $A_{n,r}$?

***Osservazione* 12.1.** Per quanto osservato in precedenza l'insieme $A_{n,r}$ è in corrispondenza biunivoca con l'insieme delle combinazioni con ripetizione di n elementi di classe r, e quindi la risposta fornisce equivalentemente il numero di tali combinazioni con ripetizione.

La risposta alla precedente domanda è : ***L'insieme $A_{n,r}$ ha cardinalità $\binom{n+r-1}{n-1}$.***

Per convincersene osserviamo che un generico elemento di $A_{n,r}$ può essere ad esempio rappresentato attraverso il seguente tipo di "disegno"

$$\underset{1^0\, sito}{\overbrace{* *\ldots * *}^{k_1}} \Big| \underset{2^0\, sito}{\overbrace{* *\ldots * *}^{k_2}} \Big| \ldots\ldots \Big| \underset{n^{simo}\, sito}{\overbrace{* *\ldots * *}^{k_n}}$$

Rappresentazione grafica di un elemento di $A_{n,r}$.

dove, guardando da sinistra verso destra, k_1 è il numero dei simboli $*$ a sinistra della prima barretta $|$, k_2 è il numero dei simboli $*$ compresi fra la prima e la seconda barretta ... e così via; k_n è il numero dei simboli $*$ a destra dell'ultima barretta (se $k_i = 0$, la $(i-1)$-esima e la i-esima barretta sono contigue, senza simboli $*$ in mezzo).

Notiamo d'altra parte che ad ognuno di tali disegni corrisponde un'unica n-upla $\mathbf{k} \in A_{n,r}$. Possiamo stabilire dunque una corrispondenza biunivoca fra l'insieme costituito dai disegni stessi e l'insieme $A_{n,r}$.

Ad esempio per $r = 4$ ed $n = 3$, il disegno $* * \ | \ * * \,|$ o più esplicitamente,

$$\underset{1^0\, sito}{\overbrace{\ * *\ }^{2}} \Big| \underset{2^0\, sito}{\overbrace{\ * *\ }^{2}} \Big| \underset{3^0\, sito}{\overbrace{\qquad}^{0}}$$

corrisponde alla tripletta dei numeri di occupazione $(2, 2, 0)$.

Analogamente, la tripletta dei numeri di occupazione $(1, 0, 3)$ corrisponde al disegno $* \ | \quad | * **$, o più esplicitamente a

$$\underset{1^0\, sito}{\overbrace{\quad * \quad}^{1}} \Big| \underset{2^0\, sito}{\overbrace{\qquad}^{0}} \Big| \underset{3^0\, sito}{\overbrace{\ * **\ }^{3}}$$

La cardinalità dell'insieme $A_{n,r}$ è quindi la stessa dell'insieme di tutti i disegni nella precedente rappresentazione grafica, con barrette e asterischi.

Ora, ciascun disegno contiene in totale $r + n - 1$ simboli, di cui r simboli sono asterischi $*$ ed $n - 1$ simboli sono barrette $|$. Inoltre ciascun disegno corrisponde ad un modo di disporre le $n - 1$ barrette sul totale degli $r + n - 1$ posti. In altre parole si tratta di scegliere $n - 1$ posti tra gli $r + n - 1$ posti disponibili, ossia, dato un insieme di cardinalità $r + n - 1$ si tratta di scegliere un sottoinsieme di cardinalità $n - 1$.

Dunque vi sono $\binom{n+r-1}{n-1}$ diversi possibili disegni e tale è anche la cardinalità dell'insieme $A_{n,r}$.

12.2 Relazioni tra disposizioni con ripetizione e combinazioni con ripetizione

Nella Lezione 3, introducendo i primissimi elementi di calcolo combinatorio, si erano inizialmente definite le nozioni di disposizioni con ripetizione e disposizioni senza ripetizione.

A partire dalle disposizioni senza ripetizione erano state quindi introdotte le combinazioni (semplici); queste ultime possono essere viste come classi di equivalenza, ove si considerino equivalenti due diverse disposizioni senza ripetizione purché costituite dagli stessi elementi, seppure disposti in ordine diverso. Con una procedura analoga vogliamo adesso definire gli elementi di $A_{n,r}$ e quindi in un certo senso, definire le combinazioni con ripetizione come classi di equivalenza a partire da disposizioni con ripetizione.

ESEMPIO BASE

Consideriamo come nell'Esempio 12.1 i vari modi in cui un cliente potrebbe chiedere 4 paste di cui tre con la panna e una al cioccolato: ad esempio potrebbe chiedere nell'ordine

Mi dia una pasta con la panna, e poi ne metta una la cioccolato, adessa un'altra con la panna e un'altra ancora con la panna.

Con le stesse convenzioni per cui $1 =$ panna, $2 =$ cioccolato e $3 =$ zabaione, il cliente avrebbe chiesto, tenendo conto dell'ordine $(1, 2, 1, 1)$. Tenendo conto dell'ordine i vari modi sono

$$\begin{matrix} (1,1,1,2) \\ (1,1,2,1) \\ (1,2,1,1) \\ (2,1,1,1) \end{matrix} \longleftrightarrow \{1,1,1,2\} \longleftrightarrow (2,1,0)$$

dove $\{1, 1, 1, 2\}$ è la corrispondente combinazione con ripetizione e $(3, 1, 0)$ è il corrispondente vettore dei numeri di occupazione.

Analogamente

$$\begin{matrix} (1,1,3,2) & (1,1,2,3) \\ (1,2,1,3) & (1,3,1,2) \\ (1,2,3,1) & (1,3,2,1) \\ (2,1,1,3) & (3,1,1,2) \\ (2,1,3,1) & (3,1,2,1) \\ (2,3,1,1) & (3,2,1,1) \end{matrix} \longleftrightarrow \{1,1,2,3\} \longleftrightarrow (2,1,1)$$

dove $\{1, 1, 2, 3\}$ è la corrispondente combinazione con ripetizione e $(2, 1, 1)$ è il corrispondente vettore dei numeri di occupazione.

Osserviamo che le due classi di equivalenza non hanno gli stessi numeri di elementi. Uno dei problemi al quale cercheremo di rispondere consiste nel cercare il numero delle disposizioni con ripetizione che corrispondo a una data combinazione con ripetizione o equivalentemente a un dato vettore di numeri di occupazione.

Più in generale, consideriamo un insieme costituito da n elementi fra loro distiguibili, che possiamo identificare con i primi n numeri naturali $1, 2, ..., n$.

Una disposizione con ripetizione di n elementi di classe r, è una r-upla ordinata $\mathbf{z} := (z_1, ..., z_r)$, con $z_h \in \{1, 2, ..., n\}$ per ogni $h = 1, 2, ..., r$, e può anche essere vista come un'applicazione

$$\mathbf{z} : \{1, 2, ..., r\} \to \{1, 2, ..., n\}.$$

L'insieme delle $\mathbf{z}$ coincide dunque con il prodotto cartesiano $\{1, 2, ..., n\}^r$ che, come sappiamo, ha cardinalità n^r.

Ad ogni disposizione con ripetizione $\mathbf{z} := (z_1, ..., z_r)$ associamo ora una n-upla $\mathbf{k} := (k_1, ..., k_n)$ dove, per $j = 1, 2, ..., n$, poniamo

$$k_j := \text{numero degli indici } h \text{ tali che } z_h = j. \tag{106}$$

Esempio 12.3. *Poniamo $r = 8$ e $n = 6$ e consideriamo la disposizione con ripetizione*

$$\mathbf{z} := (3, 4, 6, 1, 4, 6, 3, 6)\,.$$

A questa disposizione con ripetizione, associamo la combinazione con ripetizione

$$\{3, 4, 6, 1, 4, 6, 3, 6\} = \{1, 3, 3, 4, 4, 6, 6, 6\}$$

e a sua volta l'elemento di $A_{6,8}$

$$\mathbf{k} := (1, 0, 2, 2, 0, 3)\,,$$

cioè $k_1 = 1$, in quanto una sola coordinata di $\mathbf{z}$ ha il valore 1, $k_2 = 0$, in quanto nessuna coordinata di $\mathbf{z}$ ha il valore 2, e così via.

Esercizio proposto 12.1. *Una disposizione con ripetizione è data da*

$$\mathbf{z} := (4, 2, 1, 3, 1, 4, 4, 2)$$

Scrivere la combinazione con ripetizione corrispondente a $\mathbf{z}$ nei due casi
(a) $\mathbf{z}$ viene vista come un'elemento del prodotto cartesiano $\{1, 2, 3, 4\}^8$
(b) $\mathbf{z}$ viene vista come un'elemento del prodotto cartesiano $\{1, 2, 3, 4, 5\}^8$

Permutando fra loro le coordinate $z_1, ..., z_r$ otteniamo una nuova disposizione con ripetizione $\mathbf{z}'$, alla quale resta associato (secondo la logica espressa in (106)) lo stesso vettore $\mathbf{k} \in A_{n,r}$.

È evidente che a due diverse disposizioni con ripetizione corrisponde uno stesso $\mathbf{k}$ se e solo se esse si ottengono una dall'altra per permutazione delle loro coordinate.

I vettori $\mathbf{k} := (k_1, ..., k_n)$ ottenibili con tale procedura a partire dalle disposizioni di n elementi di classe r, con ripetizione, sono tutti e soli quelli le cui n coordinate soddisfano le condizioni

$$0 \leq k_j \leq r, \qquad \sum_{j=1}^{n} k_j = r.$$

Tali condizioni sono equivalenti alle condizioni in (104), ossia i vettori $\mathbf{k}$ sono elementi di $A_{n,r}$.

> I concetti appena descritti, possono sembrare astratti, ma possono esere usate come modelli per diverse situazioni reali. Osserviamo che si tratta proprio della situazione del problema del compleanno (Esempio 3.3), con $n = 365$ ed r il numero delle persone che si trovano in una stanza: $\mathbf{z}$ rappresenta il vettore r-dimensionale dei giorni di nascita delle r persone presenti (ad esempio in ordine alfabetico, o in un altro ordine prestabilito) mentre $\mathbf{k}$ rappresenta il vettore a 365 componenti: la componente j-sima indica il numero di persone che sono nate nel giorno j.

Pensiamo ad un esperimento casuale in cui lo spazio campione Ω coincide con $\{1, 2, ..., n\}^r$; cioè in cui i risultati elementari possibili sono le disposizioni, con ripetizione, di classe r.

Possiamo allora pensare una combinazione $\mathbf{k}$ (di classe r, con ripetizione) come una descrizione più grossolana, meno dettagliata, del risultato dell'esperimento; infatti, se ci viene riportato solo il risultato $\mathbf{k}$, rimaniamo con uno stato di incertezza circa quale sia stata (fra le varie disposizioni con ripetizione "equivalenti") quella $\mathbf{z}$ che si è effettivamente presentata, dando luogo alla combinazione $\mathbf{k}$.

Quale tipo di informazione abbiamo perso nel passaggio da $\mathbf{z}$ a $\mathbf{k}$?
Conoscere $\mathbf{z}$ significa sapere, per ogni $j = 1, 2, .., n$, ***quante*** e ***quali*** coordinate z_h hanno il valore j; mentre conoscere $\mathbf{k}$ significa sapere soltanto, per ogni $j = 1, 2, .., n$, ***quante*** coordinate z_h hanno il valore j.

Ci poniamo ora la seguente domanda:

Fissato un elemento $\mathbf{k} \in \mathbf{A_{n,r}}$, quante sono le diverse disposizioni $\mathbf{z}$ alle quali resta associato $\mathbf{k}$?

Come vedremo meglio nell'Esempio 12.4, questa domanda equivale a porsi il problema degli anagrammi.

In altre parole ci stiamo chiedendo quale sia la cardinalità dell'insieme

$$\mathbf{B_k} := \big\{\mathbf{z} \in \{1, 2, .., n\}^r : \ \forall\, j = 1, 2, ..., n, \ \ k_j = \#\{h \text{ tali che } z_h = j\}\big\}. \tag{107}$$

La famiglia $\{\mathbf{B_k},\ \mathbf{k} \in \mathbf{A_{n,r}}\}$ forma una partizione dell'insieme $\{1, 2, ..., n\}^r$, per cui, anche senza sapere la risposta, ossia senza conoscere i valori di $|\mathbf{B_k}|$, possiamo però affermare che

$$\sum_{\mathbf{k} \in A_{n,r}} |\mathbf{B_k}| = |\{1, 2, .., n\}^r| = n^r.$$

La risposta alla domanda è data dai ***coefficienti multinomiali***:
fissato $\mathbf{k} \in A_{n,r}$, vale

$$|\mathbf{B_k}| = \binom{r}{k_1, k_2, ..., k_n}, \tag{108}$$

dove $\binom{r}{k_1, k_2, ..., k_n}$ è il coefficiente multinomiale, definito da

$$\binom{r}{k_1, k_2, ..., k_n} := \frac{r!}{k_1!\, k_2! \ldots k_n!}. \tag{109}$$

Chiaramente i coefficienti multinomiali sono definiti solo per $\mathbf{k} \in A_{n,r}$, ovvero solo per k_1, k_2,..., k_n che soddisfano la relazione (104).

Notiamo che, essendo $0! = 1$, i valori di j per cui $k_j = 0$ non contribuiscono al calcolo del coefficiente multinomiale in (109). Anche gli j per i quali $k_j = 1$ non contribuiscono al calcolo, tuttavia conviene scriverli lo stesso, in quanto permettono di controllare che valga la relazione (104) ed in particolare che la somma dei k_j sia effettivamente uguale ad r.

Esempio* 12.4. *Data una parola, ad esempio MATEMATICA, quanti sono gli anagrammi, anche senza significato nella lingua italiana, di questa parola?*
Soluzione: *Nel caso di MATEMATICA, $r = 10$ è il numero totale di lettere di cui è composta la parola MATEMATICA:*

$$3\,A, \quad 2\,M, \quad 2\,T, \quad 1\,E, \quad 1\,I, \quad 1\,C, \quad \textit{e nessuna altra lettera.}$$

Possiamo pensare di prendere $n = 21$ simboli (le lettere dell'alfabeto italiano), ossia $k_A = 3$, $k_M = k_T = 2$, $k_E = k_I = k_C = 1$, e $k_j = 0$ per tutte le altre lettere, e che chiaramente non contano nulla.

Per convincersi che il numero degli anagrammi è proprio

$$\binom{10}{3, 2, 2, 1, 1, 1} := \frac{10!}{3!\, 2!\, 2!\, 1!\, 1!\, 1!}$$

si può procedere in due modi.

Primo modo. *Si può pensare di avere a disposizione 10 tessere:*

$$A_1,\ A_2,\ A_3,\ M_1,\ M_2,\ T_1,\ T_2,\ E_1,\ I_1,\ C_1.$$

Possiamo permutarle in $10!$ modi, ottenendo così sicuramente tutti gli anagrammi di MATEMATICA, ma alcune permutazioni corrispondono allo stesso anagramma: infatti permutando tra loro A_1, A_2, A_3 (in 3! modi) l'anagramma rimane invariato, e lo stesso accade permutando tra loro M_1, M_2 (in 2! modi), oppure T_1, T_2 (in 2! modi). E quindi il numero totale degli anagrammi si ottiene dividendo $10!$ per $3! \cdot 2! \cdot 2!$:

$$\frac{10!}{3!\, 2!\, 2!} = \frac{10!}{3!\, 2!\, 2!\, 1!\, 1!\, 1!}.$$

Ovviamente si potrebbe anche prendere $n < 21$, ma, considerando che MATEMATICA contiene (solo) le sei lettere A, M, T, E, I, C, occorre prendere n almeno uguale a 6: infatti, come già osservato, i valori per cui $k_j = 0$ non contano *nulla.*

Secondo modo. *Alternativamente si può pensare che per ottenere un anagramma si hanno $\binom{10}{3}$ scelte per posizionare nei 10 posti possibili le 3 lettere A. Una volta collocate le 3 A, rimangono 7 posti tra i quali scegliere dove mettere le 2 M, e il numero delle scelte possibili è allora $\binom{7}{2}$ e si arriva a $\binom{10}{3}\binom{7}{2}$ scelte possibili per posizionare le 3 A e le 2 M. Giunti a questo punto rimangono 5 posti in cui mettere le 2 T, e le scelte possibili sono quindi $\binom{5}{2}$. Procedendo in maniera analoga si hanno successivamente $\binom{3}{1}$ scelte per collocare la E, $\binom{2}{1}$ scelte per collocare la I, ed infine $\binom{1}{1} = 1$ scelta obbligatoria per collocare la C. Globalmente quindi*

$$\binom{10}{3}\binom{7}{2}\binom{5}{2}\binom{3}{1}\binom{2}{1}\binom{1}{1} = \frac{10!}{3!\,7!}\,\frac{7!}{2!\,5!}\,\frac{5!}{2!\,3!}\,\frac{3!}{1!\,2!}\,\frac{2!}{1!\,1!}\,1 = \frac{10!}{3!\,2!\,2!\,1!\,1!\,1!}.$$

Generalizziamo ora i due procedimenti usati nell'Esempio 12.4 per verificare che i coefficienti multinomiali sono la risposta alla nostra domanda.

Fissato un elemento $\mathbf{k} \in A_{n,r}$ e trovato un vettore $\mathbf{z} \in \{1, 2, ..., n\}^r$ ad esso associato, per trovare tutti i vettori di $\{1, 2, ..., n\}^r$ associati a $\mathbf{k}$, basta considerare le $r!$ permutazioni delle r componenti di $\mathbf{z}$ e considerare le classi di equivalenza rispetto alle $k_j!$ permutazioni delle k_j componenti con lo stesso valore j, al variare di $j = 1, 2, ..., n$, da cui la formula (109).

Alternativamente, consideriamo che, per ottenere un vettore $\mathbf{z} \in \{1, 2, ..., n\}^r$ associato a $\mathbf{k}$, si hanno a disposizione r posti e tra questi ne vanno scelti k_1 dove collocare il valore 1, e ciò si può fare in $\binom{r}{k_1}$ modi. Tra i rimanenti $r - k_1$ posti ne vanno poi scelti k_2 dove collocare il valore 2, e ciò si può fare in $\binom{r-k_1}{k_2}$ modi. Proseguendo si ottiene che il numero cercato è dato dal coefficiente multinomiale, ossia

$$\binom{r}{k_1}\binom{r-k_1}{k_2}\binom{r-(k_1+k_2)}{k_3}\cdots\binom{r-(k_1+k_2+\cdots+k_{n-2})}{k_{n-1}}\binom{r-(k_1+k_2+\cdots+k_{n-1})}{k_n}$$

$$=\binom{r}{k_1}\binom{r-k_1}{k_2}\binom{r-k_1-k_2}{k_3}\cdots\binom{k_{n-1}+k_n}{k_{n-1}}\binom{k_n}{k_n}$$

$$=\frac{r!}{k_1!\,(r-k_1)!}\,\frac{(r-k_1)!}{k_2!\,(r-k_1-k_2)!}\,\frac{(r-k_1-k_2)!}{k_3!\,(r-k_1-k_2-k_3)!}\cdots\frac{(k_{n-1}+k_n)!}{k_{n-1}!\,k_n!}\,1$$

$$=\frac{r!}{k_1!\,k_2!\,k_3!\,\cdots\,k_{n-1}!\,k_n!} = \binom{r}{k_1, k_2, ..., k_n}$$

***Osservazione* 12.2.** Terminiamo questa parte osservando che, analogamente a quanto sappiamo circa lo sviluppo delle potenze di un binomio, e riportato nella formula (23) della Lezione 3, si può verificare che i coefficienti multinomiali intervengono quali coefficienti dei diversi monomi che compongono lo sviluppo delle potenze della somma $(a_1 + a_2 + ... + a_n)$. Più esattamente, risulta

$$(a_1 + a_2 ... + a_n)^r = \sum_{\mathbf{k}\in A_{n,r}} \binom{r}{k_1, ..., k_n} a_1^{k_1}\, a_2^{k_2} \cdots a_n^{k_n}. \tag{110}$$

Ponendo $a_1 = a_2 = \cdots = a_n = 1$ nella formula (110), otteniamo immediatamente

$$\sum_{\mathbf{k} \in A_{n,r}} \binom{r}{k_1, ..., k_n} = n^r.$$

Dal momento che n^r coincide con la cardinalità dell'insieme $\{1, 2, ..., n\}^r$, ciò concorda sia con quanto risulta dalla precedente formula (108), che indica il coefficiente $\binom{r}{k_1 \ldots k_n}$ quale cardinalità di $\mathbf{B_k}$, ossia della classe di equivalenza delle $\mathbf{z} \in \{1, 2, .., n\}^r$ alle quali è associato $\mathbf{k} \in A_{n,r}$.
Si consiglia anche di rivedere la nota a pagina 172.
Riprendendo l'esempio degli anagrammi, possiamo pensare di avere un alfabeto con n caratteri, ad esempio nel caso dell'alfabeto italiano $n = 21$, mentre per l'alfabeto inglese $n = 26$. Una parola (anche senza senso) di lunghezza r è semplicemente una disposizione con ripetizione degli n caratteri, di classe r. Ovviamente si hanno n^r parole diverse (anche senza senso). Presa una parola di lunghezza r, possiamo associare la n-pla $\mathbf{k} = (k_1, k_2, ..., k_n)$, dove k_i rappresenta il numero di volte che compare il simbolo i-simo nella parola, ad esempio nella parola PROPOSIZIONE, $k_O = 3$, $k_I = k_P = 2$, $k_E = k_N = k_R = k_S = k_Z = 1$ e $k_i = 0$ per $i \notin \{E, I, N, O, P, R, S, Z\}$. Ovviamente $k_i \geq 0$ e $\sum_{i=1}^n k_i = r$, e il coefficiente multinomiale $\binom{r}{k_1, k_2, ..., k_n}$ rappresenta il numero di anagrammi della parola fissata. Ed ecco spiegata la formula appena vista, nel caso $a_1 = ...a_n = 1$.
Va infine ricordato che ad ogni parola è associata una combinazione con ripetizione, che rappresenta la composizione di ciascuno degli anagrammi della parola, ad esempio alla parola PROPOSIZIONE è associata la combinazione con ripetizione $\{E, I, I, N, O, O, O, P, P, R, S, Z\}$. Ovviamente l'$n$-pla $(k_1, k_2, ..., k_n)$ è associata alla combinazione con ripetizione, nel modo che abbiamo spiegato nel paragrafo precedente.

12.3 Esercizi di verifica

Esercizio 12.1. Sia $r = 6, n = 4$. Rappresentare la quaterna (1,2,2,1) sotto forma del disegno con barrette | e asterischi $*$, come a pagina 169.

Esercizio 12.2. Sia data la seguente rappresentazione con barrette | e asterischi $*$, come nel disegno a pagina 169,

$$* | \, | * | * * |$$

Dopo aver trovato n ed r,
(a) tradurre il disegno come vettore $\mathbf{k} \in A_{n,r}$ di numeri di occupazione;
(b) elencare tutti gli elementi di $\mathbf{B_k}$, ossia tutti i vettori $\mathbf{z}$ associati a $\mathbf{k}$.

13 Modelli di occupazione e schemi di estrazioni da urne

Come nella precedente Lezione 12, consideriamo l'insieme $A_{n,r}$ ($n, r \in \mathbb{N}$) costituito dalle n-uple ordinate $\boldsymbol{k} := (k_1, ..., k_n)$, con $k_j \geq 0$, interi, e tali che $\sum_{j=1}^{n} k_j = r$, in simboli

$$A_{n,r} = \left\{\boldsymbol{k} := (k_1, ..., k_n) : \sum_{j=1}^{n} k_j = r,\ k_j \in \mathbb{N} \cup \{0\},\ j = 1, 2, \ldots, n\right\}.$$

In questa lezione vogliamo illustrare il significato delle distribuzioni di probabilità su $A_{n,r}$ come ***modelli di occupazione*** ed esaminare alcuni casi notevoli.

Vediamo subito una naturale interpretazione di una distribuzione di probabilità su $A_{n,r}$.
Siano dati r ***soggetti*** $O_1, ..., O_r$ ed n ***diversi siti*** etichettati con i numeri $1, ..., n$.

Supponiamo che gli r soggetti si dispongano negli n siti in modo aleatorio.

> La frase ***gli r soggetti si dispongono negli n siti*** non tragga in inganno: nelle applicazioni potrebbe benissimo trattarsi di r ***oggetti che vengono disposti in n siti***.
> Si veda a questo proposito il successivo Esempio 13.1, in cui soggetto viene interpretato come prova, mentre sito come risultato o esito.
> Nel successivo Esempio 13.4, invece i soggetti sono persone e i siti sono piani di un edificio.

Consideriamo per $i = 1, 2, ..., r$ e $j = 1, 2, ..., n$ le variabili aleatorie binarie:

$$B_i^{(j)} = \begin{cases} 1 & \text{se il soggetto } O_i \text{ cade nel sito } j \\ 0 & \text{altrimenti.} \end{cases}$$

Consideriamo ora le variabili aleatorie definite come segue:

$$X_j := \sum_{i=1}^{r} B_i^{(j)}, \qquad j = 1, ..., n.$$

La variabile aleatoria X_j indica dunque il ***numero complessivo dei soggetti che si dispongono nel sito*** j.

Il vettore aleatorio

$$\boldsymbol{X} := (X_1, ..., X_n)$$

è quindi un vettore aleatorio a valori in $A_{n,r}$ e ciascuna distribuzione di probabilità su $A_{n,r}$ può essere vista come un modello probabilistico di scelta dei siti da parte dei soggetti.

Le variabili aleatorie $X_1, ..., X_n$ vengono detti ***numeri di occupazione*** e le distribuzioni di probabilità su $A_{n,r}$ sono indicate con il termine ***modelli di occupazione***.

I modelli di occupazione di cui ci occuperemo, e noti come modello di Maxwell-Boltzmann, modello di Bose-Einstein e modello di Fermi-Dirac, nascono dall'esigenza di trovare un modello per la disposizione degli elettroni attorno al nucleo di un atomo. Se il nucleo dell'atomo ha r protoni, anche il numero di elettroni è r. Gli elettroni si dispongono su vari livelli di energia: supponendo che i livelli di energia siano n, le variabili di interesse sono il numero di elettroni che si trovano al livello j, per $j = 1, 2, ..., n$. Con questa interpretazione i siti sono gli n livelli di energia, e i soggetti sono gli r elettroni.

13.1 Modello di Maxwell-Boltzmann

Si suppone che gli r soggetti siano ***distinguibili*** ossia che siano possibili tutte le configurazioni $\mathbf{z}$ di $\{1,2,...,n\}^r$, dove ciascuna delle r componenti z_h indica in quale sito si sia collocato il soggetto h-simo. Si suppone inoltre che i soggetti vengano distribuiti negli n siti in modo tale da assegnare uguale probabilità $\frac{1}{n^r}$ a ciascuna delle n^r possibili configurazioni.

Per $\boldsymbol{k} := (k_1, k_2, ..., k_n) \in A_{n,r}$ si ha che $\{\boldsymbol{X} = \mathbf{k}\} := \mathbf{B_k}$, dove è definito in (107), e dunque[48]

$$\begin{aligned}\mathbb{P}(\{\boldsymbol{X} = \mathbf{k}\}) &= \mathbb{P}(\{X_1 = k_1,\, X_2 = k_2,\, \ldots,\, X_n = k_n\})\\ &= \frac{|\mathbf{B_k}|}{|\{1,2,...,n\}^r|} = \frac{\binom{r}{k_1,k_2,...,k_n}}{n^r} = \frac{r!}{k_1!\,k_2!\,\cdots\,k_{n-1}!\,k_n!}\frac{1}{n^r}, \qquad \text{per ogni } \mathbf{k} \in A_{n,r}.\end{aligned}$$

13.2 Modello di Bose-Einstein

Si suppone che gli r soggetti siano ***indistinguibili*** e che vengano distribuiti negli n siti in modo da assegnare uguale probabilità a ciascun vettore di numeri di occupazione, cioè a ciascuna n-upla $\mathbf{k} \in A_{n,r}$.

Si pone dunque[49]

$$\mathbb{P}(\{\boldsymbol{X} = \mathbf{k}\}) = \mathbb{P}(\{X_1 = k_1,\, X_2 = k_2,\, \ldots,\, X_n = k_n\}) = \frac{1}{|A_{n,r}|} = \frac{1}{\binom{n+r-1}{n-1}}, \qquad \text{per ogni } \mathbf{k} \in A_{n,r}.$$

13.3 Modello di Fermi-Dirac

Supponiamo $r \leq n$ e poniamo

$$\widehat{A}_{n,r} := \{\mathbf{k} \in A_{n,r} :\ k_i = 0,\ \text{oppure } k_i = 1,\ i = 1,2,...,n\}$$

La cardinalità di $\widehat{A}_{n,r}$ è ovviamente uguale a $\binom{n}{r}$.

Ora si suppone che gli r soggetti siano indistinguibili e che siano distribuiti nei siti in modo da assegnare uguale probabilità a ciascuno dei disegni in $\widehat{A}_{n,r}$ (ciò in particolare implica che in ciascun sito non può cadere più di un soggetto). Si ha allora[50]

$$\begin{aligned}\mathbb{P}(\{\boldsymbol{X} = \boldsymbol{k}\}) = \mathbb{P}(\{X_1 = k_1,\, X_2 = k_2,\, \ldots,\, X_n = k_n\}) &= \frac{1}{\binom{n}{r}} \qquad \text{per } \mathbf{k} \in \widehat{A}_{n,r}\\ &= 0 \qquad \text{per } \mathbf{k} \notin \widehat{A}_{n,r}\end{aligned}$$

13.4 Schemi di estrazioni da urne

Verrà illustrato qui di seguito che ai modelli di occupazione si può anche dare, equivalentemente, un'interpretazione in termini di schemi di estrazioni casuali da urne. Tuttavia si consiglia, in prima lettura, di passare alla seguente Sezione 13.5, dedicata agli esempi, e di tornare a questa sezione subito dopo.

[48]Si noti che sarebbe più evocativo mettere al posto del simbolo $\mathbb{P}$ un simbolo del tipo $\mathbb{P}_{MB}$ per mettere in evidenza che si tratta delle probabilità relative al modello di Maxwell-Boltzmann. Tuttavia lasciamo come al solito la scrittura $\mathbb{P}$ per non appesantire le notazioni.

[49]Si noti che sarebbe più evocativo mettere al posto del simbolo $\mathbb{P}$ un simbolo del tipo $\mathbb{P}_{BE}$ per mettere in evidenza che si tratta delle probabilità relative al modello di Bose-Einstein. Tuttavia lasciamo, come al solito, la scrittura $\mathbb{P}$ per non appesantire le notazioni.

[50]Si noti che sarebbe più evocativo mettere al posto del simbolo $\mathbb{P}$ un simbolo del tipo $\mathbb{P}_{FD}$ per mettere in evidenza che si tratta delle probabilità relative al modello di Fermi-Dirac. Anche in questo caso lasciamo la scrittura $\mathbb{P}$ per non appesantire le notazioni.

Supponiamo di avere inizialmente n oggetti di n diversi tipi (va ricordato che gli n diversi tipi sono denotati come tipo 1, tipo 2, ..., tipo n) in un'urna e di eseguire in modo aleatorio r estrazioni dall'urna. Nel linguaggio urne/palline, significa che l'urna contiene n palline, ciascuna con sopra un numero tra 1 ed n.

Tale esperimento ha come risultati elementari le n^r r-uple $(i_1, i_2, ..., i_r)$ con $i_h \in \{1, 2, ..., n\}$ e poniamo, per $i = 1, ..., n$,

$$\widetilde{X}_i := \text{numero di volte in cui si presenta un oggetto di tipo } i \text{ nelle } r \text{ estrazioni}$$

Anche qui dunque $\widetilde{\boldsymbol{X}} := \left(\widetilde{X}_1, ..., \widetilde{X}_n\right)$ è un vettore aleatorio a valori in $A_{n,r}$.

Come ora vedremo, diversi tipi di estrazioni casuali corrispondono ai diversi modelli di occupazione visti sopra e ciò permette agevolmente di spiegare questi ultimi in termini del meccanismo casuale con cui ogni soggetto sceglie un sito.

L'interpretazione nel linguaggio della sezione precedente di n siti ed r soggetti è il seguente: per $h = 1, 2, \dots, r$, il soggetto O_h viene posto nel sito i_h, che corrisponde al tipo dell'oggetto estratto nella h-sima estrazione.

Ricordiamo che il termine "estrazioni casuali" si riferisce all'ipotesi che ciascuno degli oggetti, presenti nell'urna al momento di una estrazione, abbia la stessa probabilità di presentarsi, indipendentemente dal suo tipo.

Le differenze fra i diversi modelli di estrazione risiedono nelle modalità della composizione dell'urna nelle successive estrazioni.

Estrazioni casuali con reinserimento e modello di Maxwell-Boltzmann

Dopo ogni estrazione ***l'oggetto estratto viene reinserito nell'urna***. Quindi ad ogni estrazione vi sono n oggetti di diversi tipi e ciascun tipo ha la stessa probabilità di presentarsi.

Ciascuno degli risultati elementari possibili ha la stessa probabilità $\frac{1}{n^r}$.

Per ogni n-pla $\mathbf{k} \in A_{n,r}$, i casi favorevoli all'evento $\{\widetilde{\boldsymbol{X}} = \mathbf{k}\}$ sono tanti quanti gli anagrammi di lunghezza r in cui il simbolo i compare k_i volte, per ogni $i = 1, 2, ..., n$, cioè il loro numero è dato dal coefficiente multinomiale e quindi

$$\mathbb{P}(\widetilde{\boldsymbol{X}} = \mathbf{k}) = \frac{r!}{k_1!\, k_2! \cdots k_n!} \frac{1}{n^r}.$$

Ciò corrisponde dunque, per il vettore $\widetilde{\boldsymbol{X}}$, al ***modello di Maxwell-Boltzmann***.

Vale forse la pena di ricordare l'interpretazione nel linguaggio precedente di n siti ed r soggetti: il soggetto O_h viene posto nel sito i_h estratto nella h-sima estrazione (con reinserimento).

Estrazioni casuali senza reinserimento e modello di Fermi-Dirac

L'oggetto estratto non viene più reinserito nell'urna; deve dunque essere $r \leq n$.

Alla j-esima estrazione vi sono $n - (j - 1) = (n - j + 1)$ oggetti nell'urna.

Ciascun tipo può presentarsi al più una sola volta nel complesso delle estrazioni.

Tutte le r-uple $(i_1, i_2, ..., i_r)$, con $i_1 \neq i_2 \neq ... \neq i_r$ hanno la stessa probabilità

$$\frac{1}{n\,(n-1)\dots(n-r+1)}.$$

Tutte le n-uple $\boldsymbol{x} \in \widehat{A}_{n,r}$ sono equiprobabili per $\widetilde{\boldsymbol{X}}$. Ciò corrisponde dunque al ***modello di Fermi-Dirac***.

Dal punto di vista dei numeri di occupazione tuttavia ognuna delle $r!$ permutazioni $(i_1, i_2, \ldots, i_r)$ che dia luogo allo stesso insieme di cardinalità r, corrisponde allo stesso vettore di numeri di occupazione, ovvero alla stessa n-pla $\mathbf{k} \in \widehat{A}_{n,r}$. Questo è il motivo per cui si ottiene esattamente che

$$\begin{aligned}
&\mathbb{P}(\widetilde{\boldsymbol{X}} = \mathbf{k}) = \frac{r!}{n\,(n-1)\ldots(n-r+1)} = \frac{1}{\binom{n}{r}}, && \text{per } \mathbf{k} \in \widehat{A}_{n,r},\\
&\mathbb{P}(\widetilde{\boldsymbol{X}} = \mathbf{k}) = 0 && \text{per } \mathbf{k} \in A_{n,r} \backslash \widehat{A}_{n,r}
\end{aligned}$$

Anche qui vale forse la pena di ricordare l'interpretazione nel linguaggio precedente di n siti ed r soggetti: il soggetto O_h viene posto nel sito i_h estratto nella h-sima estrazione (senza reinserimento).

Estrazioni casuali con doppio reinserimento e modello di Bose-Einstein

Dopo ciascuna estrazione, ***viene inserito nell'urna, insieme all'oggetto estratto, anche un altro oggetto dello stesso tipo***, cosicché alla j-esima estrazione vi sono $(n + j - 1)$ individui nell'urna.

Vogliamo ora calcolare $\mathbb{P}(\{\widetilde{\boldsymbol{X}} = \mathbf{k}\})$ per $\mathbf{k} := (k_1, k_2, ..., k_n) \in A_{n,r}$ (ricordiamo che $\sum_{h=1}^{n} k_h = r$); cominciamo a considerare la probabilità del risultato elementare $(i_1, i_2, ..., i_r)$ definito da

$$\begin{aligned}
&i_j = 1, && \text{per } 1 \leq j \leq k_1\\
&i_j = 2, && \text{per } k_1 + 1 \leq j \leq k_1 + k_2\\
&\vdots\\
&i_j = n, && \text{per } \sum_{h=1}^{n-1} k_h + 1 \leq j \leq \sum_{h=1}^{n} k_h = r
\end{aligned}$$

(cioè : tutti i primi k_1 elementi estratti sono di tipo 1, poi segue l'estrazione di k_2 elementi tutti di tipo 2 e così via). Tale probabilità è data da

$$\begin{aligned}
&\frac{1}{n} \cdot \frac{2}{n+1} \cdot \frac{3}{n+2} \cdot \cdots\cdots \cdot \frac{k_1}{n+k_1-1}\\
&\cdot \frac{1}{n+k_1} \cdot \frac{2}{n+k_1+1} \cdot \cdots\cdots \cdot \frac{k_2}{n+k_1+k_2-1}\\
&\cdot \frac{1}{n+k_1+k_2} \cdot \frac{2}{n+k_1+k_2+1} \cdot \cdots\cdots \cdot \frac{k_3}{n+k_1+k_2+k_3-1}\\
&\vdots\\
&\cdot \frac{1}{n+k_1+k_2+\ldots+k_{n-1}} \cdot \frac{2}{n+k_1+k_2+\ldots+k_{n-1}+1} \cdot \ldots \cdot \frac{k_n-1}{n+\ \sum_{h=1}^{n} k_h - 2} \cdot \frac{k_n}{n+\ \sum_{h=1}^{n} k_h - 1}\\
&= \frac{k_1!\,k_2! \ldots k_n!}{n \cdot (n+1) \cdot \cdots\cdots \cdot (n+r-1)} = \frac{k_1!\,k_2! \ldots k_n!}{\frac{(n+r-1)!}{(n-1)!}}.
\end{aligned}$$

Possiamo ora osservare che qualunque altro risultato elementare $(i_1, i_2, ..., i_r)$, che ugualmente abbia k_1 coordinate uguali ad 1, k_2 coordinate uguali a 2, ..., k_n coordinate uguali a n, seppure in un ordine diverso, ha ancora la stessa probabilità dell'r-upla precedentemente considerata.

Tale probabilità è data cioè ancora da

$$\frac{k_1!\,k_2!...k_n!}{\frac{(n+r-1)!}{(n-1)!}} = \frac{k_1!\,k_2!...k_n!}{\frac{(n+r-1)!}{(n-1)!}\frac{r!}{r!}} = \frac{k_1!\,k_2!...k_n!}{r!\binom{n+r-1}{n-1}}.$$

Dunque, ricordando che il numero di risultati elementare $(i_1, i_2, ..., i_r)$, che abbiano k_1 coordinate uguali ad 1, k_2 coordinate uguali a 2, ..., k_n coordinate uguali a n, è dato dal coefficiente multinomiale $\binom{r}{k_1,k_2,...,k_n}$, otteniamo

$$\mathbb{P}(\{\widetilde{\boldsymbol{X}} = \mathbf{k}\}) = \binom{r}{k_1, k_2, ..., k_n} \cdot \frac{k_1!\,k_2!...k_n!}{r!\binom{n+r-1}{n-1}} = \frac{1}{\binom{n+r-1}{n-1}}.$$

Ciò corrisponde quindi al ***modello di Bose-Einstein***.

Anche qui, nell'interpretazione nel linguaggio precedente di n siti ed r soggetti, il soggetto O_h viene posto nel sito i_h estratto nella h-sima estrazione (con doppio reinserimento).

13.5 Alcuni esempi

Abbiamo dunque illustrato fin qui due linguaggi diversi ma equivalenti, per illustrare i modelli di occupazione.

Bisogna però tenere presente che, con opportune modifiche di linguaggio, gli stessi schemi si presentano in moltissimi tipi di applicazioni diverse; per tale motivo la conoscenza degli aspetti basilari sui modelli di occupazione si rivela fruttuosa nella soluzione di moltissimi tipi di problemi di probabilità nel discreto.

In particolare il modello di Maxwell-Boltzmann si ripresenta in moltissime situazioni diverse, ed è interessante in particolare capire la differenza che sussiste fra tale modello e quelli di Bose-Einstein o di Fermi-Dirac.

Ora presentiamo alcuni esempi di problemi in cui si ritrovano dei modelli di occupazione; successivamente verrà illustrato un modello (la distribuzione multinomiale) utilissimo in tutte le applicazioni e che deriva da una naturale generalizzazione del modello di Maxwell-Boltzmann.

Esempio* 13.1. *Un esperimento può dar luogo, con uguali probabilità $p = \frac{1}{3}$, a tre diversi risultati, che per semplicità indichiamo con 1, 2, 3.*

Supponiamo che l'esperimento venga condotto per 10 successive volte, sempre con le stesse modalità ed in modo indipendente una volta dall'altra e si indichi con X_1, X_2, X_3 il numero di volte in cui, rispettivamente, si verifica il risultato 1, oppure 2, oppure 3. Calcolare le probabilità dell'evento $\{X_1 = 4, X_2 = 3, X_3 = 3\}$.

Soluzione: Notiamo che ovviamente risulta

$$\mathbb{P}(\{X_1 + X_2 + X_3 = 10\}) = 1$$

e anche, ovviamente

$$\mathbb{P}(\{X_1 \geq 0,\ X_2 \geq 0,\ X_3 \geq 0,\ X_1 + X_2 + X_3 = 10\}) = 1.$$

Tale problema può essere risolto guardando a $X_1,\ X_2,\ X_3$ come a dei numeri di occupazione, con $r = 10$ ed $n = 3$. (ogni prova è vista come un soggetto ed ogni possibile risultato come un sito dentro cui si inserisce ciascuna prova/soggetto).

Infatti le terne di valori possibili per il vettore aleatorio (X_1, X_2, X_3) costituiscono l'insieme $A_{3,10}$. Per l'indipendenza e l'equiprobabilità nel comportamento delle diverse prove si ha che la distribuzione di probabilità congiunta di (X_1, X_2, X_3) coincide con il corrispondente modello di Maxwell-Boltzmann. E dunque

$$\mathbb{P}(\{X_1 = 4, X_2 = 3, X_3 = 3\}) = \frac{\binom{10}{4,\,3,\,3}}{3^{10}} = \frac{10!}{4!\,(3!)^2}\,\frac{1}{3^{10}}.$$

Esempio* 13.2. *Un dado viene lanciato 12 volte ed indichiamo con X_i ($i = 1, ..., 6$) il numero dei lanci in cui si presenta il punteggio i. Calcolare*

$$\mathbb{P}(\{X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3\}).$$

Soluzione: Si tratta nuovamente di un modello di Maxwell-Boltzmann, questa volta con $n = 6$, $r = 12$.

L'evento $\{X_1 = 3,\ X_2 = 3,\ X_3 = 3,\ X_4 = 3\}$ implica ovviamente $X_5 = 0,\ X_6 = 0$. La probabilità cercata è data da

$$\begin{aligned}&\mathbb{P}(\{X_1 = 3,\ X_2 = 3,\ X_3 = 3,\ X_4 = 3\})\\ &= \mathbb{P}(\{X_1 = 3,\ X_2 = 3,\ X_3 = 3,\ X_4 = 3,\ X_5 = 0,\ X_6 = 0\})\\ &= \frac{12!}{(3!)^4\ (0!)^2\, 6^{12}} = \frac{12!}{6^{16}}.\end{aligned}$$

Esercizio proposto 13.1 (Il problema del compleanno, come modello di occupazione)**.** *Si consideri la situazione del problema del compleanno (Esempio 3.3), in cui ci sono r persone. Si numerino da 1 a 365 i giorni dell'anno (1 corrisponde al primo gennaio, 32 al primo febbraio, etc.. fino a 365, che corrisponde al 31 dicembre). Posto X_i la variabile aleatoria che conta il numero di persone nate nel giorno i, per $i = 1, 2, ..., 365$.*
(a) A quale tipo di modello di occupazione ci si riferisce?
(b) Scrivere, in termini delle variabili aleatorie X_i, l'evento

$$\{\textit{tutte le persone sono nate in giorni diversi}\}.$$

Esercizio proposto 13.2 (Estrazioni di assi)**.** *Sia dato un mazzo di carte italiane (40 carte, 4 semi, carte numerate da 1 a 10 per ciascun seme, con 1 che corrisponde all'asso, 8 al fante, 9 al cavallo e 10 al re) ben mescolate. Si estraggano ad una ad una le carte dal mazzo fino ad esaurire tutte le carte. Sia X_1 il numero di carte estratte (diverse da un asso) prima del primo asso. Siano inoltre X_2 il numero di carte estratte (diverse da un asso) dopo il primo asso e prima del secondo asso, cioè tra il primo ed il secondo asso, X_3 il numero di carte uscite (diverse da un asso) dopo il secondo asso e prima del terzo asso, X_4, le carte (diverse da un asso) uscite tra il terzo ed il quarto asso, ed infine X_5 le carte (diverse da un asso) uscite dopo il quarto asso. Ad esempio[51] se i quattro assi sono usciti come seconda, terza, quindicesima e trentaseiesima carta, allora $X_1 = 1$, $X_2 = 0$, $X_3 = 11$, $X_4 = 20$ e $X_5 = 4$. Ovviamente*

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_4 \geq 0, X_5 \geq 0, \quad e \quad X_1 + X_2 + X_3 + X_4 + X_5 = 36$$

in quanto 36 è il numero totale di carte diverse da un asso.

Si dimostri che $(X_1, X_2, X_3, X_4, X_5)$ corrisponde al modello di occupazione di Bose-Einstein con $n = 5$ ed $r = 36$.

[51] Più in generale se i quattro assi escono come $k_1 - sima$, $k_2 - sima$, $k_3 - sima$ e $k_4 - sima$ carta, allora $X_1 = k_1 - 1$, $X_2 = k_2 - k_1 - 1$, $X_3 = k_3 - k_2 - 1$, $X_4 = k_4 - k_3 - 1$ ed infine $X_5 = 40 - k_4$.

Esempio 13.3. *Un testo contiene 20 caratteri. Supponiamo di sapere che esso contiene 5 errori (cioè 5 caratteri errati) e di valutare che ciascun carattere abbia uguale probabilità di essere errato rispetto a ciascun altro. Qual è la probabilità che gli errori si trovino nel* 3^0, 5^0, 10^0, 11^0, 12^0 *carattere?*

Questo è un esempio di modello di Fermi-Dirac, con $n = 20$ *ed* $r = 5$*, in quanto in ogni posto ci può stare al più un errore. E dunque la probabilità cercata è data da* $\frac{1}{\binom{20}{5}}$.

Esempio* 13.4. *5 persone sono in attesa dell'ascensore nella hall di un albergo di 4 piani. Poniamo, per* $i = 1, 2, 3, 4$,

$$X_i := \text{numero persone che scendono al piano } i$$

Se le persone, una volta entrate nell'ascensore, si distribuiscono a caso fra i piani, a quanto è uguale

$$\mathbb{P}(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\})?$$

Soluzione: Si tratta di un modello di occupazione con $n = 4$ ed $r = 5$. La dizione "le persone si distribuiscono a caso fra i piani" è equivoca: potrebbe trattarsi di modello di Maxwell-Boltzmann oppure di Bose-Einstein, a seconda che le persone vengano considerate o meno ***indipendenti***[52] fra di loro (o diciamo, ***distinguibili***).

Nei due diversi casi si avrebbe, indicando rispettivamente con $\mathbb{P}_{MB}$ e con $\mathbb{P}_{BE}$ le probabilità nei due modelli,

$$\mathbb{P}_{MB}(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{5!}{(0!)^3\, 4!\, 4^5} = \frac{5!}{4! 4^5} = \frac{5}{4^5} = \frac{5}{1024}$$

oppure

$$\mathbb{P}_{BE}(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{1}{\binom{8}{3}} = \frac{3!5!}{8!} = \frac{3 \cdot 2}{8 \cdot 7 \cdot 6} = \frac{1}{56}.$$

Riguardo al modello di Maxwell-Boltzmann in particolare, osserviamo che esso si rivela, per questo caso, piuttosto irrealistico: è poco ragionevole dare per scontato che vi sia indipendenza stocastica completa fra le diverse persone ma soprattutto che tutti i piani siano equiprobabili fra di loro!

In merito a quest'ultima condizione di equiprobabilità, vediamo che è opportuno estendere il modello di Maxwell-Boltzmann al caso in cui vi sia ancora indipendenza stocastica completa fra i diversi oggetti, ma non vi sia equiprobabilità fra i diversi siti. Tale estensione porta alla definizione di distribuzione multinomiale.

13.6 Distribuzione multinomiali

Consideriamo ancora un modello di occupazione con r soggetti ed n siti. Supponiamo ancora che vi sia indipendenza stocastica completa circa la scelta dei siti da parte dei diversi soggetti, ma non vi sia equiprobabilità fra i diversi siti: supponiamo che ciascun soggetto scelga il sito j con fissata probabilità p_j ($j = 1, ..., n$; $p_j \geq 0$, $\sum_{j=1}^{n} p_j = 1$).

Generalizzando quanto arguito per il modello di Maxwell-Boltzmann, possiamo concludere che, per i numeri di occupazione $X_1, ..., X_n$, risulta:

$$\mathbb{P}(\{X_1 = k_1, ..., X_n = k_n\}) = \binom{r}{k_1,\ k_2,\ ...,\ k_n} (p_1)^{k_1} \cdot (p_2)^{k_2} \cdot ... \cdot (p_n)^{k_n}\,. \tag{111}$$

$$= \frac{r!}{k_1!\ k_2!\ ...\ k_n!} p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_n^{k_n}, \qquad \text{per } (k_1, k_2, \dots k_n) \in A_{n,r}. \tag{112}$$

Diciamo in tal caso che la distribuzione congiunta di $X_1, ..., X_n$ è una *distribuzione multinomiale* di parametri $(r, n; p_1, ..., p_n)$.

[52]La frase "*le persone vengano considerate o meno indipendenti*", va intesa nel senso che eventi relativi a persone diverse (sempre riguardo la scelta del piano al quale scendere) sono eventi completamente indipendenti.

È chiaro che un modello di Maxwell-Boltzmann costituisce un caso particolare di distribuzione multinomiale con $p_j = \frac{1}{n}$.

Osserviamo anche che una distribuzione binomiale $Bin(r, p)$ è connessa ad una distribuzione multinomiale di parametri $(r, 2; p, 1-p)$; in quale modo? (Vedere l'Esercizio di verifica 13.8).

Per la comprensione della formula (111) in realtà è più comodo pensare alla distribuzione multinomiale proprio come a una generalizzazione della binomiale[53].

Osserviamo infine che essendo ovviamente $\mathbb{P}(\boldsymbol{X} \in A_{n,r}) = 1$, si ha che

$$\sum_{(k_1,k_2,\dots,k_n)\in A_{n,r}} \frac{r!}{k_1!\ k_2!\ \dots\ k_n!} p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_n^{k_n} = 1,$$

che non è altro che un caso particolare della formula della potenza del multinomio, ovvero

$$\sum_{(k_1,k_2,\dots,k_n)\in A_{n,r}} \frac{r!}{k_1!\ k_2!\ \dots\ k_n!} a_1^{k_1} \cdot a_2^{k_2} \cdot \dots \cdot a_n^{k_n} = (a_1 + a_2 + \dots + a_n)^r, \tag{113}$$

che a sua volta è una generalizzazione della formula della potenza del binomio.

Esempio* 13.5. *Un gioco fra quattro persone viene ripetuto 5 volte, sempre con gli stessi giocatori A, B, C, D. Ogni volta c'è un singolo vincitore: A, B e C hanno probabilità di vincere del 20% e D del 40%. Con X_A, X_B, X_C, X_D si indichino, rispettivamente, il numero delle vittorie di A, B, C, D sulle 5 volte. Calcolare le probabilità dell'evento*

$$\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}.$$

Soluzione: Le variabili aleatorie X_A, X_B, X_C, X_D hanno una distribuzione congiunta multinomiale di parametri $(5, 4; \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})$ e risulta

$$\mathbb{P}(\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}) = \frac{5!}{(1!)^3\, 2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{5!}{2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{48}{5^4} = \frac{48}{625}.$$

---

[53] L'idea è che ci sono r prove ad n esiti possibili, ovvero

$$E_k^{(j)}, \qquad \text{per } j = 1, 2, \dots, n,\ k = 1, 2, \dots r,$$

dove il verificarsi di $E_k^{(j)}$, significa il verificarsi dell'esito di tipo j nella k-sima prova ed esclude il verificarsi, nella prova k-sima degli altri esiti. In altre parole $\{E_k^{(j)}$, per $j = 1, 2, \dots, n, \}$ sono partizioni per ogni $k = 1, 2, \dots r$. Inoltre si suppone che, qualunque sia $k = 1, 2, \dots, r$, la probabilità $\mathbb{P}(E_k^{(j)}) = p_j$.
Infine si suppone l'indipendenza delle prove, cioè che, qualunque sia $(j_1, j_2, \dots, j_r) \in \{1, 2, \dots, n\}^r$,

$$\mathbb{P}(E_1^{(j_1)} \cap E_2^{(j_2)} \cap E_3^{(j_3)} \cap \dots \cap E_r^{(j_r)}) = \mathbb{P}(E_1^{(j_1)}) \cdot \mathbb{P}(E_2^{(j_2)}) \cdot \mathbb{P}(E_3^{(j_3)}) \cdot \dots \cdot \mathbb{P}(E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \dots \cdot p_{j_n}.$$

Il caso della binomiale $Bin(r, \theta)$ e dello schema di Bernoulli per r prove indipendenti, con probabilità di successo θ rientra in questo modello, con $n = 2$, $E_k^{(1)} = E_k$, $E_k^{(2)} = \overline{E}_k$, $p_1 = \theta$, $p_2 = 1 - \theta$.
L'evento $\{X_1 = k_1, \dots, X_n = k_n\}$, con $\mathbf{k} \in A_{n,r}$ si verifica se e solo se si verificano eventi del tipo $E_1^{(j_1)} \cap E_2^{(j_2)} \cap \dots \cap E_r^{(j_r)}$, con k_1 indici di tipo 1, k_2 indici di tipo 2, ... , k_n indici di tipo n. Per calcolare la probabilità dell'evento $\{X_1 = k_1, \dots, X_n = k_n\}$, basta allora considerare che gli eventi di questo tipo hanno tutti la stessa probabilità:

$$\mathbb{P}(E_1^{(j_1)} \cap E_2^{(j_2)} \cap \dots \cap E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \dots \cdot p_{j_r} = p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_n^{k_n},$$

e che inoltre gli eventi di questo tipo sono esattamente $\frac{r!}{k_1!\ k_2!\ \dots\ k_n!}$.

Esempio 13.6. *In una giornata del campionato di calcio si attribuisce probabilità 0.5 alla vittoria della squadra che gioca in casa (risultato 1), probabilità 0.2 alla sconfitta della squadra che gioca in casa (risultato 2), probabilità 0.3 al pareggio (risultato* x*), e il risultato di ciascuna partita è giudicato essere indipendente dai risultati delle altre. Si consideri la colonna "vincente" della schedina del totocalcio*[54]*; si ponga*

$$Z_1 := \text{numero di risultati 1 sulle tredici partite}$$

e si dia analogo significato alle variabili Z_{x}, Z_2.

Le variabili aleatorie Z_1, Z_2, Z_{x} hanno una distribuzione congiunta multinomiale di parametri $(13, 3; \frac{5}{10}, \frac{2}{10}, \frac{3}{10})$.

Esercizio proposto 13.3 (Il problema del compleanno: mesi di nascita)**.** *Si consideri la situazione del problema del compleanno (Esempio 3.3), in cui ci sono r persone. Si numerino da 1 a 12 i mesi dell'anno (1 corrisponde a gennaio, 2 a febbraio, etc.. fino a 12, che corrisponde a dicembre). Posto X_i la variabile aleatoria che conta il numero di persone nate nel mese i, per $i = 1, 2, ..., 12$.*
Chiaramente si tratta di un modello multinomiale.
(a) Calcolare p_i per $i = 1, 2, ..., 12$.
(b) Scrivere, in termini delle variabili aleatorie X_i, l'evento

$$\{\textit{tutte le persone sono nate nel mese di giugno}\}$$

e calcolarne la probabilità.
(c) Scrivere, in termini delle variabili aleatorie X_i, l'evento

$$\{\textit{tutte le persone sono nate nello stesso mese}\}$$

e calcolarne la probabilità.

13.7 Distribuzioni marginali e condizionate nei modelli di occupazione

Sia fissato ora un generico modello di occupazione, con n siti ed r soggetti. Vogliamo ricavare la distribuzione di probabilità marginale della variabile aleatoria X_1.

Si ha, per definizione di distribuzione marginale,

$$\mathbb{P}(\{X_1 = x\}) = \sum_{(k_2,\dots,k_n)\in A_{n-1,r-x}} \mathbb{P}(\{X_1 = x, X_2 = k_2, ..., X_n = k_n\}), \tag{114}$$

ossia la somma è fatta su tutti i valori di $k_2, \cdots, k_n$ tali che $k_i \geq 0$ e $\sum_{i=2}^n k_i = r - x$.

Un'analoga formula vale per le distribuzioni marginali di X_2, di X_3, ... e di X_n.

***Osservazione* 13.1.** Per motivi di simmetria, nel caso dei modelli di Maxwell-Boltzmann, Bose-Einstein e Fermi-Dirac, si ha che le distribuzioni marginali di X_i coincidono tutte con la distribuzione marginale di X_1. Infatti, ad esempio,

$$\mathbb{P}(\{X_n = x\}) = \sum_{(k_1,\dots,k_{n-1})\in A_{n-1,r-x}} \mathbb{P}(\{X_1 = k_1, X_2 = k_2, ..., , X_{n-1} = k_{n-1}, X_n = x\})$$

ed inoltre, qualunque siano i valori di$x$, y, e di $k_2, ..., k_{n-1}$, si ha

$$\mathbb{P}(\{X_1 = x, X_2 = k_2, ..., , X_{n-1} = k_{n-1}, X_n = y\}) = \mathbb{P}(\{X_1 = y, X_2 = k_2, ..., , X_{n-1} = k_{n-1}, X_n = x\}),$$

e quindi la somma precedente ha lo stesso valore della somma in (114).

[54]Una schedina del totocalcio è composta da una colonna costituita dai risultati di 13 partite fissate. I risultati possibili sono 1, 2 e x, per ogni elemento della colonna. in totale ci sono quindi 3^{13} possibili colonne, ovvero possibili schedine da giocare.

Consideriamo ora la distribuzione di probabilità della variabile aleatoria

$$Y = \sum_{j=2}^{n} X_j$$

Osserviamo che risulta

$$\{Y = y\} = \{X_1 = r - y\}$$

da cui

$$\mathbb{P}(\{Y = y\}) = \mathbb{P}(\{X_1 = r - y\})$$

e quindi, dalla (114),

$$\mathbb{P}(\{Y = y\}) = \sum_{(k_2,\ldots,k_n)\in A_{n-1,y}} \mathbb{P}(\{X_1 = r - y,\, X_2 = k_2,\, \ldots,\, X_n = k_n\}).$$

Supponiamo ora di voler calcolare la ***distribuzione condizionata di*** X_2 ***data*** X_1. Osserviamo allora che vale

$$\mathbb{P}(\{X_1 = x_1, X_2 = x_2\}) = \sum_{(k_3,\ldots,k_n)\in A_{n-2,r-x_1-x_2}} \mathbb{P}(\{X_1 = x_1,\, X_2 = x_2,\, X_3 = k_3,\, \ldots,\, X_n = k_n\}),$$

da cui

$$\mathbb{P}(\{X_2 = x_2\}|\{X_1 = x_1\}) = \frac{\displaystyle\sum_{(k_3,\ldots,k_n)\in A_{n-2,r-x_1-x_2}} \mathbb{P}(\{X_1 = x_1,\, X_2 = x_2,\, X_3 = k_3,\, \ldots, X_n = k_n\})}{\displaystyle\sum_{(\xi,\zeta_3,\ldots,\zeta_n)\in A_{n-1,r-x_1}} \mathbb{P}(\{X_1 = x_1,\, X_2 = \xi,\, X_3 = \zeta_3,\, \ldots, X_n = \zeta_n\})}.$$

Esempio 13.7. *Consideriamo il* ***modello di Maxwell-Boltzmann*** *con r soggetti ed n siti. Allora, per $x_1 = 0, 1, \ldots, r$, avremo (ricordando che $\mathbb{P}(\{\boldsymbol{X} = \mathbf{k}\}) = \frac{\binom{r}{k_1\ k_2\ \ldots\ k_n}}{n^r}$)*

$$\begin{aligned}
\mathbb{P}(\{X_1 = x\}) &= \sum_{(k_2,\ldots,k_n)\in A_{n-1,r-x}} \mathbb{P}(\{X_1 = x,\, X_2 = k_2,\, \ldots,\, X_n = k_n\}) \\
&= \frac{1}{n^r} \sum_{(k_2,\ldots,k_n)\in A_{n-1,r-x}} \frac{r!}{x!\,(r-x)!}\, \frac{(r-x)!}{k_2!\ldots k_n!} \\
&= \frac{1}{n^r}\, \frac{r!}{x!\,(r-x)!} \sum_{(k_2,\ldots,k_n)\in A_{n-1,r-x}} \frac{(r-x)!}{k_2!\ldots k_n!},
\end{aligned}$$

*da cui, ricordando l'**Osservazione 12.2** , o equivalentemente la formula della potenza del multinomio* (113)

$$\begin{aligned}
\mathbb{P}(\{X_1 = x\}) &= \binom{r}{x} \frac{1}{n^r}\, (n-1)^{r-x} = \\
&= \binom{r}{x} \left(\frac{1}{n}\right)^{x} \left(1 - \frac{1}{n}\right)^{r-x} \qquad \textit{per } x = 0, 1, \ldots, r,
\end{aligned}$$

cioè, come ci si doveva immaginare[55]*, la distribuzione marginale di X_1 è binomiale di parametri r e $\frac{1}{n}$.*

Per quanto riguarda la distribuzione condizionata di X_2 data X_1 si ha innanzitutto che

$$\begin{aligned}
\mathbb{P}(\{X_1 = x_1, X_2 = x_2\}) &= \sum_{(k_3,\ldots,k_n)\in A_{n-2,r-x_1-x_2}} \frac{\binom{r}{x_1\ x_2\ k_3\ \ldots\ k_n}}{n^r} \\
&= \frac{1}{n^r}\frac{r!}{x_1!\ x_2!\ (r-x_1-x_2)!} \sum_{(k_3,\ldots,k_n)\in A_{n-2,r-x_1-x_2}} \frac{(r-x_1-x_2)!}{k_3!\ \ldots\ k_n!} \\
&= \binom{r}{x_1,\ x_2,\ (r-x_1-x_2)} \left(\frac{1}{n}\right)^{x_1} \left(\frac{1}{n}\right)^{x_2} \left(1-\frac{2}{n}\right)^{r-x_1-x_2},
\end{aligned}$$

per $x_1 \geq 0$, $x_2 \geq 0$ e $x_1 + x_2 \leq r$.

La distribuzione condizionata di X_2 dato il valore x_1 per X_1(con ovviamente $0 \leq x_1 \leq r$) è data da

$$\begin{aligned}
\mathbb{P}(\{X_2 = x_2\}|\{X_1 = x_1\}) &= \frac{\mathbb{P}(\{X_1 = x_1, X_2 = x_2\})}{\mathbb{P}(\{X_1 = x_1\})} \\
&= \binom{r-x_1}{x_2} \left(\frac{1}{n-1}\right)^{x_2} \left(1-\frac{2}{n-1}\right)^{r-x_1-x_2};
\end{aligned}$$

per $x_2 \geq 0$, $x_1 + x_2 \leq r$, ovvero per $0 \leq x_2 \leq r - x_1$. Si tratta cioè, come ci si doveva aspettare, di una distribuzione binomiale di parametri $r - x_1$ e $\frac{1}{n-1}$.

Analoghe formule si possono facilmente ottenere per quanto riguarda le distribuzioni marginali e condizionate di più di due fra le n variabili $X_1, ..., X_n$, per le distribuzioni di probabilità delle loro somme parziali etc... ; è anche interessante vedere come tali formule si specializzino per i vari modelli di occupazione notevoli elencati in precedenza.

Non vale qui la pena di scrivere sistematicamente tali formule e tali risultati specifici, che possono invece costituire utili esercizi per il lettore.

13.8 Distribuzioni marginali e condizionate per la distribuzione multinomiale

Consideriamo una generalizzazione del modello di Maxwell-Boltzmann, ossia il caso in cui $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ segua una distribuzione multinomiale di parametri $(r, n; p_1, p_2, \ldots, p_n)$.
È facile convincersi che la distribuzione marginale di X_i è binomiale di parametri r e p_i, così come la distribuzione di $X_i + X_j$ è $Bin(r, p_i + p_j)$, e così via.

Infatti piuttosto che mettersi a fare i calcoli, basta pensare che X_i conta il numero di successi in r prove indipendenti in cui "*successo*" alla k-sima prova significa esito di tipo i alla k-sima prova (e "*insuccesso*" alla k-sima prova significa esito di tipo ℓ, per un $\ell \neq i$ alla k-sima prova). Analogamente $X_i + X_j$ conta il numero di successi in r prove indipendenti in cui però stavolta "*successo*"alla k-sima prova significa esito di tipo i oppure j alla k-sima prova (e "*insuccesso*" alla k-sima prova significa esito di tipo ℓ, per un $\ell \notin \{i, j\}$ alla k-sima prova).

[55]Il modello di Maxwell-Boltzmann è un caso particolare del modello multinomiale, con $p_j = \frac{1}{n}$, per ogni $j = 1, 2, \ldots, n$, che a sua volta deriva dal modello di prove ripetute ad n esiti. Interessarsi di X_1 significa controllare ad ogni prova solo se si è verificato l'esito di tipo 1, o no. Ovvero ci si riconduce al caso binomiale: numero di successi in r prove ripetute in cui solo due esiti sono possibili. Risulta evidente che, nel caso della distribuzione multinomiale, la distribuzione di X_1 sia allora binomiale di parametri (r, p_1).

Ancora se vogliamo calcolare invece la distribuzione condizionata di X_2 dato X_1 si procede in modo simile a quanto fatto per il modello di Maxwell-Boltzmann: si ha innanzitutto che

$$\mathbb{P}(\{X_1 = x_1, X_2 = x_2\}) = \frac{r!}{x_1!\; x_2!\; (r - x_1 - x_2)!}\, p_1^{x_1}\; p_2^{x_2}\; \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2}, \tag{115}$$

per $x_1 \geq 0$, $x_2 \geq 0$ e $x_1 + x_2 \leq r$.
Il risultato non è sorprendente: infatti basta pensare che se si considerano solo gli esiti di tipo 1, di tipo 2 e di tipo " *ne' 1 ne' 2*"e si indica con

$$W = X_3 + \ldots + X_n$$

la variabile aleatoria che conta gli esiti di tipo "*ne' 1 ne' 2*", allora immediatamente si è nel caso di una distribuzione multinomiale di parametri $(r, 3; p_1, p_2, 1 - (p_1 + p_2))$.

$$\begin{aligned}
&\mathbb{P}(\{X_1 = x_1, X_2 = x_2\}) = \mathbb{P}(\{X_1 = x_1,\ X_2 = x_2,\ W = r - (x_1 + x_2)\}) \\
&= \frac{r!}{x_1!\; x_2!\; (r - x_1 - x_2)!}\, p_1^{x_1}\; p_2^{x_2}\; \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2}.
\end{aligned}$$

Tuttavia anche senza questo discorso euristico, per dimostrare la (115) basta osservare che

$$\begin{aligned}
&\mathbb{P}(\{X_1 = x_1, X_2 = x_2\}) \\
&= \sum_{(k_3,\ldots,k_n) \in A_{n-2, r-x_1-x_2}} \binom{r}{x_1\ x_2\ k_3\ \ldots\ k_n}\, p_1^{x_1}\; p_2^{x_2}\; p_3^{k_3} \cdots\; p_n^{k_n} \\
&= \frac{r!}{x_1!\; x_2!\; (r - x_1 - x_2)!}\, p_1^{x_1}\; p_2^{x_2} \sum_{(k_3,\ldots,k_n) \in A_{n-2, r-x_1-x_2}} \frac{(r - x_1 - x_2)!}{k_3!\ \ldots\ k_n!}\, p_3^{k_3} \cdots\; p_n^{k_n} \\
&= \binom{r}{x_1,\ x_2,\ (r - x_1 - x_2)}\, p_1^{x_1}\; p_2^{x_2}\; (p_3 + \cdots + p_n)^{r - x_1 - x_2} \\
&= \frac{r!}{x_1!\; x_2!\; (r - x_1 - x_2)!}\, p_1^{x_1}\; p_2^{x_2}\; \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2},
\end{aligned}$$

per $x_1 \geq 0$, $x_2 \geq 0$ e $x_1 + x_2 \leq r$.

La distribuzione condizionata di X_2 dato il valore x_1 per X_1 (con ovviamente $0 \leq x_1 \leq r$) è data da

$$\mathbb{P}(\{X_2 = x_2\} | \{X_1 = x_1\}) = \binom{r - x_1}{x_2} \left(\frac{p_2}{1 - p_1}\right)^{x_2} \left(1 - \frac{p_2}{1 - p_1}\right)^{r - x_1 - x_2} \tag{116}$$

per per $0 \leq x_2 \leq r - x_1$. Si tratta cioè, come ci si poteva aspettare, di una distribuzione binomiale di parametri $r - x_1$ e $\frac{p_2}{1-p_1}$. Infatti le prove continuano ad essere indipendenti, ma oramai solo due tipi di esiti sono possibili l'esito di tipo 2 oppure non due. Inoltre si devono considerare solo le $r - x_1$ in cui non si è avuto esito di tipo 1. Infine la probabilità di esito 2 va valutata condizionatamente a sapere che su ciascuna di tali prove non è verificato un esito di tipo 1: in fondo questa è l'interpretazione di $\frac{p_2}{1-p_1}$.

Tuttavia anche in questo caso non è necessario questo discorso euristico, e per dimostrare la (116) basta osservare che

$$\begin{aligned}
\mathbb{P}(\{X_2 = x_2\}|\{X_1 = x_1\}) &= \frac{\mathbb{P}(\{X_1 = x_1, X_2 = x_2\})}{\mathbb{P}(\{X_1 = x_1\})} \\
&= \frac{\binom{r}{x_1,\, x_2,\, (r-x_1-x_2)}\, p_1^{x_1}\; p_2^{x_2}\; (1-(p_1+p_2))^{r-x_1-x_2}}{\binom{r}{x_1}\, p_1^{x_1}(1-p_1)^{r-x_1}} \\
&= \frac{\frac{r!}{x_1!\, x_2!\, (r-x_1-x_2)!}\, p_1^{x_1}\; p_2^{x_2}\; (1-(p_1+p_2))^{r-x_1-x_2}}{\frac{r!}{x_1!\, (r-x_1)!}\, p_1^{x_1}(1-p_1)^{r-x_1}} \\
&= \frac{\frac{1}{x_2!\, (r-x_1-x_2)!}\, p_2^{x_2}\; (1-(p_1+p_2))^{r-x_1-x_2}}{\frac{1}{(r-x_1)!}(1-p_1)^{r-x_1}} \\
&= \binom{r-x_1}{x_2} \left(\frac{p_2}{1-p_1}\right)^{x_2} \left(1 - \frac{p_2}{1-p_1}\right)^{r-x_1-x_2}
\end{aligned}$$

per $x_2 \geq 0$, $x_1 + x_2 \leq r$, ovvero per $0 \leq x_2 \leq r - x_1$.

13.9 Esercizi di verifica

Esercizio 13.1. Ricordando la rappresentazione con barrette | e asterischi $*$, come a pagina 169, calcolare la probabilità del risultato

$$*|\;|*|*|$$

assumendo rispettivamente che valga il modello di Maxwell-Boltzmann, o di Bose-Einstein o di Fermi-Dirac.

Esercizio 13.2. Consideriamo i 120 studenti del primo anno ed i 3 esami del primo semestre (chiamiamoli A, B, C). Facciamo una statistica per rilevare qual è l'esame che è stato superato per primo da ciascuno studente (supponiamo che tutti abbiano superato almeno un esame) e poniamo X_A = numero degli studenti del primo anno che hanno superato l'esame A come primo (o unico) esame; analogamente si definiscano X_B e X_C.
È ragionevole assumere un modello di Maxwell-Boltzmann per X_A, X_B, X_C? Ed un modello multinomiale?

Esercizio 13.3. I 40 membri di un dipartimento devono votare per eleggere il direttore. Vi sono i 4 candidati A, B, C e D. Ogni elettore deve esprimere un solo voto e tutti votano (non vi sono schede bianche, per tradizione). X_A, X_B, X_C e X_D sono i voti riportati dai vari candidati.
Calcolare $\mathbb{P}(\{X_A = 20, X_B = 5, X_C = 14, X_D = 1\})$ sotto l'ipotesi che ogni votante voti il candidato I con probabilità $p_I = 1/4$ ($I = A, B, C, D$), e indipendentemente dagli altri, ossia che si tratti di uno schema di Maxwell-Boltzmann.

Esercizio 13.4. Un gioco viene ripetuto 5 volte, fra i giocatori A, B, C e D, dove A, B e C hanno probabilità di vincere del 20% e D del 40%.
Indicando con X_A, X_B, X_C, X_D, rispettivamente, il numero delle vittorie di A, B, C, D sulle 5 volte, calcolare le probabilità degli eventi:
(a) $\{X_A = 2\}$
(b) $\{X_A + X_D = 5\}$
(c) $\{X_A + X_B = 3, X_C = 1, X_D = 1\}$.

Esercizio 13.5. Si consideri la colonna "vincente" della schedina del Totocalcio, nel caso in cui si attribuisce probabilità 0.5 al risultato 1, probabilità 0.2 al risultato 2 e probabilità 0.3 al risultato x, e il risultato di ciascuna partita è giudicato indipendente dai risultati delle altre (si veda il precedente Esempio 13.6).
(a) Qual è la colonna più probabile? Quanto vale la sua probabilità?
(b) Quanto vale la probabilità del risultato $(1, 1, \mathsf{x}, 2, 1, 1, 1, 1, \mathsf{x}, \mathsf{x}, 1, 2, \mathsf{x})$?
(c) Qual è la probabilità che vi siano 7 risultati 1, 2 risultati 2 e 4 risultati x?

Esercizio 13.6. Un esperimento, che può dar luogo, con uguali probabilità $p = \frac{1}{3}$ a tre diversi risultati, viene condotto per 10 successive volte, in modo indipendente una volta dall'altra e si indica con X_1, X_2, X_3 il numero di volte in cui, rispettivamente, si verifica il risultato 1, oppure 2, oppure 3.
(a) Calcolare le probabilità dell'evento $\{X_2 + X_3 = 6\}$
(b) Calcolare la probabilità condizionata $\mathbb{P}(\{X_2 = 3\}|\{X_1 = 4\})$.
(c) Calcolare la distribuzione di probabilità condizionata di X_2 dato l'evento $\{X_1 = 4\}$.

Esercizio 13.7. Calcolare la distribuzione marginale di X_1 e la distribuzione condizionata di X_2 data X_1 in un modello di Bose-Einstein con r oggetti e n siti ed in un modello di Fermi-Dirac con r oggetti e n siti (ovviamente in questi due casi si deve richiedere $r \leq n$).

Esercizio 13.8. Indichiamo con S il numero di successi in n prove bernoulliane di probabilità p e poniamo $T = n - S$.
Che cosa si ottiene come distribuzione congiunta di S e T?

Esercizio 13.9. (Urna di Polya) Si consideri un'urna con b palline bianche ed r rosse. Si estrae a caso una pallina e la si rimette nell'urna insieme a una pallina del colore estratto. L'estrazione viene ripetuta, con le stesse modalità, k volte. Si indichi con $\mathbf{P}_{b,r}$ la probabilità relativa al caso in cui inizialmente nell'urna ci sono appunto b palline bianche ed r rosse. Posto
$E_h = \{$Vengono estratte h palline bianche e $k - h$ rosse$\}$, $h = 0, \dots, k$,
(a) Calcolare $\mathbf{P}_{b,r}(E_h)$, per $h = 0, 1, ..., k$
(b) Dimostrare (ad esempio per induzione) che la probabilità che la i-esima pallina estratta, $i = 1, \dots, k$, sia bianca è $b/(b + r)$.

Suggerimento per l'Esercizio 13.7

Nel modello di Bose-Einstein la probabilità su $A_{n,r}$ è costante, e quindi per calcolare la distribuzione di X_1 basta saper calcolare il numero dei casi favorevoli all'evento $X_1 = k$, per $k = 0, 1, ..., r$. Fissato $k \in \{0, 1, ..., r\}$ si ha

$$\begin{aligned}\{X_1 = k\} &= \{X_1 = k, X_1 + X_2 + \cdots + X_n = r\}\\ &= \{X_1 = k, k + X_2 + \cdots + X_n = r\}\\ &= \big\{(X_1, X_2, ..., X_n) \in \{(k, k_2, ..., k_n :\ k + \textstyle\sum_{i=2}^n k_i = r\}\big\}\end{aligned}$$

e l'insieme delle n-ple $\{(k, k_2, ..., k_n) : k + \sum_{i=2}^n k_i = r\}$ è in corrispondenza biunivoca con l'insieme delle $(n-1)$-ple $\{(k_2, ..., k_n) : \sum_{i=2}^n k_i = r - k\}$, che a sua volta è in corrispondenza biunivoca con $A_{n-1,r-k}$
Similmente, fissati k ed h in $\{0, 1, ..., r\}$ in modo che $h + k \leq r$ si ha

$$\begin{aligned}\{X_1 = k, X_2 = h\} &= \{X_1 = k, X_2 = h, X_1 + X_2 + \cdots + X_n = r\}\\ &= \{X_1 = k, X_2 = h, k + h + X_3 + \cdots + X_n = r\}\\ &= \big\{(X_1, X_2, ..., X_n) \in \{(k, h, k_2, ..., k_n) :\ k + h + \textstyle\sum_{i=3}^n k_i = r\}\big\}\end{aligned}$$

e l'insieme delle n-ple $\{(k, h, k_3, ..., k_n) : k + h + \sum_{i=3}^n k_i = r\}$ è in corrispondenza biunivoca con l'insieme delle $(n-2)$-ple $\{(k_3, ..., k_n) : \sum_{i=3}^n k_i = r - k - h\}$, che a sua volta è in corrispondenza biunivoca con $A_{n-2,r-k-h}$.
Per il modello di Fermi-Dirac si procede in modo simile, ma usando $\widehat{A}_{n,r}$.

14 Spazi di probabilità e variabili aleatorie in casi più generali

14.1 Definizione generale di spazio di probabilità

Fin qui abbiamo trattato spazi di probabilità con un numero finito di eventi elementari e variabili aleatorie che possono assumere soltanto un numero finito di valori possibili.

Tale limitazione non consente però di trattare tutti i casi di possibile interesse e dobbiamo quindi estendere tali definizioni a casi più generali; in effetti il lettore era stato avvertito che, in attesa di riformulazioni più generali e definitive, i concetti di spazio di probabilità e di variabile aleatoria venivano introdotti in forma provvisoria.

Oramai, nelle precedenti lezioni, abbiamo familiarizzato con tali concetti e possiamo dunque passare ora a trattare tali casi più generali e a capire come e perchè le definizioni stesse vadano parzialmente modificate.

Per cominciare il discorso, rimaniamo però ancora nel caso di spazi di probabilità con un numero finito di eventi elementari e consideriamo quanto segue:

vi sono situazioni in cui non possiamo o comunque non siamo interessati ad assegnare la probabilità a tutti i sottoinsiemi dello spazio campione, in un esperimento aleatorio.

Esempio 14.1. *Lanciare m volte un dado costituisce un esperimento in cui Ω coincide con $\{1, 2, ..., 6\}^m$. Pensiamo ora ad un gioco, basato su tali m lanci, in cui un giocatore vince o perde sulla base del valore della somma degli m successivi punteggi. Allora ci interesserà soltanto assegnare le probabilità a quei sottoinsiemi di Ω definiti in termini di tale somma.*

Esempio 14.2. *Pensiamo all'esperimento che consiste nel disporre, in modo aleatorio, r oggetti in n siti. Come si era visto nella precedente Lezione 13, se gli oggetti sono numerati e quindi distringuibili è possibile considerare come spazio campione Ω l'insieme $\{1, 2, ..., n\}^r$ costituito (o, equivalentemente, l'insieme di tutte le applicazioni $\varphi : \{1, 2, ..., r\} \to \{1, 2, ..., n\}$). Supponiamo ora che gli oggetti siano indistinguibili: non vogliamo o non possiamo distinguere fra loro due diversi eventi elementari che prevedano un ugual numero di oggetti per ciascuno dei siti (che prevedano cioè uguali valori per i numeri di occupazione $X_1, ..., X_n$).*

In questo caso ciascun evento, che possa essere effettivamente osservato, è un'unione di eventi del tipo

$$\{X_1 = x_1, ..., X_n = x_n\} \quad \text{(con } \sum_j x_j = r\text{).}$$

Esempio 14.3. *Supponiamo di lanciare un dado sul quale abbiamo applicato dei bollini rossi, sopra i numeri pari, e blu, sopra i numeri dispari, in modo che sia possibile capire se è uscito un numero pari o dispari, ma non sia possibile sapere precisamente quale numero sia uscito, senza togliere il bollino.*

È naturale prendere come spazio campione $\Omega = \{1, 2, 3, 4, 5, 6\}$. Tuttavia, ***se non ci è permesso togliere i bollini****, gli unici sottoinsiemi per i quali possiamo affermare se si sono verificati oppure no sono*

$$\{esce\ un\ numero\ pari\} \quad \text{oppure} \quad \{esce\ un\ numero\ dispari\}$$

oltre ovviamente all'evento impossibile $\emptyset$ e all'evento certo Ω.

In una situazione del tipo descritto nei precedenti esempi, indichiamo con $\mathcal{F}$ la famiglia dei sottoinsiemi di Ω ai quali viene attribuita una probabilità. ***Soltanto gli elementi di $\mathcal{F}$ verranno chiamati* eventi**: ***non tutti i sottoinsiemi di Ω sono dunque degli eventi, bensì soltanto quelli ai quali viene attribuita una probabilità***; nella trattazione che abbiamo esposto all'inizio si aveva $\mathcal{F} = \mathscr{P}(\Omega)$, mentre ora poniamo $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, prevedendo anche situazioni in cui la famiglia $\mathcal{F}$ sia strettamente contenuta in $\mathscr{P}(\Omega)$.

Oltre ad essere $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, quali altre proprietà deve avere $\mathcal{F}$?

È ragionevole richiedere che, dati due arbitrari eventi A e B (cioè due sottoinsiemi di Ω, entrambi appartenenti a $\mathcal{F}$), anche $A \cup B$, $A \cap B$, $\overline{A}$, $\overline{B}$ siano degli eventi (cioè sottoinsiemi appartenenti a $\mathcal{F}$).
Si deve quindi richiedere anche che i sottoinsiemi banali di Ω, cioè $\emptyset$ (evento impossibile) e Ω (evento certo), facciano parte di $\mathcal{F}$.
Assumeremo dunque che $\mathcal{F}$ sia un'algebra (secondo la Definizione 5.5 della Lezione 5).

Passiamo ora a considerare il caso in cui lo spazio campione Ω sia un insieme infinito.
Situazioni di tale tipo si presentano necessariamente quando si debbano considerare infiniti eventi diversi, o variabili aleatorie che prendano valore in un insieme infinito, o successioni di variabili aleatorie fra loro diverse.

Estendendo quanto svolto per il caso con Ω finito e tenendo conto delle precedenti considerazioni, anche nel caso generale uno spazio di probabilità è definito come una terna $(\Omega, \mathcal{F}, \mathbb{P})$ dove Ω è un arbitrario spazio di punti, $\mathcal{F}$ è un'algebra di sottoinsiemi di Ω (gli eventi) e $\mathbb{P} : \mathcal{F} \to [0, 1]$ è una funzione di insieme con la proprietà di additività e tale che $\mathbb{P}(\Omega) = 1$.

Per comodità del lettore diamo una definizione formale di spazio di probabilità, nel caso finitamente additivo, in un modo che permette di evidenziare la connessione con la successiva Definizione 14.3.
Uno spazio di probabilità (con la probabilità finitamente additiva) è una terna $(\Omega, \mathcal{F}, \mathbb{P})$ con $\mathcal{F}$ un'algebra, cioè tale che

i) $\Omega \in \mathcal{F}$

ii) se $A \in \mathcal{F}$, allora $\overline{A} \in \mathcal{F}$

iii) se $A_k \in \mathcal{F}$, per $k = 1, ..., n$, allora $\cup_{k=1}^{n} A_k \in \mathcal{F}$

e $\mathbb{P} : \mathcal{F} \to [0, 1]$ una funzione tale che

i) per ogni $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$

ii) $\mathbb{P}(\Omega) = 1$

iii) se $A_k \in \mathcal{F}$, per $k = 1, ..., n$, e $A_h \cap A_k = \emptyset$, allora

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

Vi è però ora qualcosa da precisare in merito a degli aspetti che si presentano soltanto nel caso di Ω infinito.
Supponiamo che $E_1, E_2, ...$ sia una successione di eventi, cioè di elementi di $\mathcal{F}$; sappiamo che l'unione finita di un numero arbitrario di tali eventi deve ancora essere un evento (avendo assunto che $\mathcal{F}$ sia un'algebra). Ciò non garantisce però che anche l'unione numerabile $\bigcup_{j=1}^{\infty} E_j$ sia un elemento di $\mathcal{F}$. Se $\mathcal{F}$, oltre ad essere un'algebra, soddisfa anche tale condizione allora diremo che $\mathcal{F}$ è una σ-algebra (si legge sigma-algebra), secondo la seguente definizione.

Definizione 14.1 (σ-algebra)**.** *Una famiglia $\mathcal{F}$ di sottoinsiemi di Ω è detta **σ-algebra** se sono verificate le condizioni seguenti*

i) $\Omega \in \mathcal{F}$

ii) $E \in \mathcal{F} \Rightarrow \overline{E} \in \mathcal{F}$

iii) $E_1,\ E_2, ... \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{\infty} E_j \in \mathcal{F}$

Ovviamente l'insieme delle parti $\mathscr{P}(\Omega)$ è sempre una σ-algebra. Osserviamo inoltre che la definizione di σ-algebra differisce da quella di algebra solo nella terza condizione, ed è facile vedere che esistono algebre che non sono σ-algebre. Ad esempio, se $\Omega := \{1, 2, ..\}$ è l'insieme dei numeri naturali, allora la famiglia $\mathcal{F}$ degli insiemi finiti o cofiniti (cioè complementari di un insieme finito) è un'algebra (vedere l'Esercizio proposto 14.1), ma non una σ-algebra: i singoletti $\{i\}$ appartengono ad $\mathcal{F}$, mentre invece l'insieme $\bigcup_{k=1}^{\infty}\{2k\}$, ossia l'insieme dei numeri pari, non appartiene ad $\mathcal{F}$, in quanto non è finito e nemmeno cofinito.

Esercizio proposto 14.1. *Sia $\Omega := \{1, 2, ..\}$ l'insieme dei numeri naturali, e consideriamo la famiglia $\mathcal{F}$ degli insiemi finiti o cofiniti (cioè complementari di un insieme finito). Verificare che $\mathcal{F}$ è un'algebra.*

Inoltre se $E_1,\ E_2, ... \in \mathcal{F}$ è una successione di eventi allora $\bigcap\limits_{j=1}^{\infty} E_j \in \mathcal{F}$, come si vede subito applicando la formula di de Morgan

$$\bigcap_{j=1}^{\infty} E_j = \overline{\bigcup_{j=1}^{\infty} \overline{E}_j} \tag{117}$$

e le proprietà $ii)$ e $iii)$.

Consideriamo allora una successione di eventi $E_1, E_2, ...$ e assumiamo che $\mathcal{F}$ sia una σ-algebra; dunque $\bigcup_{j=1}^{\infty} E_j \in \mathcal{F}$ e possiamo considerare la probabilità

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} E_j\right).$$

Supponiamo ora che gli eventi $E_1, E_2, ...$ siano a due a due disgiunti e consideriamo la serie

$$\sum_{j=1}^{\infty} \mathbb{P}(E_j);$$

possiamo chiederci allora se debba o meno valere l'identità

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(E_j); \tag{118}$$

osserviamo infatti che essa non è implicata dalla proprietà di additività finita.

Possiamo assumere la (118) come un ulteriore assioma imposto alla $\mathbb{P}$, cioè che questa risulti numerabilmente additiva secondo la seguente definizione.

Definizione 14.2. *Sia $\mathbb{P} : \mathcal{F} \to [0, 1]$ una funzione di insieme. Diremo che $\mathbb{P}$ è* **numerabilmente additiva** *(o equivalentemente che è σ-**additiva**) se per qualunque successione $E_1,\ E_2, ...$, con $E_j \in \mathcal{F}$, $j \geq 1$, di eventi **incompatibili a due a due**, cioè $E_i \cap E_j = \emptyset$ per ogni $i \neq j$, risulta verificata la* (118).

A seconda che si imponga o meno tale condizione sulla $\mathbb{P}$, verremo dunque a costruire due diverse teorie della probabilità: una più "forte" (cioè in cui è possibile ricavare un numero maggiore di risultati) in cui vale la σ-additività della $\mathbb{P}$ ed una più "debole", ma più generale, in cui si considerano soltanto quei risultati che si possono ottenere imponendo solamente che $\mathbb{P}$ sia finitamente additiva.

Gran parte della attuale letteratura sulla teoria e le applicazioni della probabilità danno per scontato l'assioma dell'additività numerabile.

Anche in questi appunti il termine "misura di probabilità" verrà d'ora in poi utilizzato, salvo avviso in contrario, per designare esclusivamente il caso σ-additivo; e giungiamo così alla seguente definizione (questa volta definitiva) di spazio di probabilità.

Definizione 14.3 (spazio di probabilità). *Uno* **spazio di probabilità** *è una terna* $(\Omega, \mathcal{F}, \mathbb{P})$ *dove*

1. Ω *è un arbitrario spazio di punti,*
2. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ *è una* σ*-algebra di sottoinsiemi di* Ω
3. $\mathbb{P}\colon \mathcal{F} \to [0,1]$ *è una misura di probabilità, cioè una funzione di insieme* σ*-additiva tale che* $\mathbb{P}(\Omega) = 1$.

Spazi di probabilità (generali)

Per comodità del lettore riuniamo in una scheda la definizione di spazio di probabilità: $(\Omega, \mathcal{F}, \mathbb{P})$ è uno ***spazio di probabilità*** se e solo se

1. Ω è un arbitrario spazio di punti.
2. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è una σ-algebra di sottoinsiemi di Ω, ossia
   (i) $\Omega \in \mathcal{F}$
   (ii) se A è un evento (cioè $A \in \mathcal{F}$) allora il complementare $\overline{A}$ è un evento (cioè $\overline{A} \in \mathcal{F}$)
   (iii) se $\{A_n,\, n \geq 1\}$ è una successione di eventi (cioè $A_n \in \mathcal{F}$, per ogni $n \geq 1$) allora $A = \{\textit{si verifica almeno uno tra gli eventi } A_n\}$ è un evento (cioè $A = \bigcup_{n\geq 1} A_n \in \mathcal{F}$)
3. $\mathbb{P}\colon \mathcal{F} \to [0,1]$ è una misura di probabilità,ossia
   (a) $\mathbb{P}(\Omega) = 1$
   (b) se $\{A_n,\, n \geq 1\}$ è una successione di eventi incompatibili (cioè $A_n \in \mathcal{F}$, per ogni $n \geq 1$ e $A_i \cap A_j = \emptyset$, per ogni $i \neq j$) allora $\mathbb{P}(\bigcup_{n\geq 1} A_n) = \sum n \geq 1\, \mathbb{P}(A_n)$.

***Osservazione* 14.1.** È possibile trovare dei controesempi che mostrano quanto segue: esistono degli insiemi Ω tali che non è possibile costruire alcuna misura di probabilità (cioè σ-additiva) che risulti ben definita su tutto $\mathscr{P}(\Omega)$. Sulla stessa retta reale, la misura che associa agli intervalli la loro lunghezza non può essere coerentemente estesa a tutti i sottoinsiemi sotto il vincolo che valga la σ-additività (si veda ad esempio P. Billingsley "Probability and Measure", Ed. Wiley and Sons, 1995).

Possiamo dunque dire in altre parole che, nel caso in cui si assume la σ-additività, il fatto di limitare il dominio della probabilità ad una sotto σ-algebra $\mathcal{F} \subseteq \mathscr{P}(\Omega)$ (invece che a tutto $\mathscr{P}(\Omega)$) può diventare in molti casi una necessità piuttosto che una questione di scelta, come invece avviene nel caso finito.

***Osservazione* 14.2 (Spazi di probabilità numerabili).** Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità con $\Omega = \{\omega_i;\ i = 1, 2, ...\}$ numerabile, $\mathcal{F} = \mathscr{P}(\Omega)$ ed assumiamo che $\mathbb{P}$ sia σ-additiva.

Notiamo che, analogamente a quanto accade nel caso di uno spazio di probabilità finito, $\mathbb{P}(E)$ risulta univocamente determinata $\forall E \in \mathcal{F}$, una volta fissate le probabilità degli eventi elementari $\mathbb{P}(\{\omega_i\})$, $i = 1, 2, ...$; si ha infatti

$$\mathbb{P}(E) = \sum_{i\colon \omega_i \in E} \mathbb{P}(\{\omega_i\}).$$

Ovviamente è necessaria anche la condizione che

$$\mathbb{P}(\{\omega_i\}) \geq 0 \quad \forall i, \qquad e \qquad \sum_{i=1}^{\infty} \mathbb{P}(\{\omega_i\}) = 1.$$

È anche interessante a tale proposito considerare il seguente esempio.

Esempio 14.4. *Consideriamo il caso in cui lo spazio Ω coincida con l'insieme dei numeri naturali: $\Omega := \{1, 2, ...\}$. Non è possibile definire alcuna probabilità $\mathbb{P}$ σ-additiva e uniforme su $(\Omega, \mathcal{P}(\Omega))$, ossia in modo che risulti*

$$\mathbb{P}(\{i\}) = c, \qquad i = 1, 2, ...$$

essendo c una costante indipendente da i. Infatti, se così fosse, dovrebbe risultare

$$1 = \mathbb{P}(\Omega) = \sum_{i=1}^{\infty} c$$

Ma una serie a termini costanti $\sum_{i=1}^{\infty} c$ può convergere se e solo se si pone $c = 0$; e tale posizione implicherebbe $\mathbb{P}(\Omega) = 0 \neq 1$.

> La condizione di normalizzazione $\sum_{i=1}^{\infty} \mathbb{P}(\{\omega_i\}) = 1$ cessa invece di valere nel caso in cui si richieda che la probabilità $\mathbb{P}$ sia definita su un'algebra contenente i singoletti $\{\omega_i\}$ e inoltre che $\mathbb{P}$ sia finitamente additività finita, ma non σ-additiva. Notiamo che non è necessario che tale algebra coincida con $\mathcal{P}(\Omega)$).
> Ad esempio, si può prendere Ω come l'insieme dei numeri naturali, $\mathcal{F}$ come l'algebra degli insiemi finiti o cofiniti, e definire $\mathbb{P}(A) = 0$ se A è finito, mentre $\mathbb{P}(A) = 1$ se A è cofinito. Ovviamente $\mathbb{P}(\Omega) = 1$, ed è facile vedere che tale $\mathbb{P}$ è una probabilità finitamente additiva, tuttavia, ovviamente, non vale la condizione di normalizzazione considerata.

Nella Lezione 2 avevamo verificato, nel caso di spazi di probabilità finiti, la validità di alcune proprietà fondamentali della probabilità quali conseguenze immediate degli assiomi; fra tali proprietà, troviamo in particolare la proprietà di monotonia: per $A, B \in \mathcal{P}(\Omega)$

$$A \subseteq B \qquad \Longrightarrow \qquad \mathbb{P}(A) \leq \mathbb{P}(B).$$

Nel caso di spazi di probabilità infiniti l'aggiunta dell'assioma della σ-additività oltre a quello dell'additività finita garantisce che continuino a valere tutte le proprietà elencate[56] nella Lezione 2.

> Tutte le proprietà della probabilità dimostrate negli spazi di probabilità finiti continuano a valere anche negli spazi di probabilità più generali, in quanto si basano sulla proprietà di additività. L'unica accortezza consiste nello specificare sempre che si richiede che gli insiemi in considerazione appartengano ad $\mathcal{F}$. Ad esempio, nella precedente proprietà di monotonia va specificato che $A, B \in \mathcal{F}$, in modo che abbia senso calcolare $\mathbb{P}(A)$ e $\mathbb{P}(B)$.

In verità non occorre "aggiungere" - basta bensì sostituire - l'assioma della σ-additività a quello relativo all'additività finita, in quanto la σ-additività rende automaticamente verificata anche la additività finita,

[56]Ad esempio, la proprietà base (12), in questo contesto diviene: siano A e B eventi (ossia siano A, $B \in \mathcal{F}$) allora

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}),$$

la cui dimostrazione dipende, come prima dal fatto che $A = (A \cap B) \cup (A \cap \overline{B})$, ma nel caso generale bisogna avere l'accortezza di osservare che, se A e B sono eventi, ossia se A, $B \in \mathcal{F}$, allora anche $\overline{B} \in \mathcal{F}$ e quindi anche $A \cap B$ e $A \cap \overline{B}$ sono eventi, ossia appartengono a $\mathcal{F}$; di conseguenza si possono calcolare $\mathbb{P}(A \cap B)$ e $\mathbb{P}(A \cap \overline{B})$, e basta poi usare la proprietà di additività finita.

come verrà mostrato nella successiva ***Proposizione*** 14.2

L'assioma della σ-additività, inoltre, dà luogo ad un ulteriore importante conseguenza: **la proprietà di "continuità" delle probabilità**. Si ha cioè la seguente

***Proposizione* 14.1** (Proprietà di continuità delle probabilità)**.** *In uno spazio di probabilità* $(\Omega, \mathcal{F}, P)$ *siano dati, per* $n = 1, 2, ...,$ $A_n \in \mathcal{F}, B_n \in \mathcal{F}$ *tali che*

$$A_n \subseteq A_{n+1}; \qquad B_n \supseteq B_{n+1}$$

e poniamo

$$A := \bigcup_{n=1}^{\infty} A_n; \qquad B := \bigcap_{n=1}^{\infty} B_n.$$

Se $\mathbb{P}$ *è* σ*-additiva allora risulta*

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n); \qquad \mathbb{P}(B) = \lim_{n\to\infty} \mathbb{P}(B_n)$$

Dimostrazione. Innanzitutto notiamo quanto segue:

In virtù della proprietà di monotonia della P, $\{\mathbb{P}(A_n)\}_{n=1,2,...}$ è una successione non decrescente ed è limitata superiormente (in quanto $\mathbb{P}(A_n) \leq 1$); dunque certamente esiste $\lim_{n\to\infty} \mathbb{P}(A_n)$; analogamente esiste $\lim_{n\to\infty} \mathbb{P}(B_n)$.

Inoltre, essendo $\mathcal{F}$ una σ-algebra, risulta $A \in \mathcal{F}$ e quindi esiste $\mathbb{P}(A)$; analogamente esiste $\mathbb{P}(B)$.

Poniamo ora

$$C_1 = A_1, \qquad C_2 = A_2 \backslash A_1 = A_2 \cap \overline{A}_1, \qquad C_3 = A_3 \backslash A_2 = A_3 \cap \overline{A}_2, ...$$

È facile verificare che $\forall n \in \mathbb{N}$, $C_1, ..., C_n$ sono a due a due disgiunti e che $A_n = \bigcup_{i=1}^{n} C_i$. Ne segue quindi, per la proprietà di additività finita, $\forall n \in \mathbb{N}$,

$$\mathbb{P}(A_n) = \sum_{i=1}^{n} \mathbb{P}(C_i).$$

Inoltre possiamo scrivere $A := \bigcup_{n=1}^{\infty} C_n$ da cui segue, in virtù del fatto che $\mathbb{P}$ è σ-additiva,

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(C_i);$$

cioè, per definizione di somma di una serie,

$$\mathbb{P}(A) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(C_i) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

La relazione

$$\mathbb{P}(B) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Si verifica in modo analogo, oppure si può verificare a partire dalla relazione $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$, definendo $A_n = \overline{B}_n$ e utilizzando la formula di De Morgan (117). □

Prima di concludere questo sottoparagrafo, è opportuno prestare attenzione alla seguente osservazione, che riveste un'importanza fondamentale per la comprensione di alcuni successivi punti.

***Osservazione* 14.3.** In uno spazio di probabilità finito, si può assumere, senza effettiva perdita di generalità, che tutti gli eventi elementari abbiano probabilità positiva; infatti, se un evento elementare avesse probabilità nulla, potremmo tranquillamente escluderlo dalla lista degli eventi possibili; notiamo inoltre che, anche eliminando tutti gli eventi di probabilità nulla, la probabilità dell'unione di tutti gli eventi elementari rimasti si mantiene uguale ad 1.

Dunque, nel caso finito, è ragionevole considerare situazioni in cui tutti gli eventi, siano essi elementari o composti, hanno tutti probabilità strettamente positiva, con la sola esclusione dell'evento impossibile: baterebbe "eliminare/escludere" dal modello gli eventuali eventi di probabilità nulla.

Nel caso di uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ con $\mathcal{F}$ di potenza maggiore del numerabile ciò invece non è più necessariamente vero: in generale vi saranno degli eventi di probabilità nulla, ma che risultano possibili; infatti il generico evento elementare avrà probabilità zero, tranne al più per un insieme finito o numerabile di eventi elementari.

Notiamo che potrebbe anche accadere che tutti gli eventi elementari abbiano singolarmente probabilità zero e che gli unici eventi di probabilità strettamente positiva siano eventi composti da un'unione non numerabile di eventi elementari (ciò è connesso con il fatto che la proprietà di additività numerabile, che assicura che la probabilità di un'unione numerabile di eventi a due a due disgiunti sia uguale alla somma della serie delle loro probabilità, non è estendibile al caso di unioni più che numerabili). Inoltre, sempre nel caso in tutti gli eventi elementari abbiano probabilità nulla, se si elimininassero tutti gli eventi di probabilità nulla, resterebbe soltanto l'insieme vuoto.

Questo è un punto un po' delicato che potrebbe all'inizio apparire poco chiaro.

È opportuno in proposito pensare ad un'analogia (che in effetti è molto di più che un'analogia) con la geometria elementare: sulla retta reale l'intervallo $[0,1]$, che ha lunghezza finita uguale a 1, è composto da un'unione (più che numerabile) di punti, tutti di lunghezza 0; addirittura la stessa retta reale, che ha lunghezza infinita, è composta da un'unione (più che numerabile) di punti (tutti di lunghezza 0).

A tale stesso proposito è utile guardare con attenzione la successiva ***Osservazione* *14.6***.

Inseriamo qui, solo per lettori particolarmente interessati, la dimostrazione della proprietà di additività finita, alla quale avevamo già accennato in precedenza.

***Proposizione* 14.2.** *Se $\mathbb{P}$ è σ-additiva allora essa è anche, necessariamente, finitamente additiva.*

Dimostrazione. Siano $E_1, ..., E_n$ sottoinsiemi appartenenti a $\mathcal{F}$ e a due a due disgiunti; ponendo

$$E_{n+1} = \emptyset, \quad E_{n+2} = \emptyset, ..., , \quad E_{n+k} = \emptyset, ...$$

otteniamo che anche la successione $\{E_i\}_{i=1,2,...}$ è composta di insiemi a due a due disgiunti.

D'altra parte si ha che

$$\bigcup_{i=1}^{n} E_i = \bigcup_{i=1}^{\infty} E_i$$

e dunque

$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

La dimostrazione si conclude basandosi sull'identità $\mathbb{P}(\emptyset) = 0$, da cui possiamo far seguire

$$\sum_{i=1}^{\infty} \mathbb{P}(E_i) = \sum_{i=1}^{n} \mathbb{P}(E_i).$$

Ora osserviamo però che l'identità $\mathbb{P}(\emptyset) = 0$ era stata ottenuta, come avevamo visto nella Lezione 2. dall'assioma $\mathbb{P}(\Omega) = 1$ e da quello dell'additività finita della $\mathbb{P}$. Non possiamo quindi darla per scontata a priori in questa dimostrazione, in cui si tratta proprio di verificare la validità della additività finita.

Possiamo però notare che $\mathbb{P}(\emptyset) = 0$ segue immediatamente anche dall'assunzione che $\mathbb{P}$ sia σ-additiva.
Possiamo infatti scrivere, ponendo $\mathbb{P}(\emptyset) = c$, $O_i = \emptyset, i = 1, 2, ...,$

$$\emptyset = \bigcup_{i=1}^{\infty} O_i$$

da cui segue

$$c = \sum_{i=1}^{\infty} c$$

in quanto $O_i, i = 1, 2, ...,$ sono a due a due disgiunti e possiamo dunque concludere che deve essere $c = 0$, in quanto una serie a termini costanti può risultare convergente solo se i suoi termini sono tutti nulli. $\square$

14.1.1 Altre modifiche e/o variazioni rispetto al caso di spazi finiti

Altri cambiamenti riguardano il fatto che in uno spazio di probabilità generale si possono presentare successioni (numerabili) di eventi. Consideriamo inizialmente il caso in cui tale successione formi una ***partizione numerabile***:

$$\mathcal{H} = \{H_i,\ i \geq 1\}, \qquad , \bigcup_{i \geq 1} H_i = \Omega, \quad H_i \cap H_j = \emptyset,\ i \neq j,$$

In questo caso, sempre nell'ipotesi che $\mathbb{P}(H_i) > 0$ per ogni $i \geq 1$, vale la ***formula delle probabilità totali***, ossia per ogni evento $E \in \mathcal{F}$ si ha

$$\mathbb{P}(E) = \sum_{i \geq 1} \mathbb{P}(E \cap H_i) = \sum_{i \geq 1} \mathbb{P}(H_i)\ \mathbb{P}(E|H_i).$$

Passiamo ora a definire cosa significa una successione di eventi ***completamente/globalmente indipendenti***:

Definizione 14.4 (successione di eventi globalmente indipendenti)**.** *Una successione di eventi*

$$\{E_j, \quad j \geq\}$$

si dice che forma una famiglia di eventi globalmente indipendenti se e solo se comunque qualunche sotto famiglia che contiene un numero finito di tali eventi, forma una famiglia di eventi globalmente indipendenti, ossia, per ogni $m \geq 2$ e per ogni $k_1,\ k_2, ..., k_m$ di numeri distinti fra loro

$$\mathbb{P}(\widetilde{E}_{k_1} \cap \widetilde{E}_{k_2} \cap \cdots \widetilde{E}_{k_m}) = \mathbb{P}(\widetilde{E}_{k_1})\,\mathbb{P}(\widetilde{E}_{k_2}) \cdots \mathbb{P}(\widetilde{E}_{k_m}),$$

dove, come al solito, $\widetilde{E}_{k_i}$ può essere uguale a E_{k_i} o a $\overline{E}_{k_i}$.

14.2 Definizione generale di variabile aleatoria

Continuando secondo lo schema introdotto nella Lezione 7, consideriamo una variabile aleatoria come un'applicazione a valori reali e definita su uno spazio campione, cioè su un insieme di eventi elementari, Ω. Tuttavia, come vedremo, al contrario del caso finito, non tutte le applicazioni saranno considerate variabili aleatorie (vedere la successiva **Definizione 14.5**)

Consideriamo una generica applicazione $X : \Omega \to \mathbb{R}$.

Come abbiamo già visto, nel caso in cui Ω è un'insieme di cardinalità finita, allora necessariamente anche il codominio $X(\Omega)$ è un insieme finito, che finora abbiamo indicato, ad esempio, con il simbolo $\{x_1, ..., x_n\}$.

Nel caso in cui Ω sia un insieme di cardinalità arbitraria, $X(\Omega)$ non è più necessariamente un insieme finito: $X(\Omega)$ potrebbe anche essere costituito da una successione $\{x_1, x_2, ...\}$ o addirittura da un intervallo (limitato o illimitato) di numeri reali.

Ciò non è da vedere come un problema, piuttosto si tratta di una possibilità in accordo con le esigenze della teoria della probabilità. Difatti, come già detto, per motivi di tipo sia teorico che applicativo a noi serve poter considerare variabili aleatorie a valori in insiemi numerabili o con la potenza del continuo.

Esempio 14.5. *Sia $\{E_1, E_2, ...\}$ una successione di eventi e siano $X_1, X_2, ...$ le corrispondenti variabili indicatrici.*

Consideriamo la variabile aleatoria

$$T := \inf\{n \geq 1 : X_n = 1\},$$

per la quale si ha

$$\{T = n\} = \{X_1 = 0, X_2 = 0, ..., X_{n-1} = 0, X_n = 1\}$$

e possiamo interpretare T come il numero (aleatorio) di prove necessarie fino ad ottenere il primo successo, nella successione di prove $\{E_1,\ E_2, ...\}$.

L'insieme dei valori possibili per T coincide con l'insieme dei numeri naturali[57] $\{1, 2, ...\}$.

Esempio 14.6. *In un test di affidabilità un'apparecchiatura elettrica appena prodotta viene testata lasciandola funzionare ininterrottamente fino a quando non smetta di funzionare a causa di qualche guasto. Indicando con T la lunghezza complessiva del tempo di durata realizzato dall'apparecchiatura, otteniamo una variabile aleatoria i cui valori possibili sono in linea di principio i valori reali positivi.*

Consideriamo ora specificamente il caso di una variabile aleatoria X definita su Ω e a valori in un insieme numerabile $X(\Omega) := \{x_1, x_2, ...\}$ e sia $\mathcal{F}$ la σ-algebra costituita da quei sottoinsiemi di Ω che consideriamo come eventi, cioè su cui è definita la misura di probabilità $\mathbb{P}$.

Fissiamo ora un generico elemento $x_j \in X(\Omega)$ e guardiamo al sottoinsieme

$$X^{-1}(\{x_j\}) := \{\omega \in \Omega : X(\omega) = x_j\}$$

Naturalmente vogliamo che $X^{-1}(\{x_j\})$ sia effettivamente un evento, cioè vogliamo poter parlare della "probabilità" che X assuma il valore x_j e possiamo far questo solo se risulta

$$X^{-1}(\{x_j\}) \in \mathcal{F}. \tag{119}$$

Per tale motivo diremo che l'applicazione $X : \Omega \to X(\Omega) := \{x_1, x_2, ...\}$ costituisce effettivamente una variabile aleatoria solo se la condizione (119) risulta verificata, $\forall x_j \in X(\Omega)$. Notiamo ora che, se X è una tale variabile aleatoria, anche un sottoinsieme del tipo $\{\omega \in \Omega : X(\omega) \leq z\}$ (o, scrivendo più brevemente, $\{X \leq z\}$) risulta appartenere a $\mathcal{F}$, $\forall\, z \in \mathbb{R}$, in quanto

$$\{X \leq z\} = \bigcup_{j:\, x_j \leq z} \{X = x_j\}.$$

[57] Il lettore più attento avrà notato che c'è qualche problema nel caso in cui

$$\{X_1 = 0,\ X_2 = 0, ..., X_{n-1} = 0,\ X_n = 0,\ X_{n+1} = 0 \ldots\} = \cap_{n=1}^{\infty}\{X_n = 0\}.$$

Infatti in tale caso T va definita in modo appropriato. Un modo potrebbe essere quello di includere il valore ∞ tra i valori che puo' assumere T, e allora bisognerebbe considerare anche le variabili aleatorie che assumono valori in $\mathbb{R} \cup \{\infty\}$. Un altro modo per eliminare il problema è invece di notare che l'evento $\cap_{n=1}^{\infty}\{X_n = 0\}$ ha probabilità nulla, e quindi la definizione di T su tale evento non cambia in alcun modo la probabilità degli eventi $\{T = k\}$ per $k \in \{1, 2, ...\}$.

Esercizio proposto 14.2. *Dimostrare che se $\mathcal{F}$ è una σ-algebra allora $X : \Omega \to X(\Omega) := \{x_1, x_2, ...\}$ è una variabile aleatoria se e solo*[58] *se risulta $\{\omega \in \Omega : X(\omega) \le z\} \in \mathcal{F}, \forall z \in \mathbb{R}$.*

Consideriamo ora il caso in cui $X(\Omega)$ è un intervallo (a, b) della retta reale.

Anche in questo caso richiederemo, affinchè X sia una variabile aleatoria ben definita, che risulti $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}, \forall x \in \mathbb{R}$.

Osserviamo però che nel presente caso tale condizione non implica $\{\omega \in \Omega : X(\omega) \le z\} \in \mathcal{F}, \forall z \in \mathbb{R}$; infatti un tale insieme non è più ottenibile, in generale, attraverso un'operazione di unione numerabile a partire da insiemi del tipo $\{\omega \in \Omega : X(\omega) = x\}$.

Giungiamo quindi alla seguente definizione:

Definizione 14.5 (Variabile aleatoria). *Un'applicazione $X : \Omega \to X(\Omega) \subseteq \mathbb{R}$ è una variabile aleatoria se e solo se un qualunque sottoinsieme del tipo $\{\omega \in \Omega : X(\omega) \le z\}$ risulta appartenere a $\mathcal{F}$, per ogni $z \in \mathbb{R}$.*

14.3 Distribuzioni di probabilità, funzioni di distribuzione

È opportuno iniziare la trattazione di questo argomento con l'analisi del caso di variabili aleatorie (che diremo ***discrete***), per metterne in evidenza alcune analogie con il caso di variabili aleatorie a valori in un insieme finito.

Definizione 14.6 (Variabili aleatorie discrete). *Sia $(\Omega, \mathcal{F}, P)$ uno spazio di probabilità e sia X una variabile aleatoria*

$$X : \Omega \to X(\Omega).$$

Diremo che X è una **variabile aleatoria discreta** *se $X(\Omega)$ è un insieme discreto (finito o numerabile) di numeri reali:*

$$X(\Omega) := \{x_1, x_2, ...\}.$$

Possiamo considerare a questo punto (analogamente a quanto avevamo già fatto nella Lezione 7) un nuovo spazio di probabilità $(X(\Omega), \mathcal{G}, \mathbf{P}_X)$ dove $\mathcal{G}$ coincide con $\mathcal{P}(X(\Omega))$ e $\mathbf{P}_X$ è la misura di probabilità su $\mathcal{G}$, individuata univocamente dalla seguenti posizioni

$$\mathbf{P}_X(\{x_j\}) = \mathbb{P}(\{X = x_j\}) = \mathbb{P}\left(X^{-1}(\{x_j\})\right) = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x_j)\}$$

$$\mathbb{P}_X(E) = \sum_{i: x_i \in E} \mathbf{P}_X(\{x_j\}), \qquad \forall E \in \mathcal{G}$$

(si vedano a tale proposito anche le considerazioni svolte nella precedente ***Osservazione 14.2***).

Tale misura di probabilità $\mathbf{P}_X$ è la misura indotta da X ed è anche detta ***distribuzione di probabilità*** di X.

Dobbiamo ricordare che siamo giustificati a considerare la probabilità dell'evento $\mathbb{P}\left(X^{-1}(\{x_j\})\right)$ in quanto, proprio per definizione della nozione di variabile aleatoria, deve risultare valida la condizione (119).

Dal momento che la famiglia degli eventi della forma $X^{-1}(\{x_j\}) := \{\omega \in \Omega : X(\omega) = x_j\}$, per $j = 1, 2, ...$ costituisce una partizione numerabile di insiemi appartenenti ad $\mathcal{F}$, in virtù dell'assioma di additività numerabile di $\mathbb{P}$, deve risultare

$$\sum_{j=1}^{\infty} \mathbb{P}(X = x_j) = 1. \tag{120}$$

[58] In riferimento alla nota precedente: se $X(\Omega)$ include anche il valore ∞, allora la condizione che X sia una variabile aleatoria diviene $\{\omega \in \Omega : X(\omega) \le z\} \in \mathcal{F}, \forall z \in \mathbb{R} \cup \{+\infty\}$.

Possiamo riassumere dicendo che, analogamente al caso finito, la distribuzione di probabilità di una variabile aleatoria discreta X può essere espressa assegnando l'insieme dei valori possibili $X(\Omega) = \{x_1, x_2, ...\}$ e le loro rispettive probabilità $\mathbb{P}(X = x_1)$, $\mathbb{P}(X = x_2), ...$

Le probabilità $\mathbb{P}(X = x_1)$, $\mathbb{P}(X = x_2), ...$ devono essere delle quantità non negative e deve essere rispettata la condizione di normalizzazione (120). La probabilità $\mathbb{P}(X = x)$, vista come funzione della variabile x, $x \in X(\Omega)$, viene chiamata "*funzione di densità discreta*" (o più brevemente "*densità discreta*" della variabile aleatoria X, e analogamente al caso finito, la densità discreta viene indicata come la funzione

$$x \mapsto p_X(x) = \mathbb{P}(X = x).$$

Esempio 14.7 (Variabili aleatorie geometriche come tempi di primo successo). *Consideriamo di nuovo una successione di eventi $\{E_1, E_2, ...\}$ ed assumiamo in particolare che, $\forall n \in \mathbb{N}$, E_1, E_2,..., E_n, costituiscano delle prove bernoulliane, di probabilità θ ($0 < \theta < 1$). Siano $X_1, X_2, ...$ le corrispondenti variabili indicatrici e sia T la variabile aleatoria definita da*

$$T := \inf\{n \geq 1 : X_n = 1\},$$

e che, interpretando $X_k = 1$ come successo alla k-sima prova, rappresenta il tempo di primo successo.

Come già osservato nel'Esempio 14.5, l'evento $\{T = n\}$ si puşcrivere, usando la successione $\{X_j,\ j \geq 1\}$ o la successione degli eventi $\{E_j,\ j \geq 1\}$ come

$$\{T = n\} = \{X_1 = 0, X_2 = 0, ..., X_{n-1} = 0, X_n = 1\} = \overline{E}_1 \cap \overline{E}_1 \cap \cdots \cap \overline{E}_{n-1} \cap E_n.$$

Da ciò segue ora

$$p_T(n) = \mathbb{P}(T = n) = (1 - \theta)^{n-1}\,\theta, \qquad n = 1, 2, ...$$

Si dice che $\boldsymbol{T}$ **segue una distribuzione geometrica di parametro** θ*, in simboli $T \sim Geom(\theta)$. Osserviamo che risulta*

$$\sum_{r=1}^{\infty} \mathbb{P}(T = r) = \theta \sum_{r=0}^{\infty} (1 - \theta)^r = \theta \frac{1}{1 - (1 - \theta)} = 1. \tag{121}$$

Oltre al tempo di primo successo, si può considerare anche la variabile aleatoria T' definita come **il numero degli insuccessi prima del primo successo**. *È immediato comprendere che*

$$T' = T - 1,$$

e quindi che T' può assumere tutti i valori interi $k \geq 0$, e che

$$p_{T'}(k) = \mathbb{P}(T' = k)\Big(= \mathbb{P}(T = k + 1)\Big) = (1 - \theta)^k \theta, \qquad k = 0, 1, 2, ...$$

Nella letteratura riguardante la probabilità c'è una qualche ambiguità per quello che riguarda le variabili aleatorie con distribuzione geometrica: in alcuni testi si dice che $X \sim Geom(\theta)$ se e solo se

$$p_X(n) = (1 - \theta)^{n-1}\theta, \qquad n \in \mathbb{N},$$

mentre in altri si dice che $Y \sim Geom(\theta)$ se e solo se

$$p_Y(n) = \ (1 - \theta)^k \theta, \qquad k \in \mathbb{N} \cup \{0\}.$$

Questo fatto può generare confusione: per evitare questa ambiguità distringueremo le due distribuzioni geometriche specificando, se necessario, se la variabile aleatoria assume valori in $\mathbb{N}$ o in $\mathbb{N} \cup \{0\}$. Ad esempio diremo che X è una **v.a. geometrica a partire da** 1, e che invece Y è una **v.a. geometrica a partire da** 0.

Oltre alla famiglia delle variabili aleatorie con distribuzione geometrica, ci sono altre famiglie di variabili aleatorie molto importanti, e che vedremo nel seguito:
la famiglia delle variabili aleatorie con ***distribuzione di Pascal***. Si introducono come tempi di n-simo successo in una successione di prove indipendenti e ugualmente probabili (vedere l'Esempio 15.3);
la famiglia delle variabili aleatorie con ***distribuzione di Poisson*** (vedere l'Esercizio-proposto 14.4). Si introducono come approssimazione, in un opportuno senso, di variabili aleatorie binomiali, come vedremo succesivamente nel teorema di approssimazione di Poisson (vedere il Teorema 14.1, a pagina 216).

Passiamo ora a considerare il caso in cui si tratti con una variabile aleatoria X tale che $X(\Omega)$ è un intervallo limitato o illimitato di numeri reali (quindi $X(\Omega)$ non ha più la potenza del discreto), potremo dire che ***la distribuzione di probabilità di*** X può essere specificata assegnando la probabilità di tutti gli eventi del tipo $\{X \in I\}$ dove I è un generico intervallo della retta[59]

$$\mathbf{P}_X(I) = \mathbb{P}(\{X \in I\}).$$

Notiamo infatti che nel presente caso non ha più senso "elencare" tutti i valori possibili per X (cioè tutti i singoli elementi appartenenti a $X(\Omega)$) e specificare le loro rispettive probabilità. Fra l'altro, per $x \in X(\Omega)$, la probabilità $\mathbb{P}(X = x)$ risulta uguale a 0 tranne al più che per un insieme numerabile di valori di x; ciò verrà verificato fra breve (***Osservazione 14.6***) e può comunque essere intuito ricordando quanto accennato nella precedente ***Osservazione 14.3***.

Già sappiamo che, essendo X una variabile aleatoria (si veda la precedente Definizione 14.5), tutti i sottoinsiemi del tipo

$$\{\omega \in \Omega : X(\omega) \le x\}, \quad x \in \mathbb{R}$$

sono degli eventi ed ha senso quindi considerare $\mathbb{P}(X \le x)$.

Definizione 14.7 (funzione di distribuzione)**.** *La funzione $F_X : \mathbb{R} \to [0,1]$ definita dalla relazione*

$$F_X(x) := \mathbb{P}(X \le x)$$

*viene detta **funzione di distribuzione** della variabile aleatoria X.*

In altre parole $\boldsymbol{F_X(x)}$ **esprime *la probabilità che*** $\boldsymbol{X}$ ***cada nell'intervallo*** $(-\infty, x]$. Se conosciamo dunque la distribuzione di probabilità di X, possiamo in particolare ricavare $F_X(x), \forall x \in \mathbb{R}$.

Viceversa assegnare la funzione di distribuzione di una variabile aleatoria X equivale a conoscere interamente la distribuzione di probabilità di X, come si può verificare facilmente:

• Possiamo infatti dire di conoscere la distribuzione di probabilità di X quando sappiamo assegnare la probabilità di tutti gli eventi del tipo $\{X \in I\}$, dove I è un generico intervallo della retta.

• In effetti è possibile calcolare $\mathbf{P}_X(I) = \mathbb{P}(\{X \in I\})$ in termini della funzione $F_X(x)$, e in particolare per ogni coppia di valori $a < b \in \mathbb{R}$, risulta

$$\mathbb{P}(a < X \le b) = F_X(b) - F_X(a)\,. \tag{122}$$

[59]Gli intervalli I possono essere limitati o illimitati, aperti o chiusi, sia a destra che a sinistra:

$$(a,b], \quad (a,b), \quad [a,b), \quad [a,b],$$

$$(-\infty,b], \quad (-\infty,b), \quad (a,\infty) \quad [a,-\infty).$$

Possiamo scrivere infatti

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

dove i due eventi $\{X \leq a\}$ e $\{a < X \leq b\}$ sono, ovviamente, fra loro incompatibili; dunque

$$\mathbb{P}\left(X \leq b\right) = \mathbb{P}\left(X \leq a\right) + \mathbb{P}\left(a < X \leq b\right)$$

da cui la (122) segue immediatamente ricordando la definizione di F_X.

Analogamente si può verificare che risulta

$$\mathbb{P}(a < X < b) = F_X\left(b\right) - F_X\left(a\right) - \mathbb{P}\left(X = b\right) \tag{123}$$

$$\mathbb{P}(a \leq X < b) = F_X\left(b\right) - F_X\left(a\right) - \mathbb{P}\left(X = b\right) + \mathbb{P}\left(X = a\right) \tag{124}$$

$$\mathbb{P}(a \leq X \leq b) = F_X\left(b\right) - F_X\left(a\right) + \mathbb{P}\left(X = a\right). \tag{125}$$

È possibile anche dare delle formule per ottenere le probabilità degli eventi del tipo precedente, che dipendono solo da F. Infatti è possibile dimostrare (si veda più avanti la formula (127)) che, indicando come consueto con $F_X(x^-)$ il limite da sinistra di F, ovvero

$$F_X(x^-) := \lim_{y \to x^-} F_X(y)$$

si ha

$$\mathbb{P}(X < x) = F_X(x^-).$$

Da ciò è immediato ottenere che $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$ (si veda più avanti l'***Osservazione 14.6***) e che

$$\mathbb{P}(a < X < b) = F_X\left(b^-\right) - F_X\left(a\right)$$

$$\mathbb{P}(a \leq X < b) = F_X\left(b^-\right) - F_X\left(a^-\right)$$

$$\mathbb{P}(a \leq X \leq b) = F_X\left(b\right) - F_X\left(a^-\right).$$

D'ora in poi, quindi, la distribuzione di probabilità di una variabile aleatoria X verrà specificata assegnando la funzione $F_X\left(x\right)$.

***Osservazione* 14.4.** Nel caso di una variabile aleatoria X con $X\left(\Omega\right)$ coincidente con un intervallo, la funzione di distribuzione $F_X\left(x\right)$ è lo strumento naturale per individuare la sua distribuzione di probabilità $\mathbf{P}_X$. Ma, indipendentemente dal fatto che $X\left(\Omega\right)$ sia un insieme finito, un insieme numerabile, oppure un intervallo, è sempre possibile definire la $F_X\left(x\right)$.

Supponiamo ora in particolare che risulti $X\left(\Omega\right) := \{x_1, ..., x_n\}$, con $x_1 < x_2 < ... < x_n$ e $\mathbb{P}(X = x_1) = p_1, ..., \mathbb{P}(X = x_n) = p_n$. Si ha allora

$$\{X \leq x\} = \emptyset, \quad \text{per } x < x_1,$$

$$\{X \leq x\} = \{X = x_1\}, \quad \text{per } x_1 \leq x < x_2,$$

$$\dots\dots$$

$$\{X \leq x\} = \{X = x_1\} \cup \{X = x_2\} \cup \cdots \cup \{X = x_k\}, \quad \text{per } x_k \leq x < x_{k+1},$$

$$\dots\dots$$

$$\{X \leq x\} = \{X = x_1\} \cup \{X = x_2\} \cup \cdots \cup \{X = x_n\} = \Omega, \quad \text{per } x \geq x_n,$$

e quindi

$$F_X(x) = \begin{cases} 0 & \text{per } x < x_1, \\ p_1 & \text{per } x_1 \leq x < x_2, \\ p_1 + p_2 & \text{per } x_2 \leq x < x_3, \\ \cdots & \cdots, \\ \cdots & \cdots, \\ \sum_{j=1}^{n-1} p_j & \text{per } x_{n-1} \leq x < x_n, \\ 1 & \text{per } x \geq x_n. \end{cases}$$

Vediamo dunque che, in questo caso, $F_X(x) : \mathbb{R} \to [0,1]$ è una funzione costante a tratti, con salti positivi, rispettivamente di ampiezza $p_1, p_2, ...p_n$, nei punti $x_1, x_2, ...x_n$. Ne segue che $F_X(x)$ è una funzione continua da destra e monotona non decrescente.

Esempio 14.8. *Sia Z è una variabile aleatoria discreta finita, con*

$$Z(\Omega) := \{z_1 = -1, z_2 = 0, z_3 = 1\}$$

e con

$$\mathbb{P}(Z = -1) = \mathbb{P}(Z = 0) = \mathbb{P}(Z = 1) = \frac{1}{3}.$$

Si ha allora

$$F_Z(x) = \begin{cases} 0 & \textit{per } x < -1, \\ \frac{1}{3} & \textit{per } -1 \leq x < 0, \\ \frac{2}{3} & \textit{per } 0 \leq x < 1, \\ 1 & \textit{per } x \geq 1, \end{cases}$$

il cui grafico è riportato in Figura 1.

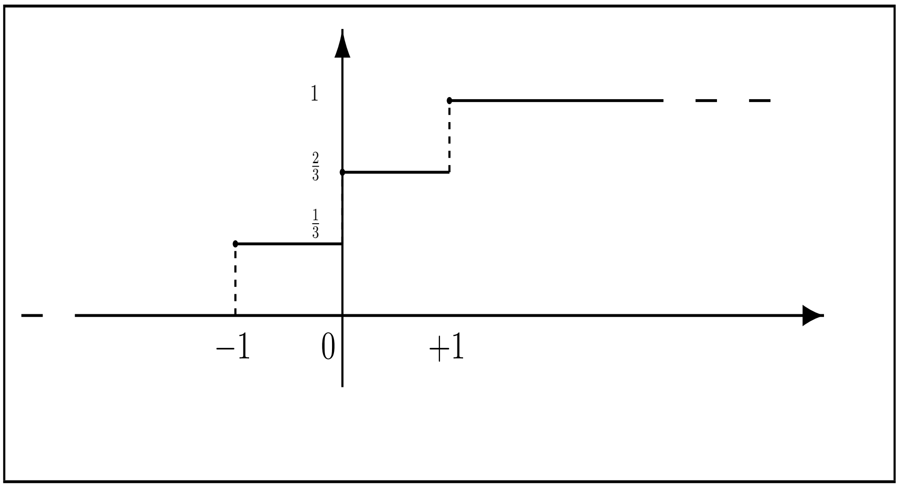

Figura 1: Grafico di $F_Z(x)$

Consideriamo ora una variabile aleatoria X arbitraria, cioè con $X(\Omega)$ non necessariamente finito. Oltre alla ovvia proprietà $0 \leq F_X(x) \leq 1$, vale in generale il seguente risultato.

***Proposizione* 14.3.** *La funzione di distribuzione F_X deve necessariamente soddisfare le seguenti proprietà:*
(i) la funzione F_X è non decrescente
(ii) la funzione $F_X(x)$ è normalizzata, ossia: $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$
(iii) F_X è continua da destra, cioè, $\forall \tilde{x} \in \mathbb{R}$, risulta

$$\lim_{x\to\tilde{x}^+} F_X(x) = F_X(\tilde{x}).$$

Dimostrazione.
(i) Questa proprietà segue immediatamente dalla relazione (122). Infatti, per $a < b \in \mathbb{R}$, risulta

$$F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b) \geq 0.$$

(ii) Osserviamo che possiamo scrivere

$$\lim_{x\to-\infty} F_X(x) = \lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}(X \leq -n);$$

ponendo

$$B_n := \{X \leq -n\}$$

otteniamo una successione tale che, per $n = 1, 2, \ldots$, $B_n \supseteq B_{n+1}$ e $\bigcap_{n=1}^{\infty} B_n = \emptyset$.
Dunque la condizione $\lim_{x\to-\infty} F_X(x) = 0$ si ottiene facilmente ricordando la ***Proposizione* 14.1**. Analogamente possiamo verificare la necessità della condizione $\lim_{x\to+\infty} F_X(x) = 1$, osservando che possiamo scrivere

$$\lim_{x\to\infty} F_X(x) = \lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} \mathbb{P}(\{X \leq n\});$$

ponendo

$$A_n := \{X \leq n\}$$

otteniamo una successione tale che $A_1 \subseteq A_2 \subseteq \ldots$ e $\bigcup_{n=1}^{\infty} A_n = \Omega$.
(iii) Notiamo innanzitutto che, essendo F_X una funzione monotona (non decrescente), i suoi eventuali punti di discontinuità possono essere soltanto punti con discontinuità di 1^a specie; si ha cioè che, $\forall \tilde{x} \in \mathbb{R}$, risultano esistere i due limiti

$$F_X\left(\tilde{x}^+\right) := \lim_{x\to\tilde{x}^+} F_X(x), \qquad F_X\left(\tilde{x}^-\right) := \lim_{x\to\tilde{x}^-} F_X(x).$$

Si tratta ora di verificare che risulta

$$F_X\left(\tilde{x}^+\right) = F_X(\tilde{x}).$$

Essendo F_X monotona non decrescente basta mostrare che risulta

$$\lim_{n\to\infty} F_X(\tilde{x} + \tfrac{1}{n}) = F_X(\tilde{x}).$$

Gli eventi $B_n = \{X \leq \tilde{x} + \frac{1}{n}\}$ costituiscono una successione non crescente e posto

$$B = \bigcap_{n=1}^{\infty} \{X \leq \tilde{x} + \tfrac{1}{n}\}$$

abbiamo, per la proprietà di continuità di P,

$$\lim_{n\to\infty} F_X(\tilde{x} + \tfrac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(X \leq \tilde{x} + \tfrac{1}{n}) = \mathbb{P}(B).$$

Bisogna ora dimostrare che risulta
$$B = \{X \leq \tilde{x}\}.$$
Notiamo allora che si ha ovviamente $\{X \leq \tilde{x}\} \subseteq B$, in quanto[60] $\{X \leq \tilde{x}\} \subseteq \{X \leq \tilde{x}+\frac{1}{n}\} = B_n$, $\forall n = 1, 2, \ldots$. D'altra parte se $\omega \in B$, vale anche[61] $X(\omega) \leq \tilde{x} + \frac{1}{n}$ per $\forall n = 1, 2, \ldots$ e dunque $X(\omega) \leq \tilde{x} + \lim_{n\to\infty} \frac{1}{n} = \tilde{x}$, cioè $\omega \in \{X \leq \tilde{x}\}$. □

Pur senza darne qui dimostrazione, è opportuno citare che, viceversa, per qualunque funzione $F : \mathbb{R} \to [0, 1]$, che soddisfi le proprietà (i), (ii), (iii) di cui nella precedente ***Proposizione* 14.3**, è possibile costruire un'opportuno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ ed una variabile aleatoria X su Ω tali che risulti
$$F_X(x) = F(x), \qquad \forall x \in \mathbb{R}.$$
L'esempio più semplice di funzione di distribuzione è la funzione
$$x \mapsto F(x) = \begin{cases} 0 & \text{per } x < 0, \\ x & \text{per } 0 \leq x < 1, \\ 1 & \text{per } x \geq 1. \end{cases}$$
Come vedremo nell'Esempio 14.14, una variabile aleatoria con questa funzione come funzione di distribuzione viene detta avere distribuzione uniforme in $(0, 1)$.

Esempio 14.9 (Variabili aleatorie di Cauchy). *Sia $F : \mathbb{R} \to [0, 1]$ definita da*
$$F(x) = \tfrac{1}{\pi} \arctan(x) + \tfrac{1}{2}. \tag{126}$$
Si vede facilmente che questa funzione soddisfa le proprietà (i), (ii), (iii) della **Proposizione** ***14.3****: la prima e la terza sono banali, e la seconda discende dal fatto che*
$$\lim_{x\to-\infty} \arctan(x) = -\tfrac{\pi}{2} \qquad e \qquad \lim_{x\to+\infty} \arctan(x) = \tfrac{\pi}{2}.$$
Sia X una variabile aleatoria, con funzione di distribuzione $F_X(x) = F(x)$, con F definita come in (126). Allora si dice che X **ha distribuzione di Cauchy** *e inoltre, ad esempio,*
$$\begin{aligned}\mathbb{P}\big(X \in (-1, 1]\big) &= F(1) - F(-1) = \tfrac{1}{\pi} \arctan(1) + \tfrac{1}{2} - \big(\tfrac{1}{\pi} \arctan(-1) + \tfrac{1}{2}\big) \\ &= \tfrac{1}{\pi}\tfrac{\pi}{4} - \tfrac{1}{\pi}\big(-\tfrac{\pi}{4}\big) = \tfrac{1}{2}.\end{aligned}$$

***Osservazione* 14.5.** Si verifica immediatamente, ancora in base alla (122), che F_X si mantiene costante in un intervallo (a, b) se e solo se risulta
$$\mathbb{P}(a < X < b) = 0.$$

***Osservazione* 14.6 (Significato dei punti di discontinuità di F_X).** Abbiamo già notato nel corso della dimostrazione della ***Proposizione* 14.3** che, essendo F_X una funzione monotona risultano esistere, $\forall \tilde{x} \in \mathbb{R}$, i due limiti
$$F_X\left(\tilde{x}^+\right), \qquad F_X\left(\tilde{x}^-\right).$$
Tenendo conto della continuità da destra di F_X, sappiamo che risulta
$$F_X\left(\tilde{x}^+\right) = F_X\left(\tilde{x}\right) = \mathbb{P}(X \leq \tilde{x}).$$

[60] Se $\omega \in \{X \leq \tilde{x}\}$, cioè se $X(\omega) \leq \tilde{x}$ allora ovviamente $X(\omega) \leq \tilde{x} + \frac{1}{n}$, cioè $\omega \in \{X \leq \tilde{x} + \frac{1}{n}\} = B_n$.

[61] Si ha $\omega \in B$ se e solo se $\omega \in B_n$ per ogni $n = 1, 2, \ldots$, ovvero se e solo se $X(\omega) \leq \tilde{x} + \frac{1}{n}$ per ogni $n = 1, 2, \ldots$. Ma allora $X(\omega) \leq \lim_{n\to\infty} \tilde{x} + \frac{1}{n} = \tilde{x}$, ovvero $\omega \in \{X(\omega) \leq \tilde{x}\}$.

Per quanto riguarda $F_X\left(\tilde{x}^-\right)$, notiamo che possiamo anche scrivere

$$F_X\left(\tilde{x}^-\right) = \lim_{x\to\tilde{x}^-} F_X(x) = \lim_{n\to\infty} F_X(\tilde{x} - \tfrac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(X \leq \tilde{x} - \tfrac{1}{n})$$

e dunque[62] che, per la proprietà di continuità di P, possiamo concludere

$$F_X\left(\tilde{x}^-\right) = \mathbb{P}(X < \tilde{x}). \tag{127}$$

La precedente osservazione è alla base della interessante proprietà:
i valori reali che una variabile aleatoria X assume con probabilità positiva sono tutti e soli i punti di discontinuità per F_X, e inoltre, se $\tilde{x}$ è un punto di discontinuità per F, allora

$$\mathbb{P}(X = \tilde{x}) = F_X(\tilde{x}) - F_X(\tilde{x}^-),$$

(si confronti anche la nota a pagina 201).
Verifichiamo immediatamente questa proprietà, osservando che

I Se esiste qualche valore $\tilde{x} \in \mathbb{R}$, tale che $\mathbb{P}(X = \tilde{x}) = \widetilde{p} > 0$, deve risultare allora

$$\mathbb{P}(X \leq \tilde{x}) = \mathbb{P}(X < \tilde{x}) + \widetilde{p} > \mathbb{P}(X < \tilde{x}) \tag{128}$$

in quanto[63], ovviamente,

$$\mathbb{P}(X = \tilde{x}) = \mathbb{P}(X \leq \tilde{x}) - \mathbb{P}(X < \tilde{x}).$$

Dalla (128) segue

$$F_X(\tilde{x}) > F_X(\tilde{x}^-),$$

cioè $\tilde{x}$ risulta essere un punto di salto (e quindi punto di discontinuità) per la funzione F_X.

II Sia ora $\tilde{x}$ un qualunque punto di discontinuità per F_X. In virtù del fatto che F_X è non decrescente (e quindi può solo avere punti di discontinuità di 1^a specie), risulta che

$$F_X(\tilde{x}^+) = F_X(\tilde{x}) > F_X(\tilde{x}^-)$$

cioè, ricordando ancora una volta la definizione di F_X e la (127)

$$\mathbb{P}(X \leq \tilde{x}) = \mathbb{P}(X < \tilde{x}) + \rho_{\tilde{x}},$$

e dunque

$$\mathbb{P}(X = \tilde{x}) = \rho_{\tilde{x}},$$

essendo $\rho_{\tilde{x}} = F_X(\tilde{x}) - F_X(\tilde{x}^-)$ una quantità strettamente positiva.

---

[62]Infatti la successione di eventi $A_n = \{X \leq \tilde{x} - \frac{1}{n}\}$ è monotona non decrescente e

$$\{X < \tilde{x}\} = \bigcup_{n=1}^{\infty} \{X \leq \tilde{x} - \tfrac{1}{n}\}$$

e quindi, per la la proprietà di continuità delle probabilità, $\mathbb{P}(X < \tilde{x}) = \lim_{n\to\infty} \mathbb{P}(X \leq \tilde{x} - \frac{1}{n})$.

[63]Si osservi che, alla luce della definizione di funzione di distribuzione e della (127), si può riscrivere la relazione $\mathbb{P}(X = \tilde{x}) = \mathbb{P}(X \leq \tilde{x}) - \mathbb{P}(X < \tilde{x})$, come

$$\mathbb{P}(X = \tilde{x}) = F_X(\tilde{x}) - F_X(\tilde{x}^-).$$

Un risultato dell'Analisi Matematica mostra che una funzione monotona (quale risulta essere la F_X) ammette, al più, un insieme numerabile di punti di discontinuità; quindi:
tutti i punti sulla retta reale hanno probabilità* 0 *per la variabile aleatoria* X*, tranne al più un insieme finito o una successione numerabile.

> Si può dare la dimostrazione del fatto che la funzione F_X ammette, al massimo, un insieme numerabile di punti di discontinuità, utilizzando l'interpretazione probabilistica dei punti di discontinuità e notando che, per ogni intero $n \geq 1$,
>
> $$\{\tilde{x} \in \mathbb{R} : \quad \mathbb{P}(X = \tilde{x}) \geq \tfrac{1}{n}\}$$
>
> può contenere al massimo n punti. Per terminare la dimostrazione bisogna poi osservare che la famiglia dei punti di discontinuità di F_X è l'unione numerabile degli insiemi di punti in cui $\mathbb{P}(X = \tilde{x}) = F_(\tilde{x}) - F_X(\tilde{x}^-) \geq \frac{1}{n}$.

Prima di passare oltre, vediamo un esempio naturale di distribuzione di probabilità, per una variabile aleatoria X, con funzione di distribuzione F_X ovunque continua (e quindi tale che $\mathbb{P}(X = x) = 0$, $\forall x \in \mathbb{R}$)

Esempio 14.10 (v.a.uniformi). *Sia (α, β) un fissato intervallo della retta reale. Per ogni $x_1 \leq x_2 \in \mathbb{R}$, indichiamo con $\big|(\alpha, \beta) \cap [x_1, x_2]\,\big|$ la lunghezza dell'intervallo $(\alpha, \beta) \cap [x_1, x_2]$:*

$$\big|(\alpha, \beta) \cap [x_1, x_2]\,\big| := \begin{cases} \min(\beta, x_2) - \max(\alpha, x_1) & \textit{se } \min(\beta, x_2) \geq \max(\alpha, x_1) \\ \\ 0 & \textit{altrimenti} \end{cases}$$

Sia ora X una variabile aleatoria tale che

$$\mathbb{P}(x_1 \leq X \leq x_2) = \frac{\big|(\alpha, \beta) \cap [x_1, x_2]\,\big|}{|(\alpha, \beta)|}.$$

Osserviamo innanzitutto che, $\forall x \in \mathbb{R}$, risulta

$$\mathbb{P}(X = x) = 0$$

(e quindi F_X risulta essere una funzione ovunque continua); inoltre si ha

$$\mathbb{P}(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{\beta - \alpha} \qquad \textit{nel caso in cui } \alpha \leq x_1 \leq x_2 \leq \beta,$$

e

$$\mathbb{P}(x_1 \leq X \leq x_2) = 0 \qquad \textit{nel caso in cui gli intervalli } (\alpha, \beta) \textit{ e } (x_1, x_2) \textit{ abbiano intersezione vuota.}$$

In un tale caso diremo che X segue **una distribuzione uniforme nell'intervallo** (α, β) *e scriveremo in simboli*

$$X \sim R(\alpha, \beta), \quad \textit{dove } R \textit{ sta per } \textbf{\textit{Random}} \qquad \textit{o anche } X \sim Unif(\alpha, \beta).$$

Determiniamo ora la funzione di distribuzione F_X. Si ha

$$\begin{aligned} F_X(x) = \mathbb{P}(X \leq x) &= \lim_{x_1 \to -\infty} \mathbb{P}(x_1 \leq X \leq x) \\ &= \lim_{x_1 \to -\infty} \frac{\big|(\alpha, \beta) \cap [x_1, x_2]\,\big|}{\beta - \alpha}. \end{aligned}$$

Si verifica facilmente che risulta

$$F_X(x) = \begin{cases} 0 & \textit{per } x \leq \alpha, \\ \dfrac{x - \alpha}{\beta - \alpha} & \textit{per } \alpha \leq x \leq \beta, \\ 1 & \textit{per } x \geq \beta. \end{cases} \tag{129}$$

Nel caso particolare $\alpha = 0$, $\beta = 1$, si ha

$$F_X(x) = \begin{cases} 0 & \textit{per } x \leq 0, \\ x & \textit{per } 0 \leq x \leq 1, \\ 1 & \textit{per } x \geq 1. \end{cases} \tag{130}$$

Il grafico di F_X, per X uniforme è riportato in Figura 2.

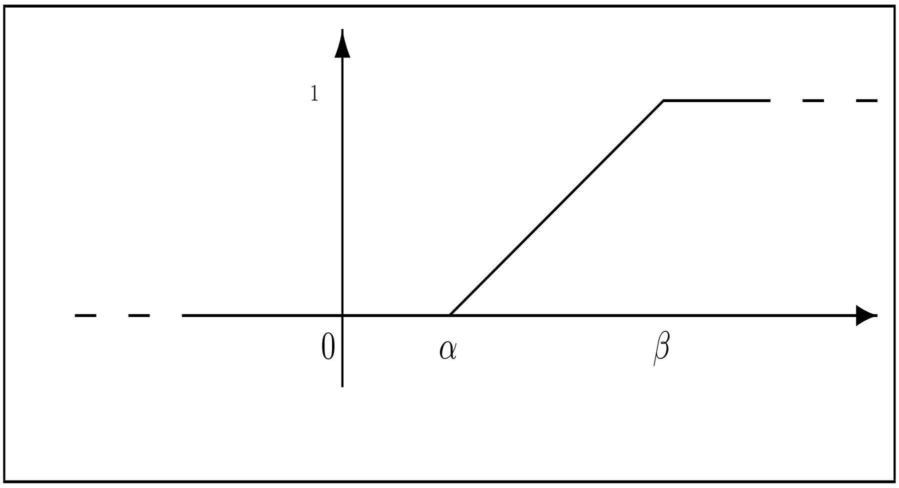

Figura 2: Grafico di $F_X(x)$, per X uniforme in (α, β)

14.4 Funzioni di distribuzione continue, funzioni di densità di probabilità

Nel caso in cui F_X risulti essere una funzione continua, diremo che la variabile aleatoria X è una ***variabile aleatoria continua***, e inoltre, come abbiamo già detto, avremo

$$\mathbb{P}(X = x) = 0, \qquad \forall x \in \mathbb{R}. \tag{131}$$

Questo fatto da una parte comporta una conseguenza piacevole:

$$\mathbb{P}(a < X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b), \tag{132}$$

(come si vede subito, tenendo conto delle relazioni (122), (123), (124) e (125)), ma d'altra parte pone un importante problema:

Come facciamo allora ad esprimere sinteticamente quali sono le zone dei valori più o meno probabili per la variabile aleatoria X?

Prima di dare la soluzione che in alcuni casi si può dare a questo problema (si veda la Definizione 14.8), proviamo a vedere come si potrebbe ragionare. Può venire spontaneo di procedere come segue: fissiamo un numero δ finito, positivo, magari abbastanza piccolo e analizziamo, al variare di $x \in \mathbb{R}$, l'andamento della probabilità

$$\mathbb{P}(x \leq X \leq x + \delta).$$

Nel caso di X distribuita in modo uniforme sull'intervallo $(0,1)$, ad esempio, otterremo

$$\mathbb{P}(x \leq X \leq x+\delta) = \begin{cases} 0 & \text{per } x \leq -\delta, \\ \delta - |x| & \text{per } -\delta \leq x \leq 0, \\ \delta & \text{per } 0 \leq x \leq 1-\delta, \\ 1-x & \text{per } 1-\delta \leq x \leq 1, \\ 0 & \text{per } x \geq 1. \end{cases}$$

Può venire anche spontaneo di vedere che cosa succede prendendo la quantità δ sempre più piccola; ma ovviamente, sempre in virtù della continuità di F_X, avremo, $\forall x \in \mathbb{R}$,

$$\lim_{\delta \to 0} \mathbb{P}(x \leq X \leq x+\delta) = \lim_{\delta \to 0} F_X(x+\delta) - F_X(x) = 0.$$

Al variare di $x \in \mathbb{R}$, studieremo allora piuttosto

$$\lim_{\delta \to 0} \frac{\mathbb{P}(x \leq X \leq x+\delta)}{\delta}. \tag{133}$$

Vediamo ora cosa possiamo ottenere in un caso abbastanza generale.

***Proposizione* 14.4.** *Sia F_X una funzione continua e derivabile in ogni x. Allora il limite in (133) esiste e risulta*

$$\lim_{\delta \to 0} \frac{\mathbb{P}(x \leq X \leq x+\delta)}{\delta} = F_X'(x),$$

essendo $F_X'(x)$ il valore in x della derivata prima di F_X.

Sia inoltre F_X', la derivata prima di F_X, una funzione continua. Allora, per ogni $a < b$, si ha

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \int_a^b F_X'(x)\, dx. \tag{134}$$

Dimostrazione. Essendo F_X continua, risulta

$$\mathbb{P}(x \leq X \leq x+\delta) = \mathbb{P}(x < X \leq x+\delta)$$

e dunque, essendo F_X anche derivabile in x,

$$\lim_{\delta \to 0} \frac{\mathbb{P}(x \leq X \leq x+\delta)}{\delta} = \lim_{\delta \to 0} \frac{F_X(x+\delta) - F_X(x)}{\delta} = F_X'(x).$$

Per dimostrare la (134) basta ricordare che, come sappiamo dall'Analisi, una funzione G continua e derivabile, con derivata prima continua, risulta essere l'integrale della sua derivata G', cioè risulta, per ogni intervallo $[a,b]$:

$$\int_a^b G'(x)\, dx = G(b) - G(a).$$

Infatti si puo' applicare la precedente relazione ad F_X, tenendo conto di (132), e del fatto che

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

La precedente ***Proposizione* 14.4**, si applica immediatamente al caso di una variabile aleatoria X con distribuzione di Cauchy (confrontare l'Esempio 14.9) per il quale si ha

$$F'_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

Tuttavia la ***Proposizione* 14.4** non si applica al caso della variabile aleatoria uniforme, la cui funzione di distribuzione è derivabile solo per $x \neq \alpha, \beta$. La derivata vale $\frac{1}{\beta-\alpha}$ per $x \in (\alpha, \beta)$, mentre vale 0 sia in $(-\infty, \alpha)$ che in (β, ∞). Tuttavia, è possibile mostrare che in questo caso la relazione (134) è valida (confrontare la ***Proposizione* 14.5**). Queste osservazioni suggeriscono la seguente definizione.

Definizione 14.8 (funzione di densità di probabilità). *Sia X una variabile aleatoria, e sia f una funzione con $f(x) \geq 0\ \forall x \in \mathbb{R}$, e per la quale, qualunque sia l'intervallo $[a, b]$, risulta*

$$\mathbb{P}\left(a \leq X \leq b\right) = \int_a^b f\left(x\right) dx.$$

Diremo allora che la distribuzione di probabilità di X **ammette densità**. *La funzione f viene detta* **funzione di densità di probabilità** *(o semplicemente* **funzione di densità***) e viene usualmente indicata con il simbolo f_X.*

Ovviamente se X ammette densità allora la sua funzione di distribuzione è continua, infatti

$$\mathbb{P}(X = a) = \mathbb{P}(X \in [a, a]) = \mathbb{P}(a \leq X \leq a) = \int_a^a f\left(x\right) dx = 0.$$

In generale, la funzione di densità non è definita univocamente in tutti i punti, come mostra la seguente ***Proposizione* 14.5**. A questo proposito si veda anche la successiva ***Osservazione 14.8***.

***Proposizione* 14.5.** *Condizione sufficiente affinché X ammetta densità è che*

(i) *$F_X(x)$ sia continua,*

(ii) *$F_X(x)$ sia derivabile, tranne al più un numero finito m di punti x_i, $i = 1, \ldots, m$,*

(iii) *la derivata prima $F'_X(x)$ sia continua in ciascuno degli $m+1$ intervalli in cui i punti x_i, $i = 1, \ldots, m$ dividono la retta, ovvero, supponendo $x_1 < x_2 < \cdots < x_m$, in ciascuno degli intervalli*

$$(-\infty, x_1), \quad (x_1, x_2), \quad \cdots, (x_k, x_{k+1}), \cdots, (x_{m-1}, x_m), \quad (x_m, \infty),$$

(iv) *la derivata prima $F'_X(x)$ sia prolungabile con continuità su ciascuno degli intervalli*

$$(-\infty, x_1], \quad [x_1, x_2], \quad \cdots, [x_k, x_{k+1}], \cdots, [x_{m-1}, x_m], \quad [x_m, \infty),$$

preso separatamente.

In tale caso X ammette come densità di probabilità qualsiasi funzione f che coincida con la derivata prima di F_X, naturalmente dove quest'ultima esiste, ossia

$$f(x) = F'_X(x), \qquad \forall x \neq x_i.$$

Invece nei punti x_i la funzione f non deve soddisfare nessuna condizione.

Dimostrazione: omessa.

Come applicazione abbiamo il caso delle variabili uniformi, ma si veda anche l'Esempio 14.15.

Esempio 14.11 (densità di una v.a. uniforme)**.** *Sia X una variabile aleatoria con distribuzione uniforme nell'intervallo (α, β), secondo quanto definito nel precedente Esempio 14.10. Derivando rispetto a x la funzione di distribuzione F_X, otteniamo la funzione di densità di probabilità:*

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ \\ 0 & x \notin [\alpha, \beta] \end{cases}$$

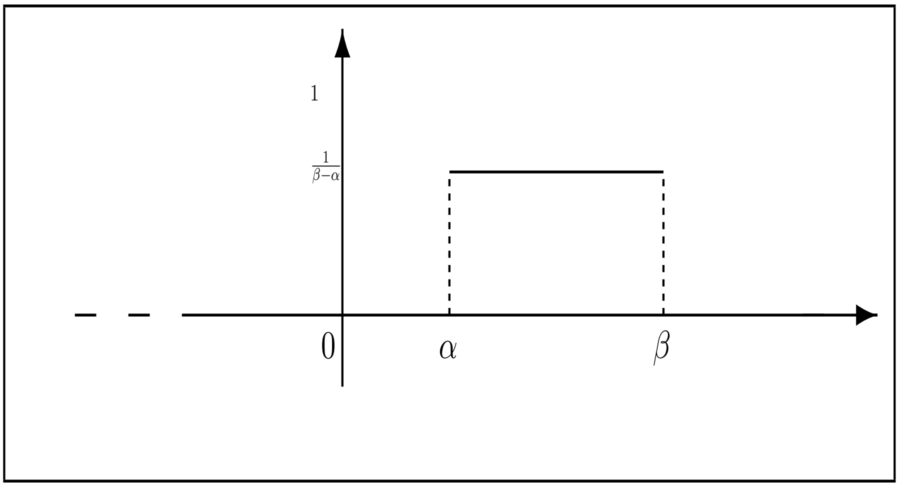

Figura 3: Grafico di $f_X(x)$, per X uniforme in (α, β)

***Osservazione* 14.7.** Se la la distribuzione di probabilità di X ammette densità $f_X(x)$ allora, per ogni valore $x \in \mathbb{R}$, risulta

$$\mathbb{P}(X \leq x) = \mathbb{P}(-\infty < X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(\xi), d\xi \tag{135}$$

come si ottiene subito facendo il limite per a che tende a $-\infty$ in

$$\mathbb{P}(a \leq X \leq x) = \int_a^b f_X(\xi)\ d\xi.$$

Inoltre, mandando x a $+\infty$ nella (135), si ottiene

$$1 = \mathbb{P}(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(\xi)\ d\xi.$$

Infine va sottolineato che, per stabilire che X ammette densità $f_X = f$ è sufficiente verificare che

$$F_X(x) = \int_{-\infty}^{x} f(\xi), d\xi$$

dal momento che

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f(x)\, dx - \int_{-\infty}^{a} f(x)\, dx = \int_a^b f(\xi)\ d\xi$$

***Osservazione* 14.8.** Sia F_X una funzione di distribuzione con densità f_X e sia g una funzione tale che

$$g(x) = f_X\,(x)$$

per tutti i punti x della retta (tranne, al più, per quelli in un insieme numerabile). Allora risulta anche per ogni intervallo $[a, b]$

$$\mathbb{P}\,(a \leq X \leq b) = \int_a^b g\,(x)\,dx.$$

In tal caso anche $g\,(x)$ viene detta *densità* per F_X; possiamo dire dunque che non esiste un'unica funzione di densità, bensì un'intera famiglia di funzioni di densità. Tale famiglia costituisce una classe di equivalenza: essa contiene la classe di tutte le funzioni non negative che si differenziano dalla f_X soltanto su un insieme di punti finito o numerabile.

Da quanto sopra deriva anche la seguente

***Osservazione* 14.9.** Una qualunque funzione di densità f gode delle seguenti proprietà

(i) $f\,(x) \geq 0$

(ii) $\displaystyle\int_{-\infty}^{+\infty} f\,(x)\,dx = 1$

Spesso quindi, invece di individuare la distribuzione di una variabile aleatoria attraverso la funzione di distribuzione si preferisce definire la distribuzione di probabilità attraverso una funzione che goda delle proprietà **(i)** e **(ii)**, come nel caso degli Esempi 14.13 e 14.14, delle v.a. esponenziali e gaussiane, rispettivamente.

Altre volte invece si definisce la densità a meno di una costante di proporzionalità, come spiegato nel seguente esempio generale.

Esempio 14.12 (densità di probabilità definite a meno di un fattore di proporzionalità). *Sia $g(x)$ una funzione assegnata a valori non negativi e sia k una costante positiva; la funzione*

$$f(x) = k \cdot g(x)$$

è una funzione di densità se e solo se esiste finito l'integrale $\int_{-\infty}^{+\infty} g\,(x)\,dx$ e risulta

$$k = \frac{1}{\int_{-\infty}^{+\infty} g\,(x)\,dx}$$

cioè

$$f(x) = \frac{g(x)}{\int_{-\infty}^{+\infty} g\,(x)\,dx}.$$

Per un'applicazione di questo modo di procedere si veda l'Esempio 14.16, alla fine di questa Lezione.

Esempio 14.13 (v.a. esponenziali). *Sia λ una costante positiva assegnata e consideriamo la funzione di densità*

$$f_X(x) = \begin{cases} \lambda \exp\{-\lambda x\} & x \geq 0 \\ \\ 0 & x < 0 \end{cases}$$

La corrispondente funzione di distribuzione è data da

$$F_X(x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi = \begin{cases} \displaystyle\int_0^x \lambda \exp\{-\lambda x\} dx & x \geq 0 \\ \\ 0 & x < 0 \end{cases}$$

e dunque

$$F_X(x) = \begin{cases} 1 - \exp\{-\lambda x\} & x \geq 0 \\ \\ 0 & x < 0 \end{cases}$$

Di una variabile aleatoria X con tale funzione distribuzione si dice che segue una ***distribuzione esponenziale di parametro*** *λ.*

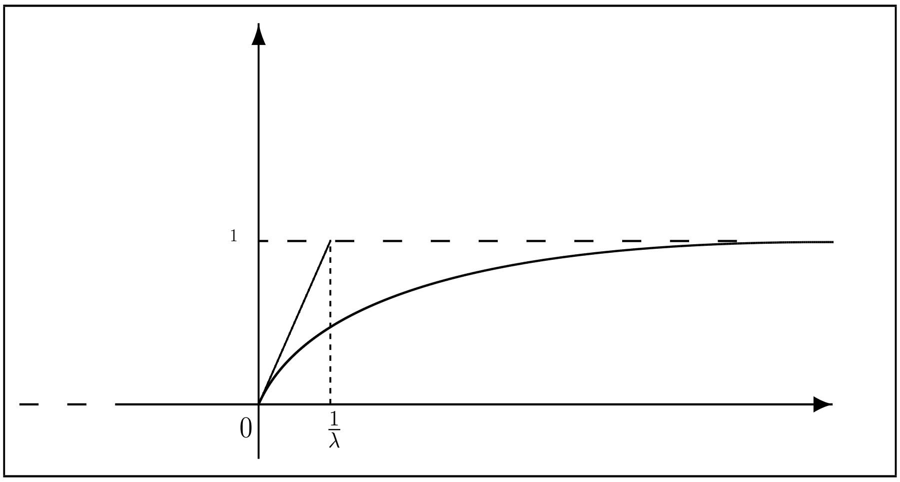

Figura 4: Grafico di $F_X(x)$, per X esponenziale di parametro λ (caso $\lambda > 1$)

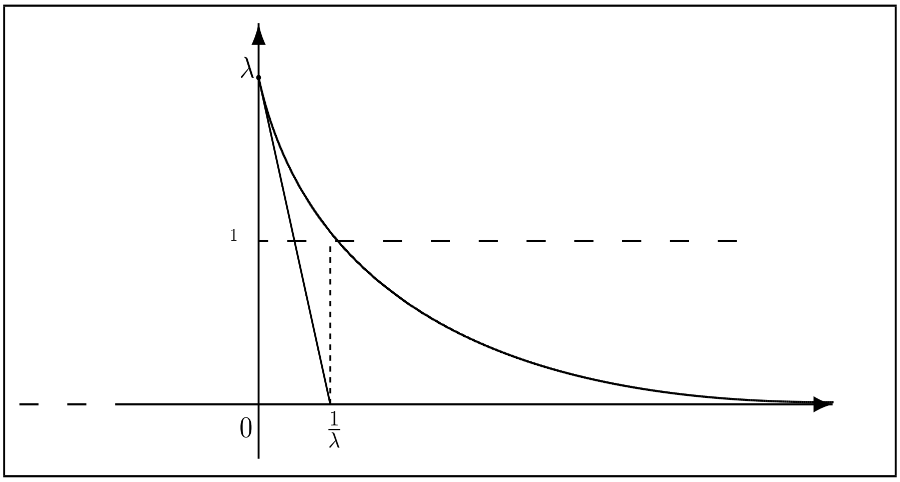

Figura 5: Grafico di $f_X(x)$, per X esponenziale di parametro λ (caso $\lambda > 1$)

Esempio 14.14 (v.a. gaussiane standard). *Una variabile aleatoria segue una distribuzione* ***gaussiana standard****, e si scrive $X \sim \mathcal{N}(0,1)$, se la sua funzione di densità è data da*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{x^2}{2}\}, x \in \mathbb{R}.$$

Tale funzione è la ben nota funzione degli errori di Gauss ed è noto che non è possibile esprimere la corrispondente funzione distribuzione

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{\xi^2}{2}\} d\xi$$

in modo esplicito, in termini di altre funzioni elementari.

Attraverso il calcolo di un opportuno integrale doppio è possibile verificare che il fattore $\frac{1}{\sqrt{2\pi}}$ ha il ruolo di costante di normalizzazione, cioè che vale l'identità

$$\int_{-\infty}^{+\infty} \exp\{-\tfrac{\xi^2}{2}\} d\xi = \sqrt{2\pi}.$$

14.5 Valori attesi per variabili aleatorie generali

Come abbiamo visto nella Lezione 9 sui valori attesi, nel caso degli spazi di probabilità finiti, ci sono diversi modi di calcolare il valore atteso di una variabile aleatoria X o di una sua funzione $h(X)$. In particolare (si veda la ***Proposizione*** **9.11** della Lezione 9) si ha che se $X(\Omega) = \{x_1, ..., x_n\}$

$$\mathbb{E}(h(X)) = \sum_{i=1}^{n} h(x_i) p_X(x_i) = \sum_{i=1}^{n} h(x_i)\, \mathbb{P}(X = x_i).$$

Questa espressione suggerisce un modo naturale di definire i valori attesi nel caso di variabili aleatorie discrete, ossia nel caso in cui $X(\Omega) = \{x_1, ..., x_n, x_{n+1}, ...\}$, come

$$\mathbb{E}(h(X)) = \sum_{i=1}^{\infty} h(x_i)\, \mathbb{P}(X = x_i).$$

Tuttavia è necessario che la serie precedente sia assolutamente convergente, per assicurarsi che la somma della serie non dipenda dall'ordine in cui la somma viene effettuata, cioè si deve richiedere che

$$\mathbb{E}(|h(X)|) = \sum_{i=1}^{\infty} |h(x_i)|\, \mathbb{P}(X = x_i) < \infty.$$

Esempi di v.a. che non ammettono valore atteso finito

Sia X una variabile aleatoria a valori in $\mathbb{N}$,e tale che $\mathbb{P}(X = k) \propto \frac{1}{k^2}, \quad k \geq 1.$
Posto $c := \frac{1}{\sum_{k=1}^{\infty} \frac{1}{k^2}}$ (è noto che $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$) si ha dunque che tale variabile aleatoria ha valore atteso infinito:

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k\, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \frac{c}{k^2} = \sum_{k=1}^{\infty} \frac{c}{k} = \infty.$$

Sia ora invece, Y, una v.a. a valori in $\mathbb{Z}\backslash\{0\}$ tale che $\mathbb{P}(Y = \pm k) \propto \frac{1}{k^2}, \quad k \geq 1.$
Si ha allora $\mathbb{P}(Y = \pm k) = \frac{c}{2\,k^2}$,

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} |k|\, \mathbb{P}(Y = k) + \sum_{k=1}^{\infty} |k|\, \mathbb{P}(Y = -k) = 2 \sum_{k=1}^{\infty} k \frac{c}{2\,k^2} = \sum_{k=1}^{\infty} \frac{c}{k} = \infty.$$

Appare anche naturale dare la definizione del valore atteso nel caso di variabili aleatorie con densità f_X come segue

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx,$$

purché

$$\mathbb{E}(|h(X)|) = \int_{-\infty}^{\infty} |h(x)| f_X(x)\, dx, < \infty.$$

Valori attesi per variabili aleatorie a valori non negativi

Nel caso di una variabile aleatoria X a valori in $\mathbb{R}^+$, oppure, più in generale, nel caso in cui $h(X)$ sia una funzione a valori in $\mathbb{R}^+$, il valore assoluto di X coincide con X ($|X| = X$), oppure $|h(X)| = h(X)$, è possibile definire sempre il valore atteso, anche quando la somma della serie o l'integrale diverge: in tale caso, però, il valore atteso viene definito uguale a $+\infty$.

È importante sottolineare che le proprietà di linearità e monotonia, viste nella Lezione 9, continuano a valere per il valore atteso definito in questo modo. Tuttavia non diamo qui la dimostrazione di questa affermazione

Si possono anche dare le definizioni della varianza come

$$Var(X) \stackrel{def}{=} \mathbb{E}\big((X - \mathbb{E}(X))^2\big),$$

ma nel caso di v.a. generali è necessario supporre che il valore atteso $\mathbb{E}(X^2)$ sia finito, a differenza del caso finito, dove non era necessario fare alcuna ipotesi.

Grazie alla proprietà di linearità si dimostra, come nel caso finito, che

$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2.$$

Esercizio proposto 14.3 (v.a. geometriche). *Sia $\theta \in (0,1)$, e siano T una v.a. geometrica di parametro θ (a partire da 1) e T' una v.a. geometrica di parametro θ (a partire da 0). Si verifichi che*

$$\mathbb{E}(T) = \frac{1}{\theta}, \quad e \quad Var(T) = \frac{1-\theta}{\theta^2}, \qquad \mathbb{E}(T') = \frac{1-\theta}{\theta}, \quad e \quad Var(T') = \frac{1-\theta}{\theta^2}.$$

Soluzione di Esercizio proposto 14.3 (valore atteso di una v.a. $Geom(\theta)$)

Per definizione si ha che

$$\mathbb{E}(T) = \sum_{k=1}^{\infty} k\, \mathbb{P}(T = k) = \sum_{k=1}^{\infty} k\theta\, (1-\theta)^{k-1}.$$

Considerando che, ovviamente, per $k = 0$ si ha $k\theta\,(1-\theta)^{k-1} = 0$, possiamo considerare la somma della serie con $k = 0$ incluso: ossia

$$\mathbb{E}(T) = \sum_{k=0}^{\infty} k\theta\, (1-\theta)^{k-1} = \sum_{k=0}^{\infty} \theta\, \frac{d}{dx} x^k \Big|_{x=1-\theta} = \theta \sum_{k=0}^{\infty} \frac{d}{dx} x^k \Big|_{x=1-\theta}$$

Per le proprietà delle serie di potenze, sappiamo che $\sum_{k=0}^{\infty} \frac{d}{dx} x^k = \frac{d}{dx} \sum_{k=0}^{\infty} x^k$ [essendo $1 - \theta \in (0,1)$ possiamo supporre che x vari in un intervallo chiuso strettamente contenuto nell'intervallo $(-1,1)$], da cui

$$\mathbb{E}(T) = \theta \frac{d}{dx} \frac{1}{(1-x)} \Big|_{x=1-\theta} = \theta \frac{1}{(1-x)^2} \Big|_{x=1-\theta} = \theta \frac{1}{(1-(1-\theta))^2} = \frac{1}{\theta}.$$

Soluzione di Esercizio proposto 14.3 (varianza di una v.a. $Geom(\theta)$)

In maniera analoga al caso del valore atteso, possiamo calcolare

$$Var(T) = \mathbb{E}\Big[\big(T - \mathbb{E}(T)\big)^2\Big] = \mathbb{E}(T^2) - \big(\mathbb{E}(T)\big)^2$$

Per calcolare $\mathbb{E}(T^2)$ osserviamo che $T^2 = T(T-1) + T$ ossia che

$$\begin{aligned}\mathbb{E}(T^2) &= \sum_{k=1}^{\infty} k^2\, \mathbb{P}(T=k) = \sum_{k=1}^{\infty} [k(k-1)+k]\theta\,(1-\theta)^{k-1} \\ &= \underbrace{\sum_{k=1}^{\infty} k(k-1)\theta\,(1-\theta)^{k-1}}_{\mathbb{E}[T(T-1)]} + \underbrace{\sum_{k=1}^{\infty} k\theta\,(1-\theta)^{k-1}}_{=\mathbb{E}(T)=\frac{1}{\theta}},\end{aligned}$$

in quanto sappiamo che $\sum_k (a_k + b_k) = \sum_k a_k + \sum_k b_k$, [inoltre, se a_k e b_k sono maggiori o uguali a zero, non ci sono problemi di convergenza]

Considerando che, ovviamente, per $k = 0$ si ha $k\theta\,(1-\theta)^{k-1} = 0$, possiamo considerare la somma della serie con $k = 0$ incluso: ossia

$$\begin{aligned}\mathbb{E}(T^2) &= \sum_{k=0}^{\infty} k(k-1)\theta\,(1-\theta)^{k-1} + \frac{1}{\theta} = \sum_{k=0}^{\infty} k(k-1)\theta\,(1-\theta)\,(1-\theta)^{k-2} + \frac{1}{\theta} \\ &= \sum_{k=0}^{\infty} \theta\,(1-\theta) \frac{d^2}{dx^2} x^k \Big|_{x=1-\theta} + \frac{1}{\theta} = \theta\,(1-\theta) \sum_{k=0}^{\infty} \frac{d^2}{dx^2} x^k \Big|_{x=1-\theta} + \frac{1}{\theta}\end{aligned}$$

Di nuovo, per le proprietà delle serie di potenze, sappiamo che

$$\sum_{k=0}^{\infty} \frac{d^2}{dx^2} x^k = \frac{d^2}{dx^2} \sum_{k=0}^{\infty} x^k$$

[essendo $1-\theta \in (0,1)$ possiamo supporre che x vari in un intervallo chiuso strettamente contenuto nell'intervallo $(-1,1)$], da cui

$$\begin{aligned}\mathbb{E}(T^2) &= \theta\,(1-\theta)\, \frac{d^2}{dx^2} \frac{1}{(1-x)} \Big|_{x=1-\theta} + \frac{1}{\theta} = \theta\,(1-\theta)\, \frac{2}{(1-x)^3} \Big|_{x=1-\theta} + \frac{1}{\theta} \\ &= \theta\,(1-\theta)\, \frac{2}{(1-(1-\theta))^3} = \frac{2(1-\theta)}{\theta^2} + \frac{1}{\theta} = \frac{2(1-\theta)+\theta}{\theta^2}.\end{aligned}$$

Quindi si ha che

$$Var(T) = \mathbb{E}(T^2) - \big(\mathbb{E}(T)\big)^2 = \frac{2(1-\theta)+\theta}{\theta^2} - \frac{1}{\theta^2} = \frac{2(1-\theta)+\theta-1}{\theta^2} = \frac{(1-\theta)}{\theta^2}.$$

Esercizio proposto 14.4 (v.a. di Poisson). *Si dimostri che, per $\lambda > 0$*

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\, e^{-\lambda} = 1.$$

suggerimento: *usare lo sviluppo di Taylor della funzione e^x e che è ricordato nella nota a pagina 37.*

Come conseguenza del precedente Esercizio proposto 14.4 ha senso considerare una v.a. X a valori in $\{0, 1, 2, ...\}$ per la quale

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, 2, ...$$

In questo caso si dice che X ***ha distribuzione di Poisson di parametro*** λ.

L'interesse della distribuzione di Poisson sta nel fatto che si può usare come approssimazione di una variabile aleatoria binomiale $Bin(n, \theta)$, quando n è grande e θ è piccolo. Si dimostra infatti che, se S_n sono v.a. binomiali $Bin(n, \theta_n)$ di parametro $\theta = \theta_n = \frac{\lambda}{n}$ (in modo che il valore atteso $\mathbb{E}(S_n) = n\theta_n = \lambda$ rimanga costante) allora,

$$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \mathbb{P}(X = k) \qquad \forall k = 0, 1, 2, ...,$$

ossia vale il seguente risultato:

Teorema 14.1 (Teorema (di approssimazione) di Poisson). *Sia $\lambda > 0$ un numero reale. Per ogni $n > \lambda$, sia S_n una variabile aleatoria binomiale $Bin(n, \lambda/n)$. Allora si ha*

$$\lim_{n\to\infty} \mathbb{P}(S_n = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad \forall\, k = 0, 1, 2, ...$$

Dimostrazione. Basta osservare che

$$\begin{aligned}\mathbb{P}(S_n = k) &= \binom{n}{k}\theta^k(1-\theta)^{n-k} = \frac{n!}{k!\,(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{\lambda^k}{k!}\,\frac{n!}{(n-k)!}\,\frac{1}{n^k}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} \frac{\lambda^k}{k!}\, e^{-\lambda}\end{aligned}$$

in quanto valgono le seguenti tre relazioni

$$(i) \qquad \frac{n!}{(n-k)!}\,\frac{1}{n^k} = \frac{n(n-1)\cdots\big(n-(k-1)\big)}{n^k} = \frac{n}{n}\times\frac{n-1}{n}\times\cdots\times\frac{n-(k-1)}{n} \underset{n\to\infty}{\longrightarrow} 1^k = 1,$$

$$(ii) \qquad \left(1-\frac{\lambda}{n}\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\lambda},$$

ed infine

$$(iii) \qquad \left(1-\frac{\lambda}{n}\right)^{-k} \underset{n\to\infty}{\longrightarrow} 1.$$

□

Esercizio proposto 14.5 (v.a. Poisson, continuazione). *Sia X una v.a. di Poisson di parametro λ. Si verifichi che*

$$\mathbb{E}(X) = \lambda, \qquad e \qquad Var(X) = \lambda.$$

Si osservi che, se $\lambda > 0$ ed n è tale che $\lambda/n \in (0, 1)$ e se $S_n := Bin(n, \frac{\lambda}{n})$, allora

$$\mathbb{E}(S_n) = \lambda = \mathbb{E}(X), \qquad e \qquad Var(S_n) = \lambda\left(1-\tfrac{\lambda}{n}\right) \to \lambda = Var(X).$$

Soluzione dell'Esercizio proposto 14.5
(valore atteso e varianza di una v.a. di Poisson):

Sia $X \sim Poiss(\lambda)$, allora $\mathbb{E}(X) = \lambda$ e $\mathbb{E}(X^2) = \lambda^2 + \lambda$, da cui

$$Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Infatti

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{\infty} k\, \mathbb{P}(X=k) = \sum_{k=0}^{\infty} k\, \frac{\lambda^k}{k!}\, e^{-\lambda} \\
&= \sum_{k=1}^{\infty} k\, \frac{\lambda^k}{k!}\, e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}\, e^{-\lambda} \\
&= \sum_{h=0}^{\infty} \frac{\lambda^{h+1}}{h!}\, e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}\, e^{-\lambda} \\
&= \lambda e^{\lambda}\, e^{-\lambda} = \lambda.
\end{aligned}$$

Analogamente si ha che $\mathbb{E}(X^2) = \lambda^2 + \lambda$: considerando che $X^2 = X + X(X-1)$, si ha

$$\begin{aligned}
\mathbb{E}\big(X^2\big) &= \mathbb{E}\big(X + X(X-1)\big) = \mathbb{E}\big(X\big) + \mathbb{E}\big(X(X-1)\big) \\
&= \lambda + \sum_{k=0}^{\infty} k(k-1)\, \mathbb{P}(X=k) = \lambda + \sum_{k=0}^{\infty} k(k-1)\, \frac{\lambda^k}{k!}\, e^{-\lambda} \\
&= \lambda + \sum_{k=2}^{\infty} k(k-1)\, \frac{\lambda^k}{k!}\, e^{-\lambda} \\
&= \lambda + \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}\, e^{-\lambda} = \lambda + \sum_{h=0}^{\infty} \frac{\lambda^{h+2}}{h!}\, e^{-\lambda} \\
&= \lambda + \lambda^2 \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}\, e^{-\lambda} = \lambda + \lambda^2\, e^{\lambda}\, e^{-\lambda} = \lambda + \lambda^2.
\end{aligned}$$

Esercizio proposto 14.6 (v.a. esponenziali). *Sia X una v.a. esponenziale di parametro λ. Si verifichi che*

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad e \qquad Var(X) = \frac{1}{\lambda^2}$$

Soluzione dell'Esercizio proposto 14.7 (valore atteso di una v.a. $N(0,1)$)

Per mostrare che $\mathbb{E}(X) = 0$, basta osservare che la funzione

$$g(x) := x\, f_X(x) \text{ è dispari, cioè } g(-x) = g(x)$$

e che, essendo $d\big(\frac{1}{2}\, x^2\big) = x\, dx$, si ha

$$\begin{aligned}
\mathbb{E}(|X|) &= \int_{-\infty}^{\infty} |x|\, f_X(x)\, dx = 2\int_0^{\infty} x\, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2}\,x^2} dx \\
&= 2\frac{1}{\sqrt{2\,\pi}} \int_0^{\infty} e^{-\frac{1}{2}\,x^2}\, d\big(\frac{1}{2}\, x^2\big) \qquad \text{(con il cambio di variabile } y = \frac{1}{2}\, x^2\text{)} \\
&= 2\frac{1}{\sqrt{2\,\pi}} \int_0^{\infty} e^{-y}\, dy = 2\frac{1}{\sqrt{2\,\pi}}\big(-e^{-y}\big)_0^{\infty} = 2\frac{1}{\sqrt{2\,\pi}}\big(-0+1\big) < \infty
\end{aligned}$$

e quindi, il valore atteso vale 0.

Esercizio proposto 14.7 (v.a. gaussiane). *Sia X una v.a. gaussiana standard. Si verifichi che*

$$\mathbb{E}(X) = 0, \qquad e \qquad Var(X) = 1$$

Soluzione dell'Esercizio Proposto 14.7 (varianza di una v.a. $N(0,1)$)

Sappiamo che $\mathbb{E}(X) = 0$, di conseguenza la varianza coincide con $\mathbb{E}(X^2)$ e sia ha

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2\, f_X(x)\, dx = \int_{-\infty}^{\infty} x^2\, \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{1}{2}\,x^2} dx\\ &= \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}\,x^2}\, d\big(\frac{1}{2}\,x^2\big)\end{aligned}$$

essendo $e^{-\frac{1}{2}\,x^2}\, d\big(\frac{1}{2}\,x^2\big) = -d\big(e^{-\frac{1}{2}\,x^2}\big)$, ed integrando per parti

$$\begin{aligned}&= \frac{1}{\sqrt{2\,\pi}} \left\{ \left[-x\, e^{-\frac{1}{2}\,x^2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}\,x^2}\, dx \right\}\\ &= \frac{1}{\sqrt{2\,\pi}} \left\{ [-0+0] + \int_{-\infty}^{\infty} e^{-\frac{1}{2}\,x^2}\, dx \right\}\\ &= \int_{-\infty}^{\infty} f_X(x)\, dx = 1\end{aligned}$$

Esercizio proposto 14.8 (v.a. di Cauchy). *Sia X una v.a. di Cauchy. Si dimostri che la condizione*

$$\mathbb{E}(|X|) = \int_{-\infty}^{\infty} |x| f_X(x)\, dx, < \infty$$

non è soddisfatta, e quindi il valore atteso non esiste.

Soluzione dell'Esercizio proposto 14.8
(una v.a. di Cauchy non ammette valore atteso)

Nel caso della distribuzione di Cauchy si ha $f_X(-x) = f_X(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}$ e quindi

$$\begin{aligned}\int_{-\infty}^{\infty} |x|\, f_X(x)\, dx &= 2\int_0^{\infty} x\, \frac{1}{\pi}\, \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_0^{\infty} \frac{1}{1+x^2}\, d(x^2)\\ &= \frac{1}{\pi} \int_0^{\infty} \frac{1}{1+y}\, dy = \frac{1}{\pi} \big(\log(1+y)\big)_0^{\infty} = \infty\end{aligned}$$

14.6 Esempi svolti

Esempio* 14.15. *Sia X una variabile aleatoria con funzione di distribuzione data da*

$$F(x) = \begin{cases} 0 & per \quad x < -1\\ \frac{1}{2}\,(x+1) & per \ -1 \le x < 0\\ \frac{1}{2} + \frac{x^2}{2} & per \quad 0 \le x < 1\\ 1 & per \quad x \ge 1 \end{cases}$$

a) ***Calcolare*** $\mathbb{P}(-\frac{1}{2} \le X \le \frac{1}{2})$.

Soluzione di a) $\mathbb{P}(-\frac{1}{2} \leq X \leq \frac{1}{2}) = \frac{3}{8}$.

Infatti, per $a \leq b$

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X = a) + \mathbb{P}(a < X \leq b) = F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a),$$

dove l'ultima uguaglianza dipende dal fatto che nel nostro caso la funzione $x \to F_X(x) = F(x)$ *è continua.* [64]
Quindi

$$\mathbb{P}(-\tfrac{1}{2} \leq X \leq \tfrac{1}{2}) = F(\tfrac{1}{2}) - F(-\tfrac{1}{2}) = \frac{1}{2} + \frac{\frac{1}{4}}{2} - \frac{1}{2}\left(-\frac{1}{2} + 1\right) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$$

b) ***Calcolare la funzione di densità di probabilità di*** $\boldsymbol{X}$**.**

Soluzione di b) *Applicando la* **Proposizione** ***14.5*** *si ha*

$$f(x) = F'(x) = \begin{cases} 0 & \textit{per} \quad x < -1 \\ \frac{1}{2} & \textit{per} \ -1 < x < 0 \\ x & \textit{per} \quad 0 < x < 1 \\ 0 & \textit{per} \quad x > 1 \end{cases}$$

Come verifica, si noti che $f(x)$ *è effettivamente una densità di probabilità, infatti* $f(x) \geq 0$ *per ogni* $x \in \mathbb{R}$*, ed inoltre*

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{-1}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx = \int_{-1}^{0} \frac{1}{2}\,dx + \int_{0}^{1} x\,dx = \frac{1}{2} + \frac{x^2}{2}\bigg|_{x=0}^{x=1} = \frac{1}{2} + \frac{1}{2} = 1.$$

Anche senza utilizzare la **Proposizione** ***14.5****, che non è stata dimostrata, si può verificare che* $f(x)$ *è la densità di* X*: infatti (si ricordi l'***Osservazione 8***)* ***basta verificare*** *che* $F(x) = \int_{-\infty}^{x} f(y)\,dy$*, per ogni* $x \in \mathbb{R}$*. Si procede intervallo per intervallo:*

- *per* $x < -1$

$$0 = F(x) \overset{?}{=} \int_{-\infty}^{x} f(y)\,dy = 0; \qquad \textit{(il punto interrogativo segnala la verifica da effettuare)}$$

- *per* $-1 \leq x < 0$

$$\frac{1}{2}(x+1) = F(x) \overset{?}{=} \int_{-\infty}^{x} f(y)\,dy = \int_{-1}^{x} f(y)\,dy = \int_{-1}^{x} \frac{1}{2}\,dy;$$

- *per* $0 \leq x < 1$

$$\frac{1}{2} + \frac{x^2}{2} = F(x) \overset{?}{=} \int_{-\infty}^{x} f(y)\,dy = \int_{-1}^{0} f(y)\,dy + \int_{0}^{x} f(y)\,dy = \frac{1}{2} + \int_{0}^{x} y\,dy = \frac{1}{2} + \frac{y^2}{2}\bigg|_{y=0}^{y=x};$$

- *per* $x \geq 1$

$$1 = F(x) \overset{?}{=} \int_{-\infty}^{x} f(y)\,dy = \int_{-1}^{0} f(y)\,dy + \int_{0}^{1} f(y)\,dy = 1.$$

[64] Per verificare la continuità basta considerare che
• $F(-1^-) = 0$, che coincide con $F(-1) = \frac{1}{2}(-1+1) = 0$; • $F(0^-) = \frac{1}{2}$, che coincide con $F(0) = \frac{1}{2} + \frac{0^2}{2}) = \frac{1}{2}$;
• $F(1^-) = \frac{1}{2} + \frac{1^2}{2}) = 1$, che coincide con $F(1) = 1$.

c) *Calcolare $\mathbb{E}(X)$.*

Soluzione di c) $\mathbb{E}(X) = \frac{1}{12}$.
Infatti

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_{-1}^{0} x\, f(x)\, dx + \int_{0}^{1} x\, f(x)\, dx \\ &= \int_{-1}^{0} x\, \frac{1}{2}\, dx + \int_{0}^{1} x\, x\, dx = \frac{1}{2}\, \frac{x^2}{2}\Big|_{x=-1}^{x=0} + \frac{x^3}{3}\Big|_{x=0}^{x=1} = -\frac{1}{4} + \frac{1}{3} = \frac{-3+4}{12} = \frac{1}{12}.\end{aligned}$$

Esempio* 14.16. *Sia X una variabile aleatoria con funzione di densità di probabilità*

$$f_X(x) = \begin{cases} kx & \textit{per } 0 \le x \le 1 \\ k(2-x) & \textit{per } 1 \le x \le 2 \\ 0 & \textit{altrimenti} \end{cases}$$

a) *Trovare il valore della costante k.*

Soluzione di a) *La costante è $k = 1$.*

Infatti si tratta di trovare k in modo che $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$. Per cui,

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)\, dx &= k\left(\int_0^1 x\, dx + \int_1^2 (2-x)\, dx\right) = k\left(\tfrac{x^2}{2}\Big|_0^1 + 2 - \tfrac{x^2}{2}\Big|_1^2\right) \\ &= k \cdot 1 = 1 \qquad \textit{se e solo se } k = 1\end{aligned}$$

b) *Trovare la funzione di distribuzione di X.*

Soluzione di b) *La funzione di distribuzione F_X è data da*

$$F_X(x) = \begin{cases} 0 & \textit{per } -\infty < x < 0 \\ \frac{x^2}{2} & \textit{per } 0 \le x < 1 \\ 2x - \frac{x^2}{2} - 1 & \textit{per } 1 \le x < 2 \\ 1 & \textit{per } 2 \le x \end{cases}$$

Infatti chiaramente per ogni $x \le 0$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy = 0,$$

per $0 \le x \le 1$

$$F_X(x) = F_X(0) + \int_0^x y\, dy = 0 + \tfrac{x^2}{2},$$

per $1 \le x < 2$

$$\begin{aligned}F_X(x) &= F_X(1) + \int_1^x (2-y)\, dy = \tfrac{1}{2} + \int_1^x (2-y)\, dy = \tfrac{1}{2} + 2\, y|_1^x - \tfrac{y^2}{2}\Big|_1^x \\ &= \tfrac{1}{2} + 2x - 2 - \tfrac{x^2}{2} + \tfrac{1}{2} = 2x - \tfrac{x^2}{2} - 1,\end{aligned}$$

ed infine per $x \ge 2$

$$F_X(x) = F_X(2) + \int_2^x f_X(y)\, dy = 1 + 0 = 1.$$

c) *Trovare il valore atteso di X (attenzione: è possibile trovarlo senza fare calcoli?)*

soluzione di c) *Il valore atteso di X vale 1*

Infatti si tratta di calcolare

$$\int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^1 x^2\, dx + \int_1^2 x\,(2-x)\, dx = \frac{x^3}{3}\Big|_0^1 + x^2\Big|_1^2 - \frac{x^3}{3}\Big|_1^2$$
$$= \tfrac{1}{3} + 4 - 1 - \tfrac{8}{3} + \tfrac{1}{3} = 1$$

Tuttavia per motivi di simmetria si ha che tale integrale deve venire 1 infatti il grafico della densità è simmetrico rispetto all'asse $x = 1$ e quindi, tenendo presente l'analogia con il baricentro, si ha immediatamente che il valore atteso deve essere 1.

ATTENZIONE: procedere con cautela

Se X ammette densità f_X, e se F_X è nota, allora è possibile trovare $f_X(x)$ derivando la funzione $F_X(x)$ (almeno nei punti in cui F_X è derivabile). Tuttavia questo procedimento, se usato senza le dovute cautele, può portare a degli errori, come mostra i seguente controesempio, in cui si vede come è possibile che una funzione di distribuzione di una v.a. possa ammettere derivata tranne in un numero finito di punti, senza che la v.a. ammetta densità.

Esempio* 14.17. *Sia Z la variabile aleatoria discreta, definita come nell'Esempio 14.8. Allora la derivata prima di F_Z esiste in ogni $x \neq -1, 0, 1$: infatti $F_Z'(x) = 0$ per ogni $x \neq -1, 0, 1$. Tuttavia, ovviamente, la funzione $f(x) := 0$ (cioè identicamente uguale a 0) non può essere la densità di Z, in quanto palesemente*

$$F_Z(x) \neq \int_{-\infty}^{x} f(y)\, dy := 0.$$

14.7 Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini

Sia X una variabile aleatoria e sia $h : \mathbb{R} \to \mathbb{R}$ una funzione reale. L'applicazione

$$\omega \mapsto Y(\omega) := h\big(X(\omega)\big)$$

è una trasformazione della variabile aleatoria X.
Nel caso degli spazi finiti Y è sempre una variabile aleatoria e la sua distribuzione è individuata da $Y(\Omega) = h\big(X(\Omega)\big) = \{y_1, y_2, ..., y_m\}$ e da

$$\mathbb{P}(Y = y_k) = \sum_{i:\, h(x_i)=y_k} \mathbb{P}(X = x_i), \qquad k = 1, 2, ..., m$$

(si veda la Sezione 7.4 della Lezione 7 e la ***Proposizione* 9.13** della Lezione 9, in cui si considera la trasformazione $W = h(X)$).

Nel caso degli spazi generali ci sono due differenze:

(i) non è sempre vero che Y sia una variabile aleatoria;

(ii) può accadere che la distribuzione di Y non si possa calcolare nel modo precedente.

Il problema si pone in particolare se la variabile aleatoria X non è discreta, mentre se la variabile aleatoria X è discreta numerabile, allora si generalizza immediatamente la precedente relazione, in quanto necessariamente Y è discreta (finita o numerabile): infatti ancora[65] si ha $Y(\Omega) = h\big(X(\Omega)\big) = \{y_k,\ k \geq 1\}$ e, similmente al caso finito

$$\{Y = y_k\} = \cup_{i:\, h(x_i)=y_k} \{X = x_i\}, \qquad k \geq 1.$$

Questo fatto implica che l'insieme $\{Y = y_k\}$ è un evento (cioè appartiene ad $\mathcal{F}$) in quanto unione finita o numerabile di eventi, e quindi

$$\mathbb{P}(Y = y_k) = \sum_{i:\, h(x_i)=y_k} \mathbb{P}(X = x_i), \qquad k \geq 1.$$

Si noti che in questa formula, a differenza della formula precedente del caso finito, la somma può essere estesa ad un insieme numerabile.

Il problema si può risolvere anche nel caso generale sotto condizioni che riguardano la funzione h, ma qui vedremo solo alcuni casi particolari. In tutti questi esempi l'idea è quella di riscrivere l'insieme

$$\{Y \leq y\}$$

in termini di un evento del tipo

$$\{X \in I\}$$

dove I è un intervallo[66]. Ad esempio se $h(x) = x^2$, allora

$$\{Y \leq y\} = \emptyset$$

[65] L'unica differenza con il caso finito è che $Y(\Omega)$ può essere un insieme infinito.

[66] Più in generale, per essere sicuri che sia $\{Y \leq y\}$ un evento, basterebbe poter scrivere

$$\{Y \leq y\} = \cup_{i=1}^{k} \{X \in I_i\}$$

oppure

$$\{Y \leq y\} = \cup_{i=1}^{\infty} \{X \in I_i\},$$

cioè poter scrivere $\{Y \leq y\}$ come unione finita o numerabile di eventi.

per $y < 0$, mentre, per $y \geq 0$

$$\{Y \leq y\} = \{X^2 \leq y\} = \{-\sqrt{y} \leq X \leq \sqrt{y}\} = \{X \in I\}, \qquad \text{dove } I = [-\sqrt{y}, \sqrt{y}].$$

Da ciò si deduce immediatamente che l'insieme $\{Y \leq y\}$ è un evento in quanto $\{X \in I\}$ lo è ed inoltre si ottiene la funzione di distribuzione di Y in quanto

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0, \\ \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) + \mathbb{P}(X = -\sqrt{y}) & \text{per } y \geq 0. \end{cases}$$

Il caso in cui $h(x) = x^3$ si ottiene in modo simile, ma è più semplice: qualunque sia y

$$\{Y \leq y\} = \{X^3 \leq y\} = \{X \leq y^{\frac{1}{3}}\}.$$

Di conseguenza, qualunque sia y,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq y^{\frac{1}{3}}) = F_X(y^{\frac{1}{3}}).$$

***Osservazione* 14.10.** Supponiamo ora che X ammetta densità. In questo caso, ci si potrebbe anche chiedere se $Y = h(X)$ ammette densità. Il candidato naturale è ovviamente la derivata rispetto ad y della funzione di distribuzione $F_Y(y)$, che nell'esempio precedente di $h(x) = x^3$ è la funzione composta $F_X(g(y))$, con $g(y) = y^{\frac{1}{3}}$, ossia

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(g(y)) = F'_X(g(y))\,\frac{d}{dy}g(y) = F'_X(g(y))\,\frac{1}{3}\,y^{-\frac{2}{3}} = f_X(g(y))\,\frac{1}{3}\,y^{-\frac{2}{3}}.$$

Tuttavia sorge qualche problema a percorrere questa strada, anche se F_X fosse continua con derivata continua, infatti la funzione $g(y) = y^{\frac{1}{3}}$ non ha derivata in zero. Questo tipo di problema si potrebbe risolvere, ma lo tralasciamo in questo corso elementare. Non sorge, però, alcun problema se, ad esempio, $\mathbb{P}(X > a) = 1$, per un $a > 0$, o, più in generale, se esiste un $a > 0$ tale che $\mathbb{P}(X \in (-a, a)) = 0$.

Il prossimo paragrafo è dedicato invece al caso delle trasformazioni affini, cioè al caso in cui $h(x) = \alpha + \beta\,x$, ed in questo caso affronteremo anche in problema della determinazione della densità di $Y = h(X)$.

14.7.1 Il caso delle trasformazioni affini

Sia X una variabile aleatoria e siano α e β due numeri reali. Si indichi con Y la seguente trasformazione affine di X

$$Y = \alpha + \beta\,X.$$

Il ***primo problema*** che ci poniamo in questo paragrafo è il seguente: ***data F_X, la funzione di distribuzione di X, calcolare F_Y, la funzione di distribuzione di Y.***

Successivamente ci occuperemo del ***secondo problema***: ***se X è una variabile aleatoria che ammette densità di probabilità f_X, la variabili aleatoria Y ammette densità di probabilità f_Y? e se (come effettivamente è) la risposta è sì, come si calcola f_Y?***

Considereremo solo il caso in cui $\beta \neq 0$, in quanto il caso $\beta = 0$ corrisponde al caso banale in cui $Y = \alpha$, cioè Y è una variabile aleatoria degenere[67].

Cominciamo dando la risposta dei procedenti problemi (la verifica che si tratta effettivamente della soluzione segue immediatamente dopo):
Risposta al primo problema.
La soluzione dipende dal segno di β e si ha che la funzione di distribuzione di $Y = \alpha + \beta\, X$ è

$$F_Y(y) = F_X\left(\tfrac{y-\alpha}{\beta}\right), \qquad \beta > 0, \tag{136}$$

$$F_Y(y) = 1 - F_X\left(\tfrac{y-\alpha}{\beta}\right) + \mathbb{P}\left(X = \tfrac{y-\alpha}{\beta}\right), \qquad \beta < 0. \tag{137}$$

Risposta al secondo problema.
Se X ammette densità f_X allora anche $Y = \alpha + \beta\, X$ ammette densità e si ha

$$f_Y(y) = f_X\left(\tfrac{y-\alpha}{\beta}\right)\frac{1}{|\beta|}, \qquad \beta \neq 0. \tag{138}$$

In particolare per $\alpha = 0$ e $\beta = -1$, ossia per $Y = -X$, si ha

$$f_{-X}(x) = f_X(-x)$$

Proseguiamo con le verifiche delle precedenti soluzioni.

Verifica della soluzione per il primo problema: funzione di distribuzione
Cominciamo con l'osservare che

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\alpha + \beta\, X \leq y).$$

A questo punto dobbiamo distinguere tra i due casi $\beta > 0$ o $\beta < 0$

caso $\boldsymbol{\beta > 0}$ In questo caso[68]

$$\{\alpha + \beta\, X \leq y\} = \left\{X \leq \tfrac{y-\alpha}{\beta}\right\}$$

e quindi

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\alpha + \beta\, X \leq y) = \mathbb{P}\left(\left\{X \leq \tfrac{y-\alpha}{\beta}\right\}\right) = F_X\left(\tfrac{y-\alpha}{\beta}\right).$$

caso $\boldsymbol{\beta < 0}$ In questo caso[69]

$$\{\alpha + \beta\, X \leq y\} = \left\{X \geq \tfrac{y-\alpha}{\beta}\right\}$$

[67] In questo caso la $F_Y(y) = 0$ per $y < \alpha$, ed $F_Y(y) = 1$ per $y \geq \alpha$. In questo caso, ovviamente, la variabile aleatoria Y non ammette densità, qualunque sia la distribuzione di X.

[68] Infatti

$$Y(\omega) \leq y \quad \Leftrightarrow \quad \alpha + \beta\, X(\omega) \leq y \quad \Leftrightarrow \quad X(\omega) \leq \tfrac{y-\alpha}{\beta}$$

e quindi

$$\{\omega:\ Y(\omega) \leq y\} = \left\{\omega \in \Omega:\ X(\omega) \leq \tfrac{y-\alpha}{\beta}\right\}$$

e quindi

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(\alpha + \beta\, X \le y) = \mathbb{P}\left(\left\{X \ge \tfrac{y-\alpha}{\beta}\right\}\right) \\ &= 1 - \mathbb{P}\left(\left\{X < \tfrac{y-\alpha}{\beta}\right\}\right) = 1 - F_X\left(\left(\tfrac{y-\alpha}{\beta}\right)^-\right) \\ &= 1 - F_X\left(\tfrac{y-\alpha}{\beta}\right) + \mathbb{P}\left(X = \tfrac{y-\alpha}{\beta}\right). \end{aligned}$$

Verifica della soluzione per il secondo problema: funzione di densità

Cominciamo con il caso in cui $F_X(x)$ è continua, derivabile con derivata continua in tutto $\mathbb{R}$, ossia

$$f_X(x) = \frac{d}{dx}\, F_X(x)$$

caso $\boldsymbol{\beta > 0}$ In questo caso, come visto in precedenza,

$$F_Y(y) = F_X\left(\tfrac{y-\alpha}{\beta}\right),$$

da cui $F_Y(y)$ è continua, derivabile, con derivata continua in ogni y, e la derivata è la densità di probabilità Inoltre si ha, utilizzando la regola della derivazione della funzione composta[70]

$$\frac{d}{dy}\, F_Y(y) = \frac{d}{dy}\, F_X\left(\tfrac{y-\alpha}{\beta}\right) = \frac{d}{dx}\, F_X(x)\Big|_{x=\frac{y-\alpha}{\beta}}\, \frac{d}{dy}\, \frac{y-\alpha}{\beta} = f_X\left(\tfrac{y-\alpha}{\beta}\right)\, \frac{1}{\beta}$$

caso $\boldsymbol{\beta < 0}$ In questo caso, essendo F_X una funzione continua,

$$F_Y(y) = 1 - F_X\left(\left(\tfrac{y-\alpha}{\beta}\right)^-\right) = 1 - F_X\left(\tfrac{y-\alpha}{\beta}\right),$$

da cui procedendo in modo simile al caso precedente

$$\frac{d}{dy}\, F_Y(y) = \frac{d}{dy}\, \left(1 - F_X\left(\tfrac{y-\alpha}{\beta}\right)\right) = -\,\frac{d}{dx}\, F_X(x)\Big|_{x=\frac{y-\alpha}{\beta}}\, \frac{d}{dy}\, \frac{y-\alpha}{\beta} = -f_X\left(\tfrac{y-\alpha}{\beta}\right)\, \frac{1}{\beta}$$

Si osservi che, in questo caso ($\beta < 0$) si ottiene il segno negativo, come del resto doveva essere: infatti così si ottiene che[71] $\frac{d}{dy}\, F_Y(y) \ge 0$, come deve essere, in quanto F_Y è una funzione crescente in senso lato (o non decrescente).

[69]Infatti

$$Y(\omega) \le y \quad \Leftrightarrow \quad \alpha + \beta\, X(\omega) \le y \quad \Leftrightarrow \quad X(\omega) \ge \tfrac{y-\alpha}{\beta}$$

e quindi

$$\{\omega:\ Y(\omega) \le y\} = \left\{\omega \in \Omega:\ X(\omega) \ge \tfrac{y-\alpha}{\beta}\right\}$$

[70]Si ricordi che, se $h(\cdot)$ è derivabile in t, e $\varphi(\cdot)$ derivabile in $h(t)$, allora

$$\frac{d}{dt}\, \varphi\,(h(t)) = \frac{d}{dx}\, \varphi(x)\Big|_{x=h(t)}\, \frac{d}{dt}\, h(t).$$

[71]Si consideri che $f_X\left(\tfrac{y-\alpha}{\beta}\right) \ge 0$ e che $-\frac{1}{\beta} \ge 0$.

Si osservi ancora che possiamo esprimere il risultato ottenuto in entrambi i casi nel seguente modo:

$$\frac{d}{dy}\, F_Y(y) = f_X\left(\tfrac{y-\alpha}{\beta}\right)\, \frac{1}{|\beta|}$$

Unificando così i due casi.

Se invece la funzione F_X non avesse derivata continua, i conti effettuati sarebbero validi solo nei punti in cui la derivata di F_X esiste. Tali conti mostrano come, se Y ammette densità, allora l'unica funzione candidata ad essere la funzione di densità di probabilità $f_Y(y)$ sia appunto

$$f_Y(y) = f_X\left(\tfrac{y-\alpha}{\beta}\right)\, \frac{1}{|\beta|}.$$

Non diamo qui la dimostrazione generale che questo è effettivamente il caso. Ricordiamo solamente che, in generale, se Z è una variabile aleatoria, con funzione di distribuzione $F_Z(x)$, allora può accadere che F_Z ammetta derivata in ogni x, esclusi un numero finito di punti, ma che la variabile aleatoria non ammetta densità, o in altre parole, che la derivata non possa essere la funzione di densità di Z, come accade nell'Esempio 14.17.

Esempio 14.18. ***Una trasformazione affine di X con distribuzione Uniforme in $(0,1)$ è ancora uniforme.***

Innanzi tutto ricordiamo che $Z \sim R(a,b)$, ovvero che Z ha distribuzione uniforme nell'intervallo (a,b) significa che la sua funzione di distribuzione $F_Z(z)$ vale

$$F_Z(z) = \begin{cases} 0 & \textit{per } z < a \\ \frac{z-a}{b-a} & \textit{per } a \le z < b \\ 1 & \textit{per } z \ge b \end{cases}$$

o equivalentemente che la sua densità di probabilità $f_Z(z)$ vale

$$f_Z(z) = \begin{cases} 0 & \textit{per } z < a \\ \frac{1}{b-a} & \textit{per } a < z < b \\ 0 & \textit{per } z > b \end{cases}$$

In particolare quindi, per $X \sim R(0,1)$, si ha

$$F_X(x) = \begin{cases} 0 & \textit{per } x < 0 \\ x & \textit{per } 0 \le x < 1 \\ 1 & \textit{per } x \ge 1 \end{cases} \qquad \textit{e} \qquad f_X(x) = \begin{cases} 0 & \textit{per } x < 0 \\ 1 & \textit{per } 0 < x < 1 \\ 0 & \textit{per } x > 1 \end{cases}$$

Sia ora

$$Y = \alpha + \beta\, X.$$

Innanzi tutto notiamo che Y assume valori tra α e $\alpha+\beta$, se $\beta > 0$, mentre assume i valori tra $\alpha+\beta$ e α, se $\beta < 0$. Quindi possiamo immediatamente capire che $Y(\Omega) = \big(\min(\alpha, \alpha+\beta), \max(\alpha, \alpha+\beta)\big)$. Ciò ci fa subito "sospettare" che Y sia appunto uniforme in tale intervallo. Ciò si può dimostrare immediatamente notando che la sua densità di probabilità $f_Y(y)$ vale

$$f_Y(y) = f_X\left(\tfrac{y-\alpha}{\beta}\right) \frac{1}{|\beta|} = \begin{cases} 0 & \textit{per } \frac{y-\alpha}{\beta} < 0 \\ 1 \cdot \frac{1}{|\beta|} & \textit{per } 0 < \frac{y-\alpha}{\beta} < 1 \\ 0 & \textit{per } \frac{y-\alpha}{\beta} > 1 \end{cases}$$

ovvero, distinguendo a seconda del segno di β, e riscrivendo le condizioni su y in modo più esplicito

se $\beta > 0$	**se $\beta < 0$**
$f_Y(y) = \begin{cases} 0 & \text{per } y < \alpha \\ \frac{1}{\lvert\beta\rvert} & \text{per } \alpha < y < \alpha+\beta \\ 0 & \text{per } y > \alpha+\beta \end{cases}$	$f_Y(y) = \begin{cases} 0 & \text{per } y > \alpha \\ \frac{1}{\lvert\beta\rvert} & \text{per } \alpha+\beta < y < \alpha \\ 0 & \text{per } y < \alpha+\beta \end{cases}$

Analogamente si può procedere con la funzione di distribuzione. Consideriamo prima il caso $\beta > 0$,

$$F_Y(y) = F_X\left(\tfrac{y-\alpha}{\beta}\right) = \begin{cases} 0 & \text{per } \frac{y-\alpha}{\beta} < 0 \\ \frac{y-\alpha}{\beta} & \text{per } 0 \leq \frac{y-\alpha}{\beta} < 1 \\ 1 & \text{per } \frac{y-\alpha}{\beta} \geq 1 \end{cases}$$

e poi il caso $\beta < 0$,

$$F_Y(y) = 1 - F_X\left(\left(\tfrac{y-\alpha}{\beta}\right)^-\right) = \begin{cases} 1 & \text{per } \frac{y-\alpha}{\beta} < 0 \\ 1 - \frac{y-\alpha}{\beta} = \frac{\alpha+\beta-y}{\beta} = \frac{y-(\alpha+\beta)}{\lvert\beta\rvert} & \text{per } 0 \leq \frac{y-\alpha}{\beta} < 1 \\ 0 & \text{per } \frac{y-\alpha}{\beta} \geq 1 \end{cases}$$

Ovvero, riscrivendo le condizioni su y in modo più esplicito,

se $\beta > 0$	**se $\beta < 0$**
$F_Y(y) = \begin{cases} 0 & \text{per } y < \alpha \\ \frac{y-\alpha}{\beta} & \text{per } \alpha \leq y < \alpha+\beta \\ 1 & \text{per } y \geq \alpha+\beta \end{cases}$	$F_Y(y) = \begin{cases} 1 & \text{per } y > \alpha \\ \frac{y-(\alpha+\beta)}{\lvert\beta\rvert} & \text{per } \alpha+\beta < y \leq \alpha \\ 0 & \text{per } y \leq \alpha+\beta \end{cases}$

Esempio 14.19. ***Una trasformazione lineare (cioè con $\alpha = 0$) di X con distribuzione esponenziale di parametro λ è ancora esponenziale, di parametro $\frac{\lambda}{\beta}$, purché $\beta > 0$.***

Innanzitutto notiamo che se $Y = \beta X$, con $X \sim Exp(\lambda)$ e se β è positivo, allora $Y(\Omega) = \mathbb{R}^+$. Allora

$$F_Y(y) = F_X\left(\tfrac{y}{\beta}\right) = \begin{cases} 0 & \text{per } \frac{y}{\beta} < 0 \\ 1 - e^{-\lambda \cdot \frac{y}{\beta}} & \text{per } \frac{y}{\beta} \geq 0 \end{cases}$$

o meglio,

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ 1 - e^{-\frac{\lambda}{\beta} y} & \text{per } y \geq 0 \end{cases}$$

che dimostra l'asserto.

Esercizio proposto 14.9. *Si ritrovi il risultato del precedente Esempio 14.19 attraverso il calcolo della densità.*

Esempio 14.20. ***Distribuzioni gaussiane come trasformazioni affini di una v.a. X con distribuzione gaussiana standard $N(0,1)$.***
Se $X \simeq N(0,1)$ ed

$$Y = \mu + \sigma X$$

allora

$$Y \sim N(\mu, \sigma^2),$$

ovvero

$$f_Y(y) = \frac{1}{\sqrt{2\pi\,\sigma^2}} \exp\{-\tfrac{1}{2}\left(\tfrac{y-\mu}{\sigma}\right)^2\}, \quad y \in \mathbb{R}.$$

Infatti, basta ricordare che

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}\,x^2\}, \quad x \in \mathbb{R},$$

in modo che, dalla formula generale, con $\alpha = \mu$ e $\beta = \sigma$ (da cui $|\beta| = \sqrt{\sigma^2}$) si ha

$$f_Y(y) = f_X\left(\tfrac{y-\mu}{\sigma}\right)\frac{1}{\sqrt{\sigma^2}}.$$

È importante sottolineare il significato dei parametri: ricordando che $\mathbb{E}(X) = 0$ e che $Var(X) = \mathbb{E}(X^2) = 1$ si ha che il parametro μ rappresenta il valore atteso di Y

$$\mathbb{E}(Y) = \mathbb{E}(\mu + \sigma X) = \mu + \sigma\mathbb{E}(X) = \mu + \sigma \cdot 0 = \mu,$$

mentre il parametro σ^2 rappresenta la varianza di Y

$$Var(Y) = \mathbb{E}\big[(Y-\mu)^2\big] = \mathbb{E}\big[(\sigma X + \mu - \mu)^2\big] = \mathbb{E}(\sigma^2 X^2) = \sigma^2\,Var(X) = \sigma^2.$$

Infine può essere utile notare che, per $\sigma > 0$, si ha

$$F_Y(y) = \Phi\left(\tfrac{y-\mu}{\sigma}\right),$$

e quindi anche la funzione di distribuzione di una gaussiana di parametri μ e σ^2, può essere calcolata attraverso l'uso delle tavole.

Invece per $\sigma < 0$, si ha

$$F_Y(y) = 1 - \Phi\left(\tfrac{y-\mu}{\sigma}\right) = \Phi\left(-\tfrac{y-\mu}{\sigma}\right), \tag{139}$$

dove l'ultima uguaglianza deriva dal fatto che, qualunque sia $x \in \mathbb{R}$, vale la relazione

$$\Phi(-x) = 1 - \Phi(x). \tag{140}$$

Tale relazione deriva immediatamente dal fatto che, se $X \sim N(0,1)$ allora anche $-X \sim N(0,1)$: infatti, dalla formula (**??**)*, per $\alpha = 0$ e $\beta = 1$, si ottiene che $f_{-X}(x) = f_X(-x) = f_X(x)$, in quanto la densità di probabilità di $X \sim N(0,1)$ è simmetrica rispetto allo zero, e di conseguenza*

$$\Phi(-x) = \mathbb{P}(X \le -x) = \mathbb{P}(-X \le -x) = \mathbb{P}(X \ge x) = \mathbb{P}(X > x) = 1 - \mathbb{P}(X \le x) = 1 - \Phi(x).$$

La relazione (140) si può anche ottenere con semplici passaggi, osservando che

$$\Phi(-x) = \int_{-\infty}^{-x} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,y^2}\,dy$$

cambiando variabile $z = -y$

$$= \int_{+\infty}^{+x} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,(-z)^2}\,(-dz) = \int_{x}^{\infty} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,z^2}\,dz$$

e d'altra parte

$$\begin{aligned} 1 - \Phi(x) &= 1 - \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,y^2}\,dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,y^2}\,dy - \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,y^2}\,dy \\ &= \int_{x}^{\infty} \frac{1}{\sqrt{2\,\pi}} e^{-\frac{1}{2}\,y^2}\,dy \end{aligned}$$

Grafico della funzione di distribuzione e della densità di una gaussiana standard

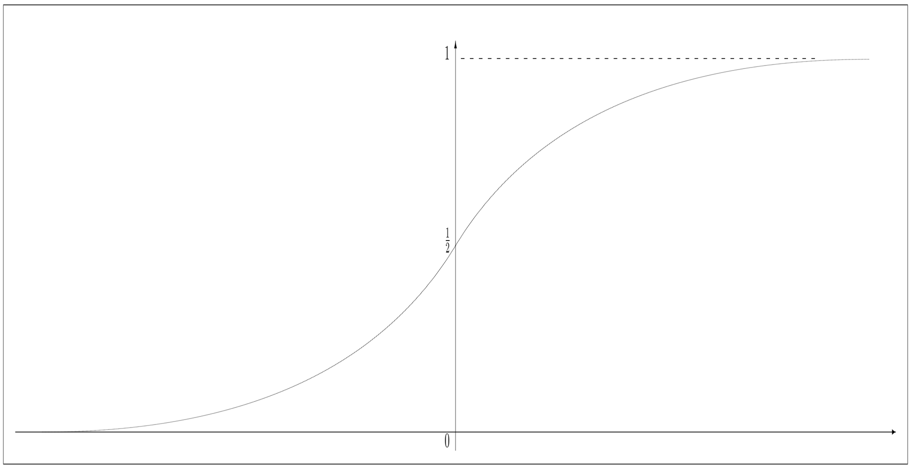

Figura 6: Grafico di $\Phi(x) = F_X(x)$, per X gaussiana standard

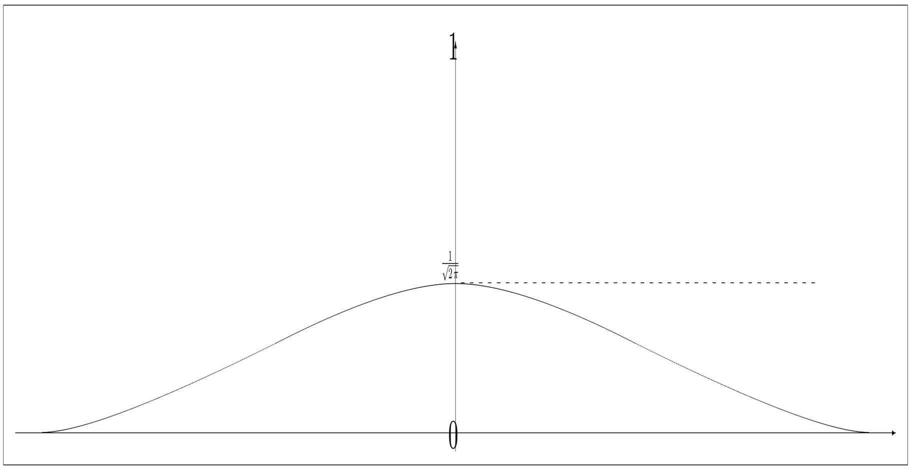

Figura 7: Grafico di $\varphi(x) = f_X(x)$, per X gaussiana standard

Studio della funzione $\varphi(x) = \frac{1}{\sqrt{2\,\pi}}\,e^{-\frac{x^2}{2}}$:

- $\varphi'(x) = \frac{1}{\sqrt{2\,\pi}}\,e^{-\frac{x^2}{2}}\,\left(-\frac{2\,x}{2}\right) = -x\,\varphi(x)$
- $\varphi''(x) = -\varphi(x) + (-x)\varphi'(x) = -\varphi(x) + (-x)(-x)\varphi(x) = \varphi(x)(x^2 - 1)$ e di conseguenza
  - $\varphi(x)$ è convessa per $x < -1$
  - $\varphi(x)$ è concava per $-1 < x < 1$
  - $\varphi(x)$ è convessa per $x > 1$

Il massimo della funzione si ha per $x = 0$ e vale $\varphi(0) = \frac{1}{\sqrt{2\,\pi}}$ $(\simeq 0,399)$.

Tavola della funzione di distribuzione gaussiana standard

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{y^2}{2}}\, dy$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Spiegazione dell'uso della tavola della gaussiana standard:

Per iniziare si noti che gli indici di riga sono i 35 numeri $\{0.0,\ 0.1,\ \dots,\ 3.3,\ 3.4\}$ che vanno da 0 a 3.4 e che differiscono tra loro di un decimo, mentre gli indici di colonna sono i 10 numeri $\{0.00,\ 0.01, \dots, 0.09\}$, che vanno da 0 a 0.09 e differiscono tra loro di un centesimo. Sommando un numero di riga, con uno di colonna si può ottenere uno tra i 350 valori di x che vanno da 0 a 3.49, e che differiscono tra loro di un centesimo. Viceversa ognuno di tali valori x, ad esempio $x = 1.43$, si può considerare come la somma della parte fino ai decimi più la parte dei centesimi, nell'esempio $x = 1.43 = 1.4 + 0.03$, individuando così un indice di riga, nell'esempio 1.4, ed uno di colonna, nell'esempio 0.03. Nella tavola, al posto di riga 1.4 e di colonna 0.03 si trova il valore di $\Phi(1.43)= 0.9236$, ovvero della funzione di distribuzione di una gaussiana standard, calcolata in $1.4 + 0.03$, e approssimato alla quarta cifra decimale.

I valori di $\Phi(x)$ per $x \geq 3.50$ si possono[72] approssimare con 1. Per quanto riguarda i valori di $\Phi(x)$ per valori negativi si usa la relazione (140), ossia

$$\Phi(-x) = 1 - \Phi(x),$$

ed in questo modo si può ottenere la funzione di distribuzione in[73] 699 valori tra $-3,49$ e $3,49$, equispaziati di un centesimo, ossia in

$$x = \frac{k}{100}, \qquad \text{per } -349 \leq k \leq 349.$$

Dalla relazione precedente si ottiene ad esempio che $\Phi(-1.43) = 1 - \Phi(1.43) = 1 - 0.9236 = 0.0764$

Infine la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria Y con distribuzione $N(\mu, \sigma^2)$, previo una trasformazione affine di Φ: questo infatti ci assicura la relazione (139), ossia

$$\mathbb{P}(Y \leq y) = \Phi\big(\tfrac{y-\mu}{\sigma}\big).$$

Ad esempio se $Z \sim N(1, 4)$ e si vuole calcolare $\mathbb{P}(Z \leq 3.86)$, dalla (139) si ottiene che, essendo $Z \sim N(\mu, \sigma^2)$, con $\mu = 1$ e $\sigma^2 = 4$,

$$\mathbb{P}(Z \leq 3.86) = \Phi\big(\tfrac{3.86-1}{2}\big) = \Phi(1.43) = 0.9236$$

Si noti infine che anche in questo la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria con distribuzione $N(\mu, \sigma^2)$ in 699 valori y, ossia i valori per i quali

$$-3.49 \leq \tfrac{y-\mu}{\sigma} \leq 3.49 \qquad \Leftrightarrow \qquad \mu - 3.49\,\sigma \leq y \leq \mu + 3.49\,\sigma$$

o più precisamente per $\frac{y-\mu}{\sigma} = \frac{k}{100}$, per $-349 \leq k \leq 349$, cioè

$$y = \mu + \frac{k}{100}\,\sigma, \qquad \text{per } -349 \leq k \leq 349.$$

Problema inverso: trovare x_α tale che $\Phi(x_\alpha) = \alpha$

$\Phi(x_\alpha) = \alpha$	.90	.95	.975	.99	.995	.999	.9995	.99995	.999995
x_α	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417

La tabella è autoesplicativa, forse vale la pena solo di sottolineare che se $x \geq x_\alpha$ allora $\Phi(x) \geq \alpha$.

[72] Ovviamente approssimare con 1 numeri maggiori o uguali a 0.9998 ha senso solo in problemi in cui la precisione non è fondamentale.

[73] Si noti che $\Phi(0) = \Phi(-0) = 1/2$ e quindi non si tratta di 700 valori, ma solo di 699

14.8 Esercizi di verifica

Esercizio 14.1. *Sia X una variabile aleatoria a valori in $\mathbb{Z}^+ = \{0\} \cup \mathbb{N}$.*
(a) Dimostrare che $\mathbb{E}(X) < \infty$ allora $n\,\mathbb{P}(X > n)$ tende a zero per n che tende ad infinito.
(b) Dimostrare che se $\mathbb{E}(X) < \infty$ allora

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k).$$

Esercizio 14.2. *Dimostrare che se X è una variabile aleatoria Geometrica di parametro θ, allora X ha la proprietà di mancanza di memoria, ovvero*

$$\mathbb{P}(X - k > h | X > k) = \mathbb{P}(X > h), \qquad \forall\, h, k \geq 0.$$

Esercizio 14.3. *Dimostrare che se X è una variabile aleatoria esponenziale di parametro λ, allora X ha la proprietà di mancanza di memoria, ovvero*

$$\mathbb{P}(X - t > s | X > t) = \mathbb{P}(X > s), \qquad \forall\, s, t \geq 0.$$

Esercizio 14.4. Sia X una variabile aleatoria con funzione di densità di probabilità data da

$$f(x) = \begin{cases} a, & 0 < x \leq 1, \\ bx, & 1 < x < 2 \\ 0 & \text{altrove} \end{cases}$$

dove a e b sono costanti non negative.
(a) Determinare l'intervallo dei possibili valori per a.
(b) Determinare l'intervallo dei possibili valori per b.
(c) Determinare, per ogni fissato valore di b, il valore di a in funzione di b.
(d) Calcolare, in funzione di b, calcolare

$$\mathbb{P}\left(\frac{1}{2} < X \leq \frac{3}{2}\right).$$

Esercizio 14.5. Sia $\lambda > 0$ e sia X una variabile aleatoria con distribuzione esponenziale di parametro λ e poniamo

$$c := \mathbb{P}\left(1 \leq X \leq 2\right).$$

(a) Determinare l'intervallo dei valori possibili per c, al variare di $\lambda > 0$.
(b) Determinare il valore di λ in funzione di c.

Esercizio 14.6. Sia X una variabile aleatoria con distribuzione simmetrica, ossia

$$\mathbb{P}(X \leq x) = \mathbb{P}(-X \leq x)\,\big(:= \mathbb{P}(X \geq -x)\big), \qquad x \in \mathbb{R}.$$

Nel caso in cui X ammetta funzione di densità di probabilità f, trovare quale condizione deve soddisfare f.

Esercizio 14.7. Consideriamo la funzione

$$F(x) = \frac{e^x}{e^x + e^{-x}}, \qquad x \in \mathbb{R}.$$

(a) Verificare che si tratta di una funzione di ripartizione.
(b) Determinare la funzione F^{-1} e la funzione di densità di probabilità f.

(c) Si tratta di una distribuzione simmetrica?

Esercizio 14.8. Sia X una variabile aleatoria con distribuzione uniforme sull'intervallo $(-2,2)$. Calcolare la probabilità condizionata

$$\mathbb{P}\left(-2 < 2X < 2 \middle| -\tfrac{5}{2} < 2X < \tfrac{5}{2}\right).$$

Esercizio 14.9. Sia X una variabile aleatoria con funzione di densità di probabilità data da

$$f(x) = 4\,x^3 \cdot \mathbf{1}_{(0,1)}(x), \qquad x \in \mathbb{R}.$$

Determinare a in modo che risulti

$$\mathbb{P}(X \le a) = \mathbb{P}(X \ge a).$$

Esercizio 14.10. Sia X una variabile aleatoria con funzione di densità di probabilità data da

$$f(x) = k\,\exp\{-|x|\}, \qquad x \in \mathbb{R}.$$

(a) Determinare il valore della costante k.
(b) Calcolare $\mathbb{P}(|X| \le 1)$.

Esercizio 14.11. Sia X una variabile aleatoria con funzione di densità di probabilità data da

$$f(x) = 2x\,\exp\{-x^2\}\,\mathbf{1}_{(0,\infty)}(x), \qquad x \in \mathbb{R}.$$

Determinare la funzione di densità di probabilità di $Y := X^2$.

Esercizio 14.12. Un punto X viene scelto a caso nell'intervallo $I = (0,2)$ e consideriamo le lunghezze dei due intervalli in cui X divide l'intervallo I. Sia Y l'area del rettangolo i cui lati hanno lunghezza uguale a questi intervalli.
(a) Dimostrare che $\mathbb{P}(Y > 1) = 0$.
(b) Determinare la distribuzione di Y.

Esercizio 14.13. Sia X una variabile aleatoria con funzione di densità di probabilità data da

$$f(x) = k\,x\,\mathbf{1}_{(0,2)}(x), \qquad x \in \mathbb{R}.$$

(a) Determinare il valore della costante k.
(b) Determinare la funzione di ripartizione di X.

Soluzione dell'Esercizio 14.1

Sia X una variabile aleatoria a valori in $\mathbb{Z}^+ = \{0\} \cup \mathbb{N}$. In tale caso la serie $\sum_{k=0}^{\infty} k\,\mathbb{P}(X=k)$ (per calcolare il valore atteso $\mathbb{E}(X)$) è a termini tutti non negativi e quindi o converge o diverge. La condizione $\mathbb{E}(X) < \infty$ equivale alla convergenza a zero della successione della serie resto:

$$\sum_{k=0}^{\infty} k\,\mathbb{P}(X=k) < \infty \quad \Leftrightarrow \quad \lim_{n\to\infty} \sum_{k=n}^{\infty} k\,\mathbb{P}(X=k) = 0.$$

Ovviamente

$$\sum_{k=n}^{\infty} k\ \mathbb{P}(X=k) \geq \sum_{k=n}^{\infty} n\ \mathbb{P}(X=k) = n \sum_{k=n}^{\infty} \mathbb{P}(X=k) = n\ \mathbb{P}(X \geq n) \geq n\ \mathbb{P}(X > n),$$

e ciò mostra che

$$\lim_{n\to\infty} n\ \mathbb{P}(X > n) = 0.$$

Procedendo come nella dimostrazione della ***Proposizione*** **9.8**, si ottiene che, per ogni variabile aleatoria X a valori in $\mathbb{Z}^+$,

$$\begin{aligned}
\sum_{k=0}^{n} k\ \mathbb{P}(X=k) &= \sum_{k=0}^{n} k\,[\mathbb{P}(X > k-1) - \mathbb{P}(X > k)] \\
&= \sum_{k=1}^{n} k\ \mathbb{P}(X > k-1) - \sum_{k=1}^{n} k\ \mathbb{P}(X > k) \\
&= \sum_{h=0}^{n-1} (h+1)\ \mathbb{P}(X > h) - \sum_{k=0}^{n-1} k\ \mathbb{P}(X > k) - n\ \mathbb{P}(X > n) \\
&= \sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\ \mathbb{P}(X > n).
\end{aligned}$$

A questo punto, per ottenere che $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$, basta osservare che, se $\mathbb{E}(X) < \infty$ allora

$$\begin{aligned}
\mathbb{E}(X) &= \lim_{n\to\infty} \sum_{k=0}^{n} k\ \mathbb{P}(X=k) = \lim_{n\to\infty} \Big[\sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\ \mathbb{P}(X > n)\Big] \\
&= \sum_{k=0}^{\infty} \mathbb{P}(X > k) - \lim_{n\to\infty} n\ \mathbb{P}(X > n) = \sum_{k=0}^{\infty} \mathbb{P}(X > k),
\end{aligned}$$

e ciò conclude la soluzione dell'Esercizio.
Tuttavia, vale al pena di osservare che, sempre per variabili aleatorie a valori in $\mathbb{Z}^+$, la condizione $\mathbb{E}(X) < \infty$ non è necessaria affinché $\mathbb{E}(X) = \sum_{k=0}^{\infty} \mathbb{P}(X > k)$ infatti (sia quando $\mathbb{E}(X) < \infty$ sia quando $\mathbb{E}(X) = \infty$)

$$\begin{aligned}
\mathbb{E}(X) &\geq \sum_{k=0}^{n} k\ \mathbb{P}(X=k) + \sum_{k=n+1}^{\infty} n\ \mathbb{P}(X > k) \\
&= \Big[\sum_{k=0}^{n-1} \mathbb{P}(X > k) - n\ \mathbb{P}(X > n)\Big] + n\ \mathbb{P}(X > n) = \sum_{k=0}^{n-1} \mathbb{P}(X > k).
\end{aligned}$$

e ciò mostra che se $\sum_{k=0}^{\infty} \mathbb{P}(X > k) = \infty$ allora anche $\mathbb{E}(X) = \infty$.

15 Variabili aleatorie in casi più generali: indipendenza, somme e somme aleatorie

Molte delle definizioni e delle proprietà delle variabili aleatorie in spazi finiti valgono anche per le variabili aleatorie generali. Ad esempio si ha ancora che il valore atteso della somma di variabili aleatorie è la somma dei valori attesi e la regola per il calcolo della varianza della somma rimane identica, ma è necessario aggiugere la condizione che tutte le variabili aleatorie ammettano valore atteso finito, nel primo caso, mentre nel secondo caso, occore aggiungere anche la condizione che il valore atteso del quadrato sia finito. Anche la definizione di covarianza di due variabili aleatorie rimane la stessa, purché i loro quadrati ammettano valore atteso finito. Volendo generalizzare il fatto che la varianza della somma coincide con la somma delle varianza, occorre prima di tutto generalizzare la nozione di indipendenza, data in precedenza solo per variabili aleatorie negli spazi di probabilità finiti, e data solo per due variabili aleatorie, con una nozione di ***indipendenza completa*** per più variabili aleatorie (Definizioni 15.3 e 15.4). Per questo motivo tecnico la lezione inizia con diverse definizioni di indipendenza. Tuttavia, in alcuni casi, è possibile, ottenere esplicitamente la distribuzione della loro somma, almeno nel caso di variabili (completamente) indipendenti. Ne vedremo un paio di esempi nella sezione 15.1.1. La lezione finisce anche con alcuni cenni alle somme di un numero aleatorio di variabili (completamente) indipendenti.

15.1 Famiglie di variabili aleatorie indipendenti

In questo paragrafo ci chiediamo come si deve definire l'indipendenza per due variabili aleatorie X ed Y, nel caso generale, e daremo anche un'ulteriore definizione di indipendenza completa (o globale) per più di due variabili aleatorie.

Tra le varie caratterizzazioni di indipendenza, sicuramente non possiamo generalizzare[74] quella per cui $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$, in quanto, ad esempio, per le variabili aleatorie con funzione di distribuzione continua, la precedente relazione sarebbe solo una banalità: infatti si ridurrebbe alla relazione[75] $0 = 0$. Possiamo invece generalizzare quella data in ***Proposizione* 8.1** della Lezione 8, nel seguente modo.

Definizione 15.1 (indipendenza di due variabili aleatorie)**.** *Due variabili aleatorie X ed Y si dicono* **indipendenti** *se e solo se comunque scelti due intervalli I e J, limitati o illimitati,*

$$\mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I) \cdot \mathbb{P}(Y \in J).$$

Come nel caso discreto, anche nel caso generale vale il risultato che l'indipendenza di due variabili aleatorie implica[76] la non correlazione, mentre non è vero il viceversa.

Strettamente collegata alla precedente definizione, c'è la seguente

Definizione 15.2 (indipendenza a due a due di n variabili aleatorie)**.** *Siano X_1, X_2, ..., X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Esse si dicono* **indipendenti a due a due** *se comunque scelti $i \neq j$, con $i, j \in \{1, 2, \ldots, n\}$, le due variabili aleatorie X_i ed X_j sono indipendenti, ovvero comunque scelti $i \neq j$, e comunque scelti I e J, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:*

$$\mathbb{P}(X_i \in I,\, X_j \in J) = \mathbb{P}(X_i \in I) \cdot \mathbb{P}(X_j \in J).$$

[74] Tuttavia nel caso delle variabili aleatorie discrete questa caratterizzazione rimane valida, infatti le dimostrazioni della equivalenza delle caratterizzazioni rimangono sostanzialmente invariate, pur di sostituire a somme finite somme infinite, per cui ad esempio due variabili aleatorie X ed Y con $X(\Omega) = \{x_1, x_2, \ldots\}$ ed $Y(\Omega) = \{y_1, y_2, \ldots\}$ sono indipendenti se e solo se $\mathbb{P}(X = x_h, Y = y_k) = \mathbb{P}(X = x_k)\,\mathbb{P}(Y = y_h)$ per ogni h e k.

[75] Se $\mathbb{P}(X = x) = 0$ per ogni $x \in \mathbb{R}$, allora anche $\mathbb{P}(X = x, Y = y) = 0$, in quanto $\{X = x, Y = y\} \subseteq \{X = x\}$.

[76] Ovviamente è necessario che le variabili aleatorie ammettano valore atteso finito.

Una condizione più forte dell'indipendenza a due a due è l'indipendenza completa o globale.

Un caso particolarmente interessante è quello in cui le variabili aleatorie X_i, per $i = 1, \dots, n$, sono completamente (o globalmente) indipendenti tra loro, ovvero

Definizione 15.3 (indipendenza di n variabili aleatorie)**.** *Siano X_1, X_2, ..., X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, P)$. Esse si dicono*[77] *completamente (o globalmente)* **indipendenti tra loro** *se comunque scelti J_1, J_2, ..., J_n, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:*

$$\mathbb{P}(X_1 \in J_1,\, X_2 \in J_2, \dots, X_n \in J_n) = \mathbb{P}(X_1 \in J_1) \cdot \mathbb{P}(X_2 \in J_2) \cdot \ldots \cdot \mathbb{P}(X_n \in J_n)$$

La precedente definizione implica l'indipendenza a due a due.

***Proposizione* 15.1.** *Se le n variabili aleatorie X_1, X_2, ..., X_n sono completamente (o globalmente) indipendenti fra loro, allora lo sono anche a due a due.*

Dimostrazione. Per semplicità di notazione mostriamo solamente che X_1 ed X_2 sono indipendenti, ma la dimostrazione è essenzialmente la stessa nel caso generale di X_i ed X_j.

Il punto essenziale da osservare è che $\mathbb{R}$ è un intervallo, e che gli eventi del tipo $\{X_k \in \mathbb{R}\}$ coincidono con l'evento certo, di conseguenza

$$\{X_1 \in J_1,\, X_2 \in J_2\} = \{X_1 \in J_1,\, X_2 \in J_2,\, X_3 \in \mathbb{R},\, \dots, X_n \in \mathbb{R}\}$$

e quindi

$$\begin{aligned}\mathbb{P}(X_1 \in J_1,\, X_2 \in J_2) &= \mathbb{P}(X_1 \in J_1,\, X_2 \in J_2, X_3 \in \mathbb{R} \dots, X_n \in \mathbb{R}) \\ &= \mathbb{P}(X_1 \in J_1) \cdot \mathbb{P}(X_2 \in J_2) \cdot \mathbb{P}(X_3 \in \mathbb{R}) \cdot \ldots \cdot \mathbb{P}(X_n \in \mathbb{R}) \\ &= \mathbb{P}(X_1 \in J_1) \cdot \mathbb{P}(X_2 \in J_2).\end{aligned}$$

□

***Osservazione* 15.1.** Come già detto, quanto visto per le variabili aleatorie discrete vale anche per le variabili aleatorie in generale: in particolare se le variabili aleatorie sono indipendenti a due a due, allora la varianza della somma è la somma delle varianze. Alla luce della precedente ***Proposizione* 15.1**, lo stesso vale nel caso in cui le variabili aleatorie sono completamente (o globalmente) indipendenti tra loro, ossia se, per ogni $n \geq 2$, comunque scelti J_1, J_2, ..., J_n, intervalli (limitati o illimitati) di $\mathbb{R}$, si ha:

$$\mathbb{P}(X_1 \in J_1,\, X_2 \in J_2, \dots, X_n \in J_n) = \mathbb{P}(X_1 \in J_1) \cdot \mathbb{P}(X_2 \in J_2) \cdot \ldots \cdot \mathbb{P}(X_n \in J_n)$$

Definizione 15.4 (indipendenza di una successione di variabili aleatorie)**.** *Sia $\{X_n;\ n = 1, 2, \dots\}$ una successione di variabili aleatorie, tutte definite sullo stesso spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$. Si dice che sono* **una successione di variabili aleatorie indipendenti** *se comunque scelto un numero finito di esse, queste risultano completamente indipendenti tra loro.*

15.1.1 Esempi di calcolo della somma di variabili aleatorie indipendenti

Sappiamo calcolare esattamente la distribuzione della somma di variabili aleatorie in alcuni casi specifici. Ad esempio quando le X_i sono le indicatrici di eventi E_i che formano uno schema di Bernoulli di parametro θ, sappiamo che la distribuzione della somma è la distribuzione binomiale $Bin(n; \theta)$.

[77] A volte il termine completamente può essere trascurato, e si può parlare semplicemente di variabili aleatorie indipendenti tra loro.

Esempio 15.1. *Ancora sappiamo che se due variabili aleatorie X_1 ed X_2 sono indipendenti e hanno distribuzione binomiale di parametri n_i e θ (attenzione n_1 può essere diverso da n_2, ma θ è lo stesso per $i = 1, 2$), allora la somma $X_1 + X_2$ ha distribuzione binomiale di parametri $n_1 + n_2$ e θ (confrontare lo svolgimento dell'Esercizio 8.4). Questo risultato si estende anche al caso di n variabili aleatorie completamente (o globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $Bin(m;\theta)$ allora S_n ha distribuzione $Bin(n \cdot m;\theta)$.*

Esempio 15.2 (Somma di variabili aleatorie di Poisson indipendenti). *Siano X_1 ed X_2 variabili aleatorie di Poisson di parametro $\lambda_1(>0)$ e $\lambda_2(>0)$ rispettivamente, ovvero per $i = 1,\ 2$*

$$\mathbb{P}(X_i = k) = \frac{\lambda_i^k}{k!}\, e^{-\lambda_i}, \qquad k = 0, 1, 2, \dots$$

Si assuma che le variabili siano indipendenti, ovvero che

$$\mathbb{P}(X_1 = h, X_2 = k) = \mathbb{P}(X_1 = h)\,\mathbb{P}(X_2 = k), \qquad \textit{per ogni } h,\ k \in \{0,\ 1\ \dots\}$$

Si vede facilmente che la variabile aleatoria $X_1 + X_2$ ha distribuzione di Poisson di parametro $\lambda_1 + \lambda_2$, ovvero che

$$\mathbb{P}(X_1 + X_2 = m) = \frac{(\lambda_1 + \lambda_2)^m}{m!}\, e^{-(\lambda_1+\lambda_2)}, \qquad m = 0, 1, 2, \dots \tag{141}$$

Infatti, per $m = 0, 1, \dots$ l'evento

$$\{X_1 + X_2 = m\} = \bigcup_{k=0}^{\infty}\{X_1 = k,\ X_2 = m-k\} = \bigcup_{k=0}^{m}\{X_1 = k,\ X_2 = m-k\},$$

in quanto $\{X_2 = m - k\} = \emptyset$ per $k = m + 1, m + 2, \dots$. Per cui

$$\begin{aligned}
\mathbb{P}(X_1 + X_2 = m) &= \sum_{k=0}^{m} \mathbb{P}(X_1 = k,\ X_2 = m-k) = \sum_{k=0}^{m} \mathbb{P}(X_1 = k)\,\mathbb{P}(X_2 = m-k) \\
&= \sum_{k=0}^{m} \frac{\lambda_1^k}{k!}\, e^{-\lambda_1}\, \frac{\lambda_2^{(m-k)}}{(m-k)!}\, e^{-\lambda_2} \\
&= \frac{1}{m!} \sum_{k=0}^{m} m!\, \frac{1}{k!}\, \frac{1}{(m-k)!}\, \lambda_1^k\, \lambda_2^{(m-k)}\, e^{-(\lambda_1+\lambda_2)}
\end{aligned}$$

dalla formula della potenza del binomio si ottiene la tesi, in quanto

$$\frac{1}{m!} \sum_{k=0}^{m} \binom{m}{k} \lambda_1^k\, \lambda_2^{(m-k)} = \frac{(\lambda_1 + \lambda_2)^m}{m!}, \qquad m = 0, 1, 2, \dots$$

da cui si ottiene immediatamente la (141).

Anche questo risultato si estende anche al caso di n variabili aleatorie completamente (o globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $Poiss(\lambda)$ allora S_n ha distribuzione $Poiss(n \cdot \lambda)$. Tuttavia va osservato che per calcolare $\mathbb{P}(S_n \le x)$, per $x \ge 0$, pur avendo a disposizione una formula esatta, ovvero

$$\mathbb{P}(S_n \le x) = \sum_{0 \le k \le \lfloor x \rfloor} \mathbb{P}(S_n = k) = \sum_{k=0}^{\lfloor x \rfloor} \frac{(n \cdot \lambda)^k}{k!}\, e^{-n \cdot \lambda},$$

se n è "grande", gli elementi della precedente sommatoria sono composti da fattori molto grandi ($(n \cdot \lambda)^k$) e molto piccoli ($e^{-n \cdot \lambda}$), e che quindi possono essere "scomodi" da calcolare. Nella successiva Lezione 16 vedremo come ottenere un valore approssimato per $\mathbb{P}(S_n \le x)$ anche in questo esempio.

Esempio 15.3 (Somma di n variabili geometriche indipendenti tutte di parametro p e distribuzione di Pascal). *Iniziamo con il caso $n = 2$. Dimostreremo che se X_1 ed X_2 sono due variabili aleatorie $Geom(p)$ (a partire da 1) indipendenti (ossia $\mathbb{P}(X_i = k) = p(1-p)^{k-1}$, per $k = 1, 2, \dots$, $i = 1, 2$ e $\mathbb{P}(X_1 = k_1, X_2 = k_2) = p(1-p)^{k_1-1}p(1-p)^{k_2-1}$, per $k_i = 1, 2, \dots$, $i = 1, 2$) allora $Z = X_1 + X_2$ è una v.a. con* **distribuzione di Pascal** *di parametri 2 e p ossia*

$$\mathbb{P}(Z = k) = \binom{k-1}{1} p^2 (1-p)^{k-2}, \quad k = 2, 3, \dots.$$

Consideriamo uno schema di Bernoulli infinito (ossia una successione di eventi $\{E_n\}_{n\geq 1}$ globalmente indipendenti) e siano T_1 uguale al tempo di primo successo e T_2 il tempo di secondo successo. Poniamo

$$\Delta_1 = T_1 \textit{ e } \Delta_2 = T_2 - T_1.$$

Dimostreremo che la distribuzione congiunta di Δ_1 e Δ_2 è la stessa di X_1 e X_2. Quindi Δ_1 e Δ_2 hanno la stesse marginali di X_1 e X_2, e sono indipendenti, e perciò Z ha la stessa distribuzione di $\Delta_1 + \Delta_2 = T_1 + (T_2 - T_1) = T_2$. Ora è facile convincersi che

$$\mathbb{P}(\Delta_1 = k_1, \, \Delta_2 = k_2) = (1-p)^{k_1-1} p \, (1-p)^{k_2-1} p, \quad k_i \geq 1, \, i = 1, \, 2,$$

(cioè la distribuzione congiunta di Δ_1, Δ_2, è la stessa di X_1, X_2) e che

$$\mathbb{P}(T_2 = k) = \binom{k-1}{1} p^2 (1-p)^{k-2}, \quad k = 2, 3, \dots,$$

in quanto l'evento $\{T_2 = k\}$ coincide con l'evento

$$\{\textit{la } k\textit{-sima prova è un successo, e tra le prime } \; k-1 \; \textit{prove c'è esattamente un successo}\} = E_k \cap \{S_{k-1} = 1\},$$

dove abbiamo posto $S_m = \sum_{k=1}^m \mathbf{1}_{E_k}$. Di conseguenza

$$\mathbb{P}(T_2 = k) = \mathbb{P}(E_k \cap \{S_{k-1} = 1\}) = \mathbb{P}(E_k) \, \mathbb{P}(S_{k-1} = 1) = p \binom{k-1}{1} p(1-p)^{(k-1)-1} = \binom{k-1}{1} p^2 (1-p)^{k-2}.$$

Consideriamo ora il caso n. Dimostreremo che se X_1, X_2, …, X_n sono variabili aleatorie $Geom(p)$ indipendenti (ossia $\mathbb{P}(X_i = k) = p(1-p)^{k-1}$, per $k = 1, 2, \dots$, $i = 1, 2, \dots$ e $\mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) = p(1-p)^{k_1-1}p(1-p)^{k_2-1} \cdots p(1-p)^{k_n-1}$, per $k_i = 1, 2, \dots$, $i = 1, 2, \dots$) allora, posto

$$Z_n = X_1 + X_2 + \cdots X_n$$

$$\mathbb{P}(Z_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k = n, n+1, \dots.$$

ossia Z_n ha una **distribuzione di Pascal** *di parametri n e p. Iniziamo considerando uno schema di Bernoulli infinito (ossia una successione di eventi globalmente indipendenti e tutti di probabilità p) e poniamo T_1 uguale al* **tempo di primo successo**, *T_2 il* **tempo di secondo successo**, …, *e T_n uguale al* **tempo di n-simo successo**. *Poniamo inoltre*

$$\Delta_1 = T_1, \; \Delta_2 = T_2 - T_1 \dots, \Delta_n = T_n - T_{n-1}.$$

Dimostreremo che la distribuzione congiunta di Δ_1, $\Delta_2, \dots \Delta_n$ è la stessa di $X_1, X_2 \dots, X_n$. Quindi Δ_1, $\Delta_2, \dots, \Delta_n$ hanno la stesse marginali di X_1, $X_2, \dots, X_n$, e sono indipendenti. Di conseguenza Z_n ha la stessa distribuzione di $\Delta_1 + \Delta_2 + \cdots + \Delta_n = T_1 + (T_2 - T_1) + \cdots + (T_n - T_{n-1}) = T_n$.
Ora è facile convincersi che

$$\mathbb{P}(\Delta_1 = k_1, \, \Delta_2 = k_2, \dots, \Delta_n = k_n) = (1-p)^{k_1-1} p \, (1-p)^{k_2-1} p \, \cdots (1-p)^{k_n-1} p, \;\; k_i = 1, 2, \dots, i = 1, \, 2, \dots, n,$$

(cioè la distribuzione congiunta di Δ_1*,* $\Delta_2, \dots \Delta_n$ *è la stessa di* $X_1,\ X_2 \dots, X_n$*) e che*

$$\mathbb{P}(T_n = k) = \binom{k-1}{1} p^n (1-p)^{k-n}, \quad k = n, n+1, \dots,$$

infatti l'evento $\{T_n = k\}$ *coincide con l'evento*

$$\{\text{la } k\text{-sima prova è un successo, e tra le prime } k-1 \text{ prove ci sono esattamente } n-1 \text{ successi}\}$$

ossia, essendo come al solito $S_m = \sum_{k=1}^m \mathbf{1}_{E_k}$*, dell'evento*

$$\{T_n = k\} = E_k \cap \{S_{k-1} = n-1\}$$

da cui

$$\mathbb{P}(T_n = k) = \mathbb{P}(E_k \cap \{S_{k-1} = n-1\}) = \mathbb{P}(E_k)\,\mathbb{P}(S_{k-1} = n-1) = p\binom{k-1}{n-1} p^{n-1}(1-p)^{(k-1)-(n-1)}$$
$$= \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

Se invece si considerano n v.a. X_1', X_2',..., X_n', indipendenti e $Geom(p)$, ma a partire da 0, ossia se $\mathbb{P}(X_1' = k_1,\ X_2' = k_2, ..., X_n' = k_n) = p(1-p)^{k_1}\,p(1-p)^{k_2} \cdots p(1-p)^{k_n}$, per $k_i = 0, 1, 2, \dots,\ i = 1,\ 2, ..., n$, allora $Z' = X_1' + X_2' + \cdots + X_n'$ è una v.a. con ***distribuzione binomiale negativa*** di parametri n e p, ossia

$$\mathbb{P}(Z' = k) = \binom{k+n-1}{n-1} p^n (1-p)^k, \quad k = 0, 1, 2, \dots.$$

Per convincersene, basta pensare che le $X_i' = X_i - 1$, dove le v.a. X_i, per $i \geq 1$, sono come nell'Esempio 15.3, e che quindi

$$Z' = \sum_{i=1}^n X_i' = \sum_{i=1}^n (X_i - 1) = \sum_{i=1}^n X_i - n = Z - n,$$

da cui

$$\mathbb{P}(Z' = k) = \mathbb{P}(Z - n = k) = \mathbb{P}(Z = k+n) = \binom{(k+n)-1}{n-1} p^n (1-p)^{(k+n)-n}$$
$$= \binom{n+k-1}{k} p^n (1-p)^{(k+n)-n}\ k \geq 0.$$

Come curiosità: il nome di ***binomiale negativa*** si spiega ricordando la definizione del coefficiente binomiale esteso ai numeri reali $\binom{\alpha}{k}$ (vedere la nota a pagina pagerefnota:coefficiente-binomiale-esteso) e osservando che, per $\alpha = -n$ si ha

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

e quindi

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

15.2 Somme aleatorie di variabili aleatorie indipendenti

In questa sezione ci occuperemo del seguente problema:

Sia data una successione $\{X_i,\, i \geq 1\}$ di variabili aleatorie globalmente indipendenti e tutte con la stessa distribuzione e una variabile aleatoria N, a valori negli interi non negativi, e indipendente dalla successione $\{X_i,\, i \geq 1\}$.

Siamo interessati alla somma delle prime N variabili aleatorie, ossia alla somma aleatoria

$$S_N := \sum_{i=1}^{N} X_i,$$

con la (solita) convenzione che $\sum_{i=1}^{0} a_k = 0$.

Un primo risultato è il seguente

***Proposizione* 15.2.** *Se le variabili aleatorie X_i sono indipendenti ed identicamente distribuite, tutte con valore atteso finito μ_X, e se la variabile aleatoria N è indipendente da $\{X_i,\ i \geq 1\}$ ed ha valore atteso finito μ_N, allora*

$$\mathbb{E}\big(S_N\big) = \mathbb{E}\Big(\sum_{i=1}^{N} X_i\Big) = \mu_N \cdot \mu_X.$$

Dimostrazione. Iniziamo osservando che, grazie alla condizione di indipendenza tra N e la successione $\{X_i,\ i \geq 1\}$, possiamo affermare che

$$\mathbb{E}\big[X_i|\{N=k\}\big] = \mathbb{E}\big[X_i\big] =: \mu_X, \quad \text{e} \quad \mathbb{E}\big[|X_i||\{N=k\}\big] = \mathbb{E}\big[|X_i|\big] =: \mu_{|X|}, \quad \forall i \geq 1 \ k \geq 0.$$

Usando la formula del valore atteso totale, rispetto alla partizione infinita

$$H_k^N = \{N=k\}, \quad k \geq 0,$$

possiamo intanto ottenere che S_N ammette valore atteso finito:

$$\begin{aligned}
\mathbb{E}\big(|S_N|\big) &= \mathbb{E}\Big(|\sum_{i=1}^{N} X_i|\Big) \leq \mathbb{E}\Big(\sum_{i=1}^{N} |X_i|\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\mathbb{E}\Big(\sum_{i=1}^{N} |X_i|\big|\{N=k\}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N=k)\mathbb{E}\Big(\sum_{i=1}^{k} |X_i|\big|\{N=k\}\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\sum_{i=1}^{k} \mathbb{E}\Big(|X_i|\big|\{N=k\}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N=k)\sum_{i=1}^{k} \mathbb{E}\Big(|X_i|\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)k\mu_{|X|} = \mu_{|X|} \sum_{k=0}^{\infty} k\ \mathbb{P}(N=k) \\
&= \mu_{|X|} \cdot \mu_N.
\end{aligned}$$

I passaggi per ottenere la tesi sono del tutto simili:

$$\begin{aligned}
\mathbb{E}\big(S_N\big) &= \sum_{k=0}^{\infty} \mathbb{P}(N=k)\mathbb{E}\Big(\sum_{i=1}^{N} X_i\big|\{N=k\}\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\mathbb{E}\Big(\sum_{i=1}^{k} X_i\big|\{N=k\}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N=k)\sum_{i=1}^{k} \mathbb{E}\Big(X_i\big|\{N=k\}\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\sum_{i=1}^{k} \mathbb{E}\Big(X_i\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N=k)k\mu_X = \mu_X \cdot \mu_N.
\end{aligned}$$

$\square$

Esempio* 15.4. *Nelle stesse ipotesi della precedente proposizione, alcolare*

$$\mathbb{E}\big(S_N|N=n\big),$$

Assumendo inoltre che le X_i abbiano varianza finita, ovvero che $Var(X_i)=\sigma_X^2$, calcolare

$$Var\big(S_N|N=n\big) := \mathbb{E}\big[\big(S_N-\mathbb{E}(S_N|N=n)\big)^2|N=n\big)\big].$$

Per la prima domanda osserviamo che chiaramente, condizionatamente a $N=n$, la variabile aleatoria coincide con la variabile aleatoria S_n e quindi

$$\mathbb{E}\big(S_N|N=n\big)=\mathbb{E}\big(S_n|N=n\big)=\mathbb{E}\big(\sum_{k=1}^{n}X_i|N=n\big)=\sum_{i=1}^{n}\mathbb{E}\big(X_i|N=n\big)=\sum_{i=1}^{n}\mathbb{E}(X_i)=n\mu_X$$

Dove l'ultimo passaggio dipende dal fatto che le variabili aleatorie X_i sono indipendenti da N.
Per la seconda domanda: si tratta di calcolare

$$\mathbb{E}\big[\big(S_N-\mathbb{E}(S_N|N=n)\big)^2|N=n\big)\big]=\mathbb{E}\big[\big(S_n-n\mu\big)^2|N=n\big)\big]=\mathbb{E}\big[\big(S_n-n\mu_X\big)^2\big)\big]$$

in quanto, essendo le v.a. X_i indipendenti da N, anche la v.a. S_n è indipendente da N. Di cosneguenza

$$\mathbb{E}\big[\big(S_N-\mathbb{E}(S_N|N=n)\big)^2|N=n\big)\big]=\mathbb{E}\big[\big(S_n-n\mu_X\big)^2\big)\big]=n\,\sigma_X^2$$

Numero di successi su un numero aleatorio di prove.

Sia N una variabile aleatoria e $\{A_i,\ i\geq 1\}$ una successione di eventi. La variabile aleatoria

$$S_N := \sum_{i=1}^{N}\mathbf{1}_{A_i}$$

rappresenta il numero di successi su un numero aleatorio di prove. Si tratta quindi dello stesso tipo di problema precedente, ma nel caso in cui le variabili aleatorie X_i coincidono con le funzioni indicatrici $\mathbf{1}_{A_i}$. Non sempre è possibile calcolare/individuare la distribuzione di S_N. Un caso abbastanza facile da capire è dato nel seguente esempio.

Esempio 15.5. *Supponiamo di lanciare n volte una moneta truccata in modo che la probabilità di testa sia p, e sia N il numero di teste ottenute. Successivamente si lancia, per N volte una seconda moneta, che è truccata in modo che la probabilità di testa sia θ. Sia X il numero di teste ottenute con la seconda moneta.*

Chiaramente N ha distribuzione $Bin(n,p)$, ma non è difficile convincersi che anche X ha distribuzione binomiale, di parametri n e $\theta\cdot p$.

L'idea è la seguente: per ottenere il numero X è equivalente supporre di lanciare per n volte contemporaneamente le due monete, e contare un successo solo se entrambe le monete presentano testa. Essendo gli eventi relativi alle due monete e a lanci diversi tutti indipendenti, è chiaro che si tratta di prove indipendenti in cui la probabilità di successo vale $\theta\cdot p$.

Esercizio proposto 15.1. *Dimostrare che $X\sim Bin(n,\theta\cdot p)$ utilizzando solo il fatto che $X:=\sum_{i=1}^{N}\mathbf{1}_{A_i}$, dove N è una variabile aleatoria $Bin(n,p)$ indipendente dalla successione di eventi A_i, con $\mathbb{P}(A_i)=\theta$.*

Nella prossima proposizione vediamo un altro caso in cui è possibile trovare la distribuzione di S_N: si tratta del caso in cui $N\sim Poiss(\lambda)$, la successione di eventi A_i, è come nel precedente Esercizio proposto 15.1, ossia una successione di eventi indipendenti con $\mathbb{P}(A_i)=\theta$, ed indipendenti da N. Alla luce del teorema di approssimazione di Poisson (Teorema 14.1) e dell'Esempio15.5, non è sorprendente che S_N abbia distribuzione di Poisson di parametro $\theta\cdot\lambda$.

***Proposizione* 15.3** (Thinning Poissoniano)**.** *Sia N una variabile aleatoria di $Poiss(\lambda)$, indipendente da una successione $\{A_i,\ i \geq 1\}$ di eventi indipendenti e tutti di probabilità θ. La variabile aleatoria*

$$S_N := \sum_{i=1}^{N} \mathbf{1}_{A_i}$$

ha distribuzione di Poisson di parametro $\theta\lambda$.

Dimostrazione. Qualunque sia $h \geq 0$

$$\begin{aligned}
\mathbb{P}(S_N = h) &= \mathbb{P}\Big(\sum_{i=1}^{N} \mathbf{1}_{A_i} = h\Big) = \sum_{k=0}^{\infty} \mathbb{P}(N=k)\, \mathbb{P}\Big(\sum_{i=1}^{N} \mathbf{1}_{A_i} = h \big| \{N=k\}\Big) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N=k)\, \mathbb{P}\Big(\sum_{i=1}^{k} \mathbf{1}_{A_i} = h \big| \{N=k\}\Big)
\end{aligned}$$

tenendo conto che se $N = k$ e $h > k$, allora è impossibile che il numero dei successi $\sum_{i=1}^{k} \mathbf{1}_{A_i}$ sia h, in quanto chiaramente $S_N \leq N$,

$$= \sum_{k=h}^{\infty} \mathbb{P}(N=k)\, \mathbb{P}\Big(\sum_{i=1}^{k} \mathbf{1}_{A_i} = h \big| \{N=k\}\Big).$$

A questo punto, sfruttando l'indipendenza tra N e $\{A_i,\ i \geq 1\}$, possiamo affermare che

$$\begin{aligned}
\mathbb{P}(S_N = h) &= \sum_{k=h}^{\infty} \mathbb{P}(N=k)\, \mathbb{P}\Big(\sum_{i=1}^{k} \mathbf{1}_{A_i} = h\Big) = \sum_{k=h}^{\infty} \mathbb{P}(N=k) \binom{k}{h} \theta^k\, (1-\theta)^{k-h} \\
&= \sum_{k=h}^{\infty} \frac{\lambda^k}{\cancel{k!}}\, e^{-\lambda}\, \frac{\cancel{k!}}{h!(k-h)!}\, \theta^k\, (1-\theta)^{k-h} = e^{-\lambda} \sum_{k=h}^{\infty} \frac{\lambda^{k-h}\lambda^h}{h!(k-h)!}\, \theta^h\, (1-\theta)^{k-h} \\
&= e^{-\lambda}\, \frac{\lambda^h\theta^h}{h!} \sum_{k=h}^{\infty} \frac{\lambda^{k-h}(1-\theta)^{k-h}}{(k-h)!} = \frac{\lambda^h\theta^h}{h!}\, e^{-\lambda} \sum_{k'=0}^{\infty} \frac{[\lambda(1-\theta)^{k'}]}{k'!} \\
&= \frac{\lambda^h\theta^h}{h!}\, e^{-\theta\lambda}.
\end{aligned}$$

□

16 Variabili aleatorie in casi più generali: Legge dei Grandi Numeri e Teorema Centrale del Limite.

In questa Lezione riprenderemo il discorso iniziato nella Lezione 10 a proposito della media aritmetica Y_n di n variabili aleatorie non correlate, iniziato con la ***Proposizione* 10.9** della Lezione 10, la cui dimostrazione è basata sulla disuguaglianza di Chebyshev.

Riprendiamo questo problema nella sezione 16.1 sulla Legge (debole) dei Grandi Numeri, sempre basandoci sulla disuguaglianza di Chebyshev, ma considerando invece una successione di variabili aleatorie.

Nella Lezione 10 ci siamo posti anche una importante domanda su quanto grande si dovesse prendere n in modo che la probabilità dell'evento

$$\{ \textit{media aritmetica } Y_n \textit{ e valore atteso differiscono di } \textbf{\textit{poco}} \}$$

sia ***vicina ad uno***. La risposta era anch'essa basata sulla disuguaglianza di Chebyshev. Sempre nella sezione 16.1 si trova qualche approfondimento su questo problema.

Tuttavia, come viene osservato all'inizio della sezione 16.2, risultati più precisi alla domanda posta nella Lezione 10 si potrebbero ottenere se fosse nota la funzione di distribuzione della somma S_n delle variabili aleatorie, dato che $Y_n = \frac{1}{n} S_n$. In alcuni casi la distribuzione di S_n si può calcolare esplicitamente, come visto nella sezione 16.2 per la somma di variabili aleatorie completamente indipendenti. Tuttavia anche nel caso di variabili (completamente) indipendenti, spesso non è facile, o addirittura non è possibile, ottenere esplicitamente la distribuzione della somma. Per la soluzione approssimata di questo problema ci aiuta il Teorema Centrale del Limite (***Proposizione* 16.3**), come è illustrato nella ***Proposizione* 16.2**.

Terminiamo questo discorso introduttivo ricordando che, come preannunciato (sempre nella Lezione 10, e precisamente nella ***Proposizione* 10.10**) il Teorema Centrale del Limite, che riguarda successioni di variabili aleatorie ***completamente indipendenti***, è connesso con una proprietà che riguarda la somma standardizzata di n variabili aleatorie. La ***Proposizione* 10.10** della Lezione 10, riguarda solo il caso di uno schema di n prove bernoulliane, in cui la media aritmetica diviene la frequenza relativa dei successi, ma si generalizza immediatamente al caso di variabili aleatorie più generali.

16.1 Legge dei Grandi Numeri

Il risultato più importante di questa Lezione è noto come la ***Legge debole dei grandi numeri***. Tale risultato è enunciato alla fine di questo paragrafo ***Proposizione* 16.1**, e riguarda le successioni di variabili aleatorie indipendenti a due a due.

Prima di arrivare ad enunciare e dimostrare la legge debole dei grandi numeri, riprendiamo la diseguaglianza di Chebyshev della Lezione 10 (***Proposizione* 10.8**), ma allargando un poco la prospettiva. Prima di tutto va detto che, trattando ora variabili aleatorie in generale, per una data variabile aleatoria X, potrebbe non ammettere valore atteso finito (e in tal caso non avrebbe senso parlare di varianza) oppure ammettere valore atteso finito, ma avere varianza uguale a $+\infty$, la diseguaglianza di Chebyshev continua a valere anche nel caso di variabili aleatorie generali, con l'unica accortezza che ***nel caso generale bisogna ipotizzare che esistano finiti valore atteso e varianza***[78] ***della variabile aleatoria*** X. Per cui, indicando come al solito $\mu = \mathbb{E}(X)$ e $\sigma^2 = Var(X)$ si ha

$$\mathbb{P}(\{|X - \mu| > \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

[78]Sappiamo che ci sono variabili aleatorie per le quali valore atteso e/o varianza non esistono, o valgono infinito (vedere la nota a pagina 213). Questo problema non si pone nel caso finito in quanto in quel caso il calcolo del valore atteso e della varianza si riduce ad una somma finita e non presenta quindi nessun tipo di problema.

Anche la ***Proposizione* 10.9** continua a valere, ***pur di assumere che esistano finiti valore atteso*** $\mathbb{E}(X_i)$ ***e varianza*** $Var(X_i)$, che come al solito poniamo uguali rispettivamente a μ e σ^2.

Teorema 16.1 (***Proposizione* 10.9**, versione generale, ossia **verso la Legge dei Grandi Numeri**). *Se X_1, X_2, ..., X_n sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, e se esistano finiti valore atteso $\mathbb{E}(X_i) = \mu$ e varianza $Var(X_i) = \sigma^2$ allora*

$$\mathbb{P}(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2},$$

dove Y_n è la media aritmetica $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

Anche se la dimostrazione è sostanzialmente la stessa della ***Proposizione* 10.9**, vale la pena riportarla qui:
Prima di tutto $\mathbb{E}(Y_n) = \mu$, come si vede subito:

$$\mathbb{E}(Y_n) = \mathbb{E}\big(\frac{X_1 + ... + X_n}{n}\big) = \frac{1}{n}\mathbb{E}(X_1 + ... + X_n) = \frac{1}{n}\left[\mathbb{E}(X_1) + ... + \mathbb{E}(X_n)\right] = \frac{1}{n}\, n\mu = \mu.$$

Inoltre $Var(Y_n) = \frac{\sigma^2}{n}$, come si vede subito:

$$Var(Y_n) = Var\big(\frac{X_1 + ... + X_n}{n}\big) = \frac{1}{n^2}\, Var(X_1 + ... + X_n) = \frac{1}{n^2}[Var(X_1) + ... + Var(X_n)] = \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}.$$

Basta poi applicare la disuguaglianza di Chebyshev alla v.a. Y_n.

Riportiamo inoltre alcune osservazioni sull'esistenza del valore atteso finito:

Proprietà 1. In uno spazio di probabilità generale se X ed Y hanno valore atteso finito, allora esiste finito anche il valore atteso di $X + Y$.
(sugg: $|X + Y| \leq |X| + |Y|$, poi si usa la monotonia del valore atteso e la linearità del valore atteso)

Proprietà 2. In uno spazio di probabilità generale se X ed Y hanno momento secondo finito, ovvero X^2 ed Y^2 hanno valore atteso finito, allora (a) anche $X + Y$ ha momento secondo finito, ovvero esiste finito anche il valore atteso di $(X + Y)^2$, ed inoltre (b) esiste finito anche il valore atteso di XY.
(sugg: (a) $(X + Y)^2 \leq 2X^2 + 2Y^2$ (b) $|XY| \leq (X^2 + Y^2)/2$)

16.1.1 Approfondimenti sull'utilizzo della disuguaglianza di Chebyshev

In prima lettura, si consiglia di saltare questo paragrafo

Se X_1, X_2, ..., X_n sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, nell'***Osservazione 10.6*** della Lezione 10 abbiamo visto come trovare il numero n di prove per cui la probabilità dell'evento "*il valore atteso μ e la media aritmetica $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ differiscono meno di una quantità prefissata ε*"sia almeno $1 - \delta$, (nell'Esempio 10.6 ciò è stato applicato al caso di variabili binarie).

Infatti sappiamo che se

$$n \geq \frac{\sigma^2}{\delta\, \varepsilon^2} \tag{142}$$

allora

$$\mathbb{P}(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} \leq \delta \quad \Leftrightarrow \quad \mathbb{P}(\{-\varepsilon \leq Y_n - \mu \leq \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} \geq 1 - \delta.$$

Nel caso particolare in cui le variabili X_i siano variabili binarie, con $\mathbb{P}(X_i = 1) = \theta$ e $\mathbb{P}(X_i = 0) = 1 - \theta$, allora $\mu = \theta$, $\sigma^2 = \theta(1-\theta)$ e basta prendere

$$n \geq \frac{\theta(1-\theta)}{\delta\, \varepsilon^2}$$

per ottenere che la probabilità dell'evento "*la frequenza relativa dei successi*[79] Y_n *differisce dalla probabilità di successo* θ *meno di* ε" sia maggiore di $1 - \delta$.
Ovvero

$$n \geq \frac{\theta(1-\theta)}{\delta\, \varepsilon^2} \quad \Rightarrow \quad \mathbb{P}(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1 - \delta. \tag{143}$$

Nelle applicazioni si usa la frequenza relativa per "*stimare*" la probabilità θ: ovvero possiamo considerare il caso in cui possiamo osservare gli esiti di diversi esperimenti di uno stesso fenomeno, gli esperimenti sono condotti nelle stesse condizioni, per cui la probabilità di successo dell'esperimento è la stessa in tutte le prove, e infine ***si assume che le prove siano stocasticamente indipendenti tra loro***, tuttavia ***non si assume che sia noto esattamente il valore della probabilità di successo*** θ.

In questo contesto la misura di probabilità dipende dal parametro $\boldsymbol{\theta}$ ***ed è quindi più opportuno indicarla con*** $\mathbb{P}_{\boldsymbol{\theta}}$***, invece che con*** $\mathbb{P}$.

Riprendendo quanto detto nell'***Osservazione 10.6*** della Lezione 10 in questo contesto possiamo riscrivere

$$\mathbb{P}_\theta(\{\theta - \varepsilon \leq Y_n \leq \theta + \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2},$$

ma anche

$$\mathbb{P}_\theta(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2}.$$

Questo secondo modo di scrivere è più interessante, in quanto, in questo contesto, mentre possiamo osservare Y_n, invece non conosciamo affatto θ. L'idea è che vorremmo poter "*valutare*" la probabilità θ con Y_n, con un errore al più di ε. Ovviamente in nessun caso, facendo degli esperimenti, avremo la garanzia che la frequenza relativa Y_n e la probabilità di successo θ differiscano meno di ε, tuttavia la disuguaglianza di Chebyshev ci permette di affermare che ciò accade con probabilità elevata, e permette anche di trovare delle limitazioni inferiori a tale probabilità.

A prima vista però sorge una difficoltà: sembra che per adoperare la disuguaglianza di Chebyshev sia necessario conoscere θ, mentre abbiamo assunto che θ non sia noto. Ma a questo problema si può ovviare osservando che la funzione $h(x) = x(1-x)$ vale al massimo[80] $\frac{1}{4}$ e quindi si ha

$$\mathbb{P}_\theta(\{|Y_n - \theta| > \varepsilon\}) \leq \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \leq \frac{1}{4\,\varepsilon^2\, n},$$

$$\Updownarrow$$

$$\mathbb{P}_\theta(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\,\varepsilon^2\, n}.$$

Ciò permette di affermare che, ***qualunque sia la probabilità di successo*** θ, la probabilità che θ e la frequenza relativa Y_n differiscano meno di ε vale almeno $1 - \frac{1}{4\,\varepsilon^2\, n}$.

[79] Successo all'i-esima prova significa $X_i = 1$.
[80] La funzione $h(x) = x(1-x)$ ha il suo punto di massimo in $x = \frac{1}{2}$ come si vede subito, e quindi $h(x) \leq h(\frac{1}{2}) = \frac{1}{4}$.

Più interessante ancora, dal punto di vista operativo, è tuttavia il fatto che siamo in grado di rispondere ***alla domanda***:

Quante prove si devono effettuare, ovvero quanto si deve prendere grande n, affinché, con probabilità almeno $1-\delta$, la frequenza relativa differisca dalla probabilità di successo meno di ε?

La risposta alla precedente domanda è molto semplice: ***è sufficiente prendere***[81]

$$n \geq \frac{1}{4\,\delta\,\varepsilon^2}, \tag{145}$$

in altre parole

$$n \geq \frac{1}{4\,\delta\,\varepsilon^2} \quad \Rightarrow \quad \mathbb{P}_\theta(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1-\delta \qquad \forall\,\theta \in (0,1). \tag{146}$$

Infatti in tale caso (145) è equivalente a $\delta \geq \frac{1}{4\,\varepsilon^2\,n}$ e quindi, ***qualunque sia*** θ

$$\mathbb{P}_\theta(|Y_n - \theta| > \varepsilon\}) \leq \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \leq \frac{1}{4\,\varepsilon^2\,n} \leq \delta,$$

$$\Updownarrow$$

$$\mathbb{P}_\theta(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\,\varepsilon^2\,n} \geq 1-\delta.$$

Esempio* 16.1. *Sia Y_n la frequenza relativa dei successi in uno schema di Bernoulli di parametro θ. Si determini un n in modo che, **qualunque sia il valore di** θ, l'errore assoluto tra Y_n e θ sia minore di 0.1, con probabilità almeno 0.99.*

Soluzione: Siamo nel caso precedente con $\varepsilon = 0.1 = \frac{1}{10}$ e con $1-\delta = 0.99$, ovvero $\delta = \frac{1}{100}$. Quindi se

$$n \geq \frac{1}{4 \cdot \left(\frac{1}{10}\right)^2 \cdot \frac{1}{100}} = \frac{10000}{4} = 2500,$$

allora

$$\mathbb{P}_\theta(-\varepsilon \leq Y_n - \theta \leq \varepsilon) \geq 0.99$$

E quindi, qualunque sia il valore di θ, è sufficiente prendere $n = 2500$.

Esempio* 16.2. *Calcolare il minimo valore di n per il quale, in uno schema di Bernoulli con probabilità θ,* **in base alla disuguaglianza di Chebyshev**, *si possa scrivere*

$$\mathbb{P}_\theta\left(\left\{\left|\frac{S_n}{n} - \theta\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{10},$$

***qualunque sia il valore di** θ.*

[81] Si deve prendere

$$n \geq n(\varepsilon,\delta) \tag{144}$$

dove

$$n(\varepsilon,\delta) := \left\lceil \frac{1}{4\,\varepsilon^2\,\delta} \right\rceil,$$

cioè la parte intera superiore di $\frac{1}{4\,\varepsilon^2\,\delta}$. Si ricordi che la parte intera superiore di un numero reale a è l'intero k tale che $k-1 < a \leq k$, ed è indicata appunto con $\lceil a \rceil$.

Soluzione: Si può procedere considerando che

$$\mathbb{P}_\theta\left(\left\{\left|\frac{S_n}{n}-\theta\right|>\frac{1}{30}\right\}\right)\leq\frac{\theta(1-\theta)}{n\,\left(\frac{1}{30}\right)^2}\leq\frac{1}{4\,n\,\left(\frac{1}{30}\right)^2}=\frac{900}{4\,n}\leq\frac{1}{10},$$

$$\Updownarrow$$

$$\frac{900}{4\,\frac{1}{10}}=\frac{9000}{4}=2250\leq n,$$

oppure direttamente utilizzando la formula (146)

$$n\geq\frac{1}{4\,\varepsilon^2\,\delta}=\frac{1}{4\,\left(\frac{1}{30}\right)^2\,\frac{1}{10}}=\frac{900}{4\,\frac{1}{10}}=\frac{9000}{4}=2250.$$

***Osservazione* 16.1.** Si suggerisce di confrontare il risultato con quello dell'Esempio 10.6, in cui invece il valore di θ era dato, e quindi si era ottenuto, sempre utilizzando la disuguaglianza di Chebyshev[82], che bastava prendere $n=2223$.

***Osservazione* 16.2.** Si faccia attenzione al fatto che queste limitazioni inferiori sono date in base alla disuguaglianza di Chebyshev. I valori ottenuti per n sono sicuramente validi, ma sono eccessivamente grandi ed in genere più elevati del necessario. In realtà bastano valori di n più piccoli (daremo un'idea del motivo per cui i valori trovati sono eccessivi nella Lezione sul Teorema centrale del limite).

***Osservazione* 16.3 (Errore relativo).** Va anche sottolineato che finora abbiamo valutato solo l'errore assoluto, tra Y_n e θ, mentre avrebbe più interesse l'errore relativo, ovvero $\left|\frac{Y_n-\theta}{\theta}\right|$: infatti se θ fosse dell'ordine di un centesimo, stimare θ con un errore assoluto dell'ordine di un decimo non sarebbe molto ragionevole[83]. In questo caso la maggiorazione della disuguaglianza di Chebyshev permette di affermare che, per ogni θ

$$\mathbb{P}_\theta\left(\left|\frac{Y_n-\theta}{\theta}\right|>\varepsilon\}\right)=\mathbb{P}_\theta\left(|Y_n-\theta|>\varepsilon\theta\}\right)\leq\frac{1}{n}\frac{\theta(1-\theta)}{(\theta\varepsilon)^2}=\frac{1-\theta}{\theta}\,\frac{1}{n\,\varepsilon^2}\leq\delta,$$

per cui

$$n\geq\frac{1-\theta}{\theta}\,\frac{1}{\delta\,\varepsilon^2}\quad\Rightarrow\quad\mathbb{P}_\theta\left(\left|\frac{Y_n-\theta}{\theta}\right|>\varepsilon\}\right)\leq\delta$$

Purtroppo, se θ non è noto, questa limitazione inferiore non è molto utile in quanto la funzione $h_1(x)=\frac{1-x}{x}=\frac{1}{x}-1$ converge ad infinito per $x\to 0^+$, ed è quindi impossibile[84] trovare un valore di n che sia valido qualunque sia θ.

16.1.2 Formulazione della Legge dei Grandi Numeri

Nel formulare la domanda con la richiesta di scegliere n abbastanza grande in modo che sia sufficientemente piccola la probabilità che media aritmetica e valore atteso differiscono di poco, c'è un punto che abbiamo volutamente trascurato fin qui. La possibilità di scegliere n presuppone di avere a disposizione un numero di eventi, (o di variabili aleatorie) completamente indipendenti potenzialmente

[82] In realtà nell'Esempio citato si è utilizzata la (143).

[83] Se nel misurare la distanza fra due città si commette un errore dell'ordine di un metro, ci possiamo dichiarare completamente soddisfatti, mentre certamente non lo saremmo se l'errore dell'ordine di un metro riguardasse la misura di un tavolo da mettere in cucina!!!

[84] Diverso è il caso in cui, pur non conoscendo esattamente θ si sappia che $\theta\geq\theta_0$ con $\theta_0>0$: allora basta prendere $n\geq\frac{1-\theta_0}{\theta_0}\,\frac{1}{\delta\,\varepsilon^2}$.

grande a piacere[85].

Dal punto di vista matematico è più comodo poter affermare direttamente di avere a disposizione una successione di eventi completamente indipendenti e tutti con la stessa probabilità θ, o una successione di variabili aleatorie completamente indipendenti. Ciò presuppone uno spazio di probabilità Ω infinito, e quindi solo dopo aver introdotto gli spazi di probabilità generali e la nozione di successioni di variabili aleatorie, riformuliamo la ***Proposizione* 10.9** della Lezione 10 per successioni di variabili aleatorie. Tale formulazione è nota con il nome di Legge Debole dei Grandi Numeri.

***Proposizione* 16.1 (Legge Debole dei Grandi Numeri).** *Sia $\{X_i,\ i \geq 1\}$ una successione di v.a. indipendenti a due a due ed identicamente distribuite[86], per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$ e $Y_n = \frac{S_n}{n}$, si ha, qualunque sia $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(|Y_n - \mu| > \varepsilon\right) = 0$$

Dimostrazione. Basta osservare che

$$0 \leq \mathbb{P}\left(|Y_n - \mu| > \varepsilon\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2},$$

mandare n all'infinito ed usare il Teorema del confronto per le successioni numeriche:

$$0 \leq \lim_{n\to\infty} \mathbb{P}\left(|Y_n - \mu| > \varepsilon\right) \leq \lim_{n\to\infty} \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} = 0.$$

□

***Osservazione* 16.4.** Dalla ***Proposizione* 15.1** appare immediato che se $\{X_i,\ i \geq 1\}$ è una successione di variabili aleatorie completamente indipendenti, allora la Legge Debole dei Grandi Numeri continua a valere. Sotto questa ulteriore ipotesi vale anche il così detto Teorema centrale del limite che è oggetto del prossimo paragrafo. Nel prossimo paragrafo vedremo anche alcune relazioni tra questi due importantissimi risultati.

16.2 Somma di variabili aleatorie indipendenti e Teorema Centrale del Limite

Come abbiamo detto la disuguaglianza di Chebyshev permette di trovare delle limitazioni inferiori alle probabilità del tipo

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right)$$

che a loro volta permettono di dedurre la legge dei grandi numeri. Tuttavia se si conoscesse la funzione di distribuzione $F_{S_n}(x)$ della variabile aleatoria S_n, tale probabilità si potrebbe calcolare esattamente come

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right) &= \mathbb{P}\left(n(\mu-\varepsilon) \leq S_n \leq n(\mu+\varepsilon)\right) = F_{S_n}\left(n(\mu-\varepsilon)\right) - F_{S_n}\left(-\left(n(\mu-\varepsilon)\right)^-\right) \\ &= F_{S_n}\left(n(\mu-\varepsilon)\right) - F_{S_n}\left(-n(\mu-\varepsilon)\right) + \mathbb{P}\left(\{S_n = -n(\mu-\varepsilon)\}\right)\end{aligned}$$

Appare quindi chiaro che calcolare la distribuzione della somma di variabili aleatorie $S_n = X_1 + X_2 + \ldots + X_n$ sia un problema interessante è, oltre che di per sé, anche per le connessioni con la legge dei grandi numeri e delle relazioni tra media aritmetica e valore atteso.

---

[85] Si potrebbe ovviare al problema supponendo di avere una successione di spazi di probabilità $(\Omega_n, \mathcal{P}(\Omega_n), P^{(n)})$ e su ciascuno spazio n eventi $E_1^{(n)}$, $E_2^{(n)}$,..., $E_n^{(n)}$ che formano uno schema di Bernoulli con probabilità $\theta^{(n)} = \theta$ per ogni n.

[86] Poiché le variabili aleatorie X_n hanno tutte la stessa distribuzione, si ha che se esistono finiti valore atteso e varianza di X_1, allora esistono finiti valore atteso e varianza di X_i e coincidono con quelli di X_1.

16.2.1 Approssimazione normale e Teorema Centrale del Limite

Più complesso risulta il calcolo della funzione di distribuzione della somma per altre variabili aleatorie[87], tuttavia si può innanzi tutto osservare come calcolare la funzione di distribuzione di S_n sia equivalente a calcolare la distribuzione di una sua trasformata affine[88] ovvero:

se a_n e b_n sono numeri reali, con $b_n > 0$, allora[89]

$$\{S_n \leq x\} = \left\{ \frac{S_n - a_n}{b_n} \leq \frac{x - a_n}{b_n} \right\}$$

Una scelta naturale per a_n e per b_n è quella che trasforma S_n in una variabile aleatoria standard, ovvero quella di prendere $a_n = \mathbb{E}(S_n)$ e $b_n = \sqrt{Var(S_n)}$.

In questo modo infatti, per la disuguaglianza di Chebyshev, sappiamo che, qualunque siano n ed $\alpha > 0$

$$\mathbb{P}(-\alpha \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} \leq \alpha) \geq 1 - \frac{1}{\alpha^2}.$$

Alla luce della seguente ***Proposizione* 16.3**, nota come Teorema Centrale del Limite (o anche Teorema del Limite Centrale), si può dimostrare il seguente risultato.

***Proposizione* 16.2 (Approssimazione normale).** *Se le variabili aleatorie X_i, per $i =, 1, 2 \ldots, n$ sono (globalmente o completamente) indipendenti, hanno la stessa distribuzione, ammettono valore atteso finito $\mu = \mu_X$, varianza finita $\sigma^2 = \sigma^2_X$ e non nulla, allora $\mathbb{E}(S_n) = n\mu$, $Var(S_n) = n\sigma^2 > 0$, e*

$$F_{S_n}(x) = \mathbb{P}(S_n \leq x) = \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq \frac{x - n\mu}{\sqrt{n\,\sigma^2}} \right) \simeq \Phi\left(\frac{x - n\mu}{\sqrt{n\,\sigma^2}} \right), \tag{147}$$

dove $\Phi(x)$ è la funzione di distribuzione di una variabile aleatoria gaussiana standard $N(0, 1)$.

A titolo di esempio riprendiamo il caso in cui le variabili aleatorie X_i hanno distribuzione di Poisson di parametro λ. Si vede facilmente che $\mathbb{E}(X_i) = \lambda$ e che $Var(X_i) = \lambda$. In particolare, per $\lambda = 1$ si ha $\mu = \lambda = 1$ e $\sigma^2 = \lambda = 1$, $n = 100$ ed $x = 100$, possiamo calcolare approssimativamente $\mathbb{P}(S_{100} \leq 100)$ attraverso $\Phi\left(\frac{x-n\mu}{\sqrt{n\,\sigma^2}}\right) = \Phi\left(\frac{100-100}{\sqrt{100}}\right) = \Phi(0) = \frac{1}{2}$.

A titolo di esempio riprendiamo il caso in cui le variabili aleatorie X_i hanno distribuzione Geometrica di parametro $p \in (0, 1)$. Sappiamo che $\mathbb{E}(X_i) = 1/p$ e che $Var(X_i) = (1 - p)/p$. In particolare, per $p = 1/2$, si ha $\mu = 1/p = 2$ e $\sigma^2 = (1 - p)/p = 1$ $n = 100$ ed $x = 200$, possiamo calcolare approssimativamente $\mathbb{P}(S_{100} \leq 200)$ attraverso $\Phi\left(\frac{x-n\mu}{\sqrt{n\,\sigma^2}}\right) = \Phi\left(\frac{200-200}{\sqrt{100}}\right) = \Phi(0) = \frac{1}{2}$.

La dimostrazione della precedente ***Proposizione* 16.2** (vedere pagina 251) si basa sul seguente risultato basilare e che svolge un ruolo "centrale" nel Calcolo delle Probabilità.

[87] Questo argomento viene svolto nel caso generale nei successivi corsi di Calcolo delle Probabilità, e richiede nozioni di Analisi, come ad esempio gli integrali di funzioni di più variabili.

[88] Più in generale, data la funzione di distribuzione di una variabile aleatoria X, è sempre possibile ottenere la distribuzione della variabile aleatoria $Y = \alpha + \beta\, X$, il caso successivo è un caso particolare di questo, con $X = S_n$, $\alpha = -\frac{a_n}{b_n}$ e $\beta = \frac{1}{b_n}$.
A questo proposito si veda la sezione sulle Trasformazioni affini di variabili aleatorie.

[89] Infatti

$$\omega \in \{S_n \leq x\} \iff S_n(\omega) \leq x \underset{b_n>0}{\iff} \frac{S_n(\omega) - a_n}{b_n} \leq \frac{x - a_n}{b_n} \iff \omega \in \left\{ \frac{S_n - a_n}{b_n} \leq \frac{x - a_n}{b_n} \right\}$$

***Proposizione* 16.3 (Teorema Centrale del Limite).** *Sia $\{X_i,\ i \geq 1\}$ una successione di v.a. indipendenti ed identicamente distribuite, per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$, si assuma che $\sigma^2 > 0$. Allora indicando con S_n^* variabile aleatoria standardizzata di S_n, si ha*

$$S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}, \tag{148}$$

e, indicando con $F_{S_n^*}(x)$ la funzione di distribuzione di S_n^*, si ha

$$\lim_{n\to\infty} F_{S_n^*}(x) = \lim_{n\to\infty} \mathbb{P}\left(S_n^* \leq x\right) = \Phi(x), \tag{149}$$

dove Φ è la funzione di distribuzione di una variabile aleatoria Gaussiana standard: in altre parole

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{y^2}{2}}\, dy. \tag{150}$$

Inoltre il limite è uniforme per $x \in \mathbb{R}$, ovvero

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{y^2}{2}}\, dy \right| = 0. \tag{151}$$

Non diamo la dimostrazione di questo risultato, ma notiamo solo che la (148) si dimostra tenendo conto che $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu$ e che, come già osservato nell'***Osservazione* *15.1***, per la completa indipendenza dalle variabili aleatorie X_i, si ha

$$Var(S_n) = Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) = n\sigma^2.$$

La precedente relazione sarebbe valida anche nel caso in cui le variabili aleatorie fossero solo indipendenti a due a due (o addirittura solo non correlate), ma <u>sottolineiamo</u> il fatto che, mentre la Legge Debole dei Grandi Numeri, vale sotto l'ipotesi di indipendenza a due a due, e non è necessario supporre $\sigma^2 > 0$, invece ***per il Teorema Centrale del Limite, serve la condizione di completa indipendenza*** e ovviamente ***è necessario supporre*** $\sigma^2 > 0$, altrimenti non si potrebbe nemmeno formulare la tesi.

Dimostrazione della **Proposizione** ***16.2*** *(Approssimazione normale).*
Fondamentale per dimostrare l'approssimazione (147) della funzione di distribuzione della somma S_n è il fatto che la convergenza in (150) sia uniforme[90]: infatti, posto

$$E_n(x) = F_{S_n^*}(x) - \Phi(x), \text{ e } x_n = \tfrac{x-n\mu}{\sqrt{n\,\sigma^2}},$$

[90]Si osservi che in generale le condizioni che

$$\lim_{n\to\infty} f_n(x) = f(x) \qquad \lim_{n\to\infty} x_n = x$$

non implicano che

$$\lim_{n\to\infty} f_n(x_n) = f(x).$$

Basta pensare al seguente **controesempio**:

$$\begin{cases} f_n(x) = 0 & x < \frac{1}{n}, \\ f_n(x) = 1 & x \geq \frac{1}{n} \end{cases} \qquad \begin{cases} f(x) = 0 & x \leq 0, \\ f(x) = 1 & x > 0 \end{cases}$$

Chiaramente se $x \leq 0$ allora $f_n(x) = 0$ e quindi $\lim_{n\to\infty} f_n(x) = f(x) = 0$, analogamente, se $x > 0$, allora per $n > \frac{1}{x}$ si ha $f_n(x) = 1$, e quindi $\lim_{n\to\infty} f_n(x) = f(x) = 1$. Inoltre, posto $x_n = \frac{1}{n}$, si ha $\lim_{n\to\infty} x_n = 0$, tuttavia ovviamente $f_n(x_n) = f_n(\frac{1}{n}) = 1$ che non converge ad $f(x) = f(0) = 0$.

si ha

$$F_{S_n}(x) = \mathbb{P}(S_n \leq x) = \mathbb{P}\left(\tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq \tfrac{x-n\mu}{\sqrt{n\,\sigma^2}}\right) = F_{S_n^*}(x_n) = \Phi(x_n) + E_n(x_n),$$

per cui

$$|F_{S_n}(x) - \Phi(x_n)| = |E_n(x_n)| \leq \sup_{x\in\mathbb{R}} |E_n(x)|.$$

Basta solo osservare che (151) garantisce che $\sup_{x\in\mathbb{R}} |E_n(x)|$ converge a zero[91] per n che tende all'infinito. □

16.2.2 Altre conseguenze del Teorema Centrale del Limite e relazioni con la legge dei grandi numeri

Si osservi che il Teorema Centrale del Limite implica che

$$\lim_{n\to\infty} \mathbb{P}(a < \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq b) = \Phi(b) - \Phi(a),$$

come si vede subito applicando la proprietà che per ogni variabile aleatoria X, con funzione di distribuzione $F(x)$, si ha $\mathbb{P}(a < X \leq b) = F(b) - F(a)$.

Il Teorema Centrale del Limite implica anche che

$$\lim_{n\to\infty} \mathbb{P}\left(a \leq \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq b\right) = \Phi(b) - \Phi(a),$$

infatti, come si vede facilmente[92],

$$\lim_{n\to\infty} \mathbb{P}\left(\tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} = a\right) = 0.$$

Dopo questa osservazione possiamo tornare indietro alle relazioni tra Legge dei Grandi Numeri e Teorema Centrale del Limite.

Indicando, come al solito, con Y_n la media aritmetica $\frac{S_n}{n}$, si ha

$$Y_n - \mu = \frac{S_n - n\mu}{n},$$

e quindi la standardizzata della media artimentica Y_n coincide con la standardizzata della somma S_n, cioè

$$\sqrt{\tfrac{n}{\sigma^2}}\,(Y_n - \mu) = \sqrt{\tfrac{n}{\sigma^2}}\,\tfrac{S_n - n\mu}{n} = \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}}$$

e inoltre

$$\{|Y_n - \mu| \leq \varepsilon\} = \left\{-\varepsilon \leq \tfrac{S_n - n\mu}{n} \leq \varepsilon\right\} = \left\{-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon \leq \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq \sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right\}.$$

[91] Pur essendo assolutamente al di fuori dell'ambito di un corso elementare di probabilità, vale la pena di ricordare che esistono delle maggiorazioni per $\sup_{x\in\mathbb{R}} |E_n(x)|$, nel caso in cui si supponga che il valore atteso $\mathbb{E}(|X|^3)$ esista e sia finito. In particolare è stato dimostrato che

$$\sup_{x\in\mathbb{R}} |E_n(x)| \leq \frac{C}{\sqrt{n}}\,\frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

con C costante. Il valore di C non è noto esattamente ma è noto che $0.4097 \leq C \leq 0.7975$, in particolare quindi vale

$$\sup_{x\in\mathbb{R}} |E_n(x)| \leq \frac{1}{\sqrt{n}}\,\frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

I primi a fornire maggiorazioni in questa direzione sono stati Berry ed Eessen all'inizio degli anni 40 dello scorso XX secolo.

[92] Si osservi che

$$\mathbb{P}\left(\tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} = a\right) \leq \mathbb{P}\left(a - \tfrac{1}{n} \leq \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq a\right) \simeq \Phi(a) - \Phi(a - \tfrac{1}{n}) \underset{n\to\infty}{\longrightarrow} 0$$

Di conseguenza

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1, \tag{152}$$

infatti

$$\begin{aligned}
\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) &= \mathbb{P}\left(-\varepsilon \leq \tfrac{S_n - n\mu}{n} \leq \varepsilon\right) = \mathbb{P}\left(-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon \leq \tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} \leq \sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right)\\
&= \mathbb{P}\left(\tfrac{S_n - n\mu}{\sqrt{n\,\sigma^2}} = -\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right)\\
&+ \Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - \Phi\left(-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) + E_n\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - E_n\left(-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right)\\
&\simeq \Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - \Phi\left(-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) = 2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1.
\end{aligned}$$

Si ottiene di nuovo la stessa tesi della Legge debole dei Grandi Numeri, (***Proposizione* 16.1**), ma sotto l'ipotesi più restrittiva che le variabili aleatorie siano completamente indipendenti. Infatti mandando n all'infinito nella precedente relazione (152) si ottiene

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\})\\
&= \lim_{n\to\infty}\left[2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1 + E_n\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - E_n\left(-\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right)\right] = 2 - 1 + 0 - 0 = 1.
\end{aligned}$$

Esempio* 16.3. *Sia X_1 una variabile aleatoria che può assumere i valori $0,\ \frac{1}{2},\ 1,\ \frac{3}{2}$ e con*

$$p_{X_1}(0) = \mathbb{P}(X_1 = 0) = \frac{1}{10}, \qquad p_{X_1}(\tfrac{1}{2}) = \mathbb{P}(X_1 = \tfrac{1}{2}) = \frac{1}{10},$$

$$p_{X_1}(1) = \mathbb{P}(X_1 = 1) = \frac{4}{10}, \qquad p_{X_1}(\tfrac{3}{2}) = \mathbb{P}(X_1 = \tfrac{3}{2}) = \frac{4}{10}.$$

Si ponga il valore atteso di X_1 uguale a μ e la sua varianza uguale a σ^2.
Siano $X_1, X_2, X_3, \ldots, X_{100}$ delle variabili aleatorie con la stessa distribuzione di X_1 e completamente (o globalmente) indipendenti tra loro e si ponga $Y_{100} := \frac{\sum_{j=1}^{100} X_j}{100}$. Utilizzando il Teorema Centrale del Limite, approssimare la probabilità

$$\mathbb{P}\left(\left\{\mu - \frac{1}{10} \leq Y_{100} \leq \mu + \frac{1}{10}\right\}\right).$$

Soluzione: Innanzi tutto come si trova facilmente si ha $\mu = \frac{21}{20}$ e $\sigma^2 = \frac{89}{400}$. Quindi la probabilità cercata è approssimata con

$$\begin{aligned}
\mathbb{P}\left(\left\{\mu - \frac{1}{10} \leq Y_{100} \leq \mu + \frac{1}{10}\right\}\right) &\simeq 2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1 = 2\Phi\left(\sqrt{\tfrac{100}{\frac{89}{400}}}\,\tfrac{1}{10}\right) - 1\\
&= 2\Phi\left(\sqrt{\tfrac{400}{89}}\right) - 1 = 2\Phi(\tfrac{20}{\sqrt{89}}) - 1\\
&\simeq 2\Phi(2,1199) - 1 \simeq 2 \cdot 0.9826 - 1 = 1,9652 - 1 = 0.9652
\end{aligned}$$

Finora ci siamo posti il problema del tipo: **fissati n (*grande*) ed $\varepsilon > 0$, quanto vale approssimativamente la probabilità che media aritmetica e valore atteso differiscano di meno di ϵ?**

Supponiamo ora di voler rispondere ***in modo approssimato*** alla domanda: ***siano* n (*grande*) *e* $\delta \in (0,1)$ *fissati*, per quale valore di $\varepsilon = \varepsilon(n, \delta)$ posso affermare che**

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 1 - \delta?$$

Il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ. Tuttavia permette di dare una buona valutazione del tipo di comportamento di ε: (trascurando E_n) andiamo a mostrare che $\varepsilon = \varepsilon(n, \delta)$ è un infinitesimo dell'ordine di $\frac{1}{\sqrt{n}}$.
Prima di tutto invece di valutare esattamente

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 1 - \delta$$

consideriamo la (152), ossia che, per n sufficientemente grande,

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1 \qquad \Leftrightarrow \qquad \mathbb{P}(\{|Y_n - \mu| \leq \sqrt{\tfrac{\sigma^2}{n}}\,x\}) \simeq 2\Phi\left(x\right) -$$

con

$$x = \sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon,$$

e quindi cerchiamo invece ε, o equivalentemente x, in modo che

$$2\Phi(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon) - 1 = 2\Phi(x) - 1 = 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) = \Phi(x) = \frac{2-\delta}{2} = 1 - \frac{\delta}{2}.$$

Sicuramente esiste[93] un valore $x = x_{1-\delta/2}$ per cui

$$\Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2};$$

Inoltre possiamo trovare un valore approssimato di $x_{1-\delta/2}$ utilizzando le tavole della gaussiana. Ad esempio per $\delta = 0.1$ si ottiene $1 - \delta/2 = 1 - 0.05 = 0.95$ ed $x_{1-\delta/2} = x_{0.95} = 1.645$.

A questo punto basta porre

$$\sqrt{\frac{n}{\sigma^2}}\,\varepsilon = x_{1-\delta/2} \quad \Leftrightarrow \quad \varepsilon = \varepsilon(n,\delta) = x_{1-\delta/2} \cdot \sqrt{\frac{\sigma^2}{n}} = \frac{x_{1-\delta/2} \cdot \sigma}{\sqrt{n}},$$

per ottenere il risultato desiderato.

***Osservazione* 16.5.** Possiamo riassumere quanto appena provato con l'affermazione che per n (grande) e $x_{1-\delta/2}$ tale che $\Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2}$, si ha

$$\mathbb{P}(\{|Y_n - \mu| \leq x_{1-\delta/2}\tfrac{\sigma}{\sqrt{n}}\}) \simeq 1 - \delta.$$

Terminiamo questa sezione tornando invece al problema[94] **in cui sia ε che δ sono fissati**, e supponiamo di voler rispondere ***in modo approssimato*** alla domanda: ***per quali n posso affermare che***

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta?$$

Anche il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ. Tuttavia permette di dare una buona valutazione del tipo di richiesta vada fatta su n per ottenere la limitazione inferiore richiesta.

[93] La funzione di distribuzione Φ è una funzione strettamente crescente e continua, e assume quindi tutti i valori $(0, 1)$.

[94] Di questo tipo di problema ci siamo occupati nella sezione degli approfondimenti sull'uso della disuguaglianza di Chebyshev, e la risposta è stata: basta prendere $n \geq n(\varepsilon, \delta)$, dove $n(\varepsilon, \delta)$ è definito in (142).

Anche in questo problema, invece di cercare una limitazione inferiore esatta

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta$$

sempre considerando che, per n sufficientemente grande,

$$\mathbb{P}(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) - 1$$

cerchiamo invece una limitazione inferiore

$$2\Phi(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon) - 1 \geq 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\tfrac{n}{\sigma^2}}\,\varepsilon\right) \geq \frac{2-\delta}{2} = 1 - \frac{\delta}{2}$$

Come nel caso precedente possiamo trovare un valore $x_{1-\delta/2}$ per cui

$$\Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2}.$$

Si osservi che, essendo Φ una funzione non decrescente[95],

$$\Phi\left(x\right) \geq \Phi\left(x_{1-\delta/2}\right) = 1 - \frac{\delta}{2}, \qquad \text{per ogni } x \geq x_{1-\delta/2},$$

A questo punto basta imporre che

$$\sqrt{\frac{n}{\sigma^2}}\,\varepsilon \geq x_{1-\delta/2} \quad \Leftrightarrow \quad \sqrt{n} \geq \frac{x_{1-\delta/2}\sqrt{\sigma^2}}{\varepsilon} \quad \Leftrightarrow \quad n \geq n_{TCL}(\varepsilon, \delta) := \frac{x_{1-\delta/2}^2\,\sigma^2}{\varepsilon^2} \tag{153}$$

per ottenere il risultato desiderato.

***Osservazione* 16.6.** Si confrontino tra loro (142) e (153): come si vede (142) e (153) sono molto simili, la seconda si ottiene sostituendo al posto di $\frac{1}{\delta}$, il valore $x_{1-\delta/2}^2$.

Quindi a parità di valori di ε e σ^2 si ottiene che la limitazione inferiore con la disuguaglianza di Chebyshev, pur essendo esatta, chiede

$$n \geq n_{Ch}(\varepsilon, \delta) = = \frac{\sigma^2}{\delta\,\varepsilon^2} = \frac{1}{\delta\,x_{1-\delta/2}^2}\, n_{TCL}(\varepsilon, \delta)$$

Per capire quindi la differenza si osservi che se $\delta = 0{,}01$, allora $\frac{1}{\delta} = 100$, mentre, essendo $x_{1-\delta/2} = x_{1-0{,}005} = x_{0{,}995} = 2{,}576$ (come si può trovare dalle tavole) si ha che $x_{1-\delta/2}^2 = 6{,}635776$.

In questo caso

$$n_{Ch}(\varepsilon, \delta) = \frac{1}{\delta\,x_{1-\delta/2}^2} n_{TCL}(\varepsilon, \delta) = \frac{16}{0{,}01 \cdot 6{,}63577}\, n_{TCL}(\varepsilon, \delta) \simeq 15{,}0698\, n_{TCL}(\varepsilon, \delta).$$

Se invece $\delta = 0.001$, allora $\frac{1}{\delta} = 1000$, mentre, essendo $x_{1-\delta/2} = x_{1-0{,}0005} = x_{0{,}9995} = 3{,}291$ (come si può trovare dalle tavole) si ha che $x_{1-\delta/2}^2 = 10{,}830681$.

In questo caso

$$n_{Ch}(\varepsilon, \delta) = \frac{1}{\delta\,x_{1-\delta/2}^2}\, n_{TCL}(\varepsilon, \delta) = \frac{1}{0{,}001 \cdot 10{,}830681}\, n_{TCL}(\varepsilon, \delta) \simeq 92{,}3302\, n_{TCL}(\varepsilon, \delta).$$

e quindi il valore di $n_{Ch}(\varepsilon, \delta)$ è circa 92 volte più grande di $n_{TCL}(\varepsilon, \delta)$, che è calcolato con il Teorema Centrale del Limite.

---

[95] In realtà basta trovare sulla tavola della gaussiana standard un valore $\overline{x}_{1-\delta/2}$ tale che

$$\Phi\left(\overline{x}_{1-\delta/2}\right) \geq 1 - \frac{\delta}{2}.$$

Il ragionamento fatto con $x_{1-\delta/2}$ si può ripetere mettendo $\overline{x}_{1-\delta/2}$ al posto di $x_{1-\delta/2}$.

16.3 Esercizi di verifica

Esercizio 16.1. *(Stessa situazione dell'Esercizio 6.1)*
Un candidato ad un'elezione ha bisogno di almeno 50 *voti per essere eletto. Prepara allora una lettera per informare i potenziali elettori circa la sua candidatura, il suo programma elettorale, etc....*
Egli valuta che ogni persona che riceve la lettera si recherà effettivamente a votare per lui con una probabilità del 40%, *indipendentemente dal comportamento degli altri (e si sottointende che egli certamente non ottiene voti da coloro cui non ha inviato la lettera).*
(a) Calcolare approssimativamente la probabilità che egli riceva esattamente 51 *voti se invia la lettera a* 200 *persone.*
(b) Calcolare approssimativamente la probabilità di essere eletto se invia la lettera a 100 *persone.*
*(c) Calcolare approssimativamente il numero minimo di persone alle quali deve inviare copia della lettera affinché la probabilità di essere eletto sia superiore all'*80%

Esercizio 16.2. *Siano* X_i *variabili aleatorie indipendenti e uniformi in* $\{1, 2, 3\}$.
Posto $S_{54} = \sum_{i=1}^{54} X_i$,
(a) Calcolare $\mathbb{E}[S_{54}]$ *e* $Var(S_{54})$.
(b) Calcolare approssimativamente $\mathbb{P}(S_{54} \leq 116)$.
(c) Calcolare approssimativamente $\mathbb{P}(S_{54} > 100)$.
(d) Calcolare approssimativamente $\mathbb{P}(100 < S_{54} \leq 116)$.

Esercizio 16.3. *Siano* X_i *variabili aleatorie indipendenti e tutte con distribuzione di Poisson di parametro* λ, *e con valore atteso* $\mathbb{E}(X_i) = 3$.
Dopo aver trovato il valore di λ *e di* $Var(X_i)$, *posto* $S_n = \sum_{i=1}^{n} X_i$, *dire che tipo di distribuzione ha* S_n *e scrivere la formula per calcolare* $\mathbb{P}(S_n = k)$, *specificando, come usuale e necessario, per quali valori di* k, $\mathbb{P}(S_n = k) > 0$.
Prendendo $n = 108$,
(a) Calcolare $\mathbb{E}[S_n]$ *e* $Var(S_n)$.
(b) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_n \leq 350)$.
(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S_n > 288)$.
(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(288 < S_n \leq 350)$.

Esercizio 16.4. *Siano* X_i *variabili aleatorie indipendenti e tutte* $Geom(p)$, *e con valore atteso* $\mathbb{E}(X_i) = 3$.
Dopo aver trovato il valore di p *e di* $Var(X_i)$, *posto* $S_n = \sum_{i=1}^{n} X_i$, *dire che tipo di distribuzione ha* S_n *e scrivere la formula per calcolare* $\mathbb{P}(S_n = k)$, *specificando, come usuale e necessario, per quali valori di* k, $\mathbb{P}(S_n = k) > 0$.
Prendendo $n = 96$
(a) Calcolare $\mathbb{E}[S_n]$ *e* $Var(S_n)$.
(b) Calcolare approssimativamente $\mathbb{P}(S_n \leq 324)$.
(c) Calcolare approssimativamente $\mathbb{P}(S_n > 252)$.
(d) Calcolare approssimativamente $\mathbb{P}(252 < S_n \leq 324)$.

Esercizio 16.5. *Sia* S *una variabile aleatoria di Poisson di parametro* $\lambda = 324$.
(a) Calcolare $\mathbb{E}[S]$ *e* $Var(S)$.
(b) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S \leq 350)$.
(c) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(S > 288)$.
(d) Scrivere l'espressione esatta e calcolare approssimativamente $\mathbb{P}(288 < S \leq 350)$.

ALFABETO GRECO

α	A	alfa
β	B	beta
γ	Γ	gamma
δ	Δ	delta
ϵ o anche ε	E	epsilon
ζ	Z	zeta
η	H	eta
θ o anche ϑ	Θ	theta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu o anche mi
ν	N	nu o anche ni
ξ	Ξ	xi (csi)
o	O	omicron
π o anche ϖ	Π	pi greco
ρ o anche ϱ	R	rho
σ o, in fine parola, ς	Σ	sigma
τ	T	tau
υ	Υo anche Y	üpsilon
ϕ o anche φ	Φ	phi (fi)
χ		chi
ψ	Ψ	psi
ω	Ω	omega

ALFABETO CORSIVO INGLESE/ITALIANO

a	$\mathcal{A}$	b	$\mathcal{B}$	c	$\mathcal{C}$
d	$\mathcal{D}$	e	$\mathscr{E}$ o $\mathcal{E}$	f	$\mathcal{F}$
g	$\mathcal{G}$	h	$\mathcal{H}$	i	$\mathcal{I}$
j	$\mathcal{J}$	k	$\mathcal{K}$	ℓ	$\mathcal{L}$
m	$\mathcal{M}$	n	$\mathcal{N}$	o	$\mathcal{O}$
p	$\mathcal{P}$ o $\mathscr{P}$	q	$\mathcal{Q}$	r	$\mathcal{R}$
s	$\mathcal{S}$	t	$\mathcal{T}$	u	$\mathcal{U}$
v	$\mathcal{V}$	w	$\mathcal{W}$	x	$\mathcal{X}$
y	$\mathcal{Y}$	z	$\mathcal{Z}$		