# Appunti di Probabilità e Statistica

Luca La Rocca

9 ottobre 2015

# Indice

**Prefazione** **ii**

**1 Probabilità elementare** **1**
1.1 Definizione soggettiva . . . . . 5
1.2 Valutazioni classiche . . . . . 12
1.3 Condizionamento . . . . . 15
1.4 Apprendimento . . . . . 21
1.5 Indipendenza . . . . . 24

**2 Probabilità avanzata** **31**
2.1 Additività numerabile . . . . . 33
2.2 Probabilità sulla retta reale . . . . . 37
2.3 Probabilità su spazi prodotto . . . . . 43
2.4 Lemmi di Borel-Cantelli . . . . . 43

# Prefazione

Questo scritto nasce dalla mia esigenza di dare alle lezioni di *Statistica ed Elementi di Probabilità* per gli studenti al secondo anno del *Corso di Laurea in Informatica* offerto dal *Dipartimento di Scienze Fisiche, Informatiche e Matematiche* dell'*Università degli Studi di Modena e Reggio Emilia* un taglio quanto più possibile corrispondente al mio punto di vista. L'ambizione era quella di affiancare prima e sostituire poi il testo di riferimento consigliato (Paolo Baldi, 1998, Calcolo delle Probabilità e Statistica, Seconda Edizione, McGraw-Hill, Milano) ereditato da Emanuele Dolera. I riscontri forniti dai primi due cicli di lezioni mi hanno spinto a consigliare un diverso testo di riferimento (Marco Boella, 2010, Probabilità e Statistica per Ingegneria e Scienze, Pearson, Milano) cui ancorare maggiormente le lezioni. Rimane il progetto di sviluppare una mia personale introduzione alla probabilità e alla statistica (quest'ultima essendo per ora presente solo sottotraccia).

Per quanto riguarda il metodo di lavoro, mi sono posto l'obiettivo di presentare un certo numero di argomenti fondamentali sulla base di alcuni esempi ragionati intorno ai quali possano ruotare le lezioni; tali esempi sono raccolti all'inizio di ogni capitolo, in modo che sia facile estrarli per averli disponibili in aula. Lo scritto si articola in capitoli (per ora solo due) ognuno dei quali suddiviso in sezioni; la numerazione di definizioni, proposizioni e teoremi (come pure quella autonoma delle equazioni) è specifica di ogni sezione, ma comune alle tre categorie. La classificazione dei risultati in proposizioni e teoremi si basa su una valutazione (necessariamente soggettiva) della loro importanza. Mi sono per il momento sottratto all'onere di inserire adeguati riferimenti bibliografici (per contenere i tempi di scrittura).

Per quanto riguarda la terminologia, resta inteso quanto segue. Se $x \leq y$, il numero $x$ è minore del numero $y$, se $x < y$, il numero $x$ è strettamente minore del numero $y$; se $x \leq 0$, il numero $x$ è negativo, se $x < 0$, il numero $x$ è strettamente negativo. Analogamente per numeri (strettamente) maggiori di altri e (strettamente) positivi, nonché per funzioni a valori numerici, quando le disuguaglianze (strette) siano soddisfatte per ogni valore dei loro argomenti. Senza entrare nel dettaglio dei singoli casi, una funzione di argomento numerico (in particolare una successione) a valori numerici sarà crescente o decrescente (in generale monotona) quando valgano le implicazioni del caso (eventualmente in senso stretto). Se $E \subseteq F$, l'insieme $E$ è incluso nell'insieme $F$; se $E \subset F$, l'insieme $E$ è strettamente incluso nell'insieme $F$. Una successione di insiemi sarà crescente o decrescente (in generale monotona ed eventualmente in senso stretto) con riferimento alla relazione di inclusione. Un insieme finito avrà cardinalità finita o nulla, un insieme numerabile avrà cardinalità numerabile, finita o nulla; analogamente saranno qualificate le intersezioni e unioni di insiemi, quando riguardino un'infinità numerabile, un numero finito o una famiglia vuota di insiemi. Una funzione numerica di insieme sarà (sub)additiva quando lo sia con riferimento a unioni finite, numerabilmente (sub)additiva quando lo sia con riferimento a unioni numerabili. La terminologia sarà naturalmente sviluppata nel corso dell'esposizione, chiarendone i dettagli e introducendo termini specifici.

Ringrazio volentieri Emanuele Dolera, dal cui programma sono partito, Mauro Leoncini, per avere discusso con me i possibili contenuti delle lezioni, Gianpaolo Scalia Tomba, per avermi raccontato cosa sia una massa aderente e molte altre cose interessanti, Giorgio Letta, per avermi introdotto agli aspetti teorici della probabilità e averlo fatto con grande eleganza. Se serve dirlo, si tratta di una lista *in fieri* quanto il resto di questo scritto e ogni responsabilità per le scelte fatte o per gli errori presenti è solo mia.

Luca La Rocca

# Capitolo 1

# Probabilità elementare

**Esempio 1 (Un destino crudele)** *Un destino crudele ci obbliga ad accettare scommesse sul lancio di un dado a sei facce (numerate da 1 a 6). Diversi giocatori possono puntare sull'eventualità che il punteggio ottenuto sia pari, divisibile per tre o entrambe le cose. Ogni giocatore decide da sé la propria condizione di vittoria e la posta che incasserà nel caso in cui essa sia soddisfatta, con la possibilità di scegliere una posta negativa (da pagare in valore assoluto). Ci è richiesto di fissare, per ogni possibile condizione di vittoria, la quota parte della posta in palio che dovrà essere pagata (incassata in valore assoluto nel caso di posta negativa) per scommettere su tale condizione. Sulla base delle quote che avremo fissato, saremo tenuti ad accettare tutte le scommesse che i giocatori ci proporranno. Come ci orientiamo?*

**Esempio 2 (Un tempo da lupi)** *Oggi lasceremo l'auto dal meccanico per farla revisionare e poi saremo appiedati. Nel decidere se portare o meno con noi l'ombrello e il cappotto ci ritroviamo a scommettere sul tempo che farà. Un nostro amico ha previsto pioggia e vento. Che modello avrà usato?*

**Esempio 3 (Una carta alla volta)** *Da un mazzo di carte italiane abbiamo estratto quattro carte: l'asso, il due e il tre di spade, assieme al quattro di bastoni. Le facciamo mischiare per bene e poi le voltiamo una alla volta, contando ad alta voce da uno a quattro. Diremo di avere una coincidenza quando il numero chiamato a voce coincida con il numero raffigurato sulla carta. Con quale probabilità avremo quattro coincidenze? zero coincidenze? Con quale probabilità avremo una coincidenza nel voltare le prime due carte? due coincidenze nel voltare le prime due carte? Con quale probabilità sulle prime due carte voltate ci saranno tre spade? cinque spade?*

**Esempio 4 (Un uomo prudente)** *Un uomo prudente, al compimento del trentesimo anno di età, decide di assicurarsi contro il rischio di ritrovarsi a ottant'anni senza alcun capitale con cui affrontare spese straordinarie. Dopo qualche ricerca, il nostro uomo trova una compagnia assicuratrice che* *gli garantisce diecimila euro (attuali) al compimento dell'ottantesimo anno di età, a fronte di un versamento annuale da ripetere sino al pensionamento (compimento del sessantacinquesimo anno di età). Qualora l'uomo dovesse morire prima di andare in pensione, l'intero importo da lui versato potrà essere riscattato dai suoi eredi ed egli, da lassù, lo considererà recuperato. Se invece dovesse morire dopo essere andato in pensione, ma prima di avere compiuto ottant'anni, l'importo versato resterà alla compagnia assicuratrice. Qual è il valore equo (attuale) di ognuno dei trentacinque versamenti? Nel momento in cui andasse in pensione, quanto valuterebbe il nostro uomo la probabilità di arrivare a compiere ottant'anni?*

**Esempio 5 (Una scatola di costruzioni)** *Una scatola di costruzioni è stata riempita con sessanta mattoncini quadrati, di cui trenta gialli e trenta rossi, e novanta mattoncini rettangolari, di cui sessanta gialli e trenta rossi. Estraiamo senza guardare un mattoncino dalla scatola: quanto valutiamo la probabilità che sia giallo? Dopo avere verificato, al tatto, che il mattoncino che stiamo estraendo è quadrato, come cambia la nostra valutazione? Se in palio ci sono 100 euro per chi estragga un mattoncino giallo, quanto varrà il diritto di effettuare un'estrazione? Come cambiano le nostre valutazioni nel caso in cui quaranta dei mattoncini quadrati siano gialli e venti rossi?*

**Esempio 6 (Una mano innocente)** *Un bimbo bendato estrae dieci biglie, in sequenza, senza reimmissione, da un'urna opaca che ne contiene ottanta bianche e quaranta rosse, tutte alla stessa temperatura. Con quale probabilità la prima biglia estratta sarà rossa? la seconda biglia estratta sarà rossa? le prime due biglie estratte saranno entrambe rosse? avranno lo stesso colore? sarà rossa almeno una due? Cambia qualcosa se consideriamo le ultime due biglie estratte invece delle prime due? Quanto valutiamo la probabilità che sia rossa almeno una delle prime tre biglie estratte? almeno una delle prime quattro biglie estratte? almeno una tra tutte e dieci le biglie estratte?*

**Esempio 7 (Una diagnosi ponderata)** *Nell'ambito di una campagna di screening volta a identificare precocemente una malattia rara, della quale sappiamo che colpisce due persone su cento, un nostro amico è purtroppo risultato positivo al test. Sappiamo che il test funziona nel* $99.9\%$ *dei casi, quando il soggetto è malato, e nel* $97.5\%$ *dei casi, quando il soggetto è sano. Quanto valutiamo la probabilità che il nostro amico sia malato?*

**Esempio 8 (Un lancio alla cieca)** *Un astuccio contiene tre dadi a sei facce (numerate da 1 a 6): il primo è equilibrato; il secondo è sbilanciato in modo da favorire l'uscita del punteggio 6 (con probabilità pari a 0.64) e penalizzare l'uscita del punteggio 1 (con probabilità pari a 0.04) lasciando gli altri quattro punteggi sullo stesso piano (equiprobabili); il terzo è sbilanciato come il primo, ma con le facce 1 e 6 invertite. I tre dadi sono indistinguibili alla vista. Una nostra amica ne estrae uno a caso dall'astuccio e lo lancia,* *senza darci modo di osservarne il rotolamento. Se ottiene un 6, quanto varrà la probabilità che il dado sia equilibrato? Dopo un secondo 6, in un secondo lancio nelle stesse condizioni, come cambierà la nostra valutazione?*

**Esempio 9 (Una carta sola)** *Un mazzo di carte francesi è formato da cinquantadue carte: tredici valori (asso, due, tre, quattro, cinque, sei, sette, otto, nove, dieci, fante, donna, re) per ognuno di quattro semi (cuori, quadri, fiori, picche). Dopo averlo ben mischiato, ne voltiamo la prima carta. Con quale probabilità sarà di cuori? sarà una donna? entrambe le cose? almeno una delle due cose? Cosa cambia se dal mazzo manca la donna di cuori?*

**Esempio 10 (Una mamma per due)** *Una donna ha due figli. Con quale probabilità sono tutti e due femmina? Sapendo che almeno uno dei due è femmina, con quale probabilità lo è anche l'altro? Sapendo invece che il più grande è femmina, con quale probabilità lo sarà anche il più piccolo?*

**Esempio 11 (Una partita di calcio)** *Stiamo per assistere a una partita di calcio e ci interessano i seguenti eventi: i) vince la squadra in trasferta; ii) il primo tempo finisce uno a uno; iii) al termine della partita sono stati segnati esattamente uno/due/tre gol. Quanti costituenti saranno possibili?*

**Esempio 12 (Un cacciatore di teste)** *Un selezionatore di personale si affida al lancio di un'ordinaria moneta da due euro per decidere se assumere o meno i candidati intervistati. Oggi effettuerà sei interviste. Se lancia la moneta onestamente, con quale probabilità finirà per assumere esattamente due candidati? Quanti candidati assumerà nell'eventualità più probabile?*

**Esempio 13 (Un'urna peculiare)** *Estraiamo casualmente una biglia da un'urna che ne contiene otto: due gialle, due rosse, due verdi e due nere. Posto* $G =$ *"la biglia estratta è gialla",* $R =$ *"la biglia estratta è rossa" e* $V =$ *"la biglia estratta è verde", che relazione c'è tra i tre eventi* $G \cup R$, $R \cup V$ *e* $G \cup V$*? Cosa cambia se le biglie nere nell'urna sono quattro?*

*Questa pagina è stata lasciata intenzionalmente in bianco.*

## 1.1 Definizione soggettiva

Nell'Esempio 1 un destino crudele ci obbliga ad accettare scommesse sul lancio di un dado a sei facce. Si tratta di uno scenario concreto, semplice ma non banale, nel quale ci caleremo per iniziare il nostro viaggio nel reame dell'incerto. In estrema sintesi, quel che faremo è adottare l'interpretazione soggettiva della probabilità per motivarne gli assiomi.

Il lancio di un dado è un *esperimento aleatorio*, in quanto il rotolamento del dado è un fenomeno la cui evoluzione non siamo in grado di prevedere con certezza (fenomeno aleatorio); si tratta di un esperimento il cui risultato è incerto. Si noti la differenza rispetto al caso della caduta di un grave in condizioni controllate, quando sia nota la legge di gravitazione e si conoscano la posizione e la velocità iniziale del grave. In questo caso l'evoluzione del fenomeno è prevedibile con certezza (fenomeno deterministico) e la misura del tempo di caduta, per esempio, può avere un unico risultato (esperimento deterministico). Non deve però sfuggire l'importanza dell'osservatore nel determinare il carattere aleatorio o deterministico di un fenomeno: un osservatore che debba combinare la legge di gravitazione con il funzionamento di uno strumento di misura impreciso o conosca solo in misura approssimata le condizioni inziali del grave ne vedrà la caduta come un fenomeno aleatorio, mentre un alieno che sappia controllare con perfetta maestria il lancio di un dado ne vedrà il rotolamento come un fenomeno deterministico. In effetti, un esperimento deterministico non è altro che un esperimento aleatorio in cui l'osservatore ha ridotto praticamente a zero l'incertezza.

Il primo passo che conviene compiere per modellare un esperimento aleatorio è individuare un insieme $\Omega$ che ne rappresenti i possibili *esiti*. Con riferimento all'Esempio 1, viene naturale prendere $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dopo di che scommettere sull'eventualità che l'esito $\omega \in \Omega$ abbia certe caratteristiche, per esempio sia pari o divisibile per tre, corrisponderà a scommettere sull'eventualità che esso appartenga a una certa parte (sottoinsieme) di $\Omega$, per esempio $B = \{2, 4, 6\}$ o $D = \{3, 6\}$. Identificheremo quindi gli *eventi* oggetto di scommessa, enti logici il cui valore di verità dipende dalle caratteristiche dell'esito, con le parti di $\Omega$ che li realizzano: diremo che $E \subseteq \Omega$ si verifica, se $\omega \in E$; altrimenti diremo che $E$ non si verifica. Si noti la differenza concettuale tra l'esito $\omega$ e l'*evento elementare* $\{\omega\}$ realizzato dal solo $\omega$. Due eventi speciali, definiti in ogni esperimento aleatorio, sono l'*evento certo* $\Omega$ e l'*evento impossibile* $\emptyset$ (corrispondente all'insieme vuoto). Diremo *evento incerto* un evento $E$ che non sia né certo né impossibile ($\emptyset \subset E \subset \Omega$).

Il destino crudele dell'Esempio 1 ci obbliga ad accettare scommesse sugli eventi $B = \{2, 4, 6\}$ e $D = \{3, 6\}$, nonché sulla loro intersezione o *prodotto logico* $B \cap D = \{\omega \in \Omega : \omega \in B \text{ e } \omega \in D\} = \{6\}$, corrispondente al verificarsi simultaneo di $B$ e $D$. Più in generale, come secondo passo nella modellazione di un esperimento aleatorio, ci conviene individuare una classe $\mathcal{A}$ di eventi "interessanti". Quindi, per ogni $E \in \mathcal{A}$, fisseremo la quota parte $\mathbb{P}(E)$

della posta in palio $S$ da pagare per scommettere su $E$; assegneremo cioè una *funzione di evento* $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ (unica per tutti i valori della posta $S$) che esprima le nostre previsioni sull'esito dell'esperimento aleatorio. Compiuto questo terzo passo, l'esperimento aleatorio sarà modellato dalla terna $(\Omega, \mathcal{A}, \mathbb{P})$ che chiameremo *spazio di probabilità* non appena soddisfi alcune proprietà che ci accingiamo a far emergere come ragionevoli.

Chi scommetta su $E$, scelta la posta $S$, pagherà $\mathbb{P}(E)S$ per avere diritto a $S$ nel caso in cui $E$ si verifichi: se $S > 0$, pagherà con certezza a fronte di un incasso incerto; se $S < 0$, incasserà con certezza a fronte di un pagamento incerto; il caso $S = 0$ corrisponde alla decisione di non scommettere. In ogni caso il *guadagno* dello scommettitore sarà $G(\omega) = S - \mathbb{P}(E)S = \{1 - \mathbb{P}(E)\}S$, se $\omega \in E$, $G(\omega) = -\mathbb{P}(E)S$, se $\omega \notin E$, in funzione dell'esito $\omega \in \Omega$. Se le quantità $\{1 - \mathbb{P}(E)\}S$ e $-\mathbb{P}(E)S$ sono entrambe strettamente positive, la scommessa $(E, S)$ offrirà un profitto certo; se entrambe sono strettamente negative, un profitto certo sarà offerto dalla scommessa opposta $(E, -S)$. Il prodotto $-\mathbb{P}(E)\{1-\mathbb{P}(E)\}S^2$ delle due quantità dovrà quindi essere negativo per qualunque scelta di $S$. Prendendo $S = 1$ troviamo $0 \leq \mathbb{P}(E) \leq 1$ e resta confermato che $\mathbb{P}(E)$ debba essere una quota parte. In particolare, visto che $\Omega$ si verifica certamente, il guadagno certo $\{1 - \mathbb{P}(\Omega)\}S$ dovrà annullarsi per ogni $S$ e quindi porremo $\mathbb{P}(\Omega) = 1$. Analogamente porremo $\mathbb{P}(\emptyset) = 0$ per annullare il guadagno certo $-\mathbb{P}(\emptyset)S$. Eventuali scommesse su $\Omega$ e $\emptyset$ in pratica non avranno effetti e dunque possiamo senz'altro supporre che questi due eventi appartengano alla classe $\mathcal{A}$. Dopo di che sembra ragionevole richiedere che neanche scommesse multiple possano offrire profitti certi.

**Definizione 1.1.1 (coerenza)** *Sia $\mathcal{A}$ una classe di eventi definita su un insieme di esiti $\Omega$. Una funzione di evento $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ è incoerente, su $\mathcal{A}$, se esistono $n$ eventi $E_1, \ldots, E_n$ in $\mathcal{A}$ ed $n$ poste $S_1, \ldots, S_n$ in $\mathbb{R}$ tali che il guadagno totale $G_\star = G_1 + \cdots + G_n$ delle scommesse $(E_1, S_1), \ldots, (E_n, S_n)$ sia strettamente positivo per ogni $\omega \in \Omega$; in caso contrario $\mathbb{P}$ è coerente.*

Equivalentemente la coerenza esclude la possibilità di perdite certe, le quali si trasformerebbero in profitti certi passando alle scommesse opposte.

Consideriamo ora il complementare o *negazione* di un evento $E$, definito come $\bar{E} = \{\omega \in \Omega : \omega \notin E\}$. Per esempio $\bar{B} = \{1, 3, 5\}$ e $\bar{D} = \{1, 2, 4, 5\}$. Si noti che $\bar{\bar{E}} = E$, qualunque sia $E$, mentre $\bar{\Omega} = \emptyset$. Visto che $\bar{E}$ si verifica esattamente quando $E$ non si verifica, il guadagno della scommessa $(E, S)$ sarà $G(\omega) = \{1 - \mathbb{P}(E)\}S$, se $\omega \notin \bar{E}$, $G(\omega) = -\mathbb{P}(E)S$, se $\omega \in \bar{E}$. Sarà pertanto come se la scommessa fosse proposta su $\bar{E}$ con posta $-S$ e quota $1 - \mathbb{P}(E)$. Questa considerazione ha due conseguenze: i) se $E \in \mathcal{A}$, tanto vale che $\bar{E} \in \mathcal{A}$; ii) per coerenza dobbiamo assegnare $\mathbb{P}$ in modo che $\mathbb{P}(\bar{E}) = 1 - \mathbb{P}(E)$. Infatti, se per fissare le idee $0.2 = \mathbb{P}(\bar{E}) < 1 - \mathbb{P}(E) = 0.4$, due scommettitori associati che propongano la posta $S = 100$ euro sia su $E$ che su $\bar{E}$ pagheranno in totale $60 + 20 = 80$ euro per la certezza di riceverne 100, realizzando così un profitto certo di 20 euro. Se invece $\mathbb{P}(\bar{E}) > 1 - \mathbb{P}(E)$, essi otterranno un profitto certo con $S < 0$. In entrambi i casi potranno garantirsi un profitto arbitrariamente elevato aumentando $|S|$.

Se $E \subseteq F$ (equivalentemente $E \cap F = E$) diremo che $E$ *implica* $F$. Per esempio, posto $C_{11} = B \cap D$, avremo $C_{11} \subseteq B$ e $C_{11} \subseteq D$ (essendo per altro l'implicazione stretta in entrambi i casi). La *differenza* tra due eventi $F$ ed $E$ con $E \subseteq F$ è definita come $F \setminus E = \{\omega \in F : \omega \notin E\}$ e si verifica quando si verifichi $F$ ma non $E$. Per esempio abbiamo $B \setminus C_{11} = \{2,4,6\} \setminus \{6\} = \{2,4\}$ e $D \setminus C_{11} = \{3,6\} \setminus \{6\} = \{3\}$. Si noti che $\Omega \setminus E = \bar{E}$ ed $F \setminus E = F \cap \bar{E}$ (comunque scelti gli eventi $E$ ed $F$ con $E \subseteq F$). Consideriamo ora le scommesse $(F, S)$ ed $(E, -S)$. Il loro guadagno totale sarà

$$G_\star(\omega) = \begin{cases} -\mathbb{P}(F)S + \mathbb{P}(E)S, & \text{se } \omega \in \bar{F}, \\ \{1 - \mathbb{P}(F)\}S + \mathbb{P}(E)S, & \text{se } \omega \in F \setminus E, \\ \{1 - \mathbb{P}(F)\}S - \{1 - \mathbb{P}(E)\}S, & \text{se } \omega \in E, \end{cases}$$

cioè $G_\star(\omega) = -\{\mathbb{P}(F) - \mathbb{P}(E)\}S$, se $\omega \notin F \setminus E$, $G_\star(\omega) = [1 - \{\mathbb{P}(F) - \mathbb{P}(E)\}]S$, se $\omega \in F \setminus E$. Le due scommesse saranno quindi equivalenti a un'unica scommessa su $F \setminus E$ con posta $S$ e quota $\mathbb{P}(F) - \mathbb{P}(E)$. Ragionando come nel caso della negazione, se ne deduce che possiamo supporre $F \setminus E \in \mathcal{A}$, quando $F$ ed $E$ appartengano ad $\mathcal{A}$, e dobbiamo assegnare $\mathbb{P}$ in modo che $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(E)$, se vogliamo essere coerenti. Per esempio dovremo accettare scommesse su $C_{10} = B \cap \bar{D} = B \setminus C_{11} = \{2,4\}$ e $C_{01} = \bar{B} \cap D = D \setminus C_{11} = \{3\}$, con quote $\mathbb{P}(C_{10}) = \mathbb{P}(B) - \mathbb{P}(C_{11})$ e $\mathbb{P}(C_{01}) = \mathbb{P}(D) - \mathbb{P}(C_{11})$. Osserviamo che da $\mathbb{P}(F \setminus E) \geq 0$ segue $\mathbb{P}(E) \leq \mathbb{P}(F)$, se $E \subseteq F$; la funzione di evento $\mathbb{P}$ deve essere *isotona* sul suo dominio $\mathcal{A}$.

L'unione o *somma logica* di due eventi $E$ ed $F$ è definita come l'evento $E \cup F = \{\omega \in \Omega : \omega \in E \text{ o } \omega \in F\}$ corrispondente al verificarsi di almeno uno tra $E$ ed $F$ (eventualmente di tutti e due). Per esempio $B \cup D = \{2,4,6\} \cup \{3,6\} = \{2,3,4,6\}$. Ci interessa in particolare il caso in cui $E$ ed $F$ non possano verificarsi assieme ($E \cap F = \emptyset$) nel qual caso diremo che $E$ ed $F$ sono *incompatibili* (disgiunti) e denoteremo con $E \uplus F$ la loro unione, qualificandola come *diretta*. Per esempio $E \cap \bar{E} = \emptyset$ ed $E \uplus \bar{E} = \Omega$, qualunque sia l'evento $E$. Se $E$ ed $F$ sono incompatibili, le scommesse $(E, S)$ ed $(F, S)$ equivalgono alla scommessa $(E \uplus F, S)$ con quota $\mathbb{P}(E) + \mathbb{P}(F)$. Potremo quindi considerare $E \uplus F$ appartenente ad $\mathcal{A}$ e dovremo assegnare $\mathbb{P}$ in modo che $\mathbb{P}(E \uplus F) = \mathbb{P}(E) + \mathbb{P}(F)$. Per esempio avremo $C_{01} \uplus C_{10} = \{2,3,4\} \in \mathcal{A}$ e, per coerenza, $\mathbb{P}(\{2,3,4\}) = \mathbb{P}(\{3\}) + \mathbb{P}(\{2,4\})$. Più in generale, comunque presi $n$ eventi $E_1, \ldots, E_n$ incompatibili (a coppie) avremo $E_1 \uplus \cdots \uplus E_n \in \mathcal{A}$ e dovremo scegliere $\mathbb{P}$ in modo che $\mathbb{P}(E_1 \uplus \cdots \uplus E_n) = \mathbb{P}(E_1) + \cdots + \mathbb{P}(E_n)$, dove $E_1 \uplus \cdots \uplus E_n$ è l'evento che si verifica quando si verifichi almeno uno degli eventi $E_1, \ldots, E_n$ (supposti incompatibili); diremo che $\mathcal{A}$ è *stabile* per unioni finite di eventi incompatibili e che $\mathbb{P}$ deve essere *additiva* su $\mathcal{A}$. Per esempio accetteremo scommesse su $C_{01} \uplus C_{10} \uplus C_{11} = \{2,3,4,6\}$ con $\mathbb{P}(\{2,3,4,6\}) = \mathbb{P}(\{3\}) + \mathbb{P}(\{2,4\}) + \mathbb{P}(\{6\})$.

Alla luce di quanto visto, la classe degli eventi su cui dobbiamo accettare scommesse nell'Esempio 1 include, di fatto, non solo $B$, $D$ e $C_{11} = B \cap D$, ma anche $\emptyset$, $\Omega$, $\bar{B}$, $\bar{D}$ e $C_{10} = B \cap \bar{D} = B \setminus C_{11}$, nonché $C_{01} = \bar{B} \cap D = D \setminus C_{11}$; essa inoltre include $C_{00} = \bar{B} \cap \bar{D} = \overline{B \cup D} = \Omega \setminus (C_{01} \uplus C_{10} \uplus C_{11})$. Gli eventi della classe $\mathcal{C} = \{C_{00}, C_{01}, C_{10}, C_{11}\}$, incompatibili per costruzione, si dicono *costituenti* degli eventi $B$ e $D$. In generale i costituenti di $n$ eventi $E_1, \ldots, E_n$ sono le loro intersezioni dopo eventuale negazione: $\mathcal{C} = \{C_\alpha \mid \alpha \in \{0,1\}^n\}$, dove $C_\alpha = E_1^{\alpha_1} \cap \cdots \cap E_n^{\alpha_n}$, $E_i^0 = \bar{E}_i$, $E_i^1 = E_i$, $i = 1, \ldots, n$, l'intersezione di $n$ eventi essendo l'evento che si verifica quando essi si verifichino tutti. Ogni volta che saremo interessati a un numero finito di eventi ci interesseremo senz'altro ai loro costituenti. Dopo di che, dal momento che i costituenti sono in numero finito, oltre che incompatibili, saranno in $\mathcal{A}$ anche le loro unioni. Per esempio, nel caso $n = 2$, avremo $C_{00} \uplus C_{11} \in \mathcal{A}$; si noti come in questo caso $C_{ij}$ abbrevi $C_{(i,j)}$ con notazione matriciale.

La classe $\mathcal{U}$ delle unioni di costituenti, nella quale metteremo anche $\emptyset$, in quanto unione banale di zero costituenti, ha la proprietà di essere stabile per tutte le operazioni su eventi che abbiamo discusso (operazioni finite): se $E_1 = \biguplus_{\alpha \in A_1} C_\alpha$ ed $E_2 = \biguplus_{\alpha \in A_2} C_\alpha$ appartengono a $\mathcal{U}$, dove $A_1$ e $A_2$ sono parti di $\{0,1\}^n$, avremo $E_1 \cap E_2 = \biguplus_{\alpha \in A_1 \cap A_2} C_\alpha$, mentre $E_1 \cup E_2 = \biguplus_{\alpha \in A_1 \cup A_2} C_\alpha$ ed $E_2 \setminus E_1 = \biguplus_{\alpha \in A_2 \setminus A_1} C_\alpha$, se $E_1 \subseteq E_2$. Per esempio, nel caso $n = 2$, se $E_1 = C_{00} \cup C_{01} = \bar{B}$ ed $E_2 = C_{01} \cup C_{11} = D$, allora $E_1 \cap E_2 = C_{01}$. Si noti che $\Omega \in \mathcal{U}$, perché i costituenti sono *esaustivi*: uno di essi deve necessariamente verificarsi e dunque l'unione di tutti loro è $\Omega$. Quindi $\bar{E} = \Omega \setminus E$ appartiene a $\mathcal{U}$, se $E \in \mathcal{U}$. Osserviamo che la classe $\mathcal{C}$ è una *partizione* dell'evento certo, essendo formata da eventi incompatibili ed esaustivi. La stabilità di $\mathcal{U}$ per le operazioni finite su eventi ci lascia senza ragioni per cercare una classe più ampia, quando interessi un numero finito di eventi. Pertanto, in questo caso, porremo senz'altro $\mathcal{A} = \mathcal{U}$.

In generale richiederemo che la classe $\mathcal{A}$ di uno spazio di probabilità goda della stessa proprietà di stabilità che abbiamo verificato per $\mathcal{U}$.

**Definizione 1.1.2 (algebra)** *Una classe $\mathcal{A}$ di eventi, definita su un insieme di esiti $\Omega$, forma un'algebra, su $\Omega$, se:*

*i) $\Omega \in \mathcal{A}$;*

*ii) $\bar{E} \in \mathcal{A}$ ogni qual volta $E$ appartenga ad $\mathcal{A}$;*

*iii) $E \cap F \in \mathcal{A}$ ogni qual volta $E$ ed $F$ appartengano ad $\mathcal{A}$.*

Le formule di De Morgan $\overline{E \cup F} = \bar{E} \cap \bar{F}$ ed $\overline{E \cap F} = \bar{E} \cup \bar{F}$, valide comunque si scelgano gli eventi $E$ ed $F$, mostrano come la iii) equivalga, data la ii), a richiedere che $E \cup F \in \mathcal{A}$ ogni qual volta $E$ ed $F$ appartengano ad $\mathcal{A}$. Si vede poi subito (per induzione) che la iii) e l'analoga per l'unione si estendono al caso di tre o più eventi. Infine, se $E$ ed $F$ sono in $\mathcal{A}$, anche la loro *differenza simmetrica* $E \triangle F = (E \cap \bar{F}) \uplus (\bar{E} \cap F)$ è in $\mathcal{A}$; in particolare

$F \setminus E \in \mathcal{A}$, se $E \subseteq F$. Un'algebra è dunque una classe di eventi all'interno della quale possiamo operare liberamente con le operazioni finite. Si noti che necessariamente $\emptyset \in \mathcal{A}$, come si verifica combinando la i) e la ii).

Si dice che l'algebra delle unioni di costituenti è *generata* dalla classe dei costituenti, in quanto è la più piccola (meno numerosa) algebra su $\Omega$ che includa $\mathcal{C}$; infatti le unioni di costituenti dovranno appartenere a qualsiasi algebra che includa $\mathcal{C}$. La stessa algebra su $\Omega$ è generata dalla classe $\{B, D\}$, dai cui elementi abbiamo ottenuto i costituenti per negazione e intersezione. In generale, l'algebra $\alpha(\mathcal{E})$ generata da una classe di eventi $\mathcal{E}$ è definita come l'intersezione di tutte le algebre su $\Omega$ che includano $\mathcal{E}$ ed è pertanto inclusa in ciascuna di esse. La definizione è ben posta, comunque si prenda $\mathcal{E}$, perché qualsiasi intersezione di algebre è ancora un'algebra e c'è sempre almeno un'algebra che includa $\mathcal{E}$ (la classe $\wp(\Omega)$ di tutte le parti di $\Omega$).

Proviamo ora a ripetere la costruzione dell'algebra delle unioni di costituenti nel caso dell'Esempio 2. In questo caso ci interessa il tempo che farà domani, il cui manifestarsi è l'esperimento aleatorio modellato dal nostro amico. Qui non sembra agevole scegliere $\Omega$ in modo da rappresentare concretamente i possibili esiti sperimentali; si può però pensare di definirlo in modo astratto come un quadrato del piano euclideo, immaginando che una dea bendata determini l'esito dell'esperimento scagliando un dardo contro un bersaglio. Si individueranno quindi, nel quadrato $\Omega$, due regioni $B$ e $D$ corrispondenti agli eventi di interesse "pioggia" e "vento", facendo in modo che le regioni $B \cap D$, $\bar{B} \cap D$, $B \cap \bar{D}$ e $\bar{B} \cap \bar{D}$ siano tutte non vuote, visto che non vi sono ragioni logiche perché uno o più costituenti siano impossibili. Un grafico che rappresenti questa costruzione sarà detto *diagramma di Venn* e potrà essere utile, indicativamente, anche quando si sia fissato $\Omega$ in modo diverso. Se ne otterrà, prendendo le unioni di costituenti, un'algebra di eventi *isomorfa* a quella dell'Esempio 1, mentre non è possibile mettere in corrispondenza biunivoca i due insiemi degli esiti (il primo finito e il secondo infinito). Queste due coppie $(\Omega, \mathcal{A})$ sono fra loro interscambiabili, in pratica, perché la scelta di $\Omega$ è strumentale alla costruzione di $\mathcal{A}$.

Per completare l'analisi degli Esempi 1 e 2 dobbiamo assegnare una $\mathbb{P}$ con le caratteristiche che abbiamo fatto emergere sull'algebra $\mathcal{U}$ delle unioni di costituenti. In primo luogo, qualunque sia l'algebra di interesse, è sufficiente assegnare su di essa una $\mathbb{P}$ additiva, isotona e normalizzata ($\mathbb{P}(\Omega) = 1$): scrivendo $\mathbb{P}(\emptyset) + \mathbb{P}(\emptyset) = \mathbb{P}(\emptyset)$ troveremo $\mathbb{P}(\emptyset) = 0$ (neutralità di $\emptyset$ per $\mathbb{P}$) e quindi per isotonia $0 \leq \mathbb{P}(E) \leq 1$; dopo di che da $\mathbb{P}(E) + \mathbb{P}(F \setminus E) = \mathbb{P}(F)$, se $E \subseteq F$, seguirà $\mathbb{P}(F \setminus E) = \mathbb{P}(F) - \mathbb{P}(E)$ e in particolare $\mathbb{P}(\bar{E}) = 1 - \mathbb{P}(E)$ quando $F = \Omega$. In secondo luogo, per assegnare una $\mathbb{P}$ con queste tre proprietà sull'algebra $\mathcal{U}$ generata da $n$ eventi, occorre e basta assegnare $P_\alpha = \mathbb{P}(C_\alpha) \geq 0$, $\alpha \in \{0,1\}^n$, in modo che $\sum_{\alpha \in \{0,1\}^n} P_\alpha = 1$; la quota di $\biguplus_{\alpha \in A} C_\alpha$ sarà allora $\sum_{\alpha \in A} P_\alpha$. Scriveremo $\mathbb{P} = \Sigma P$ per denotare questa assegnazione. Per esempio, nel prevedere il tempo che farà domani, il nostro amico potrebbe porre $P_{00} = 0.12$, $P_{01} = 0.08$, $P_{10} = 0.16$ e $P_{11} = 0.64$, valutando quindi la probabilità che venga a piovere o si alzi il vento pari a $\mathbb{P}(B \cup D) = \mathbb{P}(C_{01} \uplus C_{10} \uplus C_{11}) = 0.08 + 0.16 + 0.64 = 0.88$. La stessa $\mathbb{P}$ potrebbe andare bene per un dado sbilanciato in modo da favorire l'uscita del punteggio 6, mentre per un dado equilibrato potremmo preferire assegnare $P_{00} = \mathbb{P}(\{1,5\}) = 1/3$, $P_{01} = \mathbb{P}(\{3\}) = 1/6$, $P_{10} = \mathbb{P}(\{2,4\}) = 1/3$ e $P_{11} = \mathbb{P}(\{6\}) = 1/6$. In questo secondo caso valuteremo la probabilità che il punteggio sia pari come $\mathbb{P}(B) = \mathbb{P}(\{2,4\} \uplus \{6\}) = 1/3 + 1/6 = 1/2$.

Assegnando $\mathbb{P} = \Sigma P$, come sopra descritto, si riesce a essere coerenti? Abbiamo visto che non si può fare altrimenti, se non si vuole offrire un'opportunità di profitto certo, ma basterà? In linea di principio, per quanto ne sappiamo, la coerenza potrebbe essere impossibile da ottenere. Non è però questo il caso. Infatti, comunque si scelgano $n$ scommesse $(E_1, S_1), \ldots, (E_n, S_n)$, il loro *ricavo* totale sarà $R_\star(\alpha) = \alpha_1 S_1 + \cdots + \alpha_n S_n$, se si verifica $C_\alpha$, mentre il pagamento totale per proporle sarà $Q_\star = \mathbb{P}(E_1)S_1 + \cdots + \mathbb{P}(E_n)S_n$. Quindi, essendo $\mathbb{P}$ additiva, avremo

$$Q_\star = \sum_{i=1}^{n} S_i \sum_{\alpha : C_\alpha \subseteq S_i} \mathbb{P}(C_\alpha) = \sum_{\alpha \in \{0,1\}^n} \mathbb{P}(C_\alpha) \sum_{i=1}^{n} \alpha_i S_i = \sum_{\alpha \in \{0,1\}^n} \mathbb{P}(C_\alpha) R_\star(\alpha)$$

e di conseguenza, essendo $\mathbb{P}$ normalizzata, il guadagno totale $G_\star = R_\star - Q_\star$ soddisferà il vincolo $\sum_{\alpha \in \{0,1\}^n} \mathbb{P}(C_\alpha) G_\star(\alpha) = 0$. In virtù di questo vincolo, essendo $\mathbb{P}$ positiva, $G_\star(\alpha)$ non potrà essere strettamente positivo per ogni $\alpha \in \{0,1\}^n$. Resta così dimostrato il seguente risultato.

**Teorema 1.1.3 (de Finetti)** *Una funzione di evento $\mathbb{P} : \mathcal{A} \to \mathbb{R}$, dove $\mathcal{A}$ è un'algebra di eventi, definita su un insieme di esiti $\Omega$, è coerente se e solo se è additiva, isotona e normalizzata.*

Si noti che, data l'additività, l'isotonia equivale alla positività: abbiamo già visto che, se $\mathbb{P}$ è isotona, allora $\mathbb{P}$ è positiva; d'altra parte, se $\mathbb{P}$ è positiva ed $E \subseteq F$, allora $\mathbb{P}(F) = \mathbb{P}(E) + \mathbb{P}(F \setminus E) \geq \mathbb{P}(E)$. Si noti inoltre che, per $n$ eventi $E_1, \ldots, E_n$ non necessariamente incompatibli, dall'additività di $\mathbb{P}$ segue $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(\bar{E}_1 \cap E_2)$ e quindi, essendo $\mathbb{P}$ isotona, $\mathbb{P}(E_1 \cup E_2) \leq \mathbb{P}(E_1) + \mathbb{P}(E_2)$; più in generale (per induzione) troveremo $\mathbb{P}(E_1 \cup \cdots \cup E_n) \leq \mathbb{P}(E_1) + \cdots + \mathbb{P}(E_n)$ e diremo che $\mathbb{P}$ è *subadditiva* (su $\mathcal{A}$).

Vale la pena sottolineare che la coerenza non ci assicura di non perdere soldi, se siamo sfortunati; ci offre piuttosto un'opportunità di guadagnarne, se siamo fortunati. In effetti un'assegnazione coerente di quote ha la caratteristica di essere *equa*, agli occhi di chi la fissa, in quanto il meccanismo della scommessa permette allo scommettitore di scambiarsi di ruolo con il banco (scegliendo $S < 0$). Data una sequenza di $n$ eventi equiprobabili, siano essi $E_1, \ldots, E_n$ con comune probabilità $p$, il guadagno di uno scommettitore che scelga un'unica posta $S$ per ciascuno di essi sarà $\{K(\omega) - np\}S$, al verificarsi di $\omega \in \Omega$, dove $K(\omega)$ è il numero di elementi $i \in \{1, \ldots, n\}$ tali che $\omega \in E_i$; tale guadagno sarà tanto più piccolo quanto più avremo scelto $p$ vicina alla frequenza relativa $K(\omega)/n$ degli eventi verificatisi. Uno scommettitore che sappia prevedere $K(\omega)/n$ meglio del banco potrà dunque legittimamente sperare di ottenere un profitto, nello scenario che stiamo considerando, ma non potrà mai esserne certo, avendo di fronte un banco coerente.

Per ragioni e con implicazioni che al momento trascuriamo, alla classe $\mathcal{A}$ di uno spazio di probabilità si richiede una forma di stabilità più forte.

**Definizione 1.1.4 (tribù)** *Una classe $\mathcal{A}$ di eventi, definita su un insieme di esiti $\Omega$, forma una tribù, su $\Omega$, se:*

*i)* $\Omega \in \mathcal{A}$;

*ii)* $\bar{E} \in \mathcal{A}$ *ogni qual volta $E$ appartenga ad $\mathcal{A}$;*

*iii)* $\bigcap_n E_n \in \mathcal{A}$ *ogni qual volta la successione $E_n$, $n \in \mathbb{N}$, sia in $\mathcal{A}$.*

Equivalentemente, una tribù (o $\sigma$-algebra) è stabile per unioni numerabili, come si vede combinando la ii) e la iii). Ogni tribù è un'algebra, in quanto si può sempre porre $E_n = \Omega$, $n \geq 3$, nella iii). L'algebra $\mathcal{U}$ delle unioni di costituenti è una tribù, banalmente, in quanto ha cardinalità finita.

Similmente, si richiede che la $\mathbb{P}$ di uno spazio di probabilità sia *numerabilmente additiva*: $\mathbb{P}(\biguplus_n E_n) = \sum_n \mathbb{P}(E_n)$ comunque si prenda una successione $E_n$, $n \in \mathbb{N}$, di elementi disgiunti di $\mathcal{A}$. Si noti che la serie $\sum_n \mathbb{P}(E_n)$ è a termini positivi con somme parziali minori di uno e dunque assolutamente convergente; non importa quindi l'ordine in cui la si somma. L'additività numerabile implica $\mathbb{P}(\emptyset) = 0$, come si vede scrivendo $\emptyset = \emptyset \uplus \emptyset \uplus \dots$, quindi diventa additività (finita) quando $E_n \equiv \emptyset$ eccetto che in un numero finito di casi. Le nostre richieste complessive su $\mathbb{P}$ sono dunque le seguenti.

**Definizione 1.1.5 (probabilità)** *Una funzione di evento $\mathbb{P} : \mathcal{A} \to \mathbb{R}$, dove $\mathcal{A}$ è una tribù di eventi, su un insieme di esiti $\Omega$, è una probabilità, su $\mathcal{A}$, quando sia numerabilmente additiva, isotona e normalizzata.*

Qualsiasi $\mathbb{P} = \Sigma P$ su $\mathcal{U}$ è una probabilità, banalmente, in quanto la cardinalità di $\mathcal{U}$ è finita.

Uno spazio di probabilità sarà dunque una terna $(\Omega, \mathcal{A}, \mathbb{P})$ dove $\Omega$ è un insieme qualsiasi, $\mathcal{A}$ una tribù su $\Omega$ e $\mathbb{P}$ una probabilità su $\mathcal{A}$. Gli elementi di $\Omega$ rappresentano i possibili esiti sperimentali (in modo più o meno astratto) e quindi sembra ragionevole richiedere $|\Omega| \geq 2$; per il resto, come abbiamo visto, la scelta di $\Omega$ è strumentale alla costruzione di $\mathcal{A}$. Gli elementi di $\mathcal{A}$ rappresentano invece gli eventi di interesse, sui quali si accettano scommesse con quote date da $\mathbb{P}$. Questo sarà il modello probabilistico fondamentale di ogni esperimento aleatorio e nel seguito cercheremo sempre di esplicitarlo.

## 1.2 Valutazioni classiche

Quando si modelli un esperimento aleatorio con uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$ il cui insieme degli esiti $\Omega$ sia *finito* è sempre possibile prendere $\mathcal{A} = \wp(\Omega)$, cioè interessarsi alla tribù di tutte le parti di $\Omega$ (tribù discreta). Basterà infatti assegnare $\mathbb{P}\{\omega\}$, $\omega \in \Omega$, per ottenere $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}\{\omega\}$, $E \subseteq \Omega$ (probabilità discreta). Considereremo cioè come costituenti i singoletti di $\Omega$. Si noti come, per alleggerire la notazione, si siano lasciate cadere le parentesi tonde della funzione di evento $\mathbb{P}$ in favore di quelle graffe del suo argomento. Adotteremo senz'altro questa convenzione, nel seguito, avendo ormai chiarito quale sia la natura di $\mathbb{P}$.

Un caso notevole, per semplicità e interesse, è quello in cui $\Omega$ sia tale da farci ritenere *equiprobabili* i suoi singoletti: $\mathbb{P}\{\omega\} \equiv p$ (probabilità discreta uniforme). In questo caso particolare, scrivendo $1 = \mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}\{\omega\} = |\Omega| p$, troviamo $p = 1/|\Omega|$ e quindi $\mathbb{P}(E) = |E|/|\Omega|$, $E \subseteq \Omega$. Ritroviamo cioè, come regola di calcolo, la definizione classica di probabilità: il rapporto tra "casi favorevoli" e "casi possibili". Tale definizione (storicamente la prima a emergere) ha le sue radici nell'ambito dei giochi d'azzardo, dove è spesso talmente naturale convenire su un insieme di esiti equiprobabili da far sì che a tale accordo intersoggettivo venga attribuito lo status di oggettività. Per esempio, se si lancia un dado a sei facce equilibrato, considerazioni di simmetria inducono "oggettivamente" a optare per una probabilità discreta uniforme (su $\Omega = \{1, 2, 3, 4, 5, 6\}$). Dopo di che, ogni qual volta si voglia calcolare la probabilità classica di un evento, sarà sufficiente contare gli esiti che lo realizzano e quelli che non lo realizzano.

La definizione classica di probabilità è insoddisfacente, da un punto di vista fondazionale, perché presuppone la nozione di equiprobabilità. Questo problema non è emerso, nell'argomento sopra svolto, perché avevamo già definito la probabilità come funzione di evento coerente (con riferimento al meccanismo della scommessa) ma sarebbe stato dirompente nel caso in cui avessimo voluto adottare la definizione classica come tale (invece che ricavarla come regola di calcolo in un caso particolare). Inoltre, è chiaro che la probabilità classica presuppone la possibilità di modellare l'esperimento aleatorio con uno *spazio di probabilità discreto uniforme* e tale presupposto non può essere sempre dato per scontato. Per esempio, nel caso in cui si lanci un dado sbilanciato in modo da favorire l'uscita della faccia 6, non è chiaro quale potrebbe essere un adeguato spazio di probabilità discreto uniforme; di certo non uno basato sulla scelta naturale $\Omega = \{1, 2, 3, 4, 5, 6\}$. D'altra parte, una valutazione soggettiva di quanto il dado sia sbilanciato, assieme alla considerazione che la faccia 6 è opposta alla faccia 1, può condurre all'assegnazione non classica $\mathbb{P}\{1\} = 0.04$, $\mathbb{P}\{2\} = \mathbb{P}\{3\} = \mathbb{P}\{4\} = \mathbb{P}\{5\} = 0.08$ e $\mathbb{P}\{6\} = 0.64$, sulla tribù $\mathcal{A} = \wp(\{1, 2, 3, 4, 5, 6\})$, senza alcuna difficoltà concettuale (ma con tutte le difficoltà pratiche legate alla quantificazione dello sbilanciamento).

Si noti che la scelta $\mathcal{A} = \wp(\Omega)$ è sempre possibile anche nel caso in cui $\Omega$ abbia cardinalità numerabile. Infatti, se $\Omega = \{\omega_1, \omega_2, \dots\}$, basterà assegnare $\mathbb{P}\{\omega_n\}$, $n \in \mathbb{N}$, per ottenere $\mathbb{P}(E) = \sum_{n=1}^{\infty} \mathbb{P}\{\omega_n\}$, $E \subseteq \Omega$. In concreto si tratterà di assegnare una serie a termini positivi che sommi a $\mathbb{P}(\Omega) = 1$. Il fatto che $\Omega$ abbia cardinalità infinita comporta però l'impossibilità di assegnare una probabilità discreta uniforme: la successione $\mathbb{P}\{\omega_n\}$, $n \in \mathbb{N}$, dovrà essere infinitesima e dunque non potrà essere costante (dovendosi escludere che sommi a zero). Gli spazi di probabilità discreti uniformi sono quindi caratteristici del caso $\Omega$ finito, mentre gli *spazi di probabilità discreti* sono tipici del caso $\Omega$ numerabile.

Nell'Esempio 3 le quattro carte che voltiamo in successione possono presentarsi in un ordine qualsiasi ed essendo state ben mischiate non vi è ragione per valutare un ordinamento più probabile di un altro. Pertanto, se le rappresentiamo con i numeri da 1 a 4, identificando ogni carta con il suo punteggio, possiamo modellare l'esperimento aleatorio con lo spazio di probabilità discreto uniforme che ha come insieme degli esiti quello delle *permutazioni* dei primi quattro numeri naturali:

$$\begin{aligned} P_4 \;=\; \{&(1,2,3,4),(1,2,4,3),(1,3,2,4),(1,3,4,2),(1,4,2,3),(1,4,3,2),\\ &(2,1,3,4),(2,1,4,3),(2,3,1,4),(2,3,4,1),(2,4,1,3),(2,4,3,1),\\ &(3,1,2,4),(3,1,4,2),(3,2,1,4),(3,2,4,1),(3,4,1,2),(3,4,2,1),\\ &(4,1,2,3),(4,1,3,2),(4,2,1,3),(4,2,3,1),(4,3,1,2),(4,3,2,1)\}; \end{aligned}$$

in generale $P_n = \{\omega \in \{1,\dots,n\}^n : i \neq j \Rightarrow \omega_i \neq \omega_j\}$ è l'insieme delle permutazioni dei primi $n$ numeri naturali, dove $\{1,\dots,n\}^n$ è il prodotto cartesiano di $\{1,\dots,n\}$ per se stesso $n$ volte e $\omega = (\omega_1,\dots,\omega_n)$. Abbiamo chiaramente $|P_4| = 24 = 4 \times 3 \times 2 \times 1 = 4!$ (quattro fattoriale) e in generale $|P_n| = n \times (n-1) \times \cdots \times 2 \times 1 = n!$ $(n \in \mathbb{N})$.

I primi due quesiti chiedono le probabilità degli eventi "quattro coincidenze" e "zero coincidenze", cioè delle parti $Q = \{(1,2,3,4)\}$ e $Z = \{(2,1,4,3), (2,3,4,1), (2,4,1,3), (3,1,4,2), (3,4,1,2), (3,4,2,1), (4,1,2,3), (4,3,1,2), (4,3,2,1)\}$ di $P_4$. Contando $|Q| = 1$ e $|Z| = 9$, troviamo $\mathbb{P}(Q) = 1/24 \approx 0.042$ e $\mathbb{P}(Z) = 9/24 = 3/8 = 0.375$. Il terzo e quarto quesito chiedono invece le probabilità degli eventi "una coincidenza nel voltare le prime due carte" e "due coincidenze nel voltare le prime due carte", cioè delle parti $U = \{(1,3,2,4), (1,3,4,2), (1,4,2,3), (1,4,3,2), (3,2,1,4), (3,2,4,1), (4,2,1,3), (4,2,3,1)\}$ e $D = \{(1,2,3,4), (1,2,4,3)\}$ di $P_4$. Contando $|U| = 8$ e $|D| = 2$, troviamo $\mathbb{P}(U) = 8/24 = 1/3 \approx 0.33$ e $\mathbb{P}(D) = 2/24 = 1/12 \approx 0.083$. Se poi si vuole calcolare la probabilità di avere *almeno* una coincidenza nel voltare le prime due carte, basterà ricordare che $\mathbb{P}$ è additiva per ottenere $\mathbb{P}(U \uplus D) = \mathbb{P}(U) + \mathbb{P}(D) = 1/3 + 1/12 = 5/12 \approx 0.42$.

Il terzo e quarto quesito riguardano solamente le prime due carte voltate. Quindi, per affrontarli direttamente, in assenza dei primi due quesiti, avremmo potuto più semplicemente considerare lo spazio di probabilità discreto uniforme che ha come esiti le *disposizioni senza ripetizione* di due numeri naturali presi dai primi quattro:

$$\begin{aligned} D_2^4 &= \{(1,2),(1,3),(1,4), \\ &\quad (2,1),(2,3),(2,4), \\ &\quad (3,1),(3,2),(3,4), \\ &\quad (4,1),(4,2),(4,3)\}; \end{aligned}$$

in generale $D_k^n = \{\omega \in \{1,\ldots,n\}^k : i \neq j \Rightarrow \omega_i \neq \omega_j\}$ è l'insieme delle disposizioni senza ripetizione di $k$ numeri naturali presi dai primi $n$, dove $\{1,\ldots,n\}^k$ è il prodotto cartesiano di $\{1,\ldots,n\}$ per se stesso $k$ volte e $\omega = (\omega_1,\ldots,\omega_k)$. Chiaramente $P_n = D_n^n$ e più in generale $D_k^n$ si ottiene da $P_n$ considerando equivalenti le permutazioni con i primi $k$ elementi uguali. Quindi $|D_k^n| = n!/(n-k)! = n \times (n-1) \times \cdots \times (n-k+2) \times (n-k+1)$, con la convenzione $0! = 1$ e $|D_2^4| = 4 \times 3 = 12$ nel caso specifico. Si noti che possiamo identificare $D_{n-1}^n$ con $D_n^n$, essendo $\omega_n$ univocamente determinato da $\omega_1,\ldots,\omega_{n-1}$. Riutilizzando le stesse lettere per gli stessi eventi, le parti di $D_2^4$ che interessano per rispondere al terzo e quarto quesito sono $U = \{(1,3),(1,4),(3,2),(4,2)\}$ e $D = \{(1,2)\}$. Anche in questo secondo modello, contando $|U| = 4$ e $|D| = 1$, troviamo $\mathbb{P}(U) = 4/12 = 1/3$ e $\mathbb{P}(D) = 1/12$.

Gli ultimi due quesiti riguardano il numero di spade presenti sulle prime due carte voltate (corrispondente alla somma dei due punteggi dopo avere sostituito 4 con 0). In particolare, interessano gli eventi "tre spade" e "cinque spade", cioè le parti $T = \{(1,2),(2,1),(3,4),(4,3)\}$ e $C = \{(2,3),(3,2)\}$ di $D_2^4$. Contando $|T| = 4$ e $|C| = 2$, troviamo $\mathbb{P}(T) = 4/12 = 1/3 \approx 0.33$ e $\mathbb{P}(C) = 2/12 = 1/6 \approx 0.17$. Poiché il numero di spade presenti sulle prime due carte non dipende dall'ordine in cui esse vengono voltate, avremmo più semplicemente potuto calcolare le probabilità richieste sulla base dello spazio di probabilità discreto uniforme che ha come esiti le *combinazioni senza ripetizione* di due numeri naturali presi dai primi quattro:

$$\begin{aligned} C_{4,2} &= \{\{1,2\},\{1,3\},\{1,4\}, \\ &\quad \{2,3\},\{2,4\},\{3,4\}\}; \end{aligned}$$

in generale $C_k^n = \{\omega \subseteq \{1,\ldots,n\} : |\omega| = k\}$ è l'insieme delle combinazioni senza ripetizione di $k$ numeri naturali presi dai primi $n$. Si vede subito che $C_k^n$ si ottiene da $D_k^n$ considerando equivalenti le disposizioni che sono l'una permutazione dell'altra. Di conseguenza $|C_k^n| = n!/\{k!(n-k)!\} = \binom{n}{k}$, *coefficiente binomiale* da leggersi "$n$ su $k$", e nello specifico $|C_2^4| = \binom{4}{2} = 4!/(2!2!) = 6$. Dopo di che le parti interessanti di $C_2^4$ (per rispondere al quinto e sesto quesito) sono $T = \{(1,2),(3,4)\}$ e $C = \{(2,3)\}$; ritroviamo $\mathbb{P}(T) = 2/6 = 1/3$ e $\mathbb{P}(C) = 1/6$, contando $|T| = 2$ e $|C| = 1$. Alle stesse probabilità saremmo giunti anche nel modello iniziale (spazio di probabilità discreto uniforme con $\Omega = P_4$) solo contando più a lungo.

## 1.3 Condizionamento

La *scommessa condizionale* sull'evento $E$ dato l'evento $F$ è definita come segue: fissata una posta in palio $S$ (eventualmente negativa) lo scommettitore paga $\mathbb{P}(E|F)S$ (incassando $|\mathbb{P}(E|F)S|$ nel caso $S < 0$) per il diritto a incassare $S$ (l'obbligo di pagare $|S|$ nel caso $S < 0$) quando $E$ si verifichi, sotto la condizione che $F$ si verifichi; se $F$ non si verifica, la scommessa è annullata e lo scommettitore recupera la puntata $\mathbb{P}(E|F)S$ (restituendo $|\mathbb{P}(E|F)S|$ nel caso $S < 0$) a prescindere dal verificarsi o meno di $E$. Il ricavo dello scommettitore sarà quindi

$$R(\omega) = \begin{cases} \mathbb{P}(E|F)S, & \text{se } \omega \in \bar{F}, \\ 0, & \text{se } \omega \in F \cap \bar{E}, \\ S, & \text{se } \omega \in F \cap E, \end{cases}$$

equivalente al ricavo totale delle due seguenti scommesse (non condizionali): una scommessa su $E \cap F$, con posta $S$, e una scommessa su $\bar{F}$, con posta $\mathbb{P}(E|F)S$. Per coerenza, la somma delle puntate di queste due scommesse dovrà uguagliare la puntata della scommessa condizionale:

$$\mathbb{P}(E \cap F)S + \mathbb{P}(\bar{F})\mathbb{P}(E|F)S = \mathbb{P}(E|F)S.$$

Prendendo $S = 1$ e ricordando il vincolo di coerenza $\mathbb{P}(\bar{F}) = 1 - \mathbb{P}(F)$, cui possiamo dare il nome di *regola della negazione*, troviamo quella che chiameremo *regola del prodotto*: $\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F)$.

In virtù dell'argomento sopra svolto, comunque si prendano due eventi $E$ ed $F$ in uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, con $\mathbb{P}(F) > 0$, l'unico modo coerente di definire la *probabilità condizionale* di $E$ dato $F$, intesa come quota parte della posta in palio da pagare per scommettere su $E$ sotto la condizione che $F$ si verifichi, è porre $\mathbb{P}(E|F) = \mathbb{P}(E \cap F)/\mathbb{P}(F)$. In questo modo, per $F$ fissato, resta definita la funzione di evento $\mathbb{P}_F : E \mapsto \mathbb{P}(E|F)$ che si verifica subito essere una probabilità su $\mathcal{A}$:

i) $\mathbb{P}(\Omega|F) = \mathbb{P}(\Omega \cap F)/\mathbb{P}(F) = \mathbb{P}(F)/\mathbb{P}(F) = 1$;

ii) $E_1 \subseteq E_2 \Rightarrow E_1 \cap F \subseteq E_2 \cap F$, quindi $E_1 \subseteq E_2 \Rightarrow \mathbb{P}(E_1|F) \leq \mathbb{P}(E_2|F)$, in conseguenza dell'isotonia di $\mathbb{P}$.

iii) $\mathbb{P}(\biguplus_n E_n|F) = \mathbb{P}(\biguplus_n E_n \cap F)/\mathbb{P}(F)$, dove l'intersezione si distribuisce rispetto all'unione, quindi $\mathbb{P}(\biguplus_n E_n|F) = \sum_n \mathbb{P}(E_n|F)$, in virtù del fatto che $\mathbb{P}$ è numerabilmente additiva.

Se invece $\mathbb{P}(F) = 0$, la regola del prodotto è soddisfatta in modo banale e quindi non possiamo sfruttarla per definire univocamente $\mathbb{P}(E|F)$; in linea di principio andrebbe bene qualsiasi funzione di evento $E \mapsto \mathbb{P}(E|F)$ che fosse una probabilità.

Merita un momento di attenzione il caso particolare in cui $F = \Omega$. In tal caso $\mathbb{P}(\Omega) = 1$ e $\mathbb{P}(E|\Omega) = \mathbb{P}(E \cap \Omega)/\mathbb{P}(\Omega) = \mathbb{P}(E)$. Ritroviamo cioè la probabilità (non condizionale) $\mathbb{P}$ come probabilità condizionale dato l'evento certo. In effetti, la scelta di un particolare $\Omega$ come insieme degli esiti esclude categoricamente ogni altra eventualità e questo, in pratica, è possibile solo se si immagina di annullare le scommesse quando la realtà superi la nostra fantasia. Per esempio, non potremo che annullare le scommesse quando il dado che abbiamo lanciato rotoli in un tombino, oppure più realisticamente si fermi di traverso contro un oggetto (senza esibire chiaramente una faccia). Ogni probabilità è dunque condizionale, se non altro in quanto subordinata al modello che abbiamo scelto per l'esperimento aleatorio.

Nell'Esempio 4 l'esperimento aleatorio consiste nell'osservare l'età alla quale l'uomo prudente muore e possiamo prendere come insieme degli esiti $T = \{30, 31, \dots\} = \{n \in \mathbb{N} : n \geq 30\}$, visto che egli ha appena compiuto il trentesimo anno di età. Ci interessano gli eventi $R = \{65, 66, \dots\}$, cioè il raggiungimento dell'età pensionabile, e $O = \{80, 81, \dots\}$, cioè il compimento dell'ottantesimo anno di età. Uno dei quattro costituenti di questi due eventi è impossibile, $\bar{R} \cap O = \emptyset$, mentre gli altri tre sono da valutare sulla base delle informazioni disponibili: $\bar{R} \cap \bar{O} = \{30, 31, \dots, 63, 64\} = \bar{R}$, $R \cap \bar{O} = \{65, 66, \dots, 78, 79\}$ ed $R \cap O = \{80, 81, \dots\} = O$. Se poniamo $\mathbb{P}(\bar{R}) = 0.20$, $\mathbb{P}(R \cap \bar{O}) = 0.42$ e $\mathbb{P}(O) = 0.38$, come è senz'altro lecito, visto che i tre valori sono positivi e sommano a uno, troveremo $\mathbb{P}(O|R) = \mathbb{P}(O \cap R)/\mathbb{P}(R) = \mathbb{P}(O)/\{1 - \mathbb{P}(\bar{R})\} = 0.38/(1 - 0.20) = 38/80 = 0.475$. In virtù di questo calcolo, il valore *equo* del totale dei versamenti sarà 4750 euro e quindi quello di ogni singolo versamento sarà $4750/35 \approx 136$ euro; l'equità di tali valori si riferisce al fatto che nel meccanismo della scommessa condizionale è possibile per lo scommettitore scegliere una posta negativa e in questo modo scambiarsi di ruolo col banco. Nella realtà, se c'è bisogno di precisarlo, la compagnia assicuratrice non sarebbe disponibile allo scambio di ruoli e chiederebbe un versamento annuale maggiore di quello equo (in modo da coprire le sue spese e garantirsi un margine di profitto).

Il secondo e ultimo quesito dell'Esempio 4 chiede di valutare la probabilità che si verifichi $O$ alla luce dell'informazione che si è verificato $R$. Sembra naturale rispondere con il valore $\mathbb{P}(O|R) = 0.475$ (precedentemente calcolato) e in effetti questa è la risposta generalmente considerata valida. Non è però logicamente necessario, a rigore, sostituire $\mathbb{P}(O)$ con $\mathbb{P}(O|R)$ a seguito del verificarsi di $R$. In effetti, una volta che $R$ si sia verificato, qualsiasi probabilità sulla tribù $\mathcal{A}_R = \{E \in \mathcal{A} : E \subseteq R\}$, detta *tribù traccia* di $\mathcal{A}$ su $R$, definirà un valido spazio di probabilità sul nuovo evento certo $R$ (garantendo la nostra coerenza). Nonostante questa considerazione, per altro ineccepibile, vi è consenso generale sul fatto che la scelta $\mathbb{P}_R(E)$, $E \in \mathcal{A}_R$, sia quella "giusta", in quanto il meccanismo della scommessa condizionale ben coglie l'idea di subordinare la valutazione soggettiva di probabilità all'informazione che l'evento $R$ si sia verificato.

Diremo che $(F, \mathcal{A}_F, \mathbb{P}_F)$ è lo *spazio di probabilità condizionale* subordinato al verificarsi di $F$, qualunque sia $F \in \mathcal{A}$ con $\mathbb{P}(F) > 0$, pertanto vedremo $(\Omega, \mathcal{A}, \mathbb{P})$ come subordinato al verificarsi di $\Omega$, prendendo $F = \Omega$. In effetti, qualora la realtà ci sorprenda con un esito non previsto da $\Omega$, dovremo per forza sostituire $(\Omega, \mathcal{A}, \mathbb{P})$ con un modello "tutto nuovo". D'altra parte, per svelare l'arcano, lo spazio di probabilità da cui siamo partiti per analizzare l'Esempio 4 nasce come spazio condizionale. Infatti, sulla base della tavola di mortalità pubblicata dall'Istat (Istituto Italiano di Statistica) per i maschi della classe 1984 (con riferimento alla regione Emilia Romagna) possiamo modellare le aspettative di vita dell'uomo prudente alla nascita prendendo $\Omega = \mathbb{N} \cup \{0\} = \mathbb{Z}_+$, $\mathcal{A}$ su $\Omega$ generata da $T$, $R$ e $O$, $\mathbb{P}(O) = 36833/10000$, $\mathbb{P}(R) = 77103/10000$ e $\mathbb{P}(T) = 96730/10000$ (le probabilità dei costituenti essendo definite per differenza, dato che $O \subset R \subset T \subset \Omega$ e $\mathbb{P}(\Omega) = 1$). Dopo di che, condizionando all'evento osservato $T$, si trova che $\mathcal{A}_T$ coincide con $\mathcal{U}$ generata da $R$ e $O$, mentre $\mathbb{P}_T(O) = 36833/96730 \approx 0.38$ e $\mathbb{P}_T(R) = 77103/96730 \approx 0.80$, di modo che $\mathbb{P}_T(R \cap \bar{O}) = 0.80 - 0.38 = 0.42$ e $\mathbb{P}(\bar{R}) = 1 - 0.80 = 0.20$ (per differenza). Il punto di partenza dell'analisi precedente era dunque il punto di arrivo di questa analisi di dati ufficiali.

Nel caso in cui si abbia a che fare con una valutazione classica di probabilità, cioè con uno spazio di probabilità discreto uniforme, la probabilità condizionale di $E$ dato $F$ (supposto $\mathbb{P}(F) > 0$, cioè $|F| > 0$) può calcolarsi come rapporto di cardinalità: $\mathbb{P}(E|F) = |E \cap F|/|F|$, visto che il comune denominatore $|\Omega|$ si semplifica. Alla stessa formula si giungerebbe prendendo direttamente $F$ come insieme degli esiti.

Nell'Esempio 5, per modellare l'estrazione casuale di un mattoncino da una scatola di costruzioni, possiamo porre $\Omega = \{1, \dots, 150\}$ e convenire che:

$\alpha$) i mattoncini da 1 a 60 siano quadrati, mentre i mattoncini da 61 a 150 siano rettangolari;

$\beta$) i mattoncini da 1 a 30 e quelli da 61 a 120 siano gialli, mentre i mattoncini da 31 a 60 e quelli da 121 a 150 siano rossi.

Avremo quindi $Q = \{1, \dots, 60\}$ e $G = \{1, \dots, 30, 61, \dots, 120\}$ come eventi di interesse e troveremo $\mathbb{P}(G) = 90/150 = 3/5 = 0.6$ in risposta al primo quesito. Dopo di che risponderemo al secondo quesito calcolando $\mathbb{P}(G|Q) = 30/60 = 0.5$ e concluderemo che, quando in palio vi siano 100 euro, il diritto di effettuare un'estrazione varrà 60 euro prima di tastare il mattoncino e 50 euro dopo averlo tastato e avere scoperto che è quadrato. Si noti che, diversamente da quanto accadeva nel caso dell'uomo prudente, l'evento $G \cap \bar{Q}$ non è impossibile e di questo si deve tenere conto nel conteggio al numeratore di $\mathbb{P}(G|Q)$: $30 = |G \cap Q| \neq |G| = 90$. Infine, per completare l'analisi, vale la pena calcolare $\mathbb{P}(G|\bar{Q}) = 60/90 = 2/3 \approx 0.67$; alla luce di questa valutazione, qualora si scoprisse che il mattoncino è rettangolare, il diritto di effettuare un'estrazione varrebbe circa 67 euro.

Cosa cambierebbe se quaranta dei mattoncini quadrati fossero gialli e venti rossi? In questo caso troveremmo $\mathbb{P}(G) = (40 + 60)/150 = 2/3 = 40/60 = \mathbb{P}(G|Q)$ e quindi l'informazione che il mattoncino è quadrato non cambierebbe la nostra valutazione di probabilità. Similmente (non per caso) troveremmo $\mathbb{P}(G|\bar{Q}) = 60/90 = 2/3 = \mathbb{P}(G)$ e pertanto il diritto di estrarre un mattoncino varrebbe 67 euro (circa) indipendentemente dalla possibilità o meno di tastarlo prima di scommettere.

A ben guardare, nell'analisi dell'Esempio 5, la scelta $\mathcal{A} = \wp(\Omega)$ ci porta a lavorare con una tribù troppo grande (numerosa): solamente gli eventi definibili in termini di forma e colore dei mattoncini sono infatti ammissibili come oggetto di scommessa, perché il verificarsi o meno di ogni altro evento dipende in modo essenziale dai numeri attribuiti ai mattoncini e questi sono arbitrari (non essendo i mattoncini materialmente numerati). Converrà quindi restringere $\mathbb{P}$ a $\mathcal{U}$ generata da $G$ e $Q$, ottenendo per i costituenti $\mathbb{P}(G \cap Q) = 30/150 = 1/5 = 0.2$, $\mathbb{P}(G \cap \bar{Q}) = 60/150 = 2/5 = 0.4$, $\mathbb{P}(\bar{G} \cap Q) = 30/150 = 1/5 = 0.2$ e $\mathbb{P}(\bar{G} \cap \bar{Q}) = 30/150 = 1/5 = 0.2$, se il contenuto della scatola è quello originale (descritto ai punti $\alpha$ e $\beta$).

Si può per altro assegnare $\mathbb{P}$ direttamente su $\mathcal{U}$ con la strategia seguente. Prima si valutano in modo classico $\mathbb{P}(Q) = 60/150 = 2/5$ (considerando tutti i mattoncini assieme), $\mathbb{P}(G|Q) = 30/60 = 1/2$ e $\mathbb{P}(G|\bar{Q}) = 60/90 = 2/3$ (considerando separatamente i mattoncini delle due forme), poi si usa la regola del prodotto in modo costruttivo per ottenere le probabilità dei costituenti:

$$
\begin{aligned}
\mathbb{P}(G \cap Q) &= \mathbb{P}(G|Q)\mathbb{P}(Q) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5}, \\
\mathbb{P}(G \cap \bar{Q}) &= \mathbb{P}(G|\bar{Q})\{1 - \mathbb{P}(Q)\} = \frac{2}{3} \times \frac{3}{5} = \frac{2}{5}, \\
\mathbb{P}(\bar{G} \cap Q) &= \{1 - \mathbb{P}(G|Q)\}\mathbb{P}(Q) = \frac{1}{2} \times \frac{2}{5} = \frac{1}{5}, \\
\mathbb{P}(\bar{G} \cap \bar{Q}) &= \{1 - \mathbb{P}(G|\bar{Q})\}\{1 - \mathbb{P}(Q)\} = \frac{1}{3} \times \frac{3}{5} = \frac{1}{5},
\end{aligned}
$$

sfruttatando il fatto che $\mathbb{P}$, $\mathbb{P}_Q$ e $\mathbb{P}_{\bar{Q}}$ sono probabilità e quindi soddisfano la regola della negazione.

La strategia sopra descritta, applicata ricorsivamente, funziona più in generale per assegnare una probabilità sulla tribù generata da un numero finito di eventi: se $E_1, \ldots, E_n$ sono gli eventi di interesse e $\alpha \in \{0,1\}^n$ identifica il loro generico costituente $C_\alpha = \bigcap_{i=1}^n E_i^{\alpha_i}$, con la convenzione $E_i^0 = \bar{E}_i$ ed $E_i^1 = E_i$, $i = 1, \ldots, n$, basterà assegnere $\mathbb{P}(E_n|\bigcap_{i=1}^{n-1} E_i^{\alpha_i})$ per trovare $\mathbb{P}(\bigcap_{i=1}^n E_i^{\alpha_i}) = \mathbb{P}(E_n^{\alpha_n}|\bigcap_{i=1}^{n-1} E_i^{\alpha_i})\mathbb{P}(\bigcap_{i=1}^{n-1} E_i^{\alpha_i})$, dove si conviene che $\bigcap_{i=1}^0 E_i^{\alpha_i} = \Omega$ e naturalmente $\mathbb{P}(\bar{E}_n|\bigcap_{i=1}^{n-1} E_i^{\alpha_i}) = 1 - \mathbb{P}(E_n|\bigcap_{i=1}^{n-1} E_i^{\alpha_i})$. Va da sé che non è necessario, né in realtà è sempre possibile, avvalersi di valutazioni classiche in ogni passo della ricorsione; la fattorizzazione delle probabilità dei costituenti può comunque aiutare a elaborare una convincente assegnazione soggettiva (garantendone in ogni caso la coerenza).

Nell'Esempio 6, se immaginiamo che le biglie nell'urna siano numerate da 1 a 120, convenendo che le prime 40 siano rosse e le rimanenti bianche, analogamente a quanto abbiamo fatto nel caso dei mattoncini nella scatola di costruzioni, possiamo prendere $\Omega = D_{10}^{120}$, $\mathcal{A}$ generata dagli eventi $R_i = \{\omega \in D_{10}^{120} : \omega_i \leq 40\}$, $i = 1, \ldots, 10$, dove $\omega = (\omega_1, \ldots, \omega_{10})$, $\mathbb{P}$ uniforme su $\wp(D_{10}^{120})$ ristretta ad $\mathcal{A}$. In questo modo, troviamo subito

$$\mathbb{P}(R_1) = \frac{40 \times 119 \times 118 \times \cdots \times 112 \times 111}{120 \times 119 \times \cdots \times 112 \times 111} = \frac{40}{120} = \frac{1}{3} \approx 0.333$$

per la probabilità che la prima biglia estratta sia rossa. Dopo di che non facciamo fatica a convincerci che $\mathbb{P}(R_i) = 1/3$, $i = 1, \ldots, 10$, perché $R_i$ può essere messo in corrispondenza biunivoca con $R_1$ scambiando $\omega_i$ e $\omega_1$.

Si noti che $\mathbb{P}(R_2) = 1/3$ è la probabilità che la seconda biglia estratta sia rossa *senza* conoscere il colore della prima biglia estratta. Qualora invece conoscessimo il colore della prima biglia estratta, troveremmo

$$\begin{aligned}
\mathbb{P}(R_2|R_1) &= \frac{40 \times 39 \times 118 \times 117 \times \cdots \times 112 \times 111}{40 \times 119 \times 118 \times \cdots \times 112 \times 111} = \frac{39}{119} \approx 0.328, \\
\mathbb{P}(R_2|\bar{R}_1) &= \frac{80 \times 40 \times 118 \times 117 \times \cdots \times 112 \times 111}{80 \times 119 \times 118 \times \cdots \times 112 \times 111} = \frac{40}{119} \approx 0.336,
\end{aligned}$$

a seconda che la prima biglia estratta sia rossa o bianca. Anche qui una breve riflessione ci convince che $\mathbb{P}(R_i|R_j) = 39/119$ e $\mathbb{P}(R_i|\bar{R}_j) = 40/119$, comunque presi $i$ e $j$ in $\{1, \ldots, 10\}$, con $i \neq j$ (non necessariamente $i > j$).

Agli stessi valori di $\mathbb{P}(R_2|R_1)$ e $\mathbb{P}(R_2|\bar{R}_1)$ si giunge valutando classicamente la seconda estrazione sulla base dell'*urna residua* dopo la prima estrazione (contenente 119 biglie di cui 39 rosse e 80 bianche o 40 rosse e 79 bianche a seconda del colore della prima biglia estratta). Similmente

$$\begin{aligned}
\mathbb{P}(R_3|R_1 \cap R_2) &= \frac{38}{118} = \frac{19}{59} \approx 0.322, \\
\mathbb{P}(R_3|R_1 \cap \bar{R}_2) &= \frac{39}{118} \approx 0.331, \\
\mathbb{P}(R_3|\bar{R}_1 \cap \bar{R}_2) &= \frac{40}{118} = \frac{20}{59} \approx 0.339,
\end{aligned}$$

e ci si convince che si ottengono gli stessi valori comunque si prendano $i$, $j$ e $k$ distinti in $\{1, \ldots, 10\}$ al posto di 1, 2 e 3; in particolare $\mathbb{P}(R_3|\bar{R}_1 \cap R_2) = \mathbb{P}(R_3|R_1 \cap \bar{R}_2) = 39/118$ (39 biglie rosse tra le 118 rimaste).

Iterando la valutazione dell'urna residua, otteniamo $\mathbb{P}(R_k|\bigcap_{i=1}^{k-1} R_i^{\alpha_i}) = (40 - \alpha_+)/(120 - k + 1)$, per $\alpha \in \{0,1\}^k$, $k = 1, \ldots, 10$, dove $\alpha_+ = \sum_{i=1}^{k-1} \alpha_i$. Possiamo allora assegnare $\mathbb{P}$ direttamente su $\mathcal{A} = \sigma(\{R_1, \ldots, R_{10}\})$ con la strategia ricorsiva prima delineata e questo ci consentirà di sostituire $D_{10}^{120}$ con $\{0,1\}^{10}$, a patto di ridefinire $R_i = \{\omega \in \{0,1\}^{10} : \omega_i = 1\}$, $i = 1, \ldots, 10$. In questo modo elimineremo la necessità di immaginare che le biglie siano numerate, i costituenti saranno i singoletti di $\{0,1\}^{10}$ e avremo $\mathcal{A} = \wp(\Omega)$.

Con questo secondo modello in mente, troviamo subito la probabilità che le prime due biglie estratte siano entrambe rosse (entrambe bianche):

$$\begin{aligned}\mathbb{P}(R_1 \cap R_2) &= \mathbb{P}(R_1)\mathbb{P}(R_2|R_1) = \frac{1}{3} \times \frac{39}{119} = \frac{13}{119} \approx 0.109;\\ \mathbb{P}(\bar{R}_1 \cap \bar{R}_2) &= \mathbb{P}(\bar{R}_1)\mathbb{P}(\bar{R}_2|\bar{R}_1) = \frac{2}{3} \times \frac{79}{119} = \frac{158}{357} \approx 0.443.\end{aligned}$$

Dopo di che, essendo $\mathbb{P}$ additiva, la probabilità che le due biglie abbiano lo stesso colore sara pari a $\mathbb{P}(R_1 \cap R_2) + \mathbb{P}(\bar{R}_1 \cap \bar{R}_2) \approx 0.109 + 0.443 = 0.552$.

Meno immediato è il calcolo della probabilità che almeno una delle prime due biglie estratte sia rossa, perché gli eventi $R_1$ ed $R_2$ non sono disgiunti; non si può quindi sfruttare direttamente l'additività di $\mathbb{P}$. Ci si può però avvalere della *regola della somma*: $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$, comunque presi $E$ ed $F$ nella tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$. Con $E = R_1$ ed $F = R_2$, troveremo $\mathbb{P}(R_1 \cup R_2) = 1/3 + 1/3 - 13/119 = 199/357 \approx 0.557$. La validità della regola della somma discende dalle decomposizioni $E = (E \cap \bar{F}) \uplus (E \cap F)$ ed $F = (\bar{E} \cap F) \uplus (E \cap F)$ di $E$ ed $F$, combinate con la decomposizione $E \cup F = (E \cap \bar{F}) \uplus (\bar{E} \cap F) \uplus (E \cap F)$ della loro unione $E \cup F$. In pratica, sommando $\mathbb{P}(E)$ e $\mathbb{P}(F)$ si aggiunge due volte la probabilità $\mathbb{P}(E \cap F)$, per cui occorre sottrarla al risultato finale.

Analogamente si può calcolare $\mathbb{P}(R_1 \cup R_2 \cup R_3)$, prima sommando $\mathbb{P}(R_1)$, $\mathbb{P}(R_2)$ e $\mathbb{P}(R_3)$, poi sottraendo $\mathbb{P}(R_1 \cap R_2)$, $\mathbb{P}(R_1 \cap R_3)$ e $\mathbb{P}(R_2 \cap R_3)$, infine sommando $\mathbb{P}(R_1 \cap R_2 \cap R_3)$, visto che quest'ultima probabilità era stata prima aggiunta e poi tolta tre volte. Osservando che la probabilità dell'intersezione di una parte degli eventi $R_1, \ldots, R_{10}$ dipende solo da quanti (e non da quali) eventi si intersecano, per quanto abbiamo visto nella prima parte dell'analisi, si trova $\mathbb{P}(R_1 \cup R_2 \cup R_3) = 3\,\mathbb{P}(R_1) - 3\,\mathbb{P}(R_1 \cap R_2) + \mathbb{P}(R_1 \cap R_2 \cap R_3)$. Con questa formula, dato che

$$\begin{aligned}\mathbb{P}(R_1 \cap R_2 \cap R_3) &= \mathbb{P}(R_1)\mathbb{P}(R_2|R_1)\mathbb{P}(R_3|R_1 \cap R_2)\\ &= \frac{1}{3} \times \frac{39}{119} \times \frac{19}{59} = \frac{247}{7021} \approx 0.035,\end{aligned}$$

si trova $\mathbb{P}(R_1 \cup R_2 \cup R_3) \approx 3 \times 0.333 - 3 \times 0.109 + 0.035 = 0.707$. Si noti l'alternarsi di termini positivi e negativi che via via raffinano il risultato.

Nel caso di quattro o più eventi si può giungere a formule simili, le cui somme parziali alternano approssimazioni per eccesso e per difetto (dette *disuguaglianze di Bonferroni*) fino a che la somma totale (detta *principio di inclusione-esclusione*) fornisce il valore esatto; la prima disuguaglianza di Bonferroni (detta anche *disuguaglianza di Boole*) non esprime altro che la subadditività di $\mathbb{P}$. Al crescere del numero di eventi, tuttavia, il principio di inclusione-esclusione diventa via via meno conveniente del "passaggio al complementare": $\mathbb{P}(R_1 \cup R_2 \cup R_3 \cup R_4) = 1 - \mathbb{P}(\bar{R}_1 \cap \bar{R}_2 \cap \bar{R}_3 \cap \bar{R}_4) = 1 - (80 \times 79 \times 78 \times 77)/(120 \times 119 \times 118 \times 117) \approx 0.807$, nel caso delle prime quattro biglie, $\mathbb{P}(\bigcup_{i=1}^{10} R_i) = 1 - \mathbb{P}(\bigcap_{i=1}^{10} \bar{R}_i) = 1 - (80!/70!)/(120!/110!) \approx 0.986$, nel caso si considerino tutte e dieci le biglie.

## 1.4 Apprendimento

Nell'Esempio 7, con riferimento alla vita di un nostro amico, ci interessano gli eventi $T$ = "positivo al test" ed $M$ = "malato", nel senso che abbiamo osservato $T$ e vogliamo valutare $\mathbb{P}(M|T)$. Possiamo prendere come insieme degli esiti $\Omega$ un quadrato del piano euclideo, ragionando in astratto, visto che non sembra esservi una scelta concreta più naturale e comunque le $\mathcal{U}$ generate da $T$ ed $M$ su due qualsiasi insiemi degli esiti sono fra loro isomorfe. Come assegneremo $\mathbb{P}$ sui costituenti di $T$ ed $M$?

Sappiamo che $\mathbb{P}(M) = 0.02$, perché questa è la *prevalenza* delle malattia nella popolazione oggetto di screening e il nostro amico si è sottoposto al test senza alcun indizio in favore o contro la malattia (semplicemente nell'ambito di una campagna di screening). Si noti che la prevalenza di una malattia dipende dalla popolazione considerata (quindi per esempio dalla nazionalità). Sappiamo inoltre che $\mathbb{P}(T|M) = 0.999$ e $\mathbb{P}(\bar{T}|\bar{M}) = 0.975$, perché queste sono le caratteristiche operative (*sensitività* e *specificità* rispettivamente) del test diagnostico impiegato per lo screening. Assegneremo quindi:

$$\begin{aligned}
\mathbb{P}(M \cap T) = \mathbb{P}(T|M)\mathbb{P}(M) = 0.999 \times 0.02 &= 0.01998, \\
\mathbb{P}(M \cap \bar{T}) = \{1 - \mathbb{P}(T|M)\}\mathbb{P}(M) = 0.001 \times 0.02 &= 0.00002, \\
\mathbb{P}(\bar{M} \cap T) = \{1 - \mathbb{P}(\bar{T}|\bar{M})\}\{1 - \mathbb{P}(M)\} = 0.025 \times 0.98 &= 0.02450, \\
\mathbb{P}(\bar{M} \cap \bar{T}) = \mathbb{P}(\bar{T}|\bar{M})\{1 - \mathbb{P}(M)\} = 0.975 \times 0.98 &= 0.95550,
\end{aligned}$$

dove si verifica subito che i quattro valori sommano a uno.

Avremo allora $\mathbb{P}(M|T) = \mathbb{P}(M \cap T)/\mathbb{P}(T)$, per definizione, dove $\mathbb{P}(T) = \mathbb{P}(M \cap T) + \mathbb{P}(\bar{M} \cap T) = 0.01998 + 0.02450 = 0.04448$. Si noti che $\mathbb{P}(T) > 0$, come richiesto dalla definizione e dal buon senso; altrimenti la campagna di screening sarebbe indifendibile. Troveremo quindi $\mathbb{P}(M|T) = 1998/4448 \approx 0.45$ e sarà ancora più probabile che il nostro amico sia in buona salute (piuttosto che malato). Anche se la *verosimiglianza* $\mathbb{P}(T|M)$ dell'ipotesi $M$, come spiegazione di $T$, è quasi quaranta volte la verosimiglianza $\mathbb{P}(T|\bar{M})$ dell'ipotesi $\bar{M}$, la *probabilità iniziale* $\mathbb{P}(\bar{M})$ di $\bar{M}$ è quarantanove volte la probabilità iniziale $\mathbb{P}(M)$ di $M$, per cui la *probabilità finale* $\mathbb{P}(M|T)$ di $M$, osservando $T$, risulta più piccola della probabilità finale $\mathbb{P}(\bar{M}|T)$ di $\bar{M}$:

$$\frac{\mathbb{P}(M|T)}{\mathbb{P}(\bar{M}|T)} = \frac{\mathbb{P}(T|M)}{\mathbb{P}(T|\bar{M})} \times \frac{\mathbb{P}(M)}{\mathbb{P}(\bar{M})}, \tag{1.4.1}$$

avendo semplificato il fattore comune $\mathbb{P}(T)$. Sulla scala dei *pronostici* (odds) la valutazione *a posteriori* si trova dunque moltiplicando quella *a priori* per il *rapporto di verosimiglianza*. Il passaggio da un pronostico 98 a 2 contro la malattia a un pronostico 55 a 45, sempre contro la malattia, fa in pratica una bella differenza e conferma le buone caratteristiche operative dello screening (oltre che naturalmente suggerire opportuni approfondimenti clinici).

Il procedimento seguito nell'analisi dell'Esempio 7 si può cristallizzare in una singola formula che esprime la probabilità finale di un'ipotesi $H$, quando si osservi il dato $D$, in funzione della sua probabilità iniziale, della sua verosimiglianza e della verosimiglianza dell'ipotesi contraria: $\mathbb{P}(H|D) = \mathbb{P}(D|H)\mathbb{P}(H)/[\mathbb{P}(D|H)\mathbb{P}(H) + \mathbb{P}(D|\bar{H})\{1 - \mathbb{P}(H)\}]$. Si tratta della *formula di Bayes* per due ipotesi ($H$ e $\bar{H}$) che possiamo generalizzare come segue.

Se $H_1, \ldots, H_m$ sono $m$ ipotesi disgiunte ed esaustive ($H_1 \uplus \cdots \uplus H_m = \Omega$) scriveremo $\mathbb{P}(H_i) = \mathbb{P}(D|H_i)\mathbb{P}(H_i)/\mathbb{P}(D)$, $i = 1, \ldots, m$, quindi calcoleremo la probabilità non condizionale del dato come $\mathbb{P}(D) = \sum_{i=1}^{n} \mathbb{P}(D|H_i)\mathbb{P}(H_i)$. Questa formula, detta *regola delle probabilità totali*, è interessante di per sé, perché aiuta ogni qual volta sia più facile valutare la probabilità di un evento condizionando ad altri eventi (di probabilità nota). Nel caso specifico ci dà, come abbiamo visto, la costante di normalizzazione del risultato seguente.

**Teorema 1.4.1 (Bayes)** *Sia $(\Omega, \mathcal{A}, \mathbb{P})$ uno spazio di probabilità. Se $D \in \mathcal{A}$ e $H_1, \ldots, H_m$ sono eventi di $\mathcal{A}$ che formano una partizione di $\Omega$, allora*

$$\mathbb{P}(H_i|D) \propto \mathbb{P}(D|H_i)\mathbb{P}(H_i), \quad i = 1, \ldots, m, \tag{1.4.2}$$

*dove la costante di normalizzazione vale $\mathbb{P}(D) = \sum_{i=1}^{m} \mathbb{P}(H_i)\mathbb{P}(D|H_i)$.*

Da un punto di vista costruttivo, per applicare il Teorema 1.4.1, possiamo prendere $\mathcal{A}$ generata (su un qualche $\Omega$) da $H_1, \ldots, H_m, D$ e usare il prodotto al secondo membro della (1.4.2) per assegnare $\mathbb{P}$ sui loro $2m$ costituenti non necessariamente vuoti; naturalmente $\mathbb{P}(\bar{D}|H_i) = 1 - \mathbb{P}(D|H_i)$, $i = 1, \ldots, m$.

Nell'Esempio 8 possiamo formulare tre ipotesi sul dado lanciato dalla nostra amica: $H_1$ = "dado equilibrato", $H_2$ = "dado sbilanciato verso il 6" e $H_3$ = "dado sbilanciato verso l'1". Dal momento che la nostra amica ha scelto "a caso" quale dado lanciare, possiamo assegnare $\mathbb{P}(H_1) = \mathbb{P}(H_2) = \mathbb{P}(H_3) = 1/3$. Dopo di che, posto $S_1$ = "6 al primo lancio", assegneremo $\mathbb{P}(S_1|H_1) = 1/6 \approx 0.17$, $\mathbb{P}(S_1|H_2) = 0.64$ e $\mathbb{P}(S_1|H_3) = 0.04$, sulla base delle caratteristiche dichiarate dei dadi. Troveremo così

$$\begin{aligned}
\mathbb{P}(H_1|S_1) &\propto \frac{1}{6} \times \frac{1}{3} &&\propto \frac{1}{6} = \frac{25}{150}, \\
\mathbb{P}(H_2|S_1) &\propto \frac{64}{100} \times \frac{1}{3} = \frac{16}{25} \times \frac{1}{3} &&\propto \frac{16}{25} = \frac{96}{150}, \\
\mathbb{P}(H_3|S_1) &\propto \frac{4}{100} \times \frac{1}{3} = \frac{1}{25} \times \frac{1}{3} &&\propto \frac{1}{25} = \frac{6}{150},
\end{aligned}$$

quindi $\mathbb{P}(H_1|S_1) = 25/127 \approx 0.20$, $\mathbb{P}(H_2|S_1) = 96/127 \approx 0.76$ e $\mathbb{P}(H_3|S_1) = 6/127 \approx 0.05$ (con somma *circa* uno per via degli arrotondamenti). Si noti che, a differenza di quanto accadeva nel caso del test di screening, qui le probabilità finali sono determinate solo dalle verosimiglianze, perché le tre ipotesi sono a priori equiprobabili. Tuttavia $\mathbb{P}(S_1|H_i) \neq \mathbb{P}(H_i|S_1)$, $i = 1, 2, 3$, quindi resta ferma l'esigenza di non confondere le due quantità.

Poniamo adesso $S_2 =$ "6 al secondo lancio" e cerchiamo $\mathbb{P}(H_i|S_1 \cap S_2)$, $i = 1, 2, 3$. A tal fine calcoliamo $\mathbb{P}(S_1 \cap S_2|H_i)$, $i = 1, 2, 3$, con la regola del prodotto per $\mathbb{P}_{H_i}$:

$$\begin{aligned}
\mathbb{P}(S_1 \cap S_2|H_1) = \mathbb{P}(S_2|S_1 \cap H_1)\mathbb{P}(S_1|H_1) &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}, \\
\mathbb{P}(S_1 \cap S_2|H_2) = \mathbb{P}(S_2|S_1 \cap H_2)\mathbb{P}(S_1|H_2) &= \frac{16}{25} \times \frac{16}{25} = \frac{256}{625}, \\
\mathbb{P}(S_1 \cap S_2|H_3) = \mathbb{P}(S_2|S_1 \cap H_3)\mathbb{P}(S_1|H_3) &= \frac{1}{25} \times \frac{1}{25} = \frac{1}{625},
\end{aligned}$$

valutando che, quando siano note le caratteristiche del dado, l'esito del primo lancio non fornisca informazioni sul secondo. Si noti che

$$\mathbb{P}(E|F \cap G) = \frac{\mathbb{P}(E \cap F \cap G)}{\mathbb{P}(F \cap G)} = \frac{\mathbb{P}(E \cap F|G)}{\mathbb{P}(F|G)} = \mathbb{P}_G(E|F), \tag{1.4.3}$$

comunque si prendano $E$, $F$ e $G$, con $\mathbb{P}(F \cap G) > 0$, quindi $\mathbb{P}(G) > 0$, nella tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$.

Tenendo presente che le tre ipotesi $H_1$, $H_2$ e $H_3$ sono equiprobabili, il Teorema 1.4.1 ci fornisce

$$\begin{aligned}
\mathbb{P}(H_1|S_1 \cap S_2) &\propto \mathbb{P}(S_1 \cap S_2|H_1) = \frac{1}{36} \propto 625, \\
\mathbb{P}(H_2|S_1 \cap S_2) &\propto \mathbb{P}(S_1 \cap S_2|H_2) = \frac{256}{625} \propto 9216, \\
\mathbb{P}(H_3|S_1 \cap S_2) &\propto \mathbb{P}(S_1 \cap S_2|H_3) = \frac{1}{625} \propto 36,
\end{aligned}$$

quindi $\mathbb{P}(H_1|S_1 \cap S_2) = 625/9877 \approx 0.063$, $\mathbb{P}(H_2|S_1 \cap S_2) = 9216/9877 \approx 0.933$ e $\mathbb{P}(H_3|S_1 \cap S_2) = 36/9877 \approx 0.004$. Il secondo lancio fornisce ulteriore evidenza in favore del dado sbilanciato verso il 6 e contro le altre due ipotesi (in particolar modo contro il dado sbilanciato verso l'1). L'evidenza fornita dai dati $S_1$ ed $S_2$, osservati in successione nelle stesse condizioni, si accumula e ci porta a propendere nettamente per l'ipotesi $H_2$.

La costruzione di $(\Omega, \mathcal{A}, \mathbb{P})$ sottintesa ai calcoli con cui abbiamo risposto al secondo quesito dell'Esempio 8 è basata sulla generazione di $\mathcal{A}$ da parte di $H_1$, $H_2$, $H_3$, $S_1$ ed $S_2$. Per completare l'assegnazione ricorsiva di $\mathbb{P}$ sui $3 \times 2 \times 2 = 12$ costituenti non vuoti di questi eventi dovremmo assegnare le quantità $\mathbb{P}(S_2|\bar{S}_1 \cap H_i)$, $i = 1, 2, 3$, ma questo non è essenziale per applicare il Teorema 1.4.1, perché $\bar{S}_1$ non è stato osservato. Sarebbe comunque quasi inevitabile porre $\mathbb{P}(S_2|\bar{S}_1 \cap H_i) = \mathbb{P}(S_1|H_i)$, $i = 1, 2, 3$, avendo già valutato che, note le caratteristiche del dado, l'esito del primo lancio non informa sul secondo. Avremmo così $\mathbb{P}(S_2|S_1) = \sum_{i=1}^{3} \mathbb{P}(S_1 \cap S_2|H_i) / \sum_{i=1}^{3} \mathbb{P}(S_1|H_i) = 0.52 > 0.28 = (1/3) \sum_{i=1}^{3} \{\mathbb{P}(S_1 \cap S_2|H_i) + \mathbb{P}(\bar{S}_1 \cap S_2|H_i)\} = \mathbb{P}(S_2)$ che ci illustra come invece il verificarsi di $S_1$ sia informativo sul secondo lancio quando le caratteristiche del dado non siano note.

## 1.5 Indipendenza

Due eventi, definiti in un qualche spazio di probabilità, si dicono *logicamente indipendenti* quando i loro quattro costituenti siano tutti possibili (non vuoti); altrimenti si dicono *logicamente dipendenti*. Per esempio, come abbiamo avuto modo di verificare, sono logicamente indipendenti gli eventi $B = \{2, 4, 6\}$ e $D = \{5, 6\}$, riferiti al lancio di un dado a sei facce, mentre sono logicamente dipendenti gli eventi $R = \{65, 66, \dots\}$ e $O = \{80, 81, \dots\}$, riferiti all'osservazione della durata di una vita umana. Nel caso specifico $O$ implica $R$, equivalentemente $\bar{R}$ implica $\bar{O}$, ma in generale la dipendenza logica fra due eventi può anche renderli incompatibli o esaustivi; per esempio $O$ ed $\bar{R}$ sono incompatibili, mentre $\bar{O}$ ed $R$ sono esaustivi. La relazione di indipendenza logica fra coppie di eventi di una tribù $\mathcal{A}$, oltre che essere simmetrica e invariante per negazione, prescinde dall'effettiva assegnazione di una probabilità su $\mathcal{A}$; il non portare informazioni deterministiche sul verificarsi o meno di un altro evento è una proprietà dei soli eventi.

Dati due eventi $E$ ed $F$, entrambi appartenenti alla tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, con $\mathbb{P}(F) > 0$, diremo che $E$ è *stocasticamente indipendente* da $F$ quando $\mathbb{P}(E|F) = \mathbb{P}(E)$. Se invece $\mathbb{P}(F) = 0$, converremo senz'altro che $E$ sia stocasticamente indipendente da $F$, visto che in questo caso è sempre coerente prendere $\mathbb{P}(E|F) = \mathbb{P}(E)$; la regola del prodotto sarà infatti banalmente soddisfatta, perché $E \cap F \subseteq F$ e quindi $\mathbb{P}(E \cap F) = 0$. In virtù di questa convenzione, l'indipendenza di $E$ da $F$ risulta equivalente alla *regola del prodotto per eventi indipendenti*: $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$. Si noti la mancata qualificazione dell'indipendenza come stocastica, qui e spesso nel seguito, per snellire il riferimento a un concetto fondamentale; resta invece inteso che l'indipendenza logica sia sempre da qualificare come tale.

Anche se abbiamo definito la relazione di indipendenza stocastica fra coppie di eventi in $\mathcal{A}$ asimmetricamente rispetto agli eventi della coppia, mediante la condizione che il verificarsi dell'uno non porti informazioni probabilistiche sul verificarsi dell'altro, l'equivalenza della condizione definitoria con la regola del prodotto per eventi indipendenti mostra che si tratta in realtà di una relazione simmetrica. Come nel caso dell'indipendenza logica, si tratta inoltre di una relazione invariante per negazione:

$$\mathbb{P}(E \cap \bar{F}) = \mathbb{P}(E) - \mathbb{P}(E \cap F) = \mathbb{P}(E)\{1 - \mathbb{P}(F)\} = \mathbb{P}(E)\mathbb{P}(\bar{F}),$$

se $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$. Diremo allora che due eventi sono indipendenti, sotto la probabilità $\mathbb{P}$, quando $\mathbb{P}$ si fattorizzi sulla classe dei loro costituenti; in simboli $E_1 \perp\!\!\!\perp E_2 \Leftrightarrow \mathbb{P}(E_1^{\alpha_1} \cap E_2^{\alpha_2}) = \mathbb{P}(E_1^{\alpha_1})\mathbb{P}(E_2^{\alpha_2})$, $\alpha \in \{0, 1\}^2$, dove abbiamo introdotto la notazione $\perp\!\!\!\perp$ per l'indipendenza. A rigore dovremmo scrivere $\perp\!\!\!\perp_{\mathbb{P}}$, perché l'indipendenza stocastica (a differenza di quella logica) non può prescindere dalla probabilità assegnata su $\mathcal{A}$, ma in pratica questa precisazione non sarà necessaria ogni qual volta $\mathbb{P}$ sia chiara dal contesto

Nell'Esempio 9 possiamo prendere come spazio di probabilità quello discreto uniforme su $\Omega = \{1, 2, \ldots, 51, 52\}$, con la convenzione che $\omega \bmod 13$ dia il valore della carta ($1 =$ asso, $2 =$ due, $\ldots$, $10 =$ dieci, $11 =$ fante, $12 =$ donna, $13 =$ re) e $\omega - 1 \operatorname{div} 13$ ne dia il seme ($0 =$ cuori, $1 =$ quadri, $2 =$ fiori, $3 =$ picche). Siamo interessati agli eventi $C = \{1, 2, \ldots, 12, 13\}$ e $D = \{12, 25, 38, 51\}$, per i quali troviamo $\mathbb{P}(C) = 13/52 = 1/4$ e $\mathbb{P}(D) = 4/52 = 1/13$. Ci interessa inoltre l'evento $C \cap D = \{12\}$, per il quale troviamo $\mathbb{P}(C \cap D) = 1/52 = 1/4 \times 1/13$. Resta dunque verificato che $C$ e $D$ sono stocasticamente indipendenti sotto $\mathbb{P}$ discreta uniforme su $\Omega$ ($C \perp\!\!\!\perp_{\mathbb{P}} D$). Dopo di che possiamo trovare la probabilità di $C \cup D$ mediante la *regola della somma per eventi indipendenti*: $\mathbb{P}(C \cup D) = \mathbb{P}(C) + \mathbb{P}(D) - \mathbb{P}(C)\mathbb{P}(D) = 1/4 + 1/13 - 1/4 \times 1/13 = 16/52 = 4/13$. Infatti $|C \cup D| = 16$, visto che $C \cup D = \{1, \ldots, 13, 25, 38, 51\}$. Se invece sappiamo che la donna di cuori manca dal mazzo, posto $\mathbb{P}_\circ = \mathbb{P}_{\overline{C\cap D}}$, avremo $\mathbb{P}_\circ(C) = 12/51$ e $\mathbb{P}_\circ(D) = 3/51$, mentre $\mathbb{P}_\circ(C \cap D) = 0 \neq 12/51 \times 3/51$. Gli eventi $C$ e $D$ saranno dunque *stocasticamente dipendenti* sotto $\mathbb{P}_\circ$, discreta uniforme su $\Omega \setminus (C \cap D)$, e scriveremo $C \not\perp\!\!\!\perp_{\mathbb{P}_\circ} D$. Infine, potremo calcolare la probabilità di $C \cup D$ dato $\overline{C \cap D}$ con la *regola della somma per eventi quasi incompatibli*: $\mathbb{P}_\circ(C \cup D) = \mathbb{P}_\circ(C) + \mathbb{P}_\circ(D) = 12/51 + 3/51 = 15/51 = 5/17$, dato che $\mathbb{P}_\circ(C \cap D) = 0$. In effetti $(C \cup D) \setminus (C \cap D) = \{1, \ldots, 11, 13, 25, 38, 51\}$ è formato da 15 elementi.

Nell'Esempio 10 ci interessano gli eventi $F_1 =$ "primogenito femmmina" ed $F_2 =$ "secondogenito femmina", definiti per esempio in un quadrato $\Omega$ del piano euclideo. Per assegnare $\mathbb{P}$ sui loro costituenti (quindi sulla loro $\mathcal{U}$) possiamo porre $\mathbb{P}(F_1) = \mathbb{P}(F_2|F_1) = \mathbb{P}(F_2|\bar{F}_1) = 1/2$, valutando che le caratteristiche dello spermatozoo nella prima fecondazione non ci informino su quelle dello spermatozoo nella seconda fecondazione e che metà degli spermatozoi porti il cromosoma Y, dando luogo a figli maschi, mentre l'altra metà porti il cromosoma X, dando luogo a figlie femmine. Si tratta di una valutazione standard, in prima approssimazione, trascurando il problema dei gemelli ed escludendo pianificazioni selettive delle nascite.

Troveremo così $\mathbb{P}(F_1 \cap F_2) = \mathbb{P}(F_1)\mathbb{P}(F_2)$, visto che per la regola delle probabilità totali $\mathbb{P}(F_2)$ deve necessariamente coincidere col comune valore di $\mathbb{P}(F_2|F_1)$ e $\mathbb{P}(F_2|\bar{F}_1)$; nello specifico avremo $\mathbb{P}(F_2) = 1/2$ e $\mathbb{P}(F_1 \cap F_2) = 1/4$. Come $C$ e $D$ nell'Esempio 9, gli eventi $F_1$ ed $F_2$ sono indipendenti. C'è però una differenza: qui l'indipendenza è stata *imposta*, in base a considerazioni genetiche, mentre prima era stata *scoperta*, dopo un'assegnazione classica. In generale, dati due eventi logicamente indipendenti, basterà assegnare le loro probabilità *marginali* (non condizionali) per determinare una probabilità sulla loro $\mathcal{U}$ sotto la quale siano stocasticamente indipendenti. Se invece abbiamo due eventi logicamente dipendenti, l'unica possibilità che siano stocasticamente indipendenti è che almeno uno dei due sia quasi certo o quasi impossibile; in questo caso la regola del prodotto per eventi indipendenti varrà banalmente, nonostante il legame logico tra i due eventi.

Adesso supponiamo di sapere che almeno uno dei figli è femmina. Questa informazione corrisponde all'evento $F_1 \cup F_2$ e verifichiamo subito che

$$\mathbb{P}(F_1 \cap F_2 | F_1 \cup F_2) = \frac{\mathbb{P}(F_1 \cap F_2)}{\mathbb{P}(F_1 \cup F_2)} = \frac{1/4}{1 - 1/4} = \frac{1}{3}.$$

Quindi $\mathbb{P}(F_1 \cap F_2 | F_1 \cup F_2) \neq \mathbb{P}(F_1 \cap F_2 | F_1) = \mathbb{P}(F_2 | F_1) = 1/2 = \mathbb{P}(F_1 | F_2)$. Sapere che almeno un figlio è femmina non è la stessa cosa che sapere che il figlio maggiore/minore è femmina.

Il risultato ottenuto è ineccepibile, da un punto di vista matematico, ma lascia un po' perplessi, da un punto di vista pratico. Proviamo allora a chiederci come facciamo a sapere che almeno uno dei figli è femmina e immaginiamo che la risposta sia di avere incontrato la mamma con la figlia in un negozio. Introdotto l'evento $F_3 =$ "figlio incontrato femmina", avremo necessariamente $\mathbb{P}(F_3 | F_1 \cap F_2) = 1$ e $\mathbb{P}(F_3 | \bar{F}_1 \cap \bar{F}_2) = 0$, mentre possiamo supporre $\mathbb{P}(F_3 | F_1 \cap \bar{F}_2) = \mathbb{P}(F_3 | \bar{F}_1 \cap F_2) = p \in [0, 1]$, se i due figli hanno età comparabili. La formula di Bayes ci fornisce allora

$$\mathbb{P}(F_1 \cap F_2 | F_3) = \frac{1 \times 1/4}{1 \times 1/4 + p \times 2/4} = \frac{1}{1 + 2p},$$

pari a 1, se $p = 0$, pari a $1/2$, se $p = 1/2$, pari a $1/3$, se $p = 1$.

La probabilità cercata può dunque valere da un minimo di $1/3$ a un massimo di 1 a seconda del tipo di negozio: un valore $p \approx 0$ potrebbe andare bene per un negozio specializzato in automobiline e trenini, un valore $p \approx 1$ potrebbe andare bene per un negozio specializzato in bambole, mentre il valore $p = 1/2$ potrebbe andare bene per un supermercato. Il *report* Istat su "Infanzia e vita quotidiana" del 18 novembre 2011 fornisce informazioni statistiche sui giochi preferiti da bambini e bambine di età compresa tra 3 e 10 anni (distinguendo tra età prescolare ed età scolare). Il valore $1/3$ fornito da $\mathbb{P}(F_1 \cap F_2 | F_1 \cup F_2)$ risulta quindi estremo rispetto allo scenario descritto.

Siamo ora pronti ad affrontare il caso generale di un numero finito di eventi, ritrovando come caso particolare quello di due eventi (e come caso banale quello di un singolo evento).

**Definizione 1.5.1 (indipendenza logica)** *Una famiglia finita di eventi $E_1, \ldots, E_n$, parte della tribù $\mathcal{A}$ di uno spazio di probabilità con insieme degli esiti $\Omega$, si dice formata da eventi logicamente indipendenti se i $2^n$ costituenti di $E_1, \ldots, E_n$ sono tutti non vuoti.*

Un singolo evento è logicamente indipendente se e solo se è incerto (né certo né impossibile). Abbiamo già illustrato il caso di due eventi. L'Esempio 11 ci permette di illustrare il caso di tre eventi (con $2^3 = 8$ costituenti). Posto $E_1 =$ "vittoria esterna" ed $E_2 =$ "primo tempo 1-1", se il terzo evento è $E_3 =$ "1 gol in tutto", i due costituenti $E_1 \cap E_2 \cap E_3$ ed $\bar{E}_1 \cap E_2 \cap E_3$ saranno vuoti e i tre eventi $E_1, E_2, E_3$ logicamente dipendenti. Se invece

$E_3$ = "2 gol in tutto", il solo costituente $E_1 \cap E_2 \cap E_3$ sarà vuoto e i tre eventi risulteranno ancora logicamente dipendenti; in questo caso saranno però logicamente indipendenti a coppie, mentre prima $E_2$ ed $E_3$ erano logicamente dipendenti. Infine, se $E_3$ = "3 gol in tutto", nessun costituente sarà vuoto e i tre eventi risulteranno logicamente indipendenti (necessariamente anche a coppie). Il caso di quattro o più eventi è del tutto analogo.

Dal punto di vista stocastico, un numero finito di eventi sono indipendenti sotto le probabilità che si fattorizzano sui loro costituenti.

**Definizione 1.5.2 (indipendenza stocastica)** *Una famiglia finita di eventi $E_1, \ldots, E_n$, parte della tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, si dice formata da eventi stocasticamente indipendenti, sotto $\mathbb{P}$, se*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} E_i^{\alpha_i}\right) = \prod_{i=1}^{n} \mathbb{P}(E_i^{\alpha_i}), \quad \alpha \in \{0,1\}^n. \tag{1.5.1}$$

Un singolo evento è indipendente sotto qualsiasi probabilità, visto che la fattorizzazione (1.5.1) vale in modo banale quando $n = 1$. Questo caso non va confuso con quello dello stesso evento considerato due volte: se $E_1 = E_2$, la regola del prodotto per eventi indipendenti si scrive $\mathbb{P}(E_1) = \mathbb{P}(E_1)\mathbb{P}(E_1)$ ed è quindi soddisfatta se e solo se $\mathbb{P}(E_1) = 0$ o $\mathbb{P}(E_1) = 1$ (coerentemente con la dipendenza logica di $E_1$ ed $E_2 = E_1$). Il significato dell'indipendenza per due eventi è chiaro, alla luce di quanto abbiamo visto prima, ma resta da estenderne l'interpretazione per coprire il caso di tre o più eventi.

Il seguente risultato, nel cui enunciato alleggeriamo delle graffe la notazione dell'algebra generata da un numero finito di eventi, la loro $\mathcal{U}$, caratterizza l'indipendenza stocastica di due o più eventi.

**Proposizione 1.5.3** *Una famiglia finita di eventi $E_1, \ldots, E_n$, con $n \geq 2$, parte della tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, è formata da eventi indipendenti, sotto $\mathbb{P}$, se e solo se, comunque si scelga $k \in \{1, \ldots, n-1\}$ e si partizionino $E_1, \ldots, E_n$ in $F_1, \ldots, F_k$ e $G_1, \ldots, G_{n-k}$, si ha $\mathbb{P}(F_\star \cap G_\star) = \mathbb{P}(F_\star)\mathbb{P}(G_\star)$, per ogni $F_\star \in \alpha(F_1, \ldots, F_k)$ e $G_\star \in \alpha(G_1, \ldots, G_{n-k})$.*

*Dimostrazione.* Se $E_1, \ldots, E_n$ sono indipendenti, posto $F_\star = \biguplus_{\phi \in \Phi} C'_\phi$ e $G_\star = \biguplus_{\gamma \in \Gamma} C''_\gamma$, dove $C'_\phi$ e $C''_\gamma$ sono costituenti di $F_1, \ldots, F_k$ e $G_1, \ldots, G_{n-k}$, rispettivamente, abbiamo $F_\star \cap G_\star = \biguplus_{(\phi,\gamma) \in \Phi \times \Gamma} C'_\phi \cap C''_\gamma$ e $\mathbb{P}(C'_\phi \cap C''_\gamma) = \mathbb{P}(C'_\phi)\mathbb{P}(C''_\gamma)$ per la (1.5.1), quindi $\mathbb{P}(F_\star \cap G_\star) = \sum_{(\phi,\gamma) \in \Phi \times \Gamma} \mathbb{P}(C'_\phi)\mathbb{P}(C''_\gamma) = \mathbb{P}(F_\star)\mathbb{P}(G_\star)$. Viceversa, possiamo ottenere la (1.5.1) per induzione finita su $k = 1, \ldots, n-1$, con $F_1 = E_1^{\alpha_1}, \ldots, F_k = E_k^{\alpha_k}$ e $G_1 = E_{k+1}^{\alpha_{k+1}}, \ldots, G_{n-k} = E_n^{\alpha_n}$, $F_\star = F_k$ e $G_\star = G_1 \cap \cdots \cap G_{n-k}$. □

Dunque, a parole, due o più eventi saranno stocasticamente indipendenti se e solo se, comunque si venga a sapere qualcosa sul verificarsi o meno di alcuni, non si riceveranno informazioni probabilistiche sul verificarsi o meno degli altri. Si noti che $\mathbb{P}(F_\star | G_\star) = \mathbb{P}(F_\star)$, nella Proposizione 1.5.3, se $\mathbb{P}(G_\star) > 0$.

Nell'Esempio 12 osserviamo sei lanci di un'ordinaria moneta da due euro. Se chiamiamo "testa" il lato con Dante Alighieri e "croce" l'altro, possiamo porre $\Omega = \{0,1\}^n$, dove $n = 6$ e per esempio $1 =$ "testa" (mentre $0 =$ "croce"). Come assegneremo $\mathbb{P}$ su $\mathcal{A} = \wp(\Omega)$? Ci interessano in particolare gli eventi $T_i =$ "testa all'$i$-esimo lancio" $= \{(\omega_1, \ldots, \omega_n) \in \Omega : \omega_i = 1\}$, $i = 1, \ldots, n$, i cui costituenti sono i singoletti di $\Omega$. Basterà allora supporre $T_1, \ldots, T_n$ indipendenti ed equiprobabili per ottenere ricorsivamente $\mathbb{P}\{\omega\} = p^{\omega_+}(1-p)^{n-\omega_+}$, $\omega \in \Omega$, dove $\omega = (\omega_1, \ldots, \omega_n)$ è il generico esito degli $n$ lanci, $\omega_+ = \omega_1 + \cdots + \omega_n$ il numero di teste osservato e $p \in [0,1]$ la probabilità di ottenere testa in ogni singolo lancio: $p = \mathbb{P}(T_i)$, $i = 1, \ldots, n$. Si noti che stiamo supponendo $p$ nota: per questo conoscere il risultato di alcuni lanci non ci informa sul risultato degli altri. Se la moneta è lanciata onestamente, sarà naturale porre $p = 1 - p = 1/2$. In questo modo troveremo $\mathbb{P}(\omega) = 1/2^n$, $\omega \in \Omega$, quindi $\mathbb{P}$ risulterà discreta uniforme su $\{0,1\}^n$. Altrimenti dovremo valutare $p$ in base alle modalità di lancio (più difficile). In alternativa possiamo vedere i lanci di una moneta con $p \neq 1/2$ nota come un'astrazione che generalizza le estrazioni *con reinserimento* da un'urna contenente una frazione $p$ di biglie con una caratteristica di interesse (ammettendo il caso di $p$ irrazionale).

Consideriamo ora gli eventi $S_k =$ "esattamente $k$ teste negli $n$ lanci" $= \{\omega \in \Omega : \omega_+ = k\}$, $k = 0, 1, \ldots, n$, i quali chiaramente costituiscono una partizione dell'evento certo. Ciascun $S_k$ è formato da $n!/\{k!(n-k!)\}$ esiti, corrispondenti ai modi di scegliere $k$ degli $n$ lanci per le teste (lasciando i restanti $n-k$ per le croci). Ognuno di questi esiti ha probabilità $p^k(1-p)^{n-k}$. Avremo quindi $\mathbb{P}(S_k) = [n!/\{k!(n-k!)\}]p^k(1-p)^{n-k}$, $k = 0, 1, \ldots, n$. Resta così definita una probabilità sull'algebra generata da $S_0, S_1, \ldots, S_n$, cui possiamo fare riferimento nel caso in cui non osserviamo (o non ci interessano) i risultati dei singoli lanci, ma osserviamo (o ci interessa) solo il numero complessivo di teste. In particolare, con riferimento al caso specifico $n = 6$, troveremo $\mathbb{P}(S_2) = \{6!/(2!4!)\}p^2(1-p)^4 = 15p^2(1-p)^4 \approx 0.23$, se $p = 1/2$.

Alla stessa $\mathbb{P}$ discreta uniforme su $\{0,1\}^n$ saremmo giunti, nel caso di lanci onesti, mediante una valutazione classica. La valutazione basata sull'indipendenza è tuttavia più stringente, oltre che valida anche nel caso di lanci disonesti, perché non è del tutto ovvio che $\mathbb{P}$ debba essere discreta uniforme su $\{0,1\}^n$. In effetti, essendo interessati al conteggio delle teste, avremmo potuto prendere $\Omega_\star = \{0, 1, \ldots, n\}$ e $\mathcal{A}_\star = \wp(\Omega_\star)$, quindi assegnare classicamente $\mathbb{P}_\star$ discreta uniforme su $\{1, \ldots, n\}$. Dopo di che, identificando $\mathcal{A}_\star$ con l'algebra su $\Omega$ generata da $S_0, S_1, \ldots S_n$ e giudicando gli $n!/\{k!(n-k!)\}$ esiti in $S_k$ equiprobabili, per simmetria, avremmo ricavato $\mathbb{P}_\star(\omega) = \{k!(n-k)!\}/(n+1)!$, $\omega \in \{0,1\}^n$, da $\mathbb{P}_\star\{S_k\} = 1/(n+1)$, $k = 0, 1, \ldots, n$. Chiaramente $\mathbb{P} \neq \mathbb{P}_\star$, se solo $n > 1$, perché $\mathbb{P}_\star(S_n) = 1/(n+1) \neq 1/2^n = \mathbb{P}(S_n)$.

Sia $\mathbb{P}$ che $\mathbb{P}_\star$ sono coerenti e assegnano probabilità $1/2$ agli eventi $T_i$, $i = 1, \ldots, n$. Infatti, laddove $\mathbb{P}(T_i) = 1/2$ per costruzione, semplici passaggi algebrici forniscono

$$\begin{aligned}\mathbb{P}_\star(T_i) &= \sum_{k=1}^n \mathbb{P}_\star(T_i \cap S_k) = \frac{1}{n+1}\sum_{k=1}^n \frac{\binom{n-1}{k-1}}{\binom{n}{k}},\\ &= \frac{1}{n+1}\sum_{k=1}^n \frac{k}{n} = \frac{1}{n(n+1)}\frac{n(n+1)}{2} = \frac{1}{2},\end{aligned}$$

dal momento che gli esiti in $S_k$ con $T_i = 1$ corrispondono ai modi di scegliere $k-1$ degli altri $n-1$ lanci per le restanti teste ($k \geq 1$). D'altra parte, mentre $T_1, \ldots, T_n$ sono indipendenti sotto $\mathbb{P}$, non lo sono sotto $\mathbb{P}_\star$, perché per quanto visto $\mathbb{P}_\star(T_1 \cap \cdots \cap T_n) = \mathbb{P}(S_n) \neq 1/2^n$ ($n > 1$). Dunque, nella misura in cui riteniamo che conoscere il risultato di alcuni lanci non ci dia informazioni sugli altri, opteremo senz'altro per $\mathbb{P}$ (scartando $\mathbb{P}_\star$).

Gli eventi $S_0, S_1, \ldots, S_n$ non sono equiprobabili sotto $\mathbb{P}$ (essendolo invece sotto $\mathbb{P}_\star$). Troviamo il numero più probabile di teste, nel caso specifico $n = 6$, calcolando $\mathbb{P}(S_0) = \mathbb{P}(S_6) = 1/2^6 \approx 0.016$, $\mathbb{P}(S_1) = \mathbb{P}(S_5) = 6/2^6 \approx 0.094$, $\mathbb{P}(S_2) = \mathbb{P}(S_4) = 15/2^6 \approx 0.234$ e $\mathbb{P}(S_3) = 20/2^6 \approx 0.312$: l'evento più probabile dei sette risulta essere $S_3$. Si noti che le sette probabilità sommano a uno, come garantito dall'assegnazione ricorsiva di $\mathbb{P}$. In generale, senza suppore $p = 1/2$, la formula per la potenza $n$-esima del binomio fornisce $\sum_{k=0}^n \mathbb{P}(S_k) = \sum_{k=0}^n [n!/\{k!(n-k!)\}]p^k(1-p)^{n-k} = \{p + (1-p)\}^n = 1$.

L'assegnazione di una $\mathbb{P}$ su $\alpha(E_1, \ldots, E_n)$ sotto la quale $E_1, \ldots, E_n$ siano indipendenti con date $\mathbb{P}(E_1), \ldots, \mathbb{P}(E_n)$, non necessariamente tutte uguali, è sempre possibile (procedendo in modo ricorsivo) quando $E_1, \ldots, E_n$ siano logicamente indipendenti. Se invece $E_1, \ldots, E_n$ sono logicamente dipendenti, il verificarsi o meno di alcuni di loro ci informerà deterministicamente sul verificarsi o meno degli altri e non sarà ragionevole richiedere che non lo faccia probabilisticamente. In effetti, questo potrà accadere solo nel caso in cui uno o più eventi siano quasi certi o quasi impossibili; per esempio nel caso banale in cui uno dei loro costituenti sia quasi certo. La (1.5.1) mostra infatti chiaramente come l'indipendenza stocastica implichi quella logica, se $0 < \mathbb{P}(E_i) < 1$, $i = 1, \ldots, n$. Se invece $\mathbb{P}(E_n) \in \{0, 1\}$, gli eventi $E_1, \ldots, E_n$ saranno indipendenti sotto $\mathbb{P}$ se e solo se lo sono gli eventi $E_1, \ldots, E_{n-1}$ ed eventualmente si potrà iterare la riduzione, fino a restare banalmente con un solo evento o ricadere nel caso precedente.

L'urna dell'Esempio 13 può essere modellata come segue: $\Omega = \{1, \ldots, 8\}$; $\mathcal{A} = \alpha(G, R, V)$, dove $G = \{1, 2\}$, $R = \{3, 4\}$ e $V = \{5, 6\}$, mentre gli esiti 7 e 8 rappresentano le biglie nere; $\mathbb{P}$ discreta uniforme su $\Omega$ (ristretta ad $\mathcal{A}$). Con questo modello troviamo $(G \uplus R) \cap (R \uplus V) \cap (G \uplus V) = R \cap (G \uplus V) = \emptyset$, per cui i tre eventi $G \uplus R, R \uplus V, G \uplus V$ risultano logicamente dipendenti. Dal momento che $\mathbb{P}(G \uplus R) = \mathbb{P}(R \uplus V) = \mathbb{P}(G \uplus V) = 4/8 = 1/2$, essi risultano anche stocasticamente dipendenti. Se invece li prendiamo a coppie, essi sono stocasticamente indipendenti: $\mathbb{P}(\{G \uplus R\} \cap \{R \uplus V\}) = \mathbb{P}(R) =$

$2/8 = 1/4$ e analogamente per le altre due coppie che possiamo considerare. Di conseguenza le tre coppie sono anche logicamente indipendenti, come pure si verifica subito dalla definizione. Abbiamo così un esempio di tre eventi indipendenti a coppie, ma non *mutuamente* indipendenti, come si dice quando si voglia sottolineare il fatto che l'indipendenza è una proprietà dell'intera famiglia di eventi e non dei singoli eventi che la formano.

Cosa succede se scopriamo che nell'urna ci sono due ulteriori biglie nere? Innanzi tutto dobbiamo ridefinire $\Omega = \{1, \ldots, 10\}$, perché si tratta di un eventualità che il nostro modello non contemplava. Poi possiamo procedere come prima, trovando in questo caso $\mathbb{P}_\star(\{G \uplus R\} \cap \{R \uplus V\}) = 2/10 \neq 16/100 = \mathbb{P}_\star(G \uplus R)\mathbb{P}_\star(R \uplus V)$, con $\mathbb{P}_\star$ discreta uniforme su $\{1, \ldots, 10\}$. Sotto $\mathbb{P}_\star$ i tre eventi $G \uplus R, R \uplus V, G \uplus V$ non sono più indipendenti a coppie; lo sono solo logicamente. Resta evidenziato come l'indipendenza stocastica sia una proprietà che sussiste o meno a seconda della probabilità assegnata.

Concludiamo osservando che, se gli eventi $E_1, \ldots, E_n$ sono indipendenti, comunque si scelgano $k \in \{1, \ldots, n\}$ e $k$ eventi $F_1, \ldots, F_k$ tra loro, si otterrà una famiglia di eventi indipendenti. Questo fatto (banale per $n = 1$) segue dalla Proposizione 1.5.3, perché una volta partizionati $F_1, \ldots, F_k$ in due sottofamiglie si possono sempre aggiungere gli eventi rimanenti a una di esse e invocare la fattorizzazione per $E_1, \ldots, E_n$. Non potremo quindi trovare tre eventi mutuamente indipendenti che non siano anche indipendenti a coppie. Infatti, troveremo sempre $\mathbb{P}(F_1 \cap \cdots \cap F_k) = \mathbb{P}(F_1) \times \cdots \times \mathbb{P}(F_k)$. Questa condizione, priva di negazioni, estesa a tutte le possibili sottofamiglie (non vuote) di $E_1, \ldots, E_n$, equivale di fatto alla (1.5.1).

**Proposizione 1.5.4** *Una famiglia finita di eventi $E_1, \ldots, E_n$, con $n \geq 2$, parte della tribù $\mathcal{A}$ di uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbb{P})$, è formata da eventi indipendenti, sotto $\mathbb{P}$, se e solo se, comunque si scelgano $k \in \{1, \ldots, n\}$ e $k$ eventi $F_1, \ldots, F_k$ della famiglia, si ha $\mathbb{P}(F_1 \cap \cdots \cap F_k) = \mathbb{P}(F_1) \times \cdots \times \mathbb{P}(F_k)$.*

*Dimostrazione.* Il "solo se" discende dall'indipendenza di $F_1, \ldots, F_k$ che come abbiamo visto è implicata dall'indipendenza di $E_1, \ldots, E_n$. Abbiamo già verificato che il "se" vale nel caso $n = 2$; completiamone la dimostrazione per induzione finita su $n$. La (1.5.1) vale per ipotesi in assenza di negazioni e, se vale con $h \in \{0, 1, \ldots, n\}$ negazioni, vale anche con $h + 1$ negazioni:

$$\mathbb{P}\left(\bar{E}_1 \cap \bigcap_{i=2}^{n} E_i^{\alpha_i}\right) = \mathbb{P}\left(\bigcap_{i=2}^{n} E_i^{\alpha_i}\right) - \mathbb{P}\left(E_1 \cap \bigcap_{i=2}^{n} E_i^{\alpha_i}\right),$$

avendo per fissare le idee supposto $\alpha_1 = 0$; il minuendo si fattorizza in conseguenza dell'ipotesi induttiva, il sottraendo perché c'è una negazione in meno, quindi si può raccogliere $1 - \mathbb{P}(E_1) = \mathbb{P}(\bar{E}_1)$ e fattorizzare la differenza. Si vede allora che la (1.5.1) vale con qualsiasi numero di negazioni. □

# Capitolo 2

# Probabilità avanzata

**Esempio 14 (Uno strano soggetto)** *Uno strano soggetto, con indosso una tunica di colori cangianti, accetta scommesse sulla temperatura che farà domani mattina alle nove (in gradi Celsius). Gli parliamo e ci spiega come consideri tutti e soli i punti della retta reale quali possibili esiti di questo esperimento aleatorio e accetti scommesse sull'eventualità che l'esito osservato sia minore di una soglia scelta dallo scommettitore. La quota offerta è pari a zero se la soglia scelta è minore di* 4 *(gradi Celsius) e pari a uno altrimenti. Siamo disposti a metterci nei panni di questo strano soggetto?*

**Esempio 15 (Un postino sorprendente)** *Aspettiamo un pacco per domani mattina, ma non più tardi delle undici e mezza dovremo uscire per prendere un treno; saremo di ritorno il giorno successivo. Il postino che serve la nostra zona è molto ligio al proprio dovere, perciò siamo quasi certi che tra le otto e mezzogiorno suonerà il nostro campanello, ma è anche molto indisciplinato, nel senso che ogni volta fa un giro di consegne diverso, perciò nel prevedere il suo orario di arrivo riteniamo che due qualsiasi intervalli di tempo della stessa durata siano ugualmente probabili. Quanto valutiamo la probabilità di riuscire a portare con noi, in viaggio, il contenuto del pacco?*

*Questa pagina è stata lasciata intenzionalmente in bianco.*

## 2.1 Additività numerabile

Lo strano soggetto dell'Esempio 14 prende $\Omega = \mathbb{R}$ e accetta scommesse sugli eventi della classe $\mathcal{I} = \{]a, b] \,|\, (a, b) \in \ddot{\mathbb{R}}^2\}$, dove $\ddot{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ e $]a, b] = \{\omega \in \mathbb{R} : a < \omega \leq b\}$. Se $a = -\infty$ e $b = +\infty$, allora $]a, b] = ]-\infty, +\infty]$ è l'evento certo. Se invece $-\infty = a < b < +\infty$, allora $]a, b] = ]-\infty, b]$ è una semiretta sinistra chiusa. Analogamente, se $-\infty < a < b = +\infty$, l'evento $]a, b] = ]a, +\infty]$ è una semiretta destra aperta. Se poi $-\infty < a < b < +\infty$, l'evento $]a, b]$ è un intervallo limitato aperto a sinistra e chiuso a destra. Infine, se $a = b$, allora $]a, b] = ]a, a]$ è l'evento impossibile; analogamente $]a, b] = \emptyset$, se $a > b$, cosicché non lede la generalità supporre $a \leq b$. In realtà lo strano soggetto dichiara di accettare scommesse solo sulle semirette sinistre chiuse $]-\infty, b]$, $b \in \mathbb{R}$, ma abbiamo visto che si può sempre immaginare di scommettere su $\Omega = ]-\infty, +\infty]$ con quota unitaria, mentre $]a, b] = ]-\infty, b] \setminus ]-\infty, a]$ per ogni $a \leq b$, quindi di fatto lo strano soggetto accetta scommesse su tutti gli eventi di $\mathcal{I}$. In generale, quando l'evento certo sia $\mathbb{R}$, supporremo che $\mathcal{I} \subset \mathcal{A}$, dove l'inclusione sarà stretta perché la classe $\mathcal{I}$ non è stabile per complementazione (pur essendolo per intersezioni finite).

Per assegnare una $\mathbb{P}$ su $\mathcal{I}$ in modo che sia additiva, isotona e normalizzata è sufficiente, oltre che necessario, specificare la funzione $F(b) = \mathbb{P}(]-\infty, b])$, $b \in \mathbb{R}$, in modo che la sua estensione a $\ddot{\mathbb{R}}$ con la convenzione $F(-\infty) = 0$ e $F(+\infty) = 1$ sia crescente. Infatti, data $F : \mathbb{R} \to [0, 1]$ con la proprietà di essere crescente, porremo $\mathbb{P}(]a, b]) = F(b) - F(a)$ per ogni $]a, b] \in \mathcal{I}$, in modo che $\mathbb{P}(]-\infty, b] \setminus ]-\infty, a]) = \mathbb{P}(]-\infty, b]) - \mathbb{P}(]-\infty, a])$. Scriveremo $\mathbb{P} = \Delta F$ per denotare questa assegnazione. Si verifica subito che $\Delta F$ è normalizzata, in virtù della convenzione che estende $F$ a $\ddot{\mathbb{R}}$, e isotona, in quanto l'estensione di $F$ a $\ddot{\mathbb{R}}$ è crescente. Inoltre $\Delta F$ è additiva:

$$\mathbb{P}(]a, b]) = F(b) - F(a) = \sum_{i=1}^{n} \{F(b_i) - F(a_i)\} = \sum_{i=1}^{n} \mathbb{P}(]a_i, b_i]),$$

se $-\infty \leq a = a_1 < b_1 = a_2 < \cdots < b_{n-1} = a_n < b_n = b \leq +\infty$, come si può supporre nel caso in cui $]a, b] = ]a_1, b_1] \uplus \cdots \uplus ]a_n, b_n]$, eventualmente riordinando gli intervalli a destra del segno uguale e trascurando quelli vuoti, per i quali si verifica subito che $\Delta F$ è nulla.

Lo strano soggetto banalmente pone $F(b) = 0$, se $b \leq 4$, $F(b) = 1$, se $b > 4$. Egli assegna cioè $F(b) = \mathbb{I}_{]4,+\infty]}(b)$, $b \in \mathbb{R}$, dove $\mathbb{I}_{]4,+\infty]}$ è la *funzione indicatrice* dell'evento $]4, +\infty]$; in generale $\mathbb{I}_E(\omega) = 1$, se $\omega \in E$, $\mathbb{I}_E(\omega) = 0$, se $\omega \notin E$. Ne segue la valutazione $\mathbb{P}(]a, b]) = \mathbb{I}_{[a,b[}(4)$, $]a, b] \in \mathcal{I}$, per $\mathbb{P} = \Delta F$, dove $[a, b[ = \{\omega \in \mathbb{R} : a \leq \omega < b\}$. La $\mathbb{P}$ dello strano soggetto assume quindi valori in $\{0, 1\}$: si potrà scommettere o gratis o pagando/incassando l'intera posta in palio (a seconda dell'evento scelto). In effetti, comunque presi $a$ e $b$ in $\mathbb{R}$, lo strano soggetto deve essere *quasi certo* che: i) $\omega \in ]a, b]$, se $a \leq 4 < b$; ii) $\omega \notin ]a, b]$, altrimenti. Per esempio, presi $a = -\infty$ e $b = 4$, egli esclude che domani mattina la temperatura possa essere minore di 4 gradi Celsius.

In mancanza di informazioni molto dettagliate sul fenomeno meteorologico oggetto di previsione, tanta certezza basta a non volere vestire i suoi panni. Ugualmente, proviamo a chiarire di che certezza si tratta. Presi $a = -\infty$ e $b = 4+1/n$, $n \in \mathbb{N}$, troviamo $\mathbb{P}(]-\infty, 4+1/n]) = 1$: lo strano soggetto esclude qualsiasi valore strettamente maggiore di 4 (per un'opportuna scelta di $n$). Essendo i valori minori di 4 già tutti esclusi, quanto varrà la temperatura?

In condizioni di certezza avremmo una contraddizione. Forse lo strano soggetto è incoerente? La sua $\mathbb{P}$ è additiva, isotona e normalizzata, ma non è definita su un'algebra. La classe di eventi $\mathcal{I}$, su cui $\mathbb{P}$ è definita, ha però le proprietà "giuste" perché si possa estendere $\mathbb{P}$ ad $\alpha(\mathcal{I})$ preservandone additività, isotonia e normalizzazione.

**Definizione 2.1.1 (semialgebra)** *Una classe $\mathcal{S}$ di eventi, definita su un insieme di esiti $\Omega$, forma una semialgebra, su $\Omega$, se:*

*i)* $\Omega \in \mathcal{S}$*;*

*ii)* $\bar{E} = \biguplus_{i=1}^{n} E_i$*, con* $E_1, \ldots, E_n$ *in* $\mathcal{S}$*, per ogni* $E \in \mathcal{S}$*;*

*iii)* $E \cap F \in \mathcal{S}$ *ogni qual volta* $E$ *ed* $F$ *appartengano a* $\mathcal{S}$*.*

Ogni algebra è una semialgebra, evidentemente, ma rispetto alla definizione di algebra la condizione ii) è meno restrittiva: non è necessario che $\bar{E}$ appartenga a $\mathcal{S}$, quando $E \in \mathcal{S}$, ma basta che si possa scrivere come unione di un numero finito di elementi disgiunti di $\mathcal{S}$. Si noti che la ii), assieme alla i), implica $\emptyset \in \mathcal{S}$. La classe $\mathcal{I}$ degli intervalli aperti a sinistra e chiusi a destra è una semialgebra su $\mathbb{R}$, perché $\overline{]a,b]} = ]-\infty, a] \cup ]b, +\infty]$ nel caso in cui $-\infty < a < b < +\infty$ e $\overline{]a,b]} \in \mathcal{I}$ negli altri casi. Anche la classe $\mathcal{C} \cup \{\emptyset, \Omega\}$ è una semialgebra (in modo un po' banale) se $\mathcal{C}$ è la classe dei costituenti introdotta nel capitolo precedente.

In effetti la classe $\mathcal{I}$ gioca, nell'analisi dell'Esempio 14, lo stesso ruolo giocato dalla classe $\mathcal{C} \cup \{\emptyset, \Omega\}$ nell'analisi dell'Esempio 1: le unioni finite di elementi disgiunti della classe danno l'algebra di interesse. Poiché le unioni di costituenti distinti sono necessariamente finite e costituenti distinti sono necessariamente disgiunti, non era servito precisare questo aspetto.

**Proposizione 2.1.2** *Se $\mathcal{S}$ è una semialgebra di eventi, definita su un insieme di esiti $\Omega$, la classe $\mathcal{A}$ delle unioni finite di elementi disgiunti di $\mathcal{S}$ è un'algebra, su $\Omega$.*

*Dimostrazione.* In primo luogo, è evidente che $\Omega \in \mathcal{S}$ implica $\Omega \in \mathcal{A}$. In secondo luogo, siano $E = \biguplus_{i=1}^{n} E_i$ ed $F = \biguplus_{j=1}^{m} F_j$ due elementi di $\mathcal{A}$, dove $E_i$ ed $F_j$ appartengono a $\mathcal{S}$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Distribuendo l'intersezione rispetto all'unione troviamo $E \cap F = \biguplus_{(i,j)} (E_i \cap F_j)$, dove la coppia $(i, j)$ varia nel prodotto cartesiano $\{1, \ldots, n\} \times \{1, \ldots, m\}$ e si è tenuto conto del fatto che $(E_i \cap F_j) \cap (E_h \cap F_k) = (E_i \cap E_h) \cap (F_j \cap F_k) = \emptyset$, se $(i, j) \neq (h, k)$. Vediamo allora che $E \cap F \in \mathcal{A}$, in quanto $\mathcal{S}$ è stabile per intersezioni finite. In terzo luogo, possiamo scrivere $\bar{E} = \bigcap_{i=1}^{n} \bar{E}_i$ e osservare che $\bar{E}_i \in \mathcal{A}$, $i = 1, \ldots, n$, perché $E_i \in \mathcal{S}$ implica che $\bar{E}_i$ sia unione finita di elementi disgiunti di $\mathcal{S}$. Ne seguirà $\bar{E} \in \mathcal{A}$, per quanto già dimostrato. □

La classe di eventi $\mathcal{A}$ introdotta nella Proposizione 2.1.2 è evidentemente l'algebra generata dalla semialgebra $\mathcal{S}$ e quindi considerando le unioni finite di elementi di $\mathcal{S}$, senza richiedere che questi siano disgiunti, si ottiene la stessa classe $\mathcal{A}$. Con riferimento alla retta reale, chiameremo *plurintervalli* (aperti a sinistra e chiusi a destra) le unioni finite di elementi (disgiunti) di $\mathcal{I}$ e denoteremo con $\mathcal{P}$ la loro algebra; in simboli $\mathcal{P} = \alpha(\mathcal{I})$.

Il risultato seguente mostra che lo strano soggetto dell'Esempio 14 non ci sta offrendo alcuna opportunità di profitto certo.

**Proposizione 2.1.3** *Se $\mathcal{S}$ è una semialgebra di eventi, definita su un insieme di esiti $\Omega$, e $\mathbb{P} : \mathcal{S} \to \mathbb{R}$ è una funzione di evento additiva, isotona e normalizzata, esiste un'unica estensione di $\mathbb{P}$ all'algebra $\mathcal{A}$ generata da $\mathcal{S}$ che sia additiva, isotona e normalizzata.*

*Dimostrazione.* Sia $E = \biguplus_{i=1}^{n} E_i$ un qualsiasi elemento di $\mathcal{A}$, dove $E_i \in \mathcal{S}$, $i = 1, \ldots, n$. Poniamo $\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E_i)$ e mostriamo che la definizione è ben posta. Se $E = \biguplus_{j=1}^{m} F_j$ è un'altra rappresentazione di $E$ come unione finita di elementi disgiunti di $\mathcal{S}$, allora $F_j = F_j \cap E = \biguplus_{i=1}^{n} (F_j \cap E_i)$, $j = 1, \ldots, m$, con $E_i \cap F_j \in \mathcal{S}$, $i = 1, \ldots, n$, in quanto $\mathcal{S}$ è una semialgebra. Poiché $\mathbb{P}$ è additiva su $\mathcal{S}$ avremo $\mathbb{P}(F_j) = \sum_{i=1}^{n} \mathbb{P}(E_i \cap F_j)$, $j = 1, \ldots, n$, quindi $\sum_{j=1}^{m} \mathbb{P}(F_j) = \sum_{i=1}^{n} \mathbb{P}(E_i)$, dal momento che il ragionamento svolto si può ripetere con $E_1, \ldots, E_n$ al posto di $F_1, \ldots, F_m$. A questo punto si verifica subito che abbiamo esteso $\mathbb{P}$ da $\mathcal{S}$ ad $\mathcal{A}$ preservandone l'additività e che altre estensioni additive non sono possibili. Dopo di che è chiaro che l'estensione è positiva, quindi isotona, oltre che normalizzata. □

L'additività di $\mathbb{P}$ su $\mathcal{S}$ nella Proposizione 2.1.3 è da intendersi riferita a unioni finite di elementi disgiunti di $\mathcal{S}$ *quando* tali unioni appartengano a $\mathcal{S}$ (non essendo in generale $\mathcal{S}$ stabile per unioni finite). Allo stesso modo avevamo inteso l'additività dell'assegnazione $\mathbb{P} = \Delta F$ su $\mathcal{I}$. La Proposizione 2.1.3, assieme al Teorema 1.1.3, mostra come $\Delta F$ definisca una $\mathbb{P}$ coerente sull'algebra $\mathcal{P}$ dei plurintervalli di $\mathbb{R}$. Analogamente $\Sigma P$ definiva una $\mathbb{P}$ coerente sull'algebra $\mathcal{U}$ delle unioni di costituenti, con il vincolo di somma a uno su $P$ a garantire l'additività di $\mathbb{P} = \Sigma P$ su $\mathcal{C} \cup \{\emptyset, \Omega\}$, una volta convenuto che $\mathbb{P}(\emptyset) = 0$ e $\mathbb{P}(\Omega) = 1$. Per il resto i due casi sono del tutto analoghi.

Se lo strano soggetto dell'Esempio 14 è coerente, qual è il problema? Il problema è l'esistenza di una successione descrescente di eventi $D_n = ]-\infty, 4 + 1/n]$, $n \in \mathbb{N}$, per ognuno dei quali $\mathbb{P}(D_n) = 1$, mentre per la loro intersezione $D_\infty = \bigcap_n D_n = \{\omega \in \Omega : \omega \in D_n \text{ per ogni } n \in \mathbb{N}\} = ]-\infty, 4]$ si ha $\mathbb{P}(D_\infty) = 0$. Questa sparizione di massa nel limite per $n \to \infty$, dando a $\mathbb{P}$ un'interpretazione fisica, rende la $\mathbb{P}$ dello strano soggetto, cui possiamo dare il nome di *massa aderente*, un'espressione irragionevole di conoscenze certe. Se la sparizione fosse parziale, il problema sarebbe meno evidente, ma faremmo comunque fatica a giustificare la scelta di $\mathbb{P}$. Per questo chiederemo senz'altro che $\mathbb{P}$ sia continua lungo successioni decrescenti di eventi: $\mathbb{P}(E_\infty) = \lim_n \mathbb{P}(E_n)$ comunque si scelgano $E_1 \supseteq E_2 \supseteq \dots$ in $\mathcal{A}$ con $E_\infty = \bigcap_n E_n \in \mathcal{A}$. Si noti che $\lim_n \mathbb{P}(E_n)$ è sempre ben definito, per una successione decrescente di eventi, perché $\mathbb{P}(E_n)$, $n \in \mathbb{N}$, è una successione decrescente di numeri reali positivi. Scriveremo in breve che $E_n \downarrow E_\infty$ implica $\mathbb{P}(E_n) \downarrow \mathbb{P}(E_\infty)$, per $n \to \infty$, quando vorremo ricordare che $\mathbb{P}$ è continua lungo successioni decrescenti di eventi.

Similmente scriveremo che $E_n \uparrow E_\infty$ implica $\mathbb{P}(E_n) \uparrow \mathbb{P}(E_\infty)$, per $n \to \infty$, quando vorremo affermare che $\mathbb{P}$ è continua lungo successioni crescenti di eventi; in questo caso si avrà $E_1 \subseteq E_2 \subseteq \dots$ ed $E_\infty = \bigcup_n E_n = \{\omega \in \Omega : \omega \in E_n \text{ per almeno un } n \in \mathbb{N}\}$ con $\mathbb{P}(E_n)$, $n \in \mathbb{N}$, successione crescente di numeri reali minori di uno. Un esempio di successione crescente è quella degli eventi $\bar{D}_n = ]4 + 1/n, +\infty]$, $n \in \mathbb{N}$, per ognuno dei quali $\mathbb{P}(\bar{D}_n) = 0$ mentre $\mathbb{P}(\bar{D}_\infty) = 1$ per la loro unione $\bar{D}_\infty = ]4, +\infty]$; un altro modo di guardare alla "patologia" della massa aderente. In effetti le due nozioni di continuità sono equivalenti, per una $\mathbb{P}$ coerente su un'algebra, come mostra il passaggio ai complementari, per cui diremo semplicemente che $\mathbb{P}$ è continua su $\mathcal{A}$ lungo successioni monotone di eventi, quando sarà il caso.

La visione di $E_\infty$ come evento limite della successione monotona $E_n$, $n \in \mathbb{N}$, si motiva definendo

$$\begin{aligned}
\liminf_{n\to\infty} E_n &= \bigcup_{n\geq 1} \bigcap_{k\geq n} E_k &&= \{\omega \in \Omega : \omega \in E_n \text{ definitivamente}\}, \\
\limsup_{n\to\infty} E_n &= \bigcap_{n\geq 1} \bigcup_{k\geq n} E_k &&= \{\omega \in \Omega : \omega \in E_n \text{ frequentemente}\},
\end{aligned}$$

e osservando che, per via della monotonia, $\limsup_n E_n = \liminf_n E_n = E_\infty$. In generale (per una successione non necessariamente monotona) avremo $\liminf_n E_n \subseteq \limsup_n E_n$ (definiti allo stesso modo) e scriveremo $\lim_n E_n = E_\infty$ quando accada che $\liminf_n E_n = \limsup_n E_n = E_\infty$.

La continuità lungo successioni monotone di una funzione di evento $\mathbb{P}$ coerente su un'algebra $\mathcal{A}$ equivale alla sua additività numerabile, intesa come $\mathbb{P}(\biguplus_k F_k) = \sum_k \mathbb{P}(F_k)$ *quando* $\biguplus F_k \in \mathcal{A}$, comunque presa una successione $F_k$, $k \in \mathbb{N}$, di elementi disgiunti di $\mathcal{A}$. Se $\mathbb{P}$ è numerabilmente additiva ed $E_n \uparrow E_\infty$, per $n \to \infty$, con $E_\infty \in \mathcal{A}$, basterà porre $F_1 = E_1$ ed $F_k = E_k \setminus E_{k-1}$, $k \geq 2$, per potere scrivere $E_\infty = \biguplus_k F_k$ con $F_k$, $k \in \mathbb{N}$, elementi disgiunti di $\mathcal{A}$; ne seguirà $\mathbb{P}(E_\infty) = \sum_k \mathbb{P}(F_k) = \lim_n \mathbb{P}(E_n)$, visto che l'$n$-esima somma parziale di $\sum_k \mathbb{P}(F_k)$ è pari a $\mathbb{P}(E_n)$. Viceversa, se $\mathbb{P}$ è continua lungo successioni monotone ed $F_k$, $k \in \mathbb{N}$, è una successione di elementi disgiunti di $\mathcal{A}$ con $\biguplus F_k \in \mathcal{A}$, si potrà prendere $E_n = F_1 \uplus \dots \uplus F_n$, $n \in \mathbb{N}$, per ottenere $E_n \uparrow \biguplus_k F_k$ con $\mathbb{P}(E_n) = \mathbb{P}(F_1) + \cdots + \mathbb{P}(F_n) \uparrow \mathbb{P}(\biguplus_k F_k)$, per $n \to \infty$, di modo che $\mathbb{P}$ sarà numerabilmente additiva. Diremo allora che $\mathbb{P}$ è una probabilità sull'algebra $\mathcal{A}$ quando sia coerente e continua lungo successioni monotone, anche nel caso in cui $\mathcal{A}$ non sia una tribù. Dopo di che la $\mathbb{P}$ dello strano soggetto nell'Esempio 14 non rientra nella definizione, perché non è continua lungo successioni monotone di eventi in $\mathcal{P}$.

## 2.2 Probabilità sulla retta reale

Quali condizioni deve soddisfare $F$ affinché $\mathbb{P} = \Delta F$ definisca una probabilità sui plurintervalli della retta reale? Per rispondere a questa domanda osserviamo preliminarmente che, essendo $F$ crescente, sono ben definiti il limite destro $F_+(b) = \inf_{x>b} F(x)$ e sinistro $F_-(b) = \sup_{x<b} F(x)$ di $F$ in $b$, $b \in \mathbb{R}$, con $F_-(b) \leq F_+(b)$. Analogamente sono ben definiti i due limiti di $F$ all'infinito, vale a dire $F_+(-\infty) = \inf_{x\in\mathbb{R}} F(x)$ ed $F_-(+\infty) = \sup_{x\in\mathbb{R}} F(x)$. Tutti questi limiti possono essere valutati lungo opportune successioni; per esempio $F_+(b) = \lim_n F(b+1/n)$, qualunque sia $b \in \mathbb{R}$. Dopo di che notiamo come $]-\infty, b+1/n] \downarrow ]-\infty, b]$, per $n \to \infty$, implichi $F_+(b) = F(b)$, $b \in \mathbb{R}$, se $\Delta F$ è continua lungo successioni monotone. Pertanto $F$ deve coincidere con la sua regolarizzata destra $F_+$; altrimenti detto, $F$ deve essere continua da destra. Inoltre $]-\infty, -n] \downarrow \emptyset$ e $]-\infty, n] \uparrow \mathbb{R}$, per $n \to \infty$, implicano $F_+(-\infty) = \lim_n F(-n) = 0$ e $F_-(+\infty) = \lim_n F(n) = 1$; diremo che $F$ deve essere *propria* per intendere che deve soddisfare queste due ulteriori condizioni. In definitiva, una $F$ candidata a definire una probabilità su $\mathcal{P}$ dovrà necessariamente essere continua da destra e propria.

Andrà bene una qualsiasi $F : \mathbb{R} \to [0, 1]$ che sia crescente, continua da destra e propria per assegnare una probabilità su $\mathcal{P}$? Il risultato seguente mostra come queste condizioni siano un buon punto di partenza.

**Proposizione 2.2.1** *Se $F : \mathbb{R} \to [0, 1]$ è una funzione crescente, continua da destra e propria, allora la funzione di evento $\mathbb{P} = \Delta F$, definita su $\mathcal{I}$ da $\mathbb{P}(]a, b]) = F(b) - F(a)$, $-\infty \leq a \leq b \leq +\infty$, con la convenzione $F(-\infty) = 0$ e $F(+\infty) = 1$, è numerabilmente additiva su $\mathcal{I}$.*

*Dimostrazione.* Abbiamo visto come ogni $F : \mathbb{R} \to [0, 1]$ crescente definisca una $\mathbb{P} = \Delta F$ additiva, isotona e normalizzata. Vediamo ora come tale $\mathbb{P}$ sia numerabilmente additiva su $\mathcal{I}$, se $F$ è continua da destra e propria. Sia $]a_n, b_n]$, $n \in \mathbb{N}$, una successione di intervalli disgiunti in $\mathcal{I}$ tale che $]a, b] = \biguplus ]a_n, b_n] \in \mathcal{I}$; si possono senz'altro supporre tutti gli intervalli non vuoti, visto che $\mathbb{P}(\emptyset) = 0$. Avremo $\sum_n \mathbb{P}(]a_n, b_n]) \leq \mathbb{P}(]a, b])$, perché la disuguaglianza vale per ogni somma parziale, essendo $\mathbb{P}$ isotona, quindi passa al limite; il seguito mostra che $\mathbb{P}(]a, b]) \leq \sum_n \mathbb{P}(]a_n, b_n])$. Sia $\epsilon > 0$. Prendiamo $a^\star > a$ e $b^\star \leq b$, $a^\star < b^\star < \infty$, con $\mathbb{P}(]a^\star, b^\star]) > \mathbb{P}(]a, b]) - \epsilon$; questo è sempre possibile perché $F$ è continua da destra e propria. In modo analogo, prendiamo $b_n^\star > b_n$ con $\mathbb{P}(]a_n, b_n^\star]) < \mathbb{P}(]a_n, b_n]) + \epsilon/2^n$, $n \in \mathbb{N}$, eventualmente $b_n^\star = b_n$ se $b_n = +\infty$. Per il teorema di Heine-Borel, possiamo trovare $n_1, \ldots, n_k$ tali che $[a^\star, b^\star] \subseteq \bigcup_{i=1}^k ]a_{n_i}, b_{n_i}^\star[ \subseteq \bigcup_{i=1}^k ]a_{n_i}, b_{n_i}^\star]$ e quindi, invocando la subadditività di $\mathbb{P}$, riusciamo a scrivere

$$\mathbb{P}(]a^\star, b^\star]) \leq \sum_{i=1}^k \mathbb{P}(]a_i, b_{n_i}^\star]) < \sum_{i=1}^k \mathbb{P}(]a_i, b_{n_i}]) + \sum_{i=1}^k \frac{\epsilon}{2^{n_i}} < \sum_{n=1}^\infty \mathbb{P}(]a_n, b_n]) + \epsilon;$$

troveremo allora $\mathbb{P}(]a, b]) < \mathbb{P}(]a^\star, b^\star]) + \epsilon < \sum_n \mathbb{P}(]a_n, b_n]) + 2\epsilon$ e la tesi seguirà osservando che $\epsilon$ è arbitrario. □

Il problema nell'assegnazione della $\mathbb{P}$ dello strano soggetto nell'Esempio 14 è che $F(b) = \mathbb{I}_{]4,+\infty]}(b)$, $b \in \mathbb{R}$, è continua da sinistra, non da destra; infatti $F(4) = 0$. Basta però una piccola modifica per rimediare: porre $F(b) = \mathbb{I}_{[4,+\infty]}(b)$, $b \in \mathbb{R}$, in modo che $F(4) = 1$. Con questo piccolo cambiamento si ottiene $\mathbb{P}(]a, b]) = \mathbb{I}_{]a,b]}(4)$, $]a, b] \in \mathcal{I}$, quindi $\mathbb{P}(F) = \mathbb{I}_F(4)$, $F \in \mathcal{P}$, come è agevole verificare. Dopo di che $\mathbb{P}$ sarà una probabilità su $\mathcal{P}$: $4 \in \biguplus_k F_k$ se e solo se esiste $k$ (necessariamente unico) tale che $4 \in F_k$, comunque si prenda una successione $F_1, F_2, \ldots$ in $\mathcal{P}$ con $\biguplus_k F_k \in \mathcal{P}$.

Il risultato seguente mostra come, in generale, la numerabile additività non possa perdersi per strada: qualsiasi assegnazione coerente di quote su un algebra $\mathcal{A}$, ottenuta per estensione da una semialgebra $\mathcal{S}$ che generi $\mathcal{A}$, sarà una probabilità, su $\mathcal{A}$, se è numerabilmente additiva su $\mathcal{S}$.

**Proposizione 2.2.2** *Sia $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ una funzione di evento coerente su un'algebra di eventi $\mathcal{A}$, definita su un insieme di esiti $\Omega$, e sia $\mathcal{S}$ una semialgebra che genera $\mathcal{A}$. Se $\mathbb{P}$ è numerabilmente additiva su $\mathcal{S}$, allora $\mathbb{P}$ è numerabilmente additiva su $\mathcal{A}$.*

*Dimostrazione.* Se $F_k$, $k \in \mathbb{N}$, è una successione di eventi disgiunti in $\mathcal{A}$ con $\biguplus_k F_k \in \mathcal{A}$, possiamo scrivere $\biguplus_k F_k = E_1 \uplus \cdots \uplus E_m$ con $E_i \in \mathcal{S}$, $i = 1, \ldots, m$, ed $F_k = G_{k1} \uplus \cdots \uplus G_{kn_k}$ con $G_{kj} \in \mathcal{S}$, $j = 1, \ldots, n_k$, $k \in \mathbb{N}$. Dopo di che avremo $E_i = E_i \cap (\biguplus_k F_k) = \biguplus_k (E_i \cap F_k)$, $i = 1, \ldots, m$, nonché $(E_i \cap F_k) = E_i \cap (G_{k1} \uplus \cdots \uplus G_{kn_k}) = \biguplus_j (G_{kj} \cap E_i)$, $k \in \mathbb{N}$, di modo che $E_i = \biguplus_k \biguplus_j (G_{kj} \cap E_i)$, dove gli eventi $(G_{kj} \cap E_i)$ appartengono a $\mathcal{S}$ e sono disgiunti. Troveremo quindi

$$\begin{aligned}\mathbb{P}\left(\biguplus_{k=1}^\infty F_k\right) &= \sum_{i=1}^m \mathbb{P}(E_i) = \sum_{i=1}^m \sum_{k=1}^\infty \sum_{j=1}^{n_k} \mathbb{P}(G_{kj} \cap E_i) \\ &= \sum_{k=1}^\infty \sum_{i=1}^m \sum_{j=1}^{n_k} \mathbb{P}(G_{kj} \cap E_i) = \sum_{k=1}^\infty \mathbb{P}(F_k),\end{aligned}$$

prima considerando come un'unica unione numerabile quella su $k$ e $j$, poi sfruttando il fatto che $\biguplus_i (E_i \cap F_k) = F_k$, $k \in \mathbb{N}$. □

Vale ovviamente anche il viceversa: ogni probabilità $\mathbb{P}$ su $\mathcal{A} = \alpha(\mathcal{S})$ è numerabilmente additiva su $\mathcal{S}$, in quanto $\mathcal{S} \subseteq \mathcal{A}$. Meno ovvia e più interessante è l'osservazione che, per effetto combinato delle Proposizioni 2.2.1 e 2.2.2, l'assegnazione $\mathbb{P} = \Delta F$ definisce una probabilità su $\mathcal{P}$, ogni qual volta $F : \mathbb{R} \to [0, 1]$ sia crescente, continua da destra e propria.

Nell'Esempio 15 il postino che serve la nostra zona suonerà alla nostra porta domani mattina, tra le otto e mezzogiorno, per consegnarci un pacco. Possiamo prendere $\Omega = \mathbb{R}$ e assegnare $\mathbb{P} = \Delta F$ sull'algebra $\mathcal{P}$. Alla luce della considerazione sulla scrupolosità del postino, porremo $F(8) = 0$ ed $F(12) = 1$. Dopo di che, in virtù della considerazione sulla sua mancanza di disciplina, porremo $F(10) = 1/2$; infatti $\mathbb{P}(]8, 10]) = \mathbb{P}(]10, 12])$ implica $\mathbb{P}(]8, 10]) = \mathbb{P}(]10, 12]) = 1/2$, visto che $\mathbb{P}(]8, 10]) + \mathbb{P}(]10, 12]) = \mathbb{P}(]8, 12])$. Analogamente troveremo $F(8 + 4/3) = 1/3$ ed $F(8 + 8/3) = 2/3$, visto che $\mathbb{P}(]8, 8+4/3]) = \mathbb{P}(]8+4/3, 8+8/3]) = \mathbb{P}(]8+8/3, 12]) = 1/3$. Più in generale troveremo $F(8 + 4m/n) = m/n$, comunque si prendando $m$ ed $n$ in $\mathbb{N}$, vale a dire $F(8 + 4q) = q$, per ogni $q$ razionale tra 0 e 1. Concluderemo che $F(b) = (b - 8)/4$, per ogni $b$ reale tra 8 e 12, perché $F$ deve essere continua da destra. Chiaramente sarà $F(b) = 0$ per $b$ minore di 8 ed $F(b) = 1$ per $b$ maggiore di 12. Si noti che $F$, così specificata, risulta continua (anche da sinistra). Per questo avere considerato intervalli aperti a sinistra e chiusi a destra, senza che questo dettaglio fosse specificato dal testo dell'Esempio 15, non ha conseguenze rilevanti. Approfondiremo questo punto più avanti.

L'algebra $\mathcal{P}$ dei plurintervalli di $\mathbb{R}$ non contiene molti eventi potenzialmente interessanti: mancano gli intervalli chiusi a sinistra e aperti a destra, quelli aperti (sia a sinistra che a destra) e quelli chiusi (sia a sinistra che a destra) inclusi i *singoletti* $[a] = \{a\}$ cioè gli eventi elementari. Si vede subito che qualsiasi algebra contenesse questi ultimi, includendo $\mathcal{I}$, conterrebbe anche tutti gli altri: $[a, b] = ]a, b] \cup \{a\}$, $]a, b[ = ]a, b] \setminus \{b\}$ ed $[a, b[ = ]a, b[ \cup \{a\}$. Inoltre $\bigcap_n ]a - 1/n, a] \downarrow \{a\}$, per $n \to \infty$, qualunque sia $a \in \mathbb{R}$. Sembra dunque naturale provare a estendere $\mathcal{P}$ in modo da ottenere un'algebra che sia stabile per limiti decrescenti. Una classe con queste proprietà sarebbe stabile anche per intersezioni numerabili non monotone, come si verifica subito scrivendo $\bigcap_n E_n = \lim_n F_n$, con $F_n = E_1 \cap \cdots \cap E_n$, $n \in \mathbb{N}$, qualunque siano $E_n$, $n \in \mathbb{N}$, appartenenti alla classe; sarebbe dunque una tribù.

Si dice tribù generata da una classe $\mathcal{E}$ di eventi su $\Omega$ l'intersezione $\sigma(\mathcal{E})$ di tutte le tribù su $\Omega$ che includano $\mathcal{E}$, fra le quali annoveriamo senz'altro la tribù $\wp(\Omega)$ di tutte la parti di $\Omega$; la definizione è ben posta, perché qualsiasi intersezione di tribù su $\Omega$ è ancora una tribù su $\Omega$. Si noti l'analogia con la definizione di $\alpha(\mathcal{E})$. In modo del tutto simile $\mathcal{E}$ genera una *classe monotona*, vale a dire una classe $\mathcal{M}$ di eventi stabile per limiti monotoni: $E_n \downarrow E_\infty \Rightarrow E_\infty \in \mathcal{M}$ ed $E_n \uparrow E_\infty \Rightarrow E_\infty \in \mathcal{M}$, se $E_n \in \mathcal{M}$, $n \in \mathbb{N}$. Sia $\mu(\mathcal{E})$ la classe monotona generata da $\mathcal{E}$. In generale $\mu(\mathcal{E}) \subseteq \sigma(\mathcal{E})$, perché ogni tribù è una classe monotona; l'inclusione non può però essere stretta nel caso particolare in cui $\mathcal{E}$ sia un'algebra.

**Proposizione 2.2.3** *La classe monotona $\mathcal{M}$ generata da un'algebra $\mathcal{A}$ di eventi, definiti su un insieme di esiti $\Omega$, è una tribù, su $\Omega$.*

*Dimostrazione.* Dimostriamo che $\mathcal{M}$ è un'algebra. In primo luogo, avremo $\Omega \in \mathcal{M}$, perché $\Omega \in \mathcal{A}$. In secondo luogo, la classe $\{E \in \mathcal{M} : \bar{E} \in \mathcal{M}\}$ è una classe monotona, perché $E_n \downarrow E_\infty$ se e solo se $\bar{E}_n \uparrow \bar{E}_\infty$; essa include $\mathcal{A}$ e dunque coincide con $\mathcal{M}$, la quale è quindi stabile per complementazione. In terzo luogo, fissato $B \in \mathcal{A}$, la classe $\{E \in \mathcal{M} : B \cap E \in \mathcal{M}\}$ è monotona, perché $E_n \downarrow E_\infty$ ($E_n \uparrow E_\infty$) implica $B \cap E_n \downarrow B \cap E_\infty$ ($B \cap E_n \uparrow B \cap E_\infty$); essa include $\mathcal{A}$ e dunque coincide con $\mathcal{M}$. Di conseguenza, fissato $E \in \mathcal{M}$, la classe monotona $\{F \in \mathcal{M} : E \cap F \in \mathcal{M}\}$ include $\mathcal{A}$ e quindi coincide con $\mathcal{M}$. Resta così stabilito che $\mathcal{M}$ è un'algebra e la sua stabilità per intersezioni numerabili segue dalla sua stabilità per limiti decrescenti. □

Nel caso specifico della retta reale denoteremo con $\mathcal{B}$ la tribù generata da $\mathcal{P}$, $\mathcal{B} = \sigma(\mathcal{P})$, che chiameremo tribù dei *boreliani* di $\mathbb{R}$. La medesima tribù sarà generata da $\mathcal{I}$, da cui abbiamo ottenuto gli elementi di $\mathcal{P} = \alpha(\mathcal{I})$ come unioni finite; quindi $\mathcal{B} = \sigma(\mathcal{I})$. In virtù di quanto abbiamo visto introducendo la nozione di tribù, i singoletti faranno parte dei boreliani e dunque $\mathcal{P} \subset \mathcal{B}$.

La Proposizione 2.2.3 caratterizza la tribù generata da un'algebra. Tale caratterizzazione non è costruttiva, nel senso che non fornisce un'espressione esplicita per gli elementi della tribù, come invece faceva la Proposizione 2.1.2 per gli elementi dell'algebra generata da una semialgebra. Ciò nonostante, essa consente di stabilire l'unicità dell'estensione di una probabilità alla tribù generata dal suo dominio; tale estensione è per altro sempre possibile, in modo costruttivo, mediante un opportuno passaggio al limite.

**Teorema 2.2.4 (Carathéodory)** *Se $\mathbb{P}$ è una probabilità sull'algebra $\mathcal{A}$, definita su un insieme di esiti $\Omega$, esiste un'unica probabilità sulla tribù generata da $\mathcal{A}$ che coincida con $\mathbb{P}$ su $\mathcal{A}$:*

$$\mathbb{P}(E) = \inf \left\{ \sum_{n=1}^{\infty} \mathbb{P}(A_n) \,\middle|\, E \subseteq \biguplus_{n=1}^{\infty} A_n \text{ e } A_n \in \mathcal{A}, n \in \mathbb{N} \right\}, \tag{2.2.1}$$

*per ogni $E$ appartenente alla tribù generata da $\mathcal{A}$.*

*Dimostrazione.* Per quanto riguarda l'unicità, siano $\mathbb{P}_1$ e $\mathbb{P}_2$ due probabilità su $\sigma(\mathcal{A})$ coincidenti su $\mathcal{A}$: $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ per ogni $A \in \mathcal{A}$. La classe $\{E \in \sigma(\mathcal{A}) : \mathbb{P}_1(E) = \mathbb{P}_2(E)\}$ è monotona, perché $\mathbb{P}_1$ e $\mathbb{P}_2$ sono continue lungo successioni monotone di eventi; essa include $\mathcal{A}$ e quindi per la Proposizione 2.2.3 coincide con $\sigma(\mathcal{A})$. Per quanto riguarda invece l'esistenza, procediamo per passi (come descritto di seguito).

Innanzi tutto mostriamo che la (2.2.1) estende $\mathbb{P}$ da $\mathcal{A}$ a $\wp(\Omega)$. Infatti, se $E \in \mathcal{A}$, il membro di destra della (2.2.1) non può fornire per $\mathbb{P}(E)$ un valore diverso da quello di partenza: non può fornire un valore strettamente maggiore, perché si può prendere $A_1 = E$ e $A_n = \emptyset$, $n \geq 2$; non può fornire un valore strettamente minore, perché si può sostituire $A_n$ con $A_n \cap E$, $n \in \mathbb{N}$, essendo $\mathbb{P}$ isotona su $\mathcal{A}$, quindi si può far valere l'additività numerabile di $\mathbb{P}$ su $\mathcal{A}$. Chiaramente $\mathbb{P}$ resta normalizzata dopo l'estensione, continuando ad annullarsi su $\emptyset$, mentre la sua isotonia su $\wp(\Omega)$ è evidente dalla (2.2.1). Dopo di che la positività seguirà dall'isotonia, tenuto conto che $\mathbb{P}(\emptyset) = 0$.

In secondo luogo mostriamo che $\mathbb{P}$ è *numerabilmente subadditiva* su $\wp(\Omega)$. Siano $E_k$, $k \in \mathbb{N}$, parti qualsiasi di $\Omega$ ed $E = \bigcup_k E_k$ la loro unione. Fissiamo $\epsilon > 0$ e, per ogni $k \in \mathbb{N}$, prendiamo $A_{kn} \in \mathcal{A}$, $n \in \mathbb{N}$, con $E_k \subseteq \biguplus_n A_{kn}$ e $\sum_n \mathbb{P}(A_{kn}) < \mathbb{P}(E_k) + \epsilon/2^k$. Chiaramente $E \subseteq \bigcup_{k,n} A_{kn}$ e di conseguenza $\mathbb{P}(E) \leq \sum_{k,n} \mathbb{P}(A_{kn}) < \sum_k \mathbb{P}(E_k) + \epsilon$, se solo mostriamo che nella (2.2.1) non occorre restringersi al caso in cui gli $A_n$, $n \in \mathbb{N}$, siano disgiunti. In effetti, dati $A_1, \ldots, A_n$ non necessariamente disgiunti, si può sempre porre $A'_1 = A_1$ ed $A'_n = A_n \cap \bar{A}_{n-1} \cap \cdots \cap \bar{A}_1$, $n > 2$, trovando $\sum_n \mathbb{P}(A'_n) \leq \sum_n \mathbb{P}(A_n)$ con $A'_n$, $n \in \mathbb{N}$, disgiunti. Poiché $\epsilon > 0$ è arbitrario, avremo $\mathbb{P}(E) \leq \sum_k \mathbb{P}(E_k)$. La disuguaglianza vale anche quando $E_1, \ldots, E_n$ siano in numero finito, essendo $\mathbb{P}(\emptyset) = 0$. Si dice che $\mathbb{P}$ è una *probabilità esterna* per riassumere le proprietà dimostrate in questi primi due passi.

Introduciamo ora la classe $\mathcal{L}$ delle parti $E$ di $\Omega$ che soddisfano la *condizione di Carathéodory*:

$$\mathbb{P}(F) = \mathbb{P}(F \cap E) + \mathbb{P}(F \cap \bar{E}), \quad \text{per ogni } F \subseteq \Omega;$$

equivalentemente $\mathbb{P}(F \cap E) + \mathbb{P}(F \cap \bar{E}) \leq \mathbb{P}(F)$, per ogni $F \subseteq \Omega$, visto che $\mathbb{P}$ è subadditiva su $\wp(\Omega)$. Mostriamo che $\mathcal{L}$ è un'algebra. In effetti, è banale verificare che $\Omega \in \mathcal{A}$ e si vede subito che $\mathcal{L}$ è stabile per complementazione, avendo $E$ ed $\bar{E}$ ruoli simmetrici nella condizione di Carathéodory. Dopo di che, comunque presi $E_1$ ed $E_2$ in $\mathcal{L}$, con $F \subseteq \Omega$ qualsiasi, si può scrivere $\mathbb{P}(F \cap (E_1 \cup E_2)) \leq \mathbb{P}(F \cap E_1) + \mathbb{P}(F \cap \bar{E}_1 \cap E_2)$, per la subadditività di $\mathbb{P}$, oltre che $\mathbb{P}(F \cap \overline{E_1 \cup E_2}) = \mathbb{P}(F \cap \bar{E}_1 \cap \bar{E}_2)$; la condizione di Carathéodory per $E_1 \cup E_2$ rispetto a $F$ segue allora da quella di $E_2$ rispetto a $F \cap \bar{E}_1$ e da quella di $E_1$ rispetto a $F$ (applicate nell'ordine). Resta così dimostrato che $\mathcal{L}$ è stabile per unioni finite e quindi che $\mathcal{L}$ è un'algebra su $\Omega$.

Mostriamo adesso che $\mathbb{P}$ è numerabilmente additiva su $\mathcal{L}$. Sia $E_n$, $n \in \mathbb{N}$, una successione di elementi disgiunti di $\mathcal{L}$ e sia $F$ una parte qualsiasi di $\Omega$. Posto $G_n = E_1 \uplus \cdots \uplus E_n$, $n \in \mathbb{N}$, di modo che $G_n \uparrow G_\infty = \biguplus_n E_n$, abbiamo $\mathbb{P}(F \cap G_n) = \mathbb{P}(F \cap G_n \cap G_{n-1}) + \mathbb{P}(F \cap G_n \cap \bar{G}_{n-1})$, perché $G_n \in \mathcal{L}$; quindi $\mathbb{P}(F \cap G_n) = \mathbb{P}(F \cap G_{n-1}) + \mathbb{P}(F \cap E_n)$. Procedendo per induzione, ne ricaviamo $\mathbb{P}(F \cap G_n) = \mathbb{P}(F \cap E_1) + \cdots + \mathbb{P}(F \cap E_n)$, $n \in \mathbb{N}$, dove il caso base $n = 1$ è banale. A questo punto, ricordando che $\mathbb{P}$ è isotona e passando al limite per $n \to \infty$, troviamo $\mathbb{P}(F \cap G_\infty) \geq \lim_n \mathbb{P}(F \cap G_n) = \sum_n \mathbb{P}(F \cap E_n)$. D'altra parte $F \cap G_\infty = \biguplus_n F \cap E_n$ e $\mathbb{P}$ è numerabilmente subadditiva su $\wp(\Omega)$, per cui necessariamente $\mathbb{P}(F \cap G_\infty) = \sum_n \mathbb{P}(F \cap E_n)$. Abbiamo così dimostrato che

$$\mathbb{P}\left(F \cap \biguplus_n E_n\right) = \sum_n \mathbb{P}(F \cap E_n), \quad \text{per ogni } F \subseteq \Omega,$$

se $E_n$, $n \in \mathbb{N}$, sono elementi disgiunti di $\mathcal{L}$. Prendendo $F = \Omega$ (e supponendo $\biguplus_n E_n \in \mathcal{L}$) resta verificato che $\mathbb{P}$ è numerabilmente additiva su $\mathcal{L}$.

In realtà $\mathcal{L}$ è una tribù. Con la notazione del passo precedente, mostriamo che $G_\infty \in \mathcal{L}$. Sappiamo che $\mathbb{P}(F) = \mathbb{P}(F \cap G_n) + \mathbb{P}(F \cap \bar{G}_n)$, $n \in \mathbb{N}$, qualunque sia $F \subseteq \Omega$. Inoltre, essendo $\mathbb{P}$ isotona, abbiamo $\mathbb{P}(F \cap \bar{G}_\infty) \leq \mathbb{P}(F \cap \bar{G}_n)$. Troveremo quindi $\sum_{i=1}^n \mathbb{P}(F \cap E_i) + \mathbb{P}(F \cap \bar{G}_\infty) \leq \mathbb{P}(F)$, $n \in \mathbb{N}$. La condizione di Carathéodory per $G_\infty$ seguirà passando al limite per $n \to \infty$. Questo mostra che $\biguplus_n E_n \in \mathcal{L}$, se $E_n$, $n \in \mathbb{N}$, sono elementi disgiunti di $\mathcal{L}$. Se invece $E_n$, $n \in \mathbb{N}$, sono elementi di $\mathcal{L}$ non necessariamente disgiunti, possiamo porre $E'_1 = E_1$ ed $E'_n = E_n \cap \bar{E}_{n-1} \cap \cdots \cap \bar{E}_1$; quindi possiamo osservare che $E'_n$, $n \in \mathbb{N}$, sono elementi disgiunti di $\mathcal{L}$ e $\bigcup_n E_n = \biguplus_n E'_n$.

Ci resta solo da mostrare che $\mathcal{A} \subseteq \mathcal{L}$, cosicché $\sigma(\mathcal{A})$ sarà inclusa in $\mathcal{L}$ e $\mathbb{P}$ definirà una probabilità su $\sigma(\mathcal{A})$ per restrizione. Siano $E \in \mathcal{A}$ ed $F \in \wp(\Omega)$. Fissiamo $\epsilon > 0$ e prendiamo $A_n$, $n \in \mathbb{N}$, disgiunti in $\mathcal{A}$ con $F \subseteq \biguplus A_n$ e $\sum_n \mathbb{P}(A_n) < \mathbb{P}(F) + \epsilon$. Avremo $F \cap E \subseteq \biguplus (A_n \cap E)$ ed $F \cap \bar{E} \subseteq \biguplus (A_n \cap \bar{E})$, di modo che $\mathbb{P}(F \cap E) \leq \sum_n \mathbb{P}(A_n \cap E)$ e $\mathbb{P}(F \cap \bar{E}) \leq \sum_n \mathbb{P}(A_n \cap \bar{E})$. Quindi troveremo $\mathbb{P}(F \cap E) + \mathbb{P}(F \cap \bar{E}) \leq \sum_n \mathbb{P}(A_n) < \mathbb{P}(F) + \epsilon$ e di conseguenza la condizione di Carathéodory per $E$, essendo $\epsilon > 0$ ed $F \in \wp(\Omega)$ arbitrari. $\square$

Osserviamo che l'additività numerabile su una tribù vale per tutte le unioni numerabili di eventi disgiunti della tribù, in quanto tali unioni non possono non appartenere alla tribù. Dopo di che, con riferimento alla retta reale, il Teorema 2.2.4 e quanto precedentemente discusso ci daranno un'unica probabilità sui boreliani $\mathcal{B}$ tale che $\mathbb{P}(]-\infty, b]) = F(b)$, $b \in \mathbb{R}$, comunque scelta $F : \mathbb{R} \to [0, 1]$ crescente, continua da destra e propria. Si rinvia al libro di Alberto Tesei intitolato *Istituzioni di Analisi Superiore*, pubblicato da Bollati Boringhieri nel 1997, per un approfondimento sull'estensione di Carathéodory e sul fatto che i boreliani non esauriscano le parti di $\mathbb{R}$. Più in generale, prendiamo nota del fatto che basta specificare $\mathbb{P}$ su una semialgebra $\mathcal{S}$ in modo che sia numerabilmente additiva, isotona e normalizzata, perché si possa considerare $\mathbb{P}$ una probabilità sulla tribù $\mathcal{A} = \sigma(\mathcal{S})$.

Completiamo, per concludere, l'analisi dell'Esempio 15. Qui $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}, \Delta F)$ con $F(\omega) = (\omega - 8)/4$, $\omega \in [8, 12]$, mentre $F(\omega) = 0$, $\omega < 8$, ed $F(\omega) = 1$, $\omega > 12$. Abbiamo visto che $F$ è continua (anche da sinistra) e anticipato che, per questo, non importa precisare se gli intervalli di interesse siano aperti o chiusi (a sinistra e a destra). Perché? Dato $\omega \in \mathbb{R}$, la probabilità del singoletto $\{\omega\}$ varrà $\mathbb{P}(\{\omega\}) = \lim_n \mathbb{P}(]\omega - 1/n, \omega]) = \lim_n \{F(\omega) - F(\omega - 1/n)\} = F(\omega) - F_-(\omega)$, visto che $]\omega - 1/n, \omega] \downarrow \{\omega\}$, per $n \to \infty$. La continuità di $F$ significa quindi $\mathbb{P}(\{\omega\}) = 0$, $\omega \in \mathbb{R}$; diremo che $\mathbb{P}$ è *diffusa* (continua). Si noti che ogni evento elementare ha probabilità nulla, anche se alla fine uno di essi si verificherà; il fatto è che i singoletti formano un continuo e quindi "sfuggono" all'additività numerabile. Quando si vogliano mettere *tutti* gli $\omega \in \mathbb{R}$ sullo stesso piano, non vi sono alternative: potremo avere al più $n-1$ valori $\omega$ con $\mathbb{P}(\{\omega\}) > 1/n$, $n \in \mathbb{N}$, quindi $\{\omega \in \mathbb{R} : \mathbb{P}(\{\omega\}) > 0\}$ sarà numerabile. Concludiamo calcolando la probabilità richiesta: $\mathbb{P}([8, 11.5[) = \mathbb{P}(]8, 11.5]) = F(11.5) = (11.5-8)/4 = 3.5/4 = 7/8 = 0.875$ (circa 88%).

## 2.3 Probabilità su spazi prodotto

*Questa sezione è ancora da scrivere.*

## 2.4 Lemmi di Borel-Cantelli

*Questa sezione è ancora da scrivere.*

*Questa pagina è stata lasciata intenzionalmente in bianco.*