Michael Skeide

Probabilità Elementare

(per l'Informatica)

Primo semestre 2014/15

(Edizione *zero*)

Michael Skeide

Dipartimento S.E.G.e S.
Università degli Studi del Molise
Via de Sanctis, 86100 Campobasso, Italia
E-mail: skeide@unimol.it
Homepage: http://web.unimol.it/skeide/

Prefazione

In questo corso diamo un'introduzione ai concetti basilari della probabilità, mentre la statistica la dobbiamo rimandare ad un corso più avanzato; si veda l'introduzione. Nostro scopo è di discutere questi concetti senza riferirci ad attrezzi della matematica più avanzata (come la teoria della misura) ma, comunque, in modo matematicamente rigoroso basato solo sulle conoscenze di analisi ed algebra dei corsi precedenti.

Conseguentemente, noi facciamo probabilità su *algebre booleane* di eventi, non su σ*–algebre*; quest ultime, nonostante la base della maggior parte dei corsi di probabilità, sono decisamente fuori nostra portata. Un prezzo da pagare c'è: La teoria che otteniamo è meno potente. Infatti, mentre le σ–algebre di eventi contengono tutti gli eventi immaginabili (e ancora più eventi non immaginabili), noi spesso ci vediamo costretti di aggiungere certi eventi "naturali" alle nostre algebre booleane. Perciò, svilupperemo diversi metodi che lo rendono possibile e - più importante - possibile in modo unico, cioè, in modo "naturale".

Alcuni fatti sulle algebre booleane e probabilità su esse sono delegati agli Appendici B e C. L'Appendice B sulle algebre booleane è solo un supplemento che contiene anche le dimostrazioni di alcuni risultati del calcolo dei insiemi i quali dobbiamo, comunque, supporre ben noti già da altri corsi. L'Appendice C, invece, contiene precisazioni sull'esistenza di probabilità su algebre booleane. Anche se il livello dell'Appendice C rimane sempre elementare, non ci rimane abbastanza tempo per discutere tutti i dettagli in lezione. Soprattutto l'Appendice C non è ancora completato. Lo facciamo in una prossima edizione.

Né l'Appendice B né l'Appendice C fanno parte dell'esame.

L'Appendice A.1, invece, presenta i risultati basilari del calcolo combinatorio ed alcuni fatti sulle serie assolutamente convergenti. Questi risultati, sebbene non le loro dimostrazioni, li supponiamo noti anche per gli esami, perché sono stati discussi nei corsi Matematica I e II.

Queste dispense contengono ampio materiale supplementare, che è stampato in caratteri più piccoli ed in larghezza ridotta (come, per esempio, alcune dimostrazioni o le sezioni indicate con "nota matematica"). Intere sezioni supplementari sono indicate con "Appendice" o con un asterisco. Il materiale supplementare non fa parte del programma ma serve solo a fornire ulteriori informazioni a chi le dovessero interessare. Un'eccezione sono gli esercizi integrati

nel testo che fanno parte integrante del corso, comunque formattati come le parti supplementari per distinguerli meglio dal testo principale.

Parti indicate “Osservazione” sono come “Proposizioni”, solo che la dimostrazione è integrata nel testo. Le conclusioni enunciate nelle osservazioni sono da considerarsi note come quelle delle proposizioni. I risultati più indispensabili sono messi, comunque, in un box per indicarne l’importanza. La collezione dei contenuti di tutti i box costituisce la *spina dorsale* delle conoscenze che vogliamo acquisire in questo corso.

Ribadisco ciò che ho detto in lezione:

La matematica non si può imparare. La matematica deve essere acquisita!

È impossibile imparare la matematica leggendo un libro o ascoltando passivamente una lezione; è solo possibile acquisirla, svolgendo con regolarità gli esercizi, testando in tal modo se gli argomenti presentati in lezione fossero davvero stati capiti! Non c’è problema se uno non può venire in lezione; basta che legga contemporaneamente le parti corrispettive di queste dispense! Chi, però, non lo fa o non si esercita con regolarità svolgendo gli esercizi, non può sperare di superare l’esame.

Iniziare di sfogliare le dispense qualche giorno prima dell’esame è un atteggiamento senza speranza!

Per avere un’impressione che ci aspetta all’esame, è consigliabile guardare i problemi del Foglio A e risolverli tutti quanti. (Chi l’ha fatto durante il corso, avrà poco da lavorare. Chi non l’ha fatto ...) Di solito, nell’esame non si va oltre i problemi del tipo come appaiono nel Foglio A.

Michael Skeide, Pesche, Ottobre 2014

Indice

Prefazione **i**

Introduzione **1**

1 Spazi di probabilità booleani **7**

1.1 Algebre booleane di eventi 7

1.2 Probabilità su algebre booleane 16

1.3 Spazi di probabilità discreti 23

2 Spazi di probabilità elementari ed altre applicazioni del calcolo combinatorio **33**

2.1 Spazi di probabilità elementari 33

2.2 Spazi di probabilità finiti e loro prodotti 39

2.3 Una costruzione di uno spazio di probabilità non discreto 45

3 Indipendenza e probabilità condizionata **53**

3.1 Probabilità condizionata 53

3.2 Indipendenza 60

3.3 Il prodotto di spazi di probabilità booleani 65

4 Variabili aleatorie discrete **67**

4.1 Variabili aleatorie discrete e le loro leggi 69

4.2 Leggi congiunti 72

4.3 Indipendenza di variabili aleatorie discrete 75

4.4 Attese di variabili aleatorie discrete 79

4.5 Varianza, covarianza e la legge dei grandi numeri 82

4.6 La funzione generatrice (di Laplace) 87

4.7 Un esempio esteso 90

5 Variabili aleatorie continue **93**

5.1 Probabilità su $\mathbb{R}$ 94

5.2 Leggi e attese di variabili aleatorie continue 100

5.3 Vettori aleatori continui e le loro funzioni 106

5.4 La legge normale ed il teorema limite centrale 110
5.5 Convoluzioni ed altri altri esempi con densità 113

Appendice **118**

A Alcuni rinfrescamenti **119**
A.1 Elementi del calcolo combinatorio 120
A.2 Richiami ai limiti di numeri reali 129
A.3 Serie assolutamente convergenti 130
A.4 Le funzioni monotone 133
A.5 Integrazione m–dimensionale 133

B Algebra booleana astratta **135**
B.1 Algebra booleana 137
B.2 Omomorfismi ed isomorfismi di algebre booleane 141
B.3 Il teorema di Stone per algebre booleane finite 143
B.4 Il lemma di Zorn ed il teorema di Stone generale 146

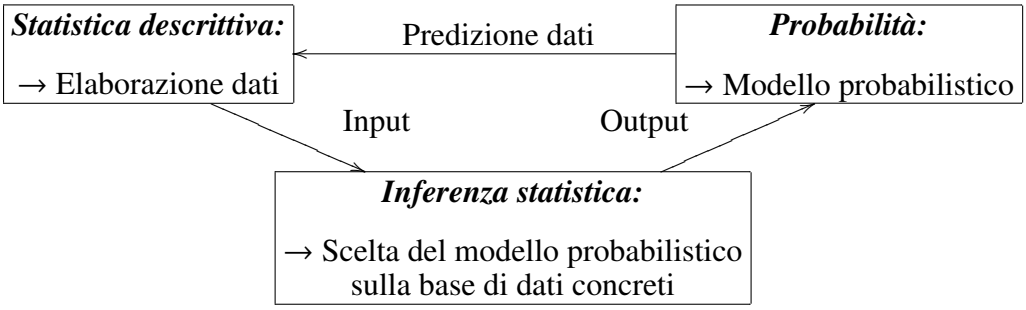
C Probabilità booleane con la funzione di ripartizione **151**
C.1 Probabilità su semialgebre booleane 151
C.2 Probabilità sugli intervalli 154
C.3 Prodotti di semialgebre booleane 154

Bibliografia **157**

Introduzione

Sia la statistica che la probabilità si occupano di fenomeni aleatori. Con ciò intendiamo un esperimento con esito non determinato. Esperimenti tipici sono il lancio di una moneta, l'estrazione del lotto, l'osservazione del sesso di neonati, il numero d'incidenti in un anno, la misurazione di quantità fisiche o economiche, eccetera. Tutti questi esperimenti hanno in comune, che l'esito di ogni singola osservazione può variare: Non è determinato. Se in un tentativo esce un risultato x, non è detto che in un altro tentativo non possa uscire un risultato y diverso da x.

Mentre la ***statistica descrittiva*** si occupa esclusivamente della raccolta di tali dati e della loro "organizzazione", la ***probabilità*** vuole fornire un *modello probabilistico* del fenomeno aleatorio, che ci permette di fare predizioni sull'esito dell'esperimento. Cioè la statistica descrittiva si limita a raccogliere dati e ad analizzare la loro struttura (rappresentazione grafica, calcolare *indici numerici* come la media o la varianza). La probabilità invece, una volta scelto il modello probabilistico, cerca di predire i risultati di una raccolta dati. In altre parole, la probabilità cerca di spiegare teoricamente perché l'elaborazione grafica o numerica di dati concreti abbia proprio l'aspetto che ha. L'***inferenza statistica***, infine, vuole rispondere alla domanda quale potrebbe essere il modello probabilistico "più giusto" per il meccanismo aleatorio che abbia prodotto un campione di dati concreti.



Tutto sommato, la statistica descrittiva si occupa esclusivamente di dati concreti rilevati in un esperimento. La probabilità invece è una teoria matematica disegnata con lo scopo di descrivere il meccanismo del fenomeno aleatorio dietro l'esperimento: Una volta scelto il modello per tale meccanismo, la teoria fa delle predizioni sui risultati di un eventuale esperimento. Se, però,

le predizioni non concordano con i dati concreti, non è mai la colpa della teoria; è sempre la colpa di chi ha scelto il modello sbagliato. L'inferenza statistica è l'unica disciplina che mette in confronto i dati concreti con il modello teorico. Una volta rilevato un mazzo di dati, l'inferenza statistica ci aiuta a scegliere un modello ragionevole (***stima***, ***intervalli di fiducia***) e di rispondere alla domanda se un modello particolare possa aver prodotto questi dati (***test***).

L'inferenza statistica è, quindi, probabilità applicata. Non è possibile comprendere i concetti dell'inferenza statistica senza prima aver acquistato una conoscenza di base solida della probabilità. In questo corso ci limitiamo ad acquistare questa conoscenza di base della probabilità, mentre l'inferenza deve rimanere riservata ad un corso avanzato. La statistica descrittiva gioca un ruolo in quanto ci serve come motivazione per le definizioni che seguono. (Concetti della statistica descrittiva come le frequenze relative e la media motiveranno gli assiomi della probabilità e la definizione dell'attesa.)

• • •

All'inizio di ogni esperimento bisogna specificare i possibili risultati che l'esperimento possa avere. È chiaro che ogni volta che eseguiamo l'esperimento, dei risultati possibili si verifica uno e uno solo. Annotiamo con Ω l'insieme di tutti i risultati con e chiameremo Ω lo ***spazio campionario*** dell'esperimento. Facciamo alcuni esempi:

I.1. Lancio di una moneta: $\Omega = \{\text{testa}, \text{croce}\}$.

I.2. Il colore di una pallina estratta da un'urna che contiene N palline, di cui n bianche e $N-n$ rosse: $\Omega = \{\text{bianco}, \text{rosso}\}$, oppure più brevemente $\Omega = \{b, r\}$.

I.3. Lancio di un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

I.4. L'altezza di un italiano al suo ventesimo compleanno nell'anno 2005: $\Omega = \mathbb{R}_+ = [0, \infty)$.

Nell'ultimo esempio abbiamo scelto un continuo di valori possibili, anche se il numero N degli italiani che hanno compiuto i vent'anni nel 2005 è decisamente finito. Se sapessimo le loro altezze, potremmo anche formare un insieme elencando semplicemente tutte le altezze $\{\omega_1, \ldots, \omega_N\}$. Il problema è che (soprattutto nella statistica) spesso vogliamo proprio evitare il compito di esaminare l'intera popolazione ancora prima dell'esperimento . (Il prezzo da pagare sono alcune difficoltà tecniche dovute a questo continuo di valori, che risolveremo solo nel Capitolo 5 e nell'Appendice C.)

L'esempio ci dice anche che la scelta dello spazio campionario non è per niente unica. Tocca a chi modella l'esperimento fare la scelta più opportuna. Consideriamo il seguente esercizio per scelte diverse, di cui effetti studieremo più tardi:

Esercizio. Consideriamo il lancio di due dadi. Siamo interessati alla somma dei due numeri. Per la scelta dello spazio campionario di tale esperimento (in un modo che ci permette di determinare la somma) ci sono parecchie possibilità. Costruiamo i seguenti spazi campionari Ω_a, Ω_b e Ω_c dove i risultati sono:

a. Le coppie ordinate (n_b, n_r) dove i dadi sono ben distinti (diciamo uno blu, uno rosso) ed i numeri n_b rispettivamente n_r corrispondono al valore del dado blu rispettivamente del dado rosso.

b. Le coppie $\{n, m\}$ non ordinate, dove non ci interessa quale numero è apparso su quale dado.

c. Le possibili somme.

Per ognuno dei casi (a), (b), e (c) scrivere una lista di tutti gli elementi dello spazio campionario. (Suggerimento: Può essere utile di rappresentare gli elementi di Ω_a in un modo che mette in evidenza la struttura di Ω_a come insieme prodotto.)

Studieremo il concetto di spazio campionario e, soprattutto, l'insieme delle domande che vogliamo porre all'esperimento (gli ***eventi***) nella Sezione 1.1.

• • •

È importante notare che alcuni esperimenti hanno come risultati numeri (forse decorati con un'unità di una quantità fisica come nell'Esempio I.4) mentre in altri esperimenti i risultati sono distinti solo qualitativamente. (Né il colore della pallina che esce nell'Esempio I.2 né il simbolo sulla moneta dell'Esempio I.2 sono numeri.) Infatti, i dati che risultano da un esperimento vanno suddivisi in ***dati quantitativi*** e ***dati qualitativi***. L'Esempio I.3 è ambiguo. Certo, a prima vista, il risultato è un numero intero fra 1 e 6. Ma veramente i numeri sono solo un modo di distinguere i sei lati del dado. Avremmo benissimo potuto scegliere sei colori diversi allo stesso scopo. Numeri hanno il vantaggio che con loro è possibile fare delle computazioni (per esempio, calcolarne la media). Perciò, la trasformazione di dati in dati numerici, con cui calcolare, tramite il concetto di *variabile aleatoria* è fra le cose più importanti. Studieremo questo concetto ampiamente nei Capitoli 4 e 5.

• • •

Se lanciamo una moneta una volta, il risultato per forza assume uno e uno solo dei valori *testa* o *croce*. Comunque, il fatto che in un solo tentativo si verifichi, diciamo, *testa* non significa che la moneta favorisca il risultato *testa*. Al contrario, noi ci aspetteremmo che la moneta *non truccata* non dimostri nessuna preferenza per nessuno dei due risultati *testa* o *croce*. Il modo in cui *quantifichiamo* questa nostra convinzione è sempre lo stesso: In una serie di, diciamo M, ripetizioni del lancio il rapporto $\frac{m_t}{m_c}$ fra il numero m_t di *teste* ed il numero $m_c = M - m_t$ di

croci non dovrebbe risultare molto lontano dalla parità, cioè, da 1. Più volte lanciamo, più ci aspettiamo che il rapporto sia vicino ad 1. Al contrario, se anche dopo un numero "alto" di lanci il rapporto continua ad essere "significativamente" diverso da 1, inizieremo a sospettare che la moneta potrebbe essere truccata. (Stabilire che cosa sia un numero "alto" o che cosa sia una deviazione "significativa" da quello che ci aspettiamo è compito dell'inferenza statistica.)

Più generalmente, se eseguiamo ripetutamente un esperimento con spazio campionario (per semplicità, finito) $\Omega = \{\omega_1, \dots, \omega_N\}$, ci aspettiamo che i risultati ottenuti $x_1, \dots, x_M \in \Omega$ in M verifiche si distribuiscano sui possibili valori $\omega \in \Omega$ in tal modo che le ***frequenze relative*** di ogni risultato possibile ω

$$h_\omega^M := \frac{\#\{m \in \{1, \dots, M\} \colon x_m = \omega\}}{M} = \frac{\text{numero di verifiche in cui appare } \omega}{\text{numero totale delle verifiche}}$$

esprimano una misura per quanto l'esperimento favorisca l'esito ω. Tutto sommato, se il numero di tentativi M è abbastanza grande, ci aspettiamo che la frequenza relativa h_ω^M ci indica quale sia la *probabilità* che esca il risultato ω. Se, per esempio, in M verifiche è sempre uscito il risultato ω, ossia se $h_\omega^M = 1$, supponiamo che il risultato ω sia ***sicuro*** perché esce con probabilità $p_\omega = 1$. Al contrario, se in M verifiche il risultato ω è uscito mai, ossia se $h_\omega^M = 0$, supponiamo che il risultato ω sia ***impossibile*** perché esce con probabilità $p_\omega = 0$.

Le frequenze relative si godono di alcune proprietà particolari che, nella Sezione 1.2, ci serviranno per scegliere gli assiomi della probabilità. Usufruiremo delle frequenze relative anche per avere un'idea su come modellare due esperimenti *indipendenti* nella trattazione del Numero 2.2.1. Infine, questa argomentazione ritorna nella Sezione 3.2 quando parleremo di *indipendenza* vera e propria nella probabilità.

• • •

Non ci dobbiamo dimenticare, però, che le frequenze relative h_ω^M calcolate in M verifiche sono e rimarranno risultati empirici di un esperimento. Le probabilità p_ω, alle quali le frequenze relative dovrebbero avvicinarsi per grandi M, sono quantità teoriche all'interno di una teoria matematica. Il nostro intuito ci dice, che la *probabilità teorica* di avere deviazioni *grandi* delle *frequenze empiriche* h_ω^M dalle *probabilità teoriche* p_ω dovrebbe essere sempre più *piccola* se aumentiamo M *abbastanza*. Quest'ultima frase si lascia formalizzare all'interno della teoria che stiamo per costruire nei seguenti capitoli. Infatti, sia $\varepsilon > 0$ un numero strettamente positivo. Annotiamo con $P\Big(|h_\omega^M - p_\omega| \geq \varepsilon\Big)$ la probabilità che in una successione di M verifiche la frequenza relativa h_ω^M risulti distante da p_ω per più di ε. Uno dei teoremi più fondamentali di tutta la teoria (che dimostreremo nella Sezione 4.5), la *legge dei grandi numeri*, afferma:

La legge dei grandi numeri (versione debole per le frequenze relative). Sia p_ω la probabilità del risultato $\omega \in \Omega$. Allora per ogni $\varepsilon > 0$ vale

$$\lim_{M\to\infty} P\Big(\, |h_\omega^M - p_\omega| \geq \varepsilon \Big) \;\; = \;\; 0.$$

Osserviamo che tutte e due le probabilità, p_ω e $P\Big(\, |h_\omega^M - p_\omega| \geq \varepsilon\Big)$, sono quantità del tutto teoriche all'interno della teoria. Il teorema mette in confronto queste quantità teoriche con le frequenze relative ottenute in un esperimento reale. È una delle cose più belle e più convincenti, che in tal modo la teoria confermerà quello che noi, intuitivamente, ci aspettiamo.

Il teorema dice che aumentando il numero di misurazioni, prima o poi la frequenza relativa di un risultato ω ci da la probabilità di ω a meno di un errore $\varepsilon > 0$ che possiamo scegliere noi. Non ci dice in nessun modo quale numero M di verifiche sarà sufficiente per un certo ε e un certo ω. Una risposta a questa domanda dipende moltissimo dal modello in considerazione e fa parte dell'inferenza statistica. Nonostante, la *legge dei grandi numeri* (con le sue variazioni a volte molto sofisticate) è alla radice di ogni tentativo dell'inferenza statistica di mettere in relazione la teoria con i dati reali.

• • •

Come possiamo notare, in tutta la storia le frequenze relative, ossia la percentuale in M verifiche di un certo risultato, giocano un ruolo cruciale. L'idea di poter fare M verifiche dello stesso esperimento, secondo l'autore, è alla base di ogni probabilità: L'esperimento, almeno idealmente, deve essere ***ripetibile***! E, almeno idealmente, ogni esperimento è ripetibile. Basta immaginarsi una ripetizione sotto condizioni uguali. Su un esperimento aleatorio (cioè con esito incerto, ossia dove nessun risultato esce con probabilità 1) reale non ripetibile, la probabilità non può dare nessuna informazione verificabile. Anche se la probabilità che da un mazzo di carte pesco la *dama di cuore* è piuttosto bassa, il fatto che in una sola verifica mi riesca l'impresa, dal punto di vista della probabilità non significa proprio niente. Anzi, parlare di un fenomeno aleatorio assumendo che ci sarà una sola verifica, non lo consideriamo proprio giustificato: Prima di fare almeno due tentativi con esiti diversi non si può neanche verificare se il fenomeno sia veramente *aleatorio* (ossia, con esito incerto) o magari *deterministico* (ossia, un esperimento che sotto condizioni uguali porti sempre allo stesso medesimo risultato).

Capitolo 1

Spazi di probabilità booleani

Nella letteratura la nozione di *spazio di probabilità* ha un significato ben specificato. Questa nozione è più rigida che quella che stiamo per usare noi, benché inevitabile nelle applicazioni più sofisticate. L'uso appropriato di *spazi di probabilità* richiede, però, una certa abilità tecnica (la *teoria della misura*) che, normalmente, possiamo aspettare solo da studenti di matematica. Noi, in questo corso, riusciamo a lavorare con una struttura più elementare che, per evitare equivoci, chiameremo *spazi di probabilità booleani*. Il nome si riferisce soprattutto al fatto che lavoriamo con *algebre booleane* di eventi (invece delle σ*–algebre* di eventi della teoria generale) e con probabilità che sono semplicemente *additive* (invece della σ*–additività* della teoria generale). Oltre ad essere più elementare, la nozione di *spazio di probabilità booleano* ha anche il vantaggio di essere perfettamente motivabile, mentre la nozione più rigida, infine, si giustifica solo dal fatto di essere più potente nelle applicazioni.

1.1 Algebre booleane di eventi

Abbiamo detto nell'introduzione che alla base della modellazione di un esperimento aleatorio è la scelta dell'insieme Ω dei possibili esiti dell'esperimento.

L'insieme Ω chiameremo lo ***spazio campionario***, i suoi elementi $\omega \in \Omega$ i ***risultati*** dell'esperimento.

Ogni volta che eseguiamo l'esperimento si verifica sempre esattamente uno dei risultati ω di Ω (naturalmente non necessariamente sempre lo stesso). Ne concludiamo che Ω debba contenere almeno un elemento.

Uno spazio campionario non è mai vuoto: $\Omega \neq \emptyset$.

La domanda "si è verificato il risultato ω?" ha la risposta "si", se e solo se si è verificato il risultato ω e, quindi, nessun altro risultato $\omega' \neq \omega$.

Una domanda che ammette solo le risposte "si" e "no" chiameremo una ***domanda binaria***.

Non tutte le domande binarie che possiamo porre all'esito dell'esperimento sono collegate alla verifica di un solo risultato. Per esempio se lanciamo un dado, possiamo chiedere se il numero che si è verificato sia dispari. Riferendoci alla descrizione dell'esperimento "lancio del dado" con lo spazio campionario $\Omega = \{1, \ldots, 6\}$ come nell'Esempio I.2, la risposta alla domanda "il numero è dispari?" è positiva se il risultato ω rilevato fa parte del sottoinsieme $\{1, 3, 5\}$ di Ω ed è negativa se ω fa parte del sottoinsieme complementare $\{1, 3, 5\}^{\complement} = \{2, 4, 6\}$. Vediamo che, in perfetta sintonia con l'uso della lingua, questo sottoinsieme complementare contiene esattamente questi risultati che rispondono alla domanda opposta "il numero non è dispari?" rispettivamente "il numero è pari?" nel senso affermativo.

Più generalmente, ad ogni domanda binaria che possiamo porre ad un esperimento, corrisponde in modo univoco un sottoinsieme di Ω, il sottoinsieme che contiene tutti i risultati ω che rispondono alla domanda nel senso affermativo:

$$\text{"domanda binaria"} \quad \rightsquigarrow \quad \{\omega \in \Omega \colon \text{per questo } \omega \text{ la risposta alla domanda è "si"}\} \subset \Omega.$$

Vice versa ad ogni sottoinsieme A di Ω corrisponde almeno la domanda (un po' sintetica) "il risultato è un elemento di A?".

Chiameremo un sottoinsieme A di uno spazio campionario Ω un ***evento*** di Ω. Un ***evento elementare*** è un evento $\{\omega\}$ che contiene un unico risultato $\omega \in \Omega$. Chiameremo Ω ***l'evento sicuro*** e chiameremo l'insieme vuoto $\emptyset$ ***l'evento impossibile***.

N.B.: Un risultato è un elemento ω di Ω, mentre un evento elementare è un sottoinsieme $\{\omega\}$ di Ω che contiene un solo elemento $\omega \in \Omega$. Ovviamente, un evento elementare $\{\omega\}$ è un evento. Un risultato ω di Ω non è un evento di Ω.

Gli eventi corrispondono alle domande binarie che possiamo porre all'esperimento. Spesso il nostro spazio campionario Ω permette di porre più domande (ossia, Ω ha più sottoinsiemi) di quelle che ci interessavano quando abbiamo deciso di modellare un esperimento su questo spazio campionario Ω. Per esempio, per la domanda che abbiamo posto al lancio del dado, dispari o pari, gli unici eventi rilevanti sono $\{1, 3, 5\}$ e $\{2, 4, 6\}$. Avremmo benissimo potuto modellare l'esperimento sullo spazio campionario di due elementi $\Omega' = \{d, p\}$, dove d sta per "dispari" e p sta per "pari". Per quanto riguarda la modellazione dell'esperimento "dispari o pari" i due spazi campionari $\{1, \ldots, 6\}$ e $\{d, p\}$ s'equivalgono.

Notiamo che l'evento $A = \{1, 3, 5\}$ di $\Omega = \{1, \dots, 6\}$ ha la descrizione verbale "il numero è dispari" o semplicemente "dispari". L'evento contiene esattamente tali risultati che rispondono nel senso affermativo alla domanda "il numero è dispari?" o semplicemente "dispari?". Tale descrizione verbale dell'evento ha perfettamente senso anche se decidessimo di modellare l'esperimento sullo spazio campionario $\{d, p\}$. Infatti, prima della modellazione bisogna sempre rendersi conto a quali domande l'esperimento debba rispondere, cioè quali eventi il modello debba ammettere. In due modellazioni diverse dello stesso esperimento un evento può avere la stessa descrizione verbale mentre la sua realizzazione come sottoinsieme degli spazi campionari diversi è diversa. (Soprattutto sono sottoinsiemi di due insiemi diversi, quindi non possono essere uguali.)

1.1.1 Esercizio. (L'esercizio dell'introduzione continuato.) Abbiamo realizzato il lancio di due dadi con l'unico scopo di individuare la somma dei due dadi su tre spazi campionari diversi Ω_a, Ω_b e Ω_c. Indicare in ogni caso in modo esplicito gli eventi "la somma è 2", "la somma è 3", "la somma è 8" e "la somma è pari".

C'è più di una ragione perché a volte conviene non ammettere tutte le domande binarie, cioè non ammettere tutti i sottoinsiemi, come eventi rilevanti. Una delle ragioni è di tipo matematico tecnico: Secondo il *teorema di Vitali* non è possibile attribuire in modo *consistente* una probabilità a tutti i sottoinsiemi A dell'intervallo $[0, 1]$ in tal modo che la probabilità di ogni sottointervallo $[a, b] \subset [0, 1]$ sia la *lunghezza* $b - a$; si veda l'Appendice C. In tal caso siamo proprio costretti a limitare gli eventi ammissibili. Un'altra ragione vedremo quando parleremo d'*indipendenza* di più di due eventi e di *variabili aleatorie*. Una terza ragione, l'*attesa condizionata*, è un concetto della probabilità già un po' avanzato che non discuteremo in queste note.

Anche se decidiamo di non ammettere tutte le domande, ci conviene comunque di ammettere almeno le domande che seguono da quelle già ammesse tramite le operazioni logiche. Abbiamo già notato nell'esempio del dado che la risposta negativa alla domanda "il numero è dispari?" risponde nel senso affermativo alla domanda opposta "il numero non è dispari?". Non ha senso escludere una delle due domande quando ammettiamo l'altra. Similmente, se $A \subset \Omega$ è l'evento associato ad una domanda, allora l'***insieme complementare***

$$A^{\complement} := \{\omega \in \Omega \colon \omega \notin A\}$$

di A è l'evento associato alla domanda opposta. Quindi, se A è un evento ammissibile, lo dovrebbe anche essere l'evento complementare $A^{\complement}$. Poi se ammettiamo le domande che corrispondono agli eventi A e B di Ω, allora conoscendo la risposta alla domanda "si è verificato A?" e la risposta all domanda "si è verificato B?", possiamo facilmente determinare la risposta alla domanda "si sono verificati tutti e due gli eventi A e B?" (logico "e") e la risposta alla domanda "si è

verificato almeno uno dei due eventi A e B?" (logico "o") e non avrebbe senso di escludere queste domande. L'evento che corrisponde alla prima domanda è

$$\{\omega \in \Omega\colon \text{"si è verificato l'evento } A\text{"} \text{ e } \text{"si è verificato l'evento } B\text{"}\} = \{\omega \in \Omega\colon \omega \in A \wedge \omega \in B\} =: A \cap B \quad (\textbf{\textit{intersezione}}),$$

mentre l'evento che corrisponde alla seconda domanda è

$$\{\omega \in \Omega\colon \text{"si è verificato l'evento } A\text{"} \text{ o } \text{"si è verificato l'evento } B\text{"}\} = \{\omega \in \Omega\colon \omega \in A \vee \omega \in B\} =: A \cup B \quad (\textbf{\textit{unione}}).$$

N.B.: Le operazioni $\cap$ e $\cup$ hanno senso anche senza specificare Ω, ma per il complementare $^{\complement}$ occorre sempre indicare l'*universo* Ω rispetto al quale il complementare va calcolato.

Tutto sommato, chiediamo che l'insieme delle domande binarie che poniamo all'esperimento sia chiuso sotto le operazioni logiche "e", "o" e "negazione". Come abbiamo visto, questo corrisponde esattamente al fatto, che l'insieme di tutti gli eventi che vogliamo ammettere sia chiuso sotto *intersezione*, *unione* e *complementare*. In altre parole, gli eventi formano un'algebra booleana di sottoinsiemi di Ω:

1.1.2 Definizione. Sia $\Omega \neq \emptyset$ un insieme non vuoto. Un'***algebra booleana su*** Ω è una famiglia $\mathcal{A}$ non vuota di sottoinsiemi di Ω che soddisfa le seguenti proprietà:

1. Se A è di $\mathcal{A}$, allora anche $A^{\complement}$ è di $\mathcal{A}$.
2. Se A e B sono di $\mathcal{A}$, allora anche $A \cap B$ e $A \cup B$ sono di $\mathcal{A}$.

Se Ω è uno spazio campionario di un esperimento, diciamo anche $\mathcal{A}$ è un'***algebra booleana di eventi*** di Ω.

Con una famiglia $\mathcal{A}$ di sottoinsiemi di Ω s'intende semplicemente un sottoinsieme $\mathcal{A} \subset \mathcal{P}(\Omega)$ dell'insieme $\mathcal{P}(\Omega)$ di ***tutte le parti*** di Ω, di cui elementi sono tutti i sottoinsiemi A di Ω, cioè $\mathcal{P}(\Omega) := \{A \subset \Omega\}$.

N.B.: Gli elementi di $\mathcal{P}(\Omega)$ (e con ciò, anche gli elementi di $\mathcal{A} \subset \mathcal{P}(\Omega)$) sono insiemi (qui: sottoinsiemi di Ω) che a loro parte possono avere degli elementi (qui: elementi ω di Ω). Occorre saper *digerire* ed interpretare nel modo corretto un'espressione come $\omega \in A \in \mathcal{P}(\Omega)$ (che equivale l'espressione $\omega \in A \subset \Omega$) oppure $\omega \in A \in \mathcal{A}$ (che implica la precedente, precisando però che non solo $A \in \mathcal{P}(\Omega)$, ma $A \in \mathcal{A} \subset \mathcal{P}(\Omega)$).

1.1.3 Esempio. Con $A, B \in \mathcal{P}(\Omega)$ (ossia $A \subset \Omega$ e $B \subset \Omega$) anche $A^{\complement}$, $A \cap B$ e $A \cup B$ sono sottoinsiemi di Ω, ossia, $A^{\complement}, A \cap B, A \cup B \in \mathcal{P}(\Omega)$. Allora $\mathcal{P}(\Omega)$ stessa è un'algebra booleana su $\Omega \neq \emptyset$.

1.1.4 Osservazione. Poiché la famiglia $\mathcal{A}$ non è vuota, esiste almeno un insieme $A \in \mathcal{A}$. Con A, secondo la Condizione (1), anche l'insieme complementare $A^{\complement}$ è di $\mathcal{A}$. Infine, secondo la Condizione (2), anche l'insieme vuoto $\emptyset = A \cap A^{\complement}$ e l'intero insieme $\Omega = A \cup A^{\complement}$ sono di $\mathcal{A}$. (Un'algebra booleana su $\Omega \neq \emptyset$ ha almeno questi due elementi e, chiaramente, $\{\emptyset, \Omega\}$ è già un'algebra booleana su Ω; **esercizio**!) Vediamo che l'ipotesi che $\mathcal{A}$ sia non vuoto può essere sostituita con una delle due (equivalenti) ipotesi $\emptyset \in \mathcal{A}$ o $\Omega \in \mathcal{A}$.

1.1.5 Esempio. Sia $\Omega \neq \emptyset$ uno spazio campionario e sia $A \subset \Omega$ un suo evento. Allora la famiglia $\beta(A) := \{\emptyset, A, A^{\complement}, \Omega\} \subset \mathcal{P}(\Omega)$ è un algebra booleana su Ω, infatti, la più piccola algebra booleana che contiene l'evento A. Se A è l'insieme vuoto o Ω allora $\beta(A)$ ha solo gli elementi $\emptyset$ ed Ω (si veda l'Osservazione 1.1.4). Altrimenti il numero di elementi di $\beta(A)$ è quattro. Lasciamo come **esercizio** la verifica che si tratti d'un'algebra booleana.

Per esempio, sia $\Omega = \{1, \ldots, 6\}$ lo spazio campionario del dado e sia $A = \{1, 3, 5\}$ l'evento "dispari". Allora la più piccola algebra booleana su Ω che contiene A è

$$\beta(A) \;=\; \Big\{ \emptyset\,,\; \{1,3,5\}\,,\; \{2,4,6\}\,,\; \{1,\ldots,6\} \Big\}.$$

Nella descrizione sullo spazio campionario $\Omega' = \{d, p\}$, l'evento "dispari" è descritto dall'evento elementare $\{d\}$. In questo caso l'algebra booleana $\beta(\{d\})$ è $\{\,\emptyset\,,\; \{d\}\,,\; \{p\}\,,\; \{d,p\}\,\}$ e, quindi, coincide con $\mathcal{P}(\Omega')$.

È noto che per gli elementi di $\mathcal{P}(\Omega)$ valgono (fra l'altro) le seguenti regole:

$$\begin{array}{rcl@{\qquad}rcl@{\qquad}l}
A \cap B &=& B \cap A & A \cup B &=& B \cup A & \text{(commutatività)},\\
A \cap (B \cup C) &=& (A \cap B) \cup (A \cap C) & A \cup (B \cap C) &=& (A \cup B) \cap (A \cup C) & \text{(distributività)},\\
A \cap (B \cup B^{\complement}) &=& A & A \cup (B \cap B^{\complement}) &=& A & \text{(elementi neutrali)},\\
(A \cap B) \cap C &=& A \cap (B \cap C) & (A \cup B) \cup C &=& A \cup (B \cup C) & \text{(associatività)},\\
(A \cap B)^{\complement} &=& A^{\complement} \cup B^{\complement} & (A \cup B)^{\complement} &=& A^{\complement} \cap B^{\complement} & \text{(De Morgan)},\\
A \cap A &=& A & A \cup A &=& A & \text{(idempotenza)},\\
A \cap (B \cap B^{\complement}) &=& B \cap B^{\complement} & A \cup (B \cup B^{\complement}) &=& (B \cup B^{\complement}) & \text{(assorbimento)}.
\end{array}$$

Le prime tre righe ci dicono che $\mathcal{P}(\Omega)$ è un'algebra booleana astratta rispetto alle operazioni $\cap, \cup, {}^{\complement}$. Con ciò, le proprietà della Definizione 1.1.2 significano semplicemente che $\mathcal{A}$ è una

sottoalgebra booleana di $\mathcal{P}(\Omega)$. In particolare, $\mathcal{A}$ stessa è un'algebra booleana astratta. Tutte le altre regole possono essere dimostrate usando solo le prime tre (si veda l'Appendice B). Inoltre si può dimostrare che gli insiemi $B \cup B^{\complement}$ e $B \cap B^{\complement}$ che soddisfano la terza riga sono già determinati univocamente (Proposizione B.1.2). Visto che Ω e $\emptyset$ sono insiemi che soddisfano queste condizioni, ne segue

$$B \cap B^{\complement} = \emptyset, \qquad B \cup B^{\complement} = \Omega$$

(come abbiamo già usato nella Osservazione 1.1.4).

1.1.6 Esercizio. Usare solo le prime tre righe e l'unicità degli elementi neutri per dimostrare l'ultima riga.

Come esempio per l'uso delle relazioni, dimostriamo le seguenti due uguaglianze molto importanti.

1.1.7 Proposizione. $(A \cup B) \cap B = B = (A \cap B) \cup B.$

Dimostrazione. Osserviamo prima che $(A \cup B) \cap B = (A \cap B) \cup (B \cap B) = (A \cap B) \cup B$. Quindi basta se dimostriamo solo la prima parità $(A \cup B) \cap B = B$. Per farlo calcoliamo $((A \cap B) \cup B) \cap ((A \cap B^{\complement}) \cup B)$ in due modi diversi. Il primo modo:

$$((A \cap B) \cup B) \cap ((A \cap B^{\complement}) \cup B) = ((A \cap B) \cap (A \cap B^{\complement})) \cup B = (A \cap B \cap B^{\complement}) \cup B = \emptyset \cup B = B.$$

Il secondo modo

$$\begin{aligned}((A \cap B) \cup B) \cap ((A \cap B^{\complement}) \cup B) &= ((A \cup B) \cap B) \cap ((A \cup B) \cap (B^{\complement} \cup B)) \\ &= (A \cup B) \cap B \cap (A \cup B) \cap \Omega = (A \cup B) \cap B. \blacksquare\end{aligned}$$

N.B.: La nostra dimostrazione, oltre alle leggi d'idempotenza e dell'assorbimento, si serve anche dell'associatività. Nell'Appendice B, dove partiamo solo con le prime tre righe, l'associatività (Proposizione B.1.9) va dimostrata **dopo** queste uguaglianze (Corollario B.1.8). Comunque, qui stiamo parlando di algebre booleane di insiemi e la lista di proprietà usate può sempre essere verificata con le tabelle di verità riferendosi alle operazioni logiche.

N.B.: Le regole dell'algebra booleana, sia assiomi che formule derivate da loro, sono simmetriche rispetto allo scambio $\cap \leftrightarrow \cup$ (più, eventualmente, $\emptyset \leftrightarrow \Omega$ dove dovessero far apparizione. Non disturbiamo tale simmetria con una convenzione che una delle operazioni fosse più vincolante dell'altra (come il $\cdot$ che sarebbe più vincolante del $+$)! Di conseguenza **non esistono** formule come $A \cup B \cap C$ o $A \cap B \cup C$. In una formula che contiene sia $\cap$ che $\cup$ bisogna sempre indicare con parentesi quale operazione deve essere eseguita per primo.

L'***inclusione*** è una *relazione* (si veda il corso di algebra [Ske04]) fra sottoinsiemi, definita come $A \subset B$ (ossia, $B \supset A$) se $\omega \in A$ implica $\omega \in B$.

1.1.8 Proposizione.

$$A \subset B \iff A \cap B = A \qquad\qquad A \subset B \iff A \cup B = B.$$

DIMOSTRAZIONE. Ci occupiamo solo della parte sinistra. La parte destra la lasciamo come **esercizio**!

"$\Longrightarrow$": Valga $A \subset B$. Dobbiamo dimostrare che $\omega \in A \cap B$ se e solo se $\omega \in A$. Sia $\omega \in A \cap B$, cioè $\omega \in A$ e $\omega \in B$, allora in particolare $\omega \in A$. *Vice versa* sia $\omega \in A$, Secondo l'ipotesi $A \subset B$ questo implica $\omega \in B$ e, infine, $\omega \in A \cap B$. Allora $A \cap B = A$.

"$\Longleftarrow$": Valga $A \cap B = A$. Dobbiamo dimostrare che $\omega \in A$ implica $\omega \in B$. Ma $A \cap B = A$ significa che $\omega \in A$ e $\omega \in B$ se e solo se $\omega \in A$. La parte "se" di questa frase è proprio l'implicazione che ci serve. ■

Notiamo che la conclusione $\omega \in A$ e $\omega \in B \implies \omega \in A$ non dipende dall'ipotesi $A \subset B$. Vale, quindi, in piena generalità:

1.1.9 Corollario. $A \cap B \subset A.$

1.1.10 Corollario. $A = B \iff A \subset B \ e \ B \subset A.$

DIMOSTRAZIONE. "$\Longleftarrow$": Secondo la Proposizione 1.1.8 la parte destra, $A \subset B$ e $B \subset A$, equivale $A \cap B = A$ e $B \cap A = B$ e, quindi, $A = B$. La direzione "$\Longrightarrow$" è ovvia. (Perché?) ■

1.1.11 Osservazione. Evidentemente, l'affermazione del corollario segue anche direttamente dalla definizione dell'inclusione e dal fatto che due insiemi sono ***uguali*** se e solo se contengono ***gli stessi elementi***. Il forte della dimostrazione precedente consiste nel fatto che non fa nessun riferimento ad insiemi ma solo alle proprietà enunciate nella Proposizione 1.1.8. Infatti, la relazione $A \subset B$ può essere definita come $A \subset B :\Leftrightarrow A \cap B = A$ senza riferimento ad insiemi. L'equivalenza $A \cap B = A \Leftrightarrow A \cup B = B$ vale, infatti, in ogni algebra booleana.

1.1.12 Esercizio. Dimostrare $A \subset B \iff B^{\complement} \subset A^{\complement}$

1. secondo la definizione insiemistica $A \subset B$ se $[\omega \in A \Rightarrow \omega \in B]$.
2. $(*)$ secondo la definizione $A \subset B \ :\Longleftrightarrow \ A \cap B = A$ (cioè usufruendo solo delle regole dell'algebra booleana astratta).

1.1.13 Osservazione. Dalle *regole di De Morgan* segue che

$$A \cap B \;=\; ((A \cap B)^{\complement})^{\complement} \;=\; (A^{\complement} \cup B^{\complement})^{\complement}.$$

Quindi, se $\mathcal{A} \subset \mathcal{P}(\Omega)$ è una famiglia di sottoinsiemi di $\Omega \neq \emptyset$ che soddisfa la Proprietà (1) della Definizione 1.1.2, allora per controllare che $\mathcal{A}$ sia un'algebra booleana su Ω basta controllare che $\mathcal{A}$ sia chiuso sotto unioni. Lasciamo come **esercizio** la dimostrazione analoga che, alternativamente, possiamo anche controllare solo chiusura sotto intersezioni.

Conviene rifare la dimostrazione che $\beta(A)$ (Esempio 1.1.5) è un'algebra booleana, prendendo in considerazione anche la presente osservazione. Il risparmio di lavoro è parecchio.

> Nell'Esempio 1.1.5 abbiamo parlato di un'algebra booleana su Ω più piccola $\beta(A)$ che contiene un sottoinsieme A di Ω. Infatti, si può dimostrare che per ogni famiglia $C \subset \mathcal{P}(\Omega)$ di sottoinsiemi di un insieme $\Omega \neq \emptyset$ esiste sempre un'unica algebra booleana $\beta(C)$ su Ω più piccola $\beta(C) \supset C$. (Esercizo: Dimostrare che
>
> $$\left\{ \bigcup_{k=1}^{n} (\, C_1^k \cap \ldots \cap C_{m_k}^k) \;\middle|\; k \in \mathbb{N}, m_k \in \mathbb{N}, C_i^k \in C \vee (C_i^k)^{\complement} \in C \right\}$$
>
> è un'algebra booleana su Ω. Evidentemente, contiene C e non può essere un'algebra booleana più piccola contente C.) Chiameremo $\beta(C)$ l'algebra booleana su $\Omega \neq \emptyset$ ***generata*** da $C \subset \mathcal{P}(\Omega)$. Ovviamente $\beta(A) = \beta(\{A\})$.

1.1.14 Esercizio.

1. Individuare l'algebra booleana $\mathcal{A}$ su $\Omega = \{1, \ldots, 6\}$ più piccola che contiene gli eventi $\{1, 3, 5\}$ e $\{2\}$.
2. $(**)$ Indicare uno spazio campionario Ω' tale che $\mathcal{A}$ è isomorfa all'algebra $\mathcal{P}(\Omega')$ di tutte le parti di Ω'.

Per completezza, ci ricordiamo ancora di due altre operazioni fra insiemi.

$$A \backslash B \;:=\; A \cap B^{\complement} \qquad (\textbf{\textit{differenza}}\ A \text{ meno } B),$$

$$A \vartriangle B \;:=\; (A \backslash B) \cup (B \backslash A) \qquad (\textbf{\textit{differenza simmetrica}}).$$

Mentre la differenza non è più che un'abbreviazione, la differenza simmetrica è di vitale importanza per tutta la teoria delle algebre booleane. Sul livello logico corrisponde al *exclusive or* che mette in relazione l'algebra booleana con il campo $\mathbb{Z}_2 = \{0, 1\}$ (con le operazioni + e $\cdot$ modulo 2). Solo in questo modo è possibile capire perché un'algebra booleana si merita il nome *algebra*; si veda l'Appendice B.

> Abbiamo visto nell'Esempio 1.1.5 che l'algebra booleana finita $\beta(A)$ è *isomorfa* (si veda la Definizione B.2.1) all'algebra booleana $\mathcal{P}(\Omega')$ di tutte le parti di un insieme opportuno Ω'. Nell'Appendice B.3 vedremo che questo è vero per tutte le algebre booleane finite (*teorema di Stone*). Non è necessariamente così per algebre booleane infinite.

1.1.15 Esempio. Sia $\Omega = \mathbb{N} = \{1, 2, \ldots\}$. Allora la famiglia

$$\mathcal{A} \;=\; \{A \subset \mathbb{N}\colon \#A < \infty \;\text{ o }\; \#A^{\complement} < \infty\}$$

è un algebra booleana su $\mathbb{N}$. Infatti, ovviamente $A \in \mathcal{A}$ se e solo se $A^{\complement} \in \mathcal{A}$. Poi, se A e B sono elementi di $\mathcal{A}$ di cui almeno uno è finito, allora anche $A \cap B$ è finito e, quindi, di $\mathcal{A}$. Se, invece, A e B sono infiniti tutti e due, allora i loro complementari $A^{\complement}$ e $B^{\complement}$ sono finiti. Quindi, anche l'unione $A^{\complement} \cup B^{\complement} = (A \cap B)^{\complement}$ è finita, cosicché l'insieme $A \cap B$ ha il complementare finito e risulta di $\mathcal{A}$. Allora $\mathcal{A}$ è chiuso anche su intersezioni e, secondo l'Osservazione 1.1.13, è un'algebra booleana.

Però, $\mathcal{A}$ non è isomorfa a nessun $\mathcal{P}(\Omega')$. Infatti, ogni $\mathcal{P}(\Omega')$ è chiuso sotto unioni arbitrarie di suoi elementi, mentre l'insieme dei numeri naturali pari

$$\{2, 4, \ldots\} \;=\; \bigcup_{n=1}^{\infty} \{2n\} \;=\; \{2\} \cup \{4\} \cup \ldots$$

non è di $\mathcal{A}$ (è infinito e ha il complementare infinito) nonostante sia l'unione degli elementi $\{2\}, \{4\}, \ldots$ di $\mathcal{A}$. (In parole, più precise: Supponiamo che $\varphi\colon \mathcal{A} \to \mathcal{P}(\Omega')$ fosse un isomorfismo di algebre booleane. Poiché in $\mathcal{P}(\Omega')$ esiste l'unione A' su tutti i $\varphi(\{2n\})$ e poiché $A' \neq \Omega'$ (infatti l'intersezione di A' e $\varphi(\{1\})$ è vuota), ne seguirebbe che $\varphi^{-1}(A') \neq \Omega$ fosse un elemento di $\mathcal{A}$ contenente tutti gli $\{2n\}$. Ma l'unico elemento di che contiene tutti gli $\{2n\}$ è Ω stesso. Contraddizione!)

Vale, però, sempre il *teorema di Stone generale* (Appendice B.4): Ogni algebra booleana astratta è isomorfa ad un'algebra booleana su un insieme Ω opportuno (benché non è necessariamente isomorfa a tutto $\mathcal{P}(\Omega)$). Vediamo che l'***esistenza di uno spazio campionario*** segue dalla ***struttura logica delle domande binarie*** a cui vogliamo che l'esperimento risponda.

Concludiamo questa sezione con un altro modo utilissimo di caratterizzare un sottoinsieme $A \subset \Omega$ tramite la sua ***funzione indicatrice*** $\chi_A\colon \Omega \to \{0, 1\}$, definita come

$$\boxed{\chi_A(\omega) \;:=\; \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A \end{cases}}$$

per ogni $\omega \in \Omega$.

1.1.16 Esercizio. Dimostrare che ogni funzione $\chi\colon \Omega \to \{0, 1\}$ è la funzione indicatrice χ_A del sottoinsieme

$$A \;:=\; \{\omega \in \Omega\colon \chi(\omega) = 1\}$$

di Ω. Dimostrare che se B è un altro sottoinsieme di Ω tale che $\chi = \chi_B$, allora $B = A$.

Vediamo che c'é un corrispondenza biunivoca fra sottinsiemi di Ω e funzioni a valori a $\{0, 1\}$. Allo stesso modo in cui gli eventi caratterittano le domande binari, la fanno anche le funzioni indicatrici. Solo che con le funzioni indicatrici a volte è più facile calcolare. Presentiamo un esempio tipico per ogni tale calcolo, e imprtantissimo per la scelta degli assiomi di probabilità nella prossima sezione.

1.1.17 Esempio. Voglimao dimostrare che

$$A \cap B = \emptyset \quad \Longrightarrow \quad \chi_{A\cup B} = \chi_A + \chi_B.$$

Ricordiamoci che due funzioni sono uguali se danno ad ogni argomento del dominio (necessariamente comune) lo stesso valore del codominio (necessariamente comune); si guardino le dispense [Ske04]. Occorre, quindi, verificare che $\chi_A(\omega) + \chi_B(\omega) = \chi_{A\cup B}(\omega)$ per ogni $\omega \in \Omega$. I valori di $\chi_A(\omega)\,\chi_B(\omega)$ variano solo in dipendenza dalle domande se ω è di A o meno e se ω è di B o meno. Ci sono i quattro casi:

	$\chi_A(\omega)$	$\chi_B(\omega)$	$\omega \in A \cup B$?	$\chi_{A\cup B}(\omega)$	$\chi_{A\cup B} = \chi_A + \chi_B$?
$\omega \notin A, \omega \notin B$	0	0	no	0	si
$\omega \notin A, \omega \in B$	0	1	si	1	si
$\omega \in A, \omega \notin B$	1	0	si	1	si

Che succede con il quarto caso $\omega \in A, \omega \in B$? Dall'ipotesi $A \cap B = \emptyset$ non può apparire! Quindi non c'è nulla da verificare. In casi dove gli eventi A e B sono generici (come nel prossimo esercizio!), bisognerebbe controllare anche questo caso. In casi con più di due eventi, la tabella diventerebbe più lunga (per esempio, $2^3 = 8$ righe per tre eventi A, B, C generici).

1.1.18 Esercizio. Dimostrare per qualsiasi sottoinsiemi $A, B \subset \Omega$:

$$\chi_{A\cap B} = \chi_A \cdot \chi_B, \qquad \chi_{A\cup B} = \chi_A + \chi_B - \chi_{A\cap B},$$
$$\chi_{A\triangle B} = \chi_A \oplus \chi_B, \qquad \chi_{A^\complement} = 1 \oplus \chi_A,$$

dove $\oplus$ è l'*exclusive or*, ossia, l'addizione modulo 2 su $\{0, 1\}$; si vedano le dispense [Ske04].

1.1.19 Esercizio. Usare il risultato $\chi_{A\triangle B} = \chi_A \oplus \chi_B$ dell'esercizio precedente per dimostrare

$$\chi_{(A\triangle B)\triangle C} = \chi_{A\triangle(B\triangle C)},$$

ossia, $(A \triangle B) \triangle C = A \triangle (B \triangle C)$.

1.2 Probabilità su algebre booleane

Nella sezione precedente abbiamo introdotto algebre booleane di eventi di uno spazio campionario, che sono un modello per le domande (binarie) che vogliamo porre all'esperimento. Ora su queste algebre booleane di eventi, cioè sull'insieme delle domande di nostro interesse, vogliamo definire delle probabilità. Più precisamente per ogni domanda binaria di cui ci interessa la risposta vogliamo sapere la *probabilità* che in **una** verifica la risposta sia "si". Come discusso

nell'introduzione, l'idea è che in un gran numero M di tentativi le frequenze relative dei risultati, o meglio degli eventi elementari (perché sono gli eventi per i quali definiremo le probabilità), si avvicinano sempre di più alle loro probabilità. (Questo nel senso della *legge dei grandi numeri*. Nell'introduzione abbiamo discusso la versione *debole*, che sarà facile da dimostrare; la versione *forte* dice addirittura *con probabilità* 1 le freqenze relative convergono verso la probabilità; sie veda **Overwork!** .) Quindi le proprietà algebriche che valgono per le frequenze relative indipendentemente da M, dovrebbero valere anche per le probabilità.

Abbiamo detto che la frequenza relativa h_ω^M di un risultato $\omega \in \Omega = \{\omega_1, \dots, \omega_N\}$ ottenuta in M tentativi, che hanno riportati gli M risultati $x_1, \dots, x_M \in \Omega$, è il numero $\#\{m\colon x_m = \omega\}$ di volte che si è verificato il risultato ω diviso per il numero totale M di tentativi. Più generalmente, definiamo la ***frequenza relativa di un evento*** $A \subset \Omega$ in M tentativi come

$$h_A^M := \frac{\#\{m \in \{1, \dots, M\}\colon x_m \in A\}}{M} = \frac{\sum_{m=1}^M \chi_A(x_m)}{M}.$$

1.2.1 Nota. Bisogna chiarire, una volta per tutte le altre volte che avverranno ancora, l'uso della funzione indicatrice allo scopo del conteggio degli elementi di un sottoinsieme. Ci interessa la cardinalità del sottoinsieme di $\{1, \dots, M\}$ che contiene tutti gli m che verificano la condizione $x_m \in A$. Per ogni $m = 1, \dots, M$ la sommatoria aggiunge 1 ogni volta che la condizione è soddisfatta, e altrimenti non aggiunge nulla. Alla fine non abbiamo fatto altro che contato proprio esattamente questi elementi di $\{1, \dots, M\}$ che soddisfanno la condizione. Una dimostrazione rigorosa si farebbe per induzione su M. La omettiamo.

Adesso vogliamo analizzare le proprietà di queste frequenze relative (come funzione $A \mapsto h_A^M$). Ci serviamo della forma $h_A^M = \frac{\sum_{m=1}^M \chi_A(x_m)}{M}$ e illustriamo così quanto sia utile l'uso delle funzioni indicatrici. La prima proprietà che notiamo è che una funzione indicatrice assume solo valori non negativi. Quindi anche ogni frequenza relativa h_A^M è non negativa: $h_A^M \geq 0$ per ogni $A \subset \Omega$. La seconda proprietà segue dall'osservazione che l'indicatrice dell'intero spazio Ω assume sempre il valore $\chi_\Omega(\omega) = 1$. Quindi, la frequenza relativa dell'intero spazio è

$$h_\Omega^M = \frac{\sum_{m=1}^M \chi_\Omega(x_m)}{M} = \frac{\sum_{m=1}^M 1}{M} = \frac{M}{M} = 1.$$

Per la terza proprietà, infine, osserviamo che, secondo la formula dell'Esempio 1.1.17, se A e B sono ***disgiunti***, cioè se $A \cap B = \emptyset$, allora $\chi_A + \chi_B = \chi_{A \cup B}$. Quindi

$$h_A^M + h_B^M = \frac{\sum_{m=1}^M \chi_A(x_m)}{M} + \frac{\sum_{m=1}^M \chi_B(x_m)}{M} = \frac{\sum_{m=1}^M \big(\chi_A(x_m) + \chi_B(x_m)\big)}{M} = \frac{\sum_{m=1}^M \chi_{A \cup B}(x_m)}{M} = h_{A \cup B}^M.$$

Queste tre proprietà vogliamo richiedere anche dalle probabilità.

1.2.2 Definizione. Sia $\Omega \neq \emptyset$ uno spazio campionario e $\mathcal{A} \subset \mathcal{P}(\Omega)$ un'algebra booleana di eventi di Ω. Una funzione $P\colon \mathcal{A} \to \mathbb{R}$ chiameremo una ***probabilità booleana*** se soddisfa le seguenti condizioni:

1. $P(A) \geq 0$ per ogni $A \in \mathcal{A}$. (P è ***positiva***.)
2. $P(\Omega) = 1$. (P è ***normalizzata***.)
3. Se $A \cap B = \emptyset$, allora $P(A \cup B) = P(A) + P(B)$. ($P$ è ***additiva***.)

Chiameremo la terna $(\Omega, \mathcal{A}, P)$ uno ***spazio di probabilità booleano***. Chiameremo un evento $A \in \mathcal{A}$ con $P(A) = 1$ ***un evento sicuro***, e chiameremo un evento $A \in \mathcal{A}$ con $P(A) = 0$ ***un evento impossibile***.

1.2.3 Esempio. Chiaramente la funzione

$$A \longmapsto h_A^M$$

che assegna ad ogni evento A di Ω la sua frequenza relativa rilevata in M tentativi, definisce una probabilità su una qualsiasi algebra booleana $\mathcal{A}$ di eventi di Ω. (Abbiamo motivato la nostra definizione di probabilità booleana proprio servendoci delle proprietà delle frequenze relative.) Chiameremo tale probabilità la ***distribuzione empirica*** dei dati $x_1, \ldots, x_M \in \Omega$. Notiamo che non è neanche necessario che Ω sia un insieme finito.

Come ci sono "strane" algebre booleane come quella dell'Esempio 1.1.15, ci sono anche "strane" probabilità booleane su esse.

1.2.4 Esempio. Nell'Esempio 1.1.15 abbiamo consideato lo spazio campionario $\Omega = \mathbb{N}$ e l'algebra booleana $\mathcal{A} = \{A \subset \mathbb{N}\colon \#A < \infty \text{ o } \#A^{\complement} < \infty\}$. Su $\mathcal{A}$ definiamo una funzione P come

$$P(A) = \begin{cases} 0 & \#A < \infty, \\ 1 & \#A^{\complement} < \infty. \end{cases}$$

L'unica proprietà di una probabilità in questione è l'additività. Siano A e B eventi di $\mathcal{A}$ disgiunti. Non ci sono problemi se almeno uno dei due è finito. (Se, per esempio, A è finito, allora $A \cup B$ e B o sono finiti tutti e due, o sono infinti tutti e due. In ogni caso vale $P(A \cup B) = P(B) = P(B) + 0 = P(B) + P(A)$.) Che succede se tutti e due sono infiniti? Supponiamo che A e B siano eventi con complementari finiti. Allora anche $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$ è un insieme finito. In particolare, il complementare $A \cap B$ di $(A \cap B)^{\complement}$ non è vuoto. Non esistono eventi A e B di $\mathcal{A}$ che siano infiniti è disgiunti. Quindi P è una probabilità booleana su $\mathcal{A}$.

Prima di discutere più esempi, traiamo subito alcune conseguenze generali direttamente dalla nostra definizione.

1.2.5 Osservazione. 1. $A \cap A^{\complement} = \emptyset$, quindi $\boxed{1 = P(\Omega) = P(A \cup A^{\complement}) = P(A) + P(A^{\complement}).}$

2. $1 = P(\Omega) = P(\emptyset) + P(\Omega) = P(\emptyset) + 1$, quindi $\boxed{0 = P(\emptyset).}$ (Oppure: $P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset)$, quindi $0 = P(\emptyset)$.)

3. $0 \leq P(A^{\complement})$, allora $P(A) \leq P(A) + P(A^{\complement}) = 1$. Ne segue $\boxed{0 \leq P(A) \leq 1}$ per ogni $A \in \mathcal{A}$.

N.B.: Chi dovesse mai calcolare una probabilità **non compresa fra** 0 **e** 1, dovrebbe almeno far notare che qualcosa è andato storto! Errori di calcolo possono succedere, errori nella comprensione dei fatti più basilari no.

N.B.: L'evento sicuro Ω è un evento sicuro; l'evento impossibile è un evento impossibile. Non vale, però, il contrario. Se un evento è impossibile o meno dipende dalla probabilità P, mentre l'unico evento di Ω che è impossibile per qualsiasi probabilità booleane su $\mathcal{A} \subset \mathcal{P}(\Omega)$, è proprio $\emptyset$. Che un evento A abbia la probabilità $P(A) = 0$, non significa che non si potesse verificare! Infatti, nella Sezione 5 vedremo molte probabilità booleane per i quali ogni evento elementare ha la probabilità zero, nonostante in ogni verifica per forza si verifica uno fra questi.

1.2.6 Proposizione. *Se $A_1, \ldots, A_m \in \mathcal{A}$ sono eventi **due a due disgiunti**, cioè se $k \neq \ell \Longrightarrow A_k \cap A_\ell = \emptyset$, allora*

$$P(A_1 \cup \ldots \cup A_m) \;=\; P(A_1) + \ldots + P(A_m). \qquad \textit{(**additività finita**)}$$

Dimostrazione. Il ragionamento è più o meno questo: Per ogni k gli eventi A_k e $A_1 \cup \ldots \cup A_{k-1}$ sono disgiunti, perché

$$A_k \cap (A_1 \cup \ldots \cup A_{k-1}) \;=\; (A_k \cap A_1) \cup \ldots \cup (A_k \cap A_{k-1}) \;=\; \emptyset \cup \ldots \cup \emptyset \;=\; \emptyset.$$

Quindi

$$\begin{aligned} P(A_1 \cup \ldots \cup A_m) &= P(A_1 \cup \ldots \cup A_{m-1}) + P(A_m) \\ &= P(A_1 \cup \ldots \cup A_{m-2}) + P(A_{m-1}) + P(A_m) \\ &= \ldots = \\ &= P(A_1) + \ldots + P(A_{m-1}) + P(A_m). \end{aligned}$$

Un'argomentazione precisa richiederebbe, in tutte e due le parti, una dimostrazione per induzione. ■

Se conosciamo $P(A)$ e $P(B)$ e se A e B sono disgiunti, allora conosciamo anche $P(A \cup B)$. Il nostro prossimo scopo è calcolare $P(A \cup B)$ per eventi arbitrari, se conosciamo anche $P(A \cap B)$. (Motivazione: *Diagramma di Venn*.) Dimostriamo prima:

1.2.7 Proposizione. $P(A\backslash B) = P(A) - P(A \cap B)$.

Dimostrazione. Notiamo che $A\backslash B = A \cap B^{\complement}$ and $A \cap B$ dividono A in due parti disgiunte. Infatti,

$$\boxed{(A \cap B^{\complement}) \cup (A \cap B) \;=\; A \cap (B^{\complement} \cup B) \;=\; A}$$

e

$$\boxed{(A \cap B^{\complement}) \cap (A \cap B) \;=\; A \cap B^{\complement} \cap B \;=\; \emptyset.}$$

In altre parole, $P(A) = P((A \cap B^{\complement}) \cup (A \cap B)) = P(A\backslash B) + P(A \cap B)$. ■

1.2.8 Corollario. $A \subset B \Longrightarrow P(A) \leq P(B)$.

Dimostrazione. Abbiamo $P(B) = P(B \cap A) + P(B \cap A^{\complement}) \geq P(B \cap A) = P(A)$. ■

Usando anche la Proposizione 1.1.7, troviamo una formula per $P(A \cup B)$.

1.2.9 Corollario. *$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ per ogni $A, B \in \mathcal{A}$.*

Dimostrazione. Abbiamo $P(A \cup B) = P((A \cup B) \cap B) + P((A \cup B) \cap B^{\complement}) = P(B) + P((A\backslash B) \cup \emptyset) = P(B) + P(A) - P(A \cap B)$. ■

1.2.10 Esercizio. Dimostrare, usando solo gli assiomi della probabilità (cioè $P(A) \geq 0$, $P(\Omega) = 1$ ed $A \cap B = \emptyset$ implica $P(A \cup B) = P(A) + P(B)$) e le loro conseguenze, che

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

La suddivisione della probabilità $P(A)$ di A in una parte $P(A \cap B)$ corrispondente alla parte di A che appartiene a B e una parte $P(A \cap B^{\complement})$ corrispondente alla parte di A appartiene a $B^{\complement}$ ammette una generalizzazione. (Motivazione: *Diagramma di Venn*.)

1.2.11 Teorema. (Prima forma della formula della probabilità totale.) *Sia $B_1, \ldots, B_m \in \mathcal{A}$ una **partizione** di Ω, cioè i B_k sono due a due disgiunti e $\Omega = B_1 \cup \ldots \cup B_m$. Allora per ogni $A \in \mathcal{A}$ vale*

$$P(A) \;=\; P(A \cap B_1) + \ldots + P(A \cap B_m).$$

N.B.: La formula $P(A) = P(A \cap B) + P(A \cap B^{\complement})$ è contenuta come caso particolare per la partizione $B, B^{\complement}$ di Ω.

Dimostrazione. Se $B_k \cap B_\ell = \emptyset$ per ogni $k \neq \ell$, allora vale anche $(A \cap B_k) \cap (A \cap B_\ell) = A \cap B_k \cap B_\ell = \emptyset$. Quindi, secondo la Proposizione 1.2.6 abbiamo

$$\begin{aligned} P(A \cap B_1) + \ldots + P(A \cap B_m) &= P\big((A \cap B_1) \cup \ldots \cup P(A \cap B_m)\big) \\ &= P\big(A \cap (B_1 \cup \ldots \cup B_m)\big) = P(A \cap \Omega) = P(A). \end{aligned}$$ ■

In una gran parte degli esempi, lo spazio campionario Ω è finito e la probabilità P è definita su tutto $\mathcal{P}(\Omega)$.

1.2.12 Proposizione. *Sia $\Omega = \{\omega_1, \ldots, \omega_N\}$ uno spazio campionario finito, e siano $p_\omega \geq 0$ ($\omega \in \Omega$) numeri positivi tali che*

$$\sum_{n=1}^{N} p_\omega = 1.$$

Allora

$$P(A) := \sum_{n=1}^{N} \chi_A(\omega_n) p_{\omega_n}$$

definisce una probabilità booleana su $\mathcal{P}(\Omega)$ (oppure su una qualsiasi sottoalgebra booleana $\mathcal{A}$ di eventi di Ω). Inoltre, P è l'unica probabilità booleana definita su $\mathcal{P}(\Omega)$ che soddisfa $P(\{\omega\}) = p_\omega$ per ogni $\omega \in \Omega$.

Dimostrazione. La dimostrazione che P soddisfi la Definizione 1.2.2 di probabilità booleana, funziona esattamente come la discussione per le frequenze relative che precede la definizione. Visto che ogni evento di uno spazio campionario finito è l'unione disgiunta di un numero finito di eventi elementari, si deduce che P è già determinata univocamente dalle probabilità $P(\{\omega\})$ di suoi eventi elementari. Ovviamente, $P(\{\omega\}) = \sum_{n=1}^{N} \chi_{\{\omega\}}(\omega_n) p_{\omega_n} = p_\omega$. ■

1.2.13 Esempio. Sia $\Omega = \{b, r\}$ lo spazio campionario dell'urna (N palline, di cui n bianche e le altre rosse) come discusso nell'Esempio (I.2) dell'introduzione. Supponiamo che che la probabilità che esca una pallina bianca sia $p_b = \frac{n}{N}$ mentre quella di trovare una pallina rossa sia $p_r = \frac{N-n}{N}$. Verifichiamo che $p_b + p_r = \frac{N-n+n}{N} = 1$. Infatti, è facile controllare che $P(\emptyset) = 0$, $P(\{b\}) = p_b$, $P(\{r\}) = p_r$ e $P(\{b, r\}) = 1$ definisce una probabilità booleana su $\mathcal{P}(\{b, r\})$. È anche facile vedere che ogni esperimento con solo due esiti, uno descritto dall'evento A e l'altro dall'evento $A^\complement$, può essere modellato in questo modo sull'algebra booleana $\beta(A)$ generata da $A \neq \emptyset, \Omega$, assegnando ad A una probabilità $P(A) := p \in [0, 1]$ ed ad $A^\complement$ la probabilità $P(A^\complement) := 1 - p$. Un esperimento con solo due esiti (non banali, ossia, nessuno degli esiti corrisponde ad un evento impossibile o ad un evento sicuro) si chiama ***esperimento bernoulliano***. Spesso usiamo l'abbreviazione $q := 1 - p$.

1.2.14 Esempio. Fra gli esempi più basilari di tutta la probabilità, sono quelli con uno spazio campionario finito $\Omega = \{\omega_1, \dots, \omega_N\}$ in cui nessun risultato $\omega \in \Omega$ appare essere favorito dall'esperimento rispetto ad ogni altro risultato $\omega' \in \Omega$. Esempi sono il lancio di una moneta (Esempio (I.1)) o di un dado (Esempio (I.3)) dove nessun lato $\omega \in \{t, c\}$ della moneta o $\omega \in \{1, \dots, 6\}$ del dado dovrebbe risultare *più probabile* di un altro. Visto che i risultati dovrebbero risultare equiprobabili (ossia, $p_\omega = p$ dove la costante p non dipende da ω), la normalizzazione $1 = \sum_{n=1}^N p_{\omega_n} = \sum_{n=1}^N p = Np$ ci costringe a porre $p = \frac{1}{N}$. Per la moneta troviamo $p_t = p_c = \frac{1}{2}$, mentre per il dado otteniamo $p_k = \frac{1}{6}$ per ogni $k \in \{1, \dots, 6\}$. In genere, per la probabilità di un evento $A \subset \Omega$ troviamo

$$P(A) = \sum_{n=1}^N \chi_A(\omega_n) p_{\omega_n} = \sum_{n=1}^N \frac{\chi_A(\omega_n)}{N} = \frac{\#A}{\#\Omega} = \frac{\text{numero dei casi favorevoli}}{\text{numero totale dei casi}}.$$

Chiamiamo uno spazio di probabilità booleano finito $(\Omega, \mathcal{P}(\Omega), P)$ probabilizzato con $P(A) = \frac{\#A}{\#\Omega}$ uno ***spazio di probabilità elementare***. Il problema di calcolare le probabilità di eventi di uno spazio di probabilità elementare è così ridotto ad un problema di conteggio, cioè ad un problema del *calcolo combinatorio*. Discuteremo più problemi di questo tipo nel Capitolo 2

1.2.15 Esempio. Torniamo all'urna dell'Esempio 1.2.13. Ma questa volta come spazio campionario scegliamo le palline individuali. In altre parole, enumeriamo le N palline, e lo facciamo in modo che prima vengono da 1 ad n le n palline bianche e poi da $n + 1$ fino ad N le $N - n$ rosse. Lo spazio campionario diventa così $\Omega' = \{1, \dots, N\}$, l'evento "bianco" diventa $A =$ "$\omega \le n$" $= \{1, \dots, n\}$ ed il suo complementare "rosso" diventa $A^\complement =$ "$\omega > n$" $= \{n + 1, \dots, N\}$. È naturale di supporre che le palline individuali sono equiprobabili (il colore non costituisce una proprietà fisica che possa influenzare l'estrazione). Quindi consideriamo Ω' probabilizzato come spazio di probabilità elementare.

Come già nell'Esempio 1.2.13 troviamo $P($"bianco"$) = P(A) = \frac{\#A}{\#\Omega'} = \frac{n}{N}$ e $P($"rosso"$) = P(A^\complement) = \frac{\#A^\complement}{\#\Omega'} = \frac{N-n}{N}$. Solo che adesso il modello esprime molto chiaramente, perché le probabilità di "bianco" e "rosso" sono proprio queste. Nessuno mette in dubbio che le palline individuali debbano essere equiprobabili. Un qualsiasi intuito, che ci fa credere che le probabilità che abbiamo indovinate nell'Esempio 1.2.13 siano quelle giuste, non reggerebbe senza un'argomentazione in cui alla fine si fa riferimento all'equiprobabilità delle palline individuali.

Nel seguente esercizio bisogna prima chiarire il modello. Quale sarebbe lo spazio campionario che possiamo considerare uno spazio di probabilità elementare? Quali sono gli eventi su

cui abbiamo informazioni, quali sono invece gli eventi su cui vorremmo acquisire informazioni? Sarebbe forse opportuno di prendere in considerazione anche sottospazi dello spazio che abbiamo deciso di considerare?

1.2.16 Esercizio. Supponiamo che in un giorno di pioggia ognuno porti o l'ombrello o l'impermeabile. 60% delle ragazze portano l'ombrello. Invece, 80% dei ragazzi portano l'impermeabile. D'avanti un'aula con 150 studenti si trovano 70 ombrelli. Qual è la probabilità che una persona scelta a caso sia una studentessa?

1.3 Spazi di probabilità discreti

Nella Proposizione 1.2.12 abbiamo visto che possiamo probabilizzare uno spazio campionario finito $\Omega = \{\omega_1, \ldots, \omega_N\}$, assegnando degli numeri positivi $p_\omega \geq 0$ come probabilità degli eventi elementari $P(\{\omega\})$. Basta che questi numeri abbiano la sommatoria $\sum_{n=1}^{N} p_{\omega_n}$ uguale ad 1. In tal caso, l'algebra booleana (contentente tutti gli eventi elementari) è necessariamente $\mathcal{P}(\Omega)$, e la probabilità di un evento $A \subset \Omega$ qualsiasi è necessariamente

$$P(A) \;=\; \sum_{n=1}^{N} \chi_A(\omega_n) p_{\omega_n}. \tag{1.3.1}$$

In questa sezione vogliamo fare lo stesso per spazi campionari arbitrari.

Prima di farlo, occorre che ci inventiamo una notazione più *agevole* e *intuitiva* per le sommatorie di tipo (1.3.1) che funzionerà anche per spazi campionari arbitrari. Intanto, la sommatoria non fa altro che addizionare addendi di tipo $\chi_A(\omega)p_\omega$ su tutti gli elementi $\omega = \omega_n$ ($n = 1, \ldots, N$), oppure su tutti gli elementi ω di $\Omega = \{\omega_1, \ldots, \omega_N\}$. Non sarebbe più opportuno di usare un notazione come

$$P(A) \;=\; \sum_{\omega \in \Omega} \chi_A(\omega) p_\omega$$

che esprime chiaramente il fatto che stiamo addizionando quantità che dipendono da $\omega \in \Omega$ su tutti gli elementi ω di Ω, liberandoci così anche dell'indice di sommazione n anche rendendo la formula più leggibile? Infatti, l'affermazione che un insieme Ω (non vuoto) abbia la cardinalità $N \in \mathbb{N}$ non significa altro che esiste una biezione $f\colon \{1, \ldots, N\} \to \Omega$. (Si veda l'inizio dell'Appendice A, dove ci sono alcuni richiami alla cardinalità di insiemi.) Scrivendo $\Omega = \{\omega_1, \ldots, \omega_N\}$ come elenco dei suoi elementi $\omega_1, \ldots, \omega_N$ non significa alto che aver fissato la biezione f in tal modo che $\omega_n = f(n)$ per ogni $n = 1, \ldots, N$. (Tacitamente, si assume sempre che in un elenco $\omega_1, \ldots, \omega_N$ degli elementi di un insieme Ω, ognun elemento di Ω appare una ed una sola volta. Altrimenti la funzione f non sarebbe biettiva, e la cardinalità di Ω sarebbe strettamente più piccola di N.)

1.3.1 Definizione. Sia $\Omega \neq \emptyset$ un insieme finito di cardinalità $\#\Omega = N \in \mathbb{N}$, e siano a_ω ($\omega \in \Omega$) numeri (o altre cose che si possono addizionare come, per esempio, matrici). Allora

$$\sum_{\omega\in\Omega} a_\omega \;:=\; \sum_{n=1}^{N} a_{f(n)},$$

dove $f\colon \{1,\dots,N\} \to \Omega$ è una biezione qualsiasi. Per l'insieme vuoto, ovviamente si pone $\sum_{\omega\in\emptyset} a_\omega := 0$.
N.B.: Un elenco $\Omega = \{\omega_1,\dots,\omega_N\}$ corrisponde alla biezione $f(n) = \omega_n$. Se ci è dato un elenco continuiamo usare la notazione $\sum_{n=1}^{N} a_{\omega_n} = \sum_{n=1}^{N} a_{f(n)}$.

È importante notare che tale definizione non dipende dalla scelta della biezione: Per ogni altra biezione $g\colon \{1,\dots,N\} \to \Omega$ vale $\sum_{n=1}^{N} a_{f(n)} = \sum_{n=1}^{N} a_{g(n)}$. La dimostrazione per induzione di questo fatto non è per niente difficile, ma è un po' fastidiosa in che richiede un certo allenamento come scrivere una dimostrazione. La lasciamo a studenti di matematica.

Notiamo che con Ω, anche $A \subset \Omega$ è un insieme finito. Osserviamo che la sommatoria nella (1.3.1), veramente, calcola la somma solo sugli elementi di A: Grazie alla indicatrice di A, i contributi degli elementi $\omega \notin A$ sono 0 tutti quanti. Ci aspettiamo, quindi, che possiamo calcolare $P(A)$ anche come

$$P(A) \;=\; \sum_{\omega\in A} p_\omega.$$

Infatti, la seguente proposizione lo conferma.

1.3.2 Proposizione. *Sia $\Omega \neq \emptyset$ un insieme finito e sia A un suo sottoinsieme. Allora*

$$\sum_{\omega\in\Omega} \chi_A(\omega) a_\omega \;=\; \sum_{\omega\in A} a_\omega.$$

Lasciamo anche la dimostrazione (altrettanto fastidiosa) di questa proposizione agli studenti di matematica.

1.3.3 Nota. Per calcolare $P(A)$ come $\sum_{\omega\in\Omega} \chi_A(\omega) p_\omega$, bisogna elencare gli elementi di Ω come $\omega_n = f(n)$ tramite una biezione $f\colon \{1,\dots,N\} \to \Omega$ dove $N = \#\Omega$, e addizionare i numeri $\chi_A(\omega_n) p_{\omega_n}$ per $n = 1,\dots,N$. Mentre per calcolare $P(A)$ come $\sum_{\omega\in A} p_\omega$, bisogna elencare gli elementi di A come $w_\ell = g(\ell)$ tramite una biezione $g\colon \{1,\dots,L\} \to A$ dove $L = \#A$, e addizionare i numeri p_{w_ℓ} per $\ell = 1,\dots,L$. A prima vista sembrerebbe più facile, addizionare subito i numeri p_{w_ℓ} giusti sull'insieme giusto senza far uso dell'indicatrice. Però, rimane il fatto che per calcolare la prima sommatoria per tutti gli eventi A, basta elencare gli elementi di Ω una sola volta, mentre per calcolare la seconda somma, occorre trovare per ogni evento A un suo elenco individuale. Inoltre, solo con l'indicatrice riusciamo dare una dimostrazione facile facile dell'additività di P.

Finora ci siamo occupati di cambiare un po' le formule con cui calcolare le probabilità di eventi di uno spazio di probabilità finito, probabilizzato secondo la Proposizione 1.2.12. Più precisamente, abbiamo introdotto la notazione di sommatoria su tutti gli elementi di un insieme, comunque ancora di un insieme finito, in cui non appare più nessun elenco degli elementi, e in cui siamo riusciti a liberarci di un indice che rendeva non troppo leggibile le formule. Adesso vogliamo passare a sommatorie su numeri indicati da elementi di insiemi non più necessariamente finiti. Dobbiamo, quindi, dare un significato ad una sommatoria come

$$\sum_{\omega\in\Omega} a_\omega$$

per un'insieme Ω arbitrario, usando le sommatorie su insiemi finiti come appena studiate.

Ci sono due possibilità di procedere. La prima possibilità riduce tutto alle serie $\sum_{n=1}^{\infty} a_n$ come discusse nell'analisi. Le Appendici A.2 e A.3 presentano alcuni richiami alle serie. Soprattutto, l'Appendice A.3 sottolinea l'importanza delle serie assolutamente convergenti. La seconda possibilità da una definizione diretta del concetto di convergenza di una sommatoria $\sum_{\omega\in\Omega} a_\omega$. Però, sin dall'inizio tale concetto è un concetto di convergenza assoluta. Solo dopo, come regola di calcolo, diamo il collegamento con le serie $\sum_{n=1}^{\infty} a_n$ assolutamente convergenti.

1.3.4 Definizione. Sia $\Omega \neq \emptyset$ e siano a_ω ($\omega \in \Omega$) numeri. Per un numero a diciamo

$$\sum_{\omega\in\Omega} a_\omega \;=\; a,$$

se per ogni $\varepsilon > 0$ esiste un sottoinsieme finito Ω_ε di Ω, tale che

$$\Big|a - \sum_{\omega\in\Omega'} a_\omega\Big| \;\leq\; \varepsilon$$

per ogni insieme finito Ω' con $\Omega_\varepsilon \subset \Omega' \subset \Omega$.

Se esiste un tale numero a, lo chiamiamo il ***limite*** della sommatoria $\sum_{\omega\in\Omega} a_\omega$. Diciamo la sommatoria ***converge*** (ad a) e la sommatoria è ***convergente***. Diciamo $\sum_{\omega\in\Omega} a_\omega$ è ***assolutamente convergente***, se $\sum_{\omega\in\Omega} |a_\omega|$ è convergente. Scriviamo anche $\sum_{\omega\in\Omega} |a_\omega| < \infty$ se converge, e $\sum_{\omega\in\Omega} |a_\omega| = \infty$ se non converge.

Come per le successioni, se esiste un limite, allora è unico. Valgono anche altre proprietà note dalle serie, come

$$\sum_{\omega\in\Omega} a_\omega \;=\; a, \quad \sum_{\omega\in\Omega} b_\omega \;=\; b \quad\Longrightarrow\quad \sum_{\omega\in\Omega} (a_\omega + c b_\omega) \;=\; a + cb. \tag{1.3.2}$$

Notiamo, che nel caso che Ω fosse già finito, le Definizioni 1.3.1 e 1.3.4 danno a $\sum_{\omega\in\Omega} a_\omega$ lo stesso valore.

1.3.5 Proposizione. *Per numeri reali a_ω le seguenti proprietà sono equivalenti:*

1. *$\sum_{\omega\in\Omega} a_\omega$ è convergente.*

2. *$\sum_{\omega\in\Omega} a_\omega$ è assolutamente convergente.*

3. *Esiste una costante M tale che*
$$\sum_{\omega\in\Omega'} |a_\omega| \;\leq\; M$$
per ogni sottoinsieme finito Ω' di Ω.

Nel caso di convergenza, il limite di $\sum_{\omega\in\Omega} |a_\omega|$ coincide con la più piccola costante M_0 che soddisfa la terza condizione.

Bozza di una dimostrazione. Se esiste una costante M come in (3), allora anche l'infimo M_0 di tutte queste costanti possibili è una costante possibile. È facile vedere che tale infimo coincide con il supremo dei numeri $\sum_{\omega\in\Omega'} |a_\omega|$ su tutti i sottoinsiemi finiti Ω' di ω, e che tale supremo è il limite della sommatoria $\sum_{\omega\in\Omega} |a_\omega|$. *Vice versa*, se esiste questo limite, tale limite è senz'altro non solo una costante possibile, ma proprio la più piccola. Abbiamo, quindi, non solo l'equivalenza di (2) e (3), ma anche l'affermazione su M_0.

Non è difficile, vedere che una sommatoria non assolutamente convergente, non può neanche essere convergente. Se, invece, la sommatoria è assolutamente convergente, allora per ogni $\varepsilon > 0$ esiste un sottoinsieme finito Ω_ε di Ω, tale $0 \leq M_0 - \sum_{\omega\in\Omega'} |a_\omega| \leq \varepsilon$, ossia $\sum_{\omega\in(\Omega'\setminus\Omega_\varepsilon)} |a_\omega| \leq \varepsilon$, per ogni Ω' finito con $\Omega_\varepsilon \subset \Omega' \subset \Omega$. Con ciò, anche

$$\Big|\sum_{\omega\in\Omega'_1} a_\omega - \sum_{\omega\in\Omega'_2} a_\omega\Big| \;\leq\; \sum_{\omega\in(\Omega'_1\triangle\Omega'_2)} |a_\omega| \;\leq\; \varepsilon \tag{1.3.3}$$

per ogni scelta $\Omega_\varepsilon \subset \Omega'_i \subset \Omega$ ($i = 1, 2$). (N.B.: Si dice che la famiglia $\left(\sum_{\omega\in\Omega'} a_\omega\right)_{\Omega'\subset\Omega,\#\Omega'<\infty}$ è una *rete di Cauchy* che, grazie alla completezza di $\mathbb{R}$, ha un limite.) Non è troppo difficile convincersi da questo, che la sommatoria $\sum_{\omega\in\Omega} a_\omega$ abbia un limite. ■

1.3.6 Corollario. *Se $\sum_{\omega\in\Omega} |a_\omega| < \infty$ e $|b_\omega| \leq |a_\omega|$, allora*

$$\Big|\sum_{\omega\in\Omega} b_\omega\Big| \;\leq\; \sum_{\omega\in\Omega} |b_\omega| \;\leq\; \sum_{\omega\in\Omega} |a_\omega|\,.$$

1.3.7 Corollario. *Se $\sum_{\omega\in\Omega} |a_\omega| < \infty$, allora*

$$\sum_{\omega\in\Omega} \chi_A(\omega) a_\omega \;=\; \sum_{\omega\in A} a_\omega$$

per ogni $A \subset \Omega$.

DIMOSTRAZIONE. Per ogni sottoinsieme finito Ω' di Ω vale

$$\sum_{\omega\in\Omega'} \chi_A(\omega)a_\omega \;=\; \sum_{\omega\in(A\cap\Omega')} a_\omega.$$

Quindi, scegliendo Ω' abbastanza grande, tutti e due i lati dell'equazione sono distanti non più di un ε dai loro limiti. Diminuendo ε, i limiti no possono essere che uguali. ■

Adesso facciamo il collegamento fra la sommatoria $\sum_{\omega\in\Omega} a_\omega$ e le serie (assolutamente convergenti) che conosciamo dall'analisi (Appendice A.3).

1.3.8 Lemma. *Sia Ω un insieme infinito e siano a, a_ω ($\omega \in \Omega$) numeri reali. Le seguenti proprietà sono equivalenti:*

1. *$\sum_{\omega\in\Omega} a_\omega = a$.*
2. *Esiste una funzione iniettiva $f\colon \mathbb{N} \to \Omega$ tale che $a_\omega = 0$ per ogni $\omega \notin f(\mathbb{N}) := \{f(n)\colon n \in \mathbb{N}\}$ e la serie $\sum_{n=1}^{\infty} a_{f(n)}$ converge assolutamente ad a.*
3. *Esiste una funzione iniettiva $f\colon \mathbb{N} \to \Omega$ tale che $a_\omega = 0$ per ogni $\omega \notin f(\mathbb{N}) := \{f(n)\colon n \in \mathbb{N}\}$, e per qualsiasi tale funzione la serie $\sum_{n=1}^{\infty} a_{f(n)}$ converge semplicemente ad a.*

N.B.: Il lemma ci dice che nel caso di convergenza di $\sum_{\omega\in\Omega} a_\omega$, è possibile scrivere una successione $\omega_1, \omega_2, \ldots$ non-ripetitiva di elementi di Ω (infatti, $\omega_n = f(n)$) tale che tutti gli a_ω sono 0 all'infuori del sottoinsieme $\{\omega_1, \omega_2, \ldots\}$ di Ω, e tale che

$$\sum_{\omega\in\Omega} a_\omega \;=\; \sum_{n=0}^{\infty} a_{\omega_n}.$$

Inoltre, ci dice che se troviamo una tale successione e la serie a destra converge assolutamente, allora esiste anche la sommatoria a sinistra, ed il suo valore non dipende, quindi, dalla successione scelta.

DIMOSTRAZIONE DEL LEMMA. Per prima cosa osserviamo che (3) implica (2). Infatti, se per una scelta f la serie $\sum_{n=1}^{\infty} a_{f(n)}$ convergesse semplicemente ma non assolutamente verso a, allora secondo il *piccolo teorema sugli riordinamenti* (Nota A.3.2) per qualsiasi $a' \neq a$ esisterebbe una biezione g su $\mathbb{N}$, tale che $\sum_{n=1}^{\infty} a_{f\circ g(n)} = a'$, contraddicendo le ipotesi di (3).

Per qualsiasi sottoinsieme Ω_0 di Ω tale che $a_\omega = 0$ per ogni $\omega \notin \Omega_0$, dal Corollario 1.3.7 segue che $\sum_{\omega\in\Omega} a_\omega = \sum_{\omega\in\Omega} \chi_{\Omega_0}(\omega)a_\omega = \sum_{\omega\in\Omega_0} a_\omega$.

Valga (2) e poniamo $\Omega_0 := f(\mathbb{N})$ e $M := \sum_{n=1}^{\infty} |a_{f(n)}|$. Per ogni sottoinsieme finito Ω' di Ω_0 si trova N, tale che $f(\{1, \ldots, N\}) = \{f(1), \ldots, f(N)\} \supset \Omega'$. (Infatti, per ogni $\omega \in \Omega'$ esiste n_ω tale che $f(n_\omega) = \omega$. Ogni $N \geq \max\{n_\omega\colon \omega \in \Omega'\}$ è, quindi, un numero N

opportuno.) Allora $\sum_{\omega\in\Omega'} |a_\omega| \le \sum_{n=1}^{N} |a_{f(n)}| \le M$. Secondo la Proposizione 1.3.5, $\sum_{\omega\in\Omega_0} a_\omega$ converge, e di conseguenza lo fa anche $\sum_{\omega\in\Omega} a_\omega$.

Poi ricordiamoci della definizione di numerabilità di un insieme; si veda Appendice A. Siano $p_\omega \ge 0$ numeri positivi tali che $\sum_{\omega\in\Omega} p_\omega = p < \infty$. Allora l'insieme

$$\Omega_0 \;:=\; \{\omega \in \Omega\colon\, p_\omega \neq 0\}$$

è numerabile. (Infatti, abbiamo $0 \le p_\omega \le M := \sum_{\omega\in\Omega} p_\omega$ per ogni $\omega \in \Omega$. Se per ogni $n \in \mathbb{N}$ definiamo l'insieme

$$\Omega_n \;=\; \Big\{\omega \in \Omega\colon\, \tfrac{M}{n+1} < p_\omega \le \tfrac{M}{n}\Big\},$$

allora ogni $\omega \in \Omega_0$ è contenuto in esattamente uno degli insiemi Ω_n. Ogni insieme Ω_n contiene al massimo un numero finito di elementi. Più precisamente, se $p_\omega \in \Omega_n$ allora $p_\omega \ge \frac{M}{n+1}$. Quindi, $M \ge \sum_{\omega\in\Omega_n} p_\omega \ge \frac{M\#\Omega_n}{n+1}$, ossia $\#\Omega_n \le n+1$. Allora Ω_0, come unione numerabile degli insiemi numerabili Ω_n, è un insieme numerabile.)

Valga (1), quindi, con $p_\omega := |a_\omega|$, l'insieme $\Omega_0 := \{\omega \in \Omega\colon\, a_\omega \neq 0\}$ è numerabile. Esiste, quindi, un'iniezione f tale che $f(\mathbb{N}) \supset \Omega_0$. Per una qualsiasi tale iniezione vale, quindi, $\sum_{n=1}^{N} p_{f(n)} \le p$. Secondo la Proposizione A.2.2, ciò significa che $\sum_{n=1}^{N} a_{f(n)}$ converge assolutamente.

Abbiamo dimostrato (3) $\Rightarrow$ (2), che la convergenza assoluta di (2) implica convergenza di (1), e che la convergenza di (1) implica convergenza di assoluta (3) par qualsiasi f. La dimostrazione che nelle ultime conclusioni i rispettivi limiti sono uguali (ciò significa, in particolare, che dato (1), il limite in (3) non dipende dalla scelta di f), funziona come nella Proposizione 1.3.5 usando una disuguaglianza come nella (1.3.3), e come li la lasciamo come bozza. ■

Adesso siamo pronti a generalizzare la probabilizzazione secondo la Proposizione 1.2.12 da spazi campionari finiti a spazi campionari arbitrari.

1.3.9 Teorema. *Sia $\Omega \neq \emptyset$ uno spazio campionario e siano $p_\omega \ge 0$ ($\omega \in \Omega$) numeri tali che $\sum_{\omega\in\Omega} p_\omega = 1$. Allora la formula*

$$P(A) \;:=\; \sum_{\omega\in\Omega} \chi_A(\omega) p_\omega \qquad \textit{(ossia}\quad P(A) \;=\; \sum_{\omega\in A} p_\omega \;\textit{)}$$

*definisce una probabilità booleana su $\mathcal{P}(\Omega)$ o su una qualsiasi algebra booleana $\mathcal{A}$ di eventi di Ω. Inoltre, $P(A)$ è l'unica probabilità booleana su $\mathcal{P}(\Omega)$ che soddisfa $P(\{\omega\}) = p_\omega$. Chiameremo uno spazio campionario numerabile probabilizzato in questo modo un **spazio di probabilità discreto**.*

Dimostrazione. L'esistenza di P come il fatto che $P(A) \ge 0$ segue da $0 \le \chi_A(\omega) p_\omega \le p_\omega$ per ogni $\omega \in \Omega$ grazie al Corollario 1.3.6. Il fatto $P(\Omega) = 1$ è parte delle ipotesi. Come tutte le altre volte, l'additività di P segue dalla formula dell'Esempio 1.1.17 grazie alla (1.3.2). Chiaramente $P(\{\omega\}) = p_\omega$ per ogni $\omega \in \Omega$.

Notiamo che fino a questo punto non abbiamo usato il Lemma 1.3.8. Adesso, però, ci tocca a dimostrare l'unicità di P. Scegliamo una biezione f come sotto Lemma 1.3.8(2) e definiamo $\omega_1, \omega_2, \ldots$ con $\omega_n = f(n)$. Il lemma ci dice che possiamo calcolare tutte sommatorie su tutti gli elementi calcolando delle serie su questi ω_n.

Supponiamo, quindi, che ci sia un'altra probabilità booleana Q su $\mathcal{P}(\Omega)$ soddisfacendo $Q(\{\omega\}) = p_\omega$ per ogni $\omega \in \Omega$. Abbiamo finito quando dimostreremo che $Q(A) \geq P(A)$ per tutti gli eventi. (Infatti, in questo caso varrebbe anche $Q(A^\complement) \geq P(A^\complement)$ ossia $Q(A) = 1 - Q(A^\complement) \leq 1 - P(A^\complement) = P(A)$). Chiaramente per ogni evento A finito vale $Q(A) = P(A)$. (Secondo il Teorema 1.2.11 se $A = \{w_1, \ldots, w_L\} = \{w_1\} \cup \ldots \cup \{w_L\}$, allora $Q(A) = p_{w_1} + \ldots + p_{w_L} = P(A)$.) Allora se A non è finito, definiamo $A_n = \{\omega_1, \ldots, \omega_n\} \cap A$. Chiaramente $A_n \subset A$. Quindi, secondo il Corollario 1.2.9

$$Q(A) \;\geq\; Q(A_n) \;=\; P(A_n) \;=\; \sum_{k=1}^{n} \chi_A(\omega_k) p_{\omega_k}.$$

Secondo il Lemma 1.3.8, $\sum_{k=1}^{\infty} \chi_A(\omega_k) p_{\omega_k} = P(A)$. La disuguaglianza precedente sussiste anche dopo il limite $n \to \infty$, quindi $Q(A) \geq P(A)$. ■

Ogni spazio di probabilità booleano $(\Omega, \mathcal{P}(\Omega), P)$ finito è probabilizzato necessariamente nel senso del Teorema 1.3.9 (oppure la Proposizione 1.2.12). Questo è dovuto al fatto che ogni evento A di uno spazio campionario finito è un'unione finita di eventi elementari. Vogliamo sapere fino a che punto lo stesso vale per uno spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$ dove Ω non è più finito. La seguente generalizzazione della formula della probabilità totale (Teorema 1.2.11) non solo ci permetterà rispondere alla domanda quale condizione deve rimpiazzare l'additività finita, ma ha anche delle applicazioni multiple in tutto il resto di questo testo.

Prima generalizziamo la nozione di partizione finita. Una ***partizione*** di $\Omega \neq \emptyset$ è una famiglia $(B_s)_{s\in S}$ di sottoinsiemi B_s di Ω tali che $s \neq s' \Rightarrow B_s \cap B_{s'} = \emptyset$ e $\bigcup_{s\in S} B_s = \Omega$. Scriviamo $(B_s)_{s\in S} \subset \mathcal{A}$ per indicare che gli elementi B_s della partizione sono di $\mathcal{A} \subset \mathcal{P}(\Omega)$.

1.3.10 Teorema. *1. Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità booleano arbitrario con una partizione $(B_s)_{s\in S} \subset \mathcal{A}$ di Ω, tale che $\sum_{s\in S} P(B_s) = 1$. Allora vale la seguente generalizzazione della formula della probabilità totale*

$$P(A) \;=\; \sum_{s\in S} P(A \cap B_s), \qquad (A \in \mathcal{A}). \tag{1.3.4}$$

2. Sia $(\Omega, \mathcal{P}(\Omega), P)$ uno spazio di probabilità discreto. Allora, qualunque partizione $(B_s)_{s\in S}$ di Ω soddisfa

$$\sum_{s\in S} P(B_s) \;=\; 1.$$

Dimostrazione. 1.) Sia S' un sottoinsieme finito di S. Come nella dimostrazione del Teorema 1.3.9, per ogni $A \in \mathcal{A}$ vale

$$\sum_{s\in S'} P(A \cap B_s) \;=\; P\Big(A \cap \bigcup_{s\in S'} B_s\Big) \;\le\; P(A),$$

di cui segue che $p_A := \sum_{s\in S} P(A \cap B_s)$ esiste e che non è più grande di $P(A)$. Allora

$$\begin{aligned} p_A + p_{A^{\complement}} \;&=\; \sum_{s\in S} P(A \cap B_s) + \sum_{s\in S} P(A^{\complement} \cap B_s) \\ &= \sum_{s\in S}\Big(P(A \cap B_s) + P(A^{\complement} \cap B_s)\Big) \;=\; \sum_{s\in S} P(B_s) \;=\; 1 \end{aligned}$$

ed, infine,

$$P(A) \;\ge\; p_A \;=\; 1 - p_{A^{\complement}} \;\ge\; 1 - P(A^{\complement}) \;=\; P(A),$$

di cui segue $p_A = P(A)$.

2.) Per ogni sottoinsieme finito S' di S, vale $\sum_{s\in S'} P(B_s) = P(\bigcup_{s\in S'} B_s) \le 1$. Quindi, la sommatoria $\sum_{s\in S} P(B_s)$ esiste è ha un limite non più grande di 1. Poi per ogni sottoinsieme finito Ω' di Ω esistono un sottoinsieme finito S' di S e per ogni $s \in S'$ un sottoinsieme finito B'_s di B_s tali che $(\bigcup_{s\in S'} B'_s) \supset \Omega'$. Ne segue che

$$P(\Omega') \;\le\; P\Big(\bigcup_{s\in S'} B'_s\Big) \;=\; \sum_{s\in S'} P(B'_s) \;\le\; \sum_{s\in S'} P(B_s) \;\le\; \sum_{s\in S} P(B_s).$$

Facendo crescere Ω', la probabilità $P(\Omega') = \sum_{\omega\in\Omega'} p_\omega$ si avvicina sempre più ad 1. Quindi, $\sum_{s\in S} P(B_s) \le P(\Omega')$ no può essere neanche più piccola di 1. ∎

1.3.11 Corollario. *In uno spazio di probabilità discreto vale la* (1.3.4) *per qualunque partizione* $(B_s)_{s\in S}$ *e per ogni sottoinsieme* $A \subset \Omega$.

Dopo la seguente definizione siamo pronti a caratterizzare gli spazi di probabilità discreti.

1.3.12 Definizione. La probabilità P di uno spazio di probabilità booleano $(\Omega, \mathcal{P}(\Omega), P)$ è ***assolutamente additiva*** (o si gode dell'***additività assoluta***), se per qualsiasi famiglia $(B_s)_{s\in S}$ di eventi due a due disgiunti vale

$$P\Big(\bigcup_{s\in S} B_s\Big) \;=\; \sum_{s\in S} P(B_s).$$

(Aggiungiamo che P si dice ***σ–additiva***, se la formula vale almeno per tutti i casi dove S è numerabile.)

1.3.13 Teorema. *Uno spazio di probabilità booleano* $(\Omega, \mathcal{P}(\Omega), P)$ *è discreto se e solo se P è assolutamente additiva.*

Dimostrazione. L'additività assoluta di P in uno spazio di probabilità discreto, segue semplicemente aggiungendo ad una famiglia di eventi disgiunti l'evento complementare dell'unione di tutti gli eventi della famiglia ed applicare il Teorema 1.3.10(2) a quella partizione.

Se, *vice versa*, P è assolutamente additiva, basta applicare il Teorema 1.3.10(1) alla partizione $(\{\omega\})_{\omega\in\Omega}$ di Ω. ■

Come nella dimostrazione del Teorema 1.3.10(2), con la riduzione ad insiemi finiti si dimostra anche il seguente risultato.

1.3.14 Corollario. *Siano $p_\omega \geq 0$ numeri positivi. Allora per qualunque partizione $(B_s)_{s\in S}$ di Ω vale $\sum_{\omega\in\Omega} p_\omega = \sum_{s\in S}(\sum_{\omega\in B_s} p_\omega)$, il valore ∞ incluso.*

Concludiamo con un teorema molto potente che, data una partizione, ci permetterà di aggiungere in modo consistente all'algebra booleana $\mathcal{A}$ tutti gli eventi che posseggano una "decomposizione in $\mathcal{A}$ rispetto alla partizione".

1.3.15 Teorema. *Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità booleano arbitrario con una partizione $\mathcal{P} = (B_s)_{s\in S} \subset \mathcal{A}$ di Ω, tale che $\sum_{s\in S} P(B_s) = 1$. Allora*

$$\mathcal{B}_\mathcal{P} \;:=\; \{A \subset \Omega\colon A \cap B_s \in \mathcal{A} (s \in S)\}$$

è un'algebra booleana su Ω e

$$Q_\mathcal{P}(A) \;:=\; \sum_{s\in S} P(A \cap B_s)$$

definisce una probabilità booleana su $\mathcal{B}_\mathcal{P}$. Inoltre, $Q_\mathcal{P}$ è l'unica probabilità booleana su $\mathcal{B}_\mathcal{P}$ che estende P (cioè, $Q_\mathcal{P}(A) = P(A)$ per ogni $A \in \mathcal{A}$).

Dimostrazione. Ovviamente, $A_1, A_2 \in \mathcal{B}_\mathcal{P}$, allora $(A_1 \cap A_2) \cap B_s = (A_1 \cap B_s) \cap (A_1 \cap B_s) \in \mathcal{A}$ per ogni $s \in S$, quindi, $A_1 \cap A_2 \in \mathcal{B}_\mathcal{P}$. Poi se $A \in \mathcal{B}_\mathcal{P}$, allora $A^\complement \cap B_s = (A \cap B_s)^\complement \cap B_s \in \mathcal{A}$, quindi, $A^\complement \in \mathcal{B}_\mathcal{P}$. Ne segue che $\mathcal{B}_\mathcal{P}$ è un'algebra booleana.

Visto che $0 \leq P(A \cap B_s) \leq P(B_s)$ per ogni $s \in S$, allora $Q_\mathcal{P}(A)$ esiste per ogni $A \in \mathcal{B}_\mathcal{P}$ e $0 \leq Q_\mathcal{P}(A) \leq 1$. Dal Teorema 1.3.10 sappiamo $P(A) = \sum_{s\in S} P(A \cap B_s) = Q_\mathcal{P}(A)$ per ogni $A \in \mathcal{A}$, in particolare, per $\Omega \in \mathcal{A}$ vale $Q_\mathcal{P}(\Omega) = P(\Omega) = 1$. Se $A_1, A_2 \in \mathcal{B}$ sono disgiunti, allora lo sono anche $A_1 \cap B_s$ e $A_1 \cap B_s$ per ogni $s \in S$. Ne segue

$$\begin{aligned} Q_\mathcal{P}(A_1 \cup A_2) \;&=\; \sum_{s\in S} P((A_1 \cup A_2) \cap B_s) \;=\; \sum_{s\in S} P((A_1 \cap B_s) \cup (A_2 \cap B_s)) \\ =\; \sum_{s\in S} \Big(P(A_1 \cap B_s) + P(A_2 \cap B_s)\Big) \;&=\; \sum_{s\in S} P(A_1 \cap B_s) + \sum_{s\in S} P(A_2 \cap B_s) \;=\; Q_\mathcal{P}(A_1) + Q_\mathcal{P}(A_2). \end{aligned}$$

In altre parole, $Q_{\mathcal{P}}$ è una probabilità booleana su $\mathcal{B}_{\mathcal{P}}$ che estende P. L'unicità, adesso, segue applicando Teorema 1.3.10 alla probabilità booleana $Q_{\mathcal{P}}$. ■

1.3.16 Corollario. *Sia $\mathcal{P}$ la famiglia di tutte le partizioni $\mathcal{P}$ di Ω che soddisfanno le ipotesi del teorema. Allora $\mathcal{B}_1 := \bigcup_{\mathcal{P}\in\mathcal{P}} \mathcal{B}_{\mathcal{P}}$ è un'algebra booleana e $Q_1(A) := Q_{\mathcal{P}}(A)$ per $A \in \mathcal{B}_{\mathcal{P}}$ definisce l'unica probabilità booleana su $\mathcal{B}_1$ che estende P.*

Dimostrazione. Basta osservare che se $(B_s)_{s\in S}$ e $(B'_{s'})_{s'\in S'}$ sono di $\mathcal{P}$, allora anche la partizione $(B_s \cap B'_{s'})_{(s,s')\in S\times S'}$ è di $\mathcal{P}$. ■

Note conclusive

Sappiamo che uno spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$ non necessariamente è probabilizzato secondo il Teorema 1.3.9, neanche se $\mathcal{A}$ contiene tutti gli eventi elementari $\{\omega\}$. Infatti, nell'Esempio 1.2.4 vale $\sum_{\omega\in\mathbb{N}} p_\omega = 0 \neq 1$. Però questa sommatoria significa soprattutto che P non è σ–additiva, perché l'insieme numerabile $\mathbb{N}$ è unione numerabile di tutti suoi eventi elementari.

Si potrebbe sospettare, che la mancata probablizzazione secondo il Teorema 1.3.9 possa dipendere dalla mancata σ–additività di P. Però non è così. Infatti, l'Esempio 1.2.4 si può modificare, passando ad $\mathcal{A} := \{A \subset \mathbb{R} \colon \#A \leq \#\mathbb{N} \text{ o } \#A^{\complement} \leq \#\mathbb{N}\} \subset \mathcal{P}(\mathbb{R})$ e $P(A) := 0$ se A è numerabile e $P(A) = 1$ se $A^{\complement}$ è numerabile. (Per dimostrare che P è σ–additiva, basta l'osservazione che l'unione numerabile di insiemi numerabili è un insieme numerabile.) Però anche questo esempio ci da $\sum_{\omega\in\mathbb{R}} p_\omega = 0 \neq 1$.

Poi si potrebbe sospettare, che mancata probablizzazione secondo il Teorema 1.3.9 possa dipendere dal fatto che l'algebra booleana $\mathcal{A}$ non è tutto $\mathcal{P}(\Omega)$. Però neanche questa è la ragione—benché è abbastanza difficile trovare esempi. Tali esempi si trovano cercando delle probabilità booleane su $\Omega = [0,1]^m$ invariate sotto *movimenti*. (Un movimento su $\mathbb{R}^m$ è, vagamente, una biezione che non "cambia le distanze" fra i punti. Si cerca una funzione positiva e additiva su tutti gli sottoinsiemi di $\mathbb{R}^m$ che non cambia valore dopo un movimento, e che da a $[0,1]^m$ il valore 1. Quindi, la restrizione è un probabilità booleana.)

- Vitali (1905): Non esiste soluzione σ–additiva per qualsiasi $m \in \mathbb{N}$.
- Hausdorff (1914): Non esiste soluzione per $m \geq 3$.
- Banach (1923): Esiste soluzione per $m < 3$, ma non è unica.

Notiamo che ogni soluzione del problema necessariamente soddisfa $p_\omega = 0$. (Se $p_\omega \neq 0$ allora $p_{\omega'} = p_\omega$ per ogni ω'. Allora ogni insieme $A \subset \Omega$ finito averebbe probabilità $\#A \cdot p_\omega$. Se A è abbastanza grande, superebbe 1.) Il teorema di Banach ci assicura, quindi, esempi.

Un'algebra booleana $\mathcal{A}$ è una ***σ–algebra*** se e solo se per ogni paritzione numerabile l'algebra booleana $\mathcal{B}$ del Teorema 1.3.15 coincide con $\mathcal{A}$. (Lo possiamo prendere come definizione.) È possibile limitarsi nell'ultimo corollario alle partizioni numerabili. All contrario di quello che si potrebbe sospettare, di solito C_1 non è una σ–algebra, perché $C_1 \supset \mathcal{A}$ può permetter più partizioni numerabili di $\mathcal{A}$. Si può dimostrare che nenanche iterando questo processo $\mathcal{A} \subset C_1 \subset C_2 \subset \ldots$, l'unione di tutti C_n sarà necessariamente una σ–algebra.

Capitolo 2

Spazi di probabilità elementari ed altre applicazioni del calcolo combinatorio

In questo capitolo discuteremo alcuni esempi di spazi di probabilità booleane che, per calcolare le loro probabilità, si servono in modo essenziale del calcolo combinatorio. In primo piano, ci sono gli spazi di probabilità elementari. Ma anche in altri spazi, alcuni di loro addirittura non più discreti, i problemi di conteggio (ossia, il calcolo combinatorio) continuano a giocare un ruolo centrale. Per i risultati del calcolo combinatorio ci riferiamo all'Appendice A.1.

2.1 Spazi di probabilità elementari

Prima di discutere degli esempi di spazi di probabilità elementari, presentiamo alcuni esercizi che dimostrano quanto sia importante la scelta dello spazio campionario.

2.1.1 Esercizio. (L'esercizio dell'introduzione e l'Esercizio 1.1.1 continuati.) Abbiamo realizzato il lancio di due dadi, con l'unico scopo di osservare la somma dei due dadi, su tre spazi campionari diversi Ω_a, Ω_b e Ω_c. Supponiamo che Ω_a sia uno spazio di probabilità elementare, cioè i dadi non sono truccati. Calcolare le probabilità degli eventi elementari nei casi (b) e (c) considerandoli eventi non necessariamente elementari di Ω_a..

2.1.2 Esercizio. Un certo Chevalier de Méré si meravigliò perché lanciando molte volte tre dadi si vedesse più spesso la somma 11 che la somma 12, nonostante il numero di combinazioni che realizzino 11 (cioè tutte le triplice 6-4-1, 6-3-2, 5-5-1, 5-4-2, 5-3-3, 4-4-3) fosse uguale a quello per realizzare 12. Dove sbaglia?

2.1.3 Esempio. L'estrazione del lotto 5 di 90. Da 90 palline enumerate verranno estratte 5. Quindi, alla fine abbiamo d'avanti a noi 5 palline che mancano al contenitore (l'*urna*). Delle

5 palline ci interessano solo i 5 numeri che portano e non l'ordine in cui sono stati estratti. Scegliamo, allora, come spazio campionario lo spazio

$$\Omega \;=\; \Big\{\, S \subset \{1,\dots,90\} \colon \#S = 5 \Big\}$$

di tutti i sottoinsiemi dell'insieme delle palline enumerate $\{1,\dots,90\}$ che hanno 5 elementi. La Proposizione A.1.10 ci dice che la cardinalità di Ω (cioè il numero di possibilità di scegliere 5 di 90 senza ripetizioni e senza prendere in considerazione l'ordine) è $\#\Omega = \binom{90}{5} = \frac{90\cdot89\cdot88\cdot87\cdot86}{5\cdot4\cdot3\cdot2\cdot1} = 6\cdot89\cdot11\cdot87\cdot86 = 43.949.268$.

Ogni quintupla (non ordinata!) dovrebbe avere la stessa probabilità. Il problema di calcolare la probabilità di un evento $A \subset \Omega$ di questo spazio di probabilità elementare è, quindi, ridotto a quello di calcolare il numero $\#A$ di elementi di A. Calcoliamo la probabilità dell'evento

$$A \;:=\; \text{“indovino con 3 giusti e 2 sbagliati”}$$

che l'estrazione ci dia 3 numeri fra quelli di nostro indovino e 2 no. Per calcolare la probabilità di questo evento dobbiamo individuare il numero di esiti favorevoli. Sia $T \in \Omega$ il nostro indovino. L'evento accade se il risultato S dell'estrazione ha esattamente 3 elementi in comune con T e, quindi, 2 con il complementare di T:

$$A \;=\; \{S \in \Omega \colon \#(S \cap T) = 3\}$$

Osserviamo che ci sono $\binom{5}{3}$ possibilità di scegliere i 3 elementi fra i 5 elementi di T che dovrebbero apparire anche in S, mentre ogni tale terna può essere completata in $\binom{85}{2}$ modi con due elementi di $T^{\complement}$. Otteniamo un numero totale di $\#A = \binom{5}{3}\binom{85}{2}$ possibilità. Quindi

$$P(A) \;=\; \frac{\#A}{\#\Omega} \;=\; \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \;=\; \frac{\frac{5\cdot4}{2}\cdot\frac{85\cdot84}{2}}{\frac{90\cdot89\cdot88\cdot87\cdot86}{5\cdot4\cdot3\cdot2\cdot1}} \;\approx\; 0,0008.$$

> N.B.: Gli ***arrotondamenti*** si fanno sempre ***alla fine*** di ogni calcolo! Se il risultato $P(A)$ dovesse servire per il calcolo successivo di un altro risultato, in questo calcolo successivo bisogna inserire il numero esatto per $P(A)$. Errori causati da arrotondamenti prematuri costano punti all'esame.

In genere, estraiamo k individui di N fra i quali n sono favorevoli ed $N-n$ sfavorevoli. (k palline dall'urna senza rimpiazzo e senza prendere in considerazione l'ordine, dove le palline bianche sono da considerarsi “favorevoli”.) Allora la probabilità di trovare fra i k individui estratti esattamente ℓ degli n favorevoli ed $k-\ell$ degli $N-n$ sfavorevoli, è

$$P(\text{“}\ell \text{ di } n \text{ fra } k \text{ di } N\text{”}) \;=\; \frac{\binom{n}{\ell}\binom{N-n}{k-\ell}}{\binom{N}{k}} \;=:\; H_{N,n,k}(\ell)\,.$$

La probabilità P su $\mathcal{P}(\mathbb{N}_0)$ che da all'evento $\{\ell\}$ di ottenere ℓ "buoni" la probabilità $P(\{\ell\}) = H_{N,n,k}(\ell)$ si chiama la ***legge ipergeometrica*** di parametri $N \in \mathbb{N}_0, n \leq N, k \leq N$.

Evidentemente per avere probabilità $H_{N,n,k}(\ell)$ diverso da zero, ℓ deve soddisfare le condizioni $0 \leq \ell \leq n$ e $0 \leq k - \ell \leq N - n$, cioè, $k \geq \ell \geq k + n - N$. Infine, $\max(0, k + n - N) \leq \ell \leq \min(n, k)$. Questi dettagli sono di interesse secondario, perché fuori questi limiti la formula da 0 automaticamente.

Sappiamo che le probabilità p_ℓ degli risultati ℓ della legge ipergeometrica debbano definire una probabilità.

2.1.4 Corollario. $\displaystyle\sum_{\ell\in\mathbb{N}_0} \frac{\binom{n}{\ell}\binom{N-n}{k-\ell}}{\binom{N}{k}} = 1$, *ossia* $\displaystyle\sum_{\ell\in\mathbb{N}_0} \binom{n}{\ell}\binom{N-n}{k-\ell} = \binom{N}{k}$.

2.1.5 Esercizio. Dimostrare la formula $\sum_{\ell\in\mathbb{N}_0} \binom{n}{\ell}\binom{N-n}{k-\ell} = \binom{N}{k}$ per induzione (su N), usando solo relazioni fra i coefficienti binomiali.

2.1.6 Esercizio. Un gioco di poker consiste di 52 carte (i tredici valori 2 a 10, fante, donna, re ed asso in quattro semi, quadri, cuori, picche e fiori) delle quali ognun giocatore riceve 5. Quante possibilità ci sono per le carte di un giocatore?

2.1.7 Esempio. Il procedimento nell'Esempio 2.1.3 per calcolare la probabilità di un tris o, più generale, per derivare la legge ipergeometrica, è tipico per molte applicazioni. Nel Poker (si veda l'esercizio precedente) le carte vanno valutate secondo le loro probabilità. (Vince la carta con la probabilità più piccola.) Ci sarà un intero foglio di esercizi solo sul Poker, calcolando fra l'altro anche le probabilità di tutte le carte. Come esempio calcoliamo adesso il numero di possibilità per avere "two pairs", cioè due coppie di due valori diversi più un'altra carta con un terzo valore diverso dagli altri.

Iniziamo con il numero delle possibilità di scegliere i due valori che appaiono in coppia. Ci sono 13 valori di cui scegliere 2. Un ordine vale l'altro, quindi, ci sono $\binom{13}{2}$ possibilità. Di un valore ci sono 4 carte dei quali ci servono solo 2. Per scegliere queste due abbiamo $\binom{4}{2}$ possibilità. Visto che questa scelta la dobbiamo fare due volte (una per ogni coppia), questo fattore appare due volte. Infine, rimane l'ultima carta. Ci sono rimasti 11 valori a disposizione e ognuno di questi valori è rimasto disponibile in 4 semi. Allora il numero totale di aver "two pairs" è

$$\binom{13}{2}\binom{4}{2}^2 \cdot 11 \cdot 4 = \frac{13 \cdot 12}{2}\left(\frac{4 \cdot 3}{2}\right)^2 \cdot 11 \cdot 4 = 6^3 \cdot 11 \cdot 13 \cdot 4 = 216 \cdot 143 \cdot 4 = 123.552.$$

2.1.8 Esercizio. Le carte del Poker (Esercizio 2.1.6), ordinate per valore in modo crescente, sono:

one pair	2 carte dello stesso valore, tutte le altre tre di valori diversi (anche fra di loro)
two pairs	2 "one pair" e la quinta carta, tutti e tre di valori diversi
triplet	3 carte dello stesso valore, le altre due di valori diversi (anche fra di loro)
straight	5 carte di valori in fila (dove un asso può apparire solo come la prima o l'ultima carta) ma non uno "flush"
flush	5 carte dello stesso seme ma non uno "straight"
full house	un "triplet" e un "pair" allo stesso tempo
four a kind	tutte e quattro le carte di un valore (più un'altra)
straight flush	uno "straight" e un "flush" allo stesso tempo
Royal flush	uno "straight flush" che finisce con un asso.

1. Calcolare le probabilità di queste carte e verificare che il valore di una carta corrisponde alla probabilità nel senso che una carta meno probabile abbia il valore più alto.
2. A volte qualche gruppo di giocatori decide di prendere in considerazione anche le seguenti carte: "Four flush" (4 carte di un seme), "blaze" (tutte e 5 le carte sono immagini) e "round the corner straight" (uno "straight" che può avere un asso al interno, per esempio re, asso, 2, 3, 4).

   Mettere in fila queste carte con le altre secondo le loro probabilità.
3. Ho sulla mano uno "straight flush" e giochiamo in due. Sotto questa informazione, quanto è alta la probabilità che anche l'altro abbia uno "straight flush". Paragonare questa probabilità (che riconosceremo nella Sezione 3.1 come probabilità condizionata) con la probabilità di uno "straight flush" senza informazioni.

2.1.9 Esempio. Torniamo alla nostra urna con N palline, $n \leq N$ di loro bianche, le altre $N - n$ rosse. L'estrazione di una pallina, nell' Esempio 1.2.15, l'avevamo modellata in modo più convincente su $\Omega_1 = \{1, \dots, N\}$ dove le palline $1, \dots, n$ sono queste bianche. "Più convincente" si riferiva al fatto che in questo spazio di probabilità i singoli risultati, cioè le palline individuali, sono equiprobabili. In altre parole, Ω_1 è uno spazio di probabilità elementare.

Questa volta vogliamo considerare l'estrazione di due palline, una dopo l'altra. Cioè, come nel lotto o nel Poker l'estrazione è senza rimpiazzo. Però adesso prendiamo nota dell'ordine. Uno spazio campionario opportuno è, quindi,

$$\Omega_2 := \{(k_1, k_2) \in \Omega_1^2 \colon k_1 \neq k_2\}.$$

Sappiamo dalla Proposizione A.1.7 che il numero di elementi di questo spazio è $\#\Omega_2 = N(N-1)$. (**Esercizio:** Dare un ragionamento diretto in questo caso molto semplice.) Non ci sono ragioni che favorissero una delle coppie di Ω_2. Quindi supponiamo che anche Ω_2 sia uno spazio di probabilità elementare.

Tutte le domande che si riferiscono solo ai colori delle due palline estratte possono essere espressi con l'aiuto dei due eventi

$$A_1 := \left\{ (k_1, k_2) \in \Omega_2 \colon k_1 \in \{1, \dots, n\} \right\} = \text{“la prima pallina è bianca’,}$$
$$A_2 := \left\{ (k_1, k_2) \in \Omega_2 \colon k_2 \in \{1, \dots, n\} \right\} = \text{“la seconda pallina è bianca’,}$$

e le operazioni $\cap, \cup, ^{\complement}$. (Si veda il seguente Esercizio 2.1.10.) Calcoliamo le probabilità degli eventi che sono intersezioni di A_1 o il suo complementare e A_2 o il suo complementare. In ogni caso, il numero di realizzazioni è il prodotto del numero di possibilità per realizzare l'evento che riguarda solo la prima pallina estratta (cioè n se ci interessa "bianco" e $M - n$ si ci interessa "rosso") per il numero per realizzare l'evento che riguarda solo la seconda pallina estratta e che dipende dal risultato della prima estrazione. Facciamo una tabella:

	$A_1 \cap A_2$	$A_1 \cap A_2^{\complement}$	$A_1^{\complement} \cap A_2$	$A_1^{\complement} \cap A_2^{\complement}$
#	$n(n-1)$	$n(N-n)$	$(N-n)n$	$(N-n)(N-n-1)$
P	$\frac{n}{N} \cdot \frac{n-1}{N-1}$	$\frac{n}{N} \cdot \frac{N-n}{N-1}$	$\frac{N-n}{N} \cdot \frac{n}{N-1}$	$\frac{N-n}{N} \cdot \frac{N-n-1}{N-1}$

Usiamo queste probabilità per calcolare $P(A_1)$ e $P(A_2)$ usufruendo dell'additività. Infatti, A_1 è l'unione dei due eventi disgiunti $A_1 \cap A_2$ e $A_1 \cap A_2^{\complement}$. (Perché è l'unione? Perché sono disgiunti?) Quindi

$$P(A_1) = P(A_1 \cap A_2) + P(A_1 \cap A_2^{\complement}) = \frac{n}{N} \cdot \frac{n-1}{N-1} + \frac{n}{N} \cdot \frac{N-n}{N-1} = \frac{n}{N},$$

che, non a sorpresa, conferma il risultato già avuto: La probabilità che in un tentativo esca "bianco" è $\frac{n}{N}$. Il secondo tentativo non può influenzare in nessun modo il risultato del primo. Ma anche per A_2 troviamo

$$P(A_2) = P(A_1 \cap A_2) + P(A_1^{\complement} \cap A_2) = \frac{n}{N} \cdot \frac{n-1}{N-1} + \frac{N-n}{N} \cdot \frac{n}{N-1} = \frac{n}{N}.$$

Non è strano? Il primo tentativo non dovrebbe influenzare il secondo? Infatti lo fa. La distribuzione di $P(A_2)$ sui due casi A_1 ed $A_1^{\complement}$ è *assimetrica* in un senso che specifichiamo meglio quando parliamo di probabilità condizionate (Sezione 3.1). Se noi prendessimo nota del risultato della prima estrazione, allora la probabilità *relativa* (ossia, condizionata) al numero dei soli casi che verifichino questo risultato della prima estrazione sarebbe diversa da $\frac{n}{N}$. (Avremmo $\frac{n-1}{N-1}$ nel caso che si fosse verificato A_1 e $\frac{n}{N-1}$ nel caso opposto.) Però la prima estrazione non ci interessa proprio se chiediamo solo dell'evento A_2. Ed, infatti, un'occhiata allo spazio campionario Ω_2 ci dice che la situazione è simmetrica rispetto allo scambio dell'ordine dei due lanci. Abbiamo esattamente le stesse coppie, se variamo prima k_2 fra gli N numeri $1, \dots, N$ e poi k_1 fra gli $N-1$ numeri rimasti dopo aver scelto k_2.

2.1.10 Esercizio. Sia $\Omega \neq \emptyset$ uno spazio campionario. Qual è l'algebra booleana su Ω più piccola che contiene gli eventi $A \subset \Omega$ e $B \subset \Omega$. (Si veda anche l'Esercizio 1.1.14.)

2.1.11 Esercizio. Nell'estrazione di due palline dall'urna dell'Esempio 2.1.9, calcolare la probabilità dell'evento $A :=$ "Almeno una pallina è bianca" in due modi:

1. Usando gli eventi A_1 e A_2 di Ω_2 come nell'Esempio 2.1.9.
2. Modellando l'intero esperimento sin dall'inizio in un modo che non prende nota dell'ordine delle due estrazioni.

2.1.12 Esempio. Il lancio, una o più volte, di una moneta equilibrata è fra gli esempi più classici e vecchi che esistano. Discuteremo la moneta "astratta" nel senso che i risultati gli annotiamo con 1 e 0 che rende più generale (ciò significa, più applicabile!) le nostre considerazioni. Normalmente 1 sta per "testa", 0 per "croce". In genere, se uno dei risultati a noi sembra favorevole (per esempio, perché noi scommettiamo sempre su "testa"), allora è quello il risultato che indichiamo con 1. Ma la scelta 1 o 0 può anche essere privo di ogni pregiudizio sulla preferibilità di uno dei risultati. Come vedremo, indicare i risultati con 1 e 0 ha dei vantaggi schiaccianti per il conteggio del "numero di successi" in più tentativi.

Un solo lancio va, quindi, modellato sullo spazio campionario $\Omega = \{0, 1\}$ probabilizzato come spazio di probabilità elementare. (A questo fatto, $p_0 = p_1 = \frac{1}{2}$, si riferisce alla parola "equilibrata".)

n lanci di seguito possiamo modellare sullo spazio campionario

$$\Omega^n = \underbrace{\Omega \times \ldots \times \Omega}_{n \text{ volte}} = \{\omega = (\omega_1, \ldots, \omega_n) \colon \omega_1, \ldots, \omega_n = 0, 1\}.$$

Chiaramente, in questo spazio con la cardinalità $\#(\Omega^n) = 2^n$ prendiamo nota dell'ordine in cui i risultati dei lanci individuali sono usciti; si veda il Corollario A.1.5. Intuitivamente, una di tali successioni (n–uple) di risultati dovrebbe valere l'altra, cosicché possiamo supporre che Ω^n sia un spazio di probabilità elementare. (Lo confermiamo più tardi, quanto discuteremo la moneta reale, ossia non equilibrata.)

Qual è la probabilità di vincere in questi n lanci esattamente k volte? Dobbiamo determinare la cardinalità dell'evento

$$A_k^n := \left\{\omega \in \Omega^n \colon \omega_1 + \ldots + \omega_n = k\right\}.$$

Ogni elemento di questo evento è un'n–upla che ha 1 su esattamente k dei n posti, e zero sugli altri $n - k$. Quindi ci serve il numero di possibilità per tali scelte. Evidentemente, l'ordine in che distribuiamo i k "1" su loro posti non centra. Ne segue, secondo la Proposizione A.1.10 che

$$\#A_k^n = \binom{n}{k}, \quad \text{ossia,} \quad P(A_k^n) = \frac{\binom{n}{k}}{2^n}.$$

2.1.13 Nota matematica.

Come sempre, per dare una dimostrazione rigorosa che la cardinalità di A_k^n sia proprio $\binom{n}{k}$, bisognerebbe procedere per induzione su n. Notiamo che A_k^{n+1} possiede la decomposizione

$$\begin{aligned} & A_k^{n+1} \\ &= \left\{\omega \in \Omega^{n+1} \colon \omega_{n+1} = 1, \omega_1 + \ldots + \omega_n = k-1\right\} \cup \left\{\omega \in \Omega^{n+1} \colon \omega_{n+1} = 0, \omega_1 + \ldots + \omega_n = k\right\} \end{aligned}$$

in due eventi disgiunti: Uno dove l'ultimo tentativo è un successo (cosicché nei precedenti ci devono essere $k-1$ successi), e uno dove l'ultimo tentativo è un insuccesso (cosicché tutti e k i successi devono capitare fra i tentativi precedenti). Per le cardinalità troviamo, quindi,

$$\#A_k^{n+1} \;=\; \#A_{k-1}^n + \#A_k^n.$$

Per l'ennesima volta (contando gli esempi della Sezione A.1) vediamo al lavoro la formula del Corollario A.1.13.

Fine della nota.

Addizionando tutti i casi $k = 0, \ldots, n$ (uno di questi numeri di teste deve pure uscire) troviamo come corollario il seguente caso particolare del Teorema A.1.14 (formula del binomio).

2.1.14 Corollario. $$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k}.$$

2.2 Spazi di probabilità finiti e loro prodotti

Prima di passare al lancio di una moneta non necessariamente equilibrata, ci occupiamo del problema come modellare la descrizione simultanea (cioè in un solo spazio di probabilità) di due esperimenti individuali che non s'influenzano l'uno l'altro. Supponiamo che il primo esperimento sia modellato su uno spazio di probabilità discreto $(\Omega_1, \mathcal{P}(\Omega_1), P)$ con le probabilità degli eventi elementari $p_{\omega_1} = P(\{\omega_1\})$ per ogni $\omega_1 \in \Omega_1$ e, similmente, lo spazio di probabilità discreto $(\Omega_2, \mathcal{P}(\Omega_2), Q)$ con le probabilità $q_{\omega_2} = Q(\{\omega_2\})$ $(\omega_2 \in \Omega_2)$ per il secondo esperimento. Evidentemente, l'esecuzione simultanea di questi due esperimenti in ogni tentativo da due risultati, uno di Ω_1, l'altro di Ω_2. Lo possiamo descrivere in modo opportuno come un risultato dello spazio campionario $\Omega := \Omega_1 \times \Omega_2$. Ma ci tocca ancora a probabilizzare questo nuovo spazio campionario Ω, e lo dobbiamo fare in tal modo che il fatto che stiamo apprendendo informazioni su due esperimenti non cambi le probabilità degli eventi (cioè, delle domande con le risposte "si" o "no") che riguardano uno solo degli esperimenti.

Il seguente ragionamento, che serve come motivazione, lo mettiamo in caratteri piccoli. Discuteremo un'altra motivazione nella Sezione 3.2. Per procedere qui ci serve solo il risultato, l'Equazione (2.2.1). La motivazione si può anche saltare.

2.2.1 Frequenze relative in due esperimenti indipendenti. Diamo un'occhiata alle frequenze relative in M verifiche dell'esperimento congiunto. Supponiamo che le M verifiche abbiano dato i risultati $(x_1, y_1), \dots, (x_M, y_M) \in \Omega$. Per $x \in \Omega_1$ annotiamo con A_x l'evento che il primo esperimento abbia avuto il risultato x. Come sottoinsieme di Ω questo evento corrisponde ad

$$A_x \;=\; \{(\omega_1, \omega_2) \in \Omega \colon \omega_1 = x\} \;=\; \{(x, \omega_2) \colon \omega_2 \in \Omega_2\} \;=\; \{x\} \times \Omega_2.$$

La frequenza relativa di questo evento è

$$h^M_{A_x} \;=\; \frac{\#\{m \colon (x_m, y_m) \in A_x\}}{M} \;=\; \frac{\#\{m \colon x_m = x\}}{M}.$$

Il termine a destra, soddisfacentemente, ci rida la frequenza relativa come se ci fosse stato eseguito solo l'Esperimento 1. Ricordiamoci che per M grande aspettiamo (con alta probabilità) che la frequenza relativa sia vicina alla probabilità vera p_x. Però se il verificarsi di un certo risultato $y \in \Omega_2$ del secondo esperimento, come supponiamo, non ha nessuna influenza sull'esito del primo esperimento, allora anche la frequenza relativa del risultato x del primo esperimento calcolata prendendo in considerazione solo questi risultati (x_m, y_m) dove il risultato y_m del secondo esperimento è uguale a y, dovrebbe essere vicina alla probabilità p_x vera, non appena il numero dei risultati con $y_m = y$ sia abbastanza grande. (Da evitare è un y con $q_y = 0$.) Per esempio, se il secondo esperimento consiste nel verificare il sesso di una persona scelta fra la popolazione di Milano mentre il primo esperimento consiste nel controllo di qualità (guasto o no?) di un certo prodotto di una fabbrica in Alaska, noi crediamo che il sesso della persona di Milano non possa avere nessuna influenza sulla qualità dell'attrezzo prodotto in Alaska. Allora se ci limitassimo a guardare solo le frequenze relative rispetto ai risultati congiunti dove nel secondo esperimento si è verificato una donna, allora questa frequenza relativa dovrebbe essere simile a quella che prende in considerazione tutti i risultati. Questa frequenza relativa rispetto al verificarsi di un certo valore $y \in \Omega_2$ (il termine giusto sarebbe *frequenza relativa condizionata*; si veda la Sezione 3.1 sulla probabilità condizionata) è

$$\frac{\#\{m \colon x_m = x, y_m = y\}}{\#\{m \colon y_m = y\}}.$$

Se il numero M e il numero $\#\{m \colon y_m = y\}$ di verifiche del risultato y sono abbastanza grande, allora queste due frequenze relative dovrebbero essere più o meno uguali. Troviamo

$$\frac{\#\{m \colon x_m = x\}}{M} \;\approx\; \frac{\#\{m \colon x_m = x, y_m = y\}}{\#\{m \colon y_m = y\}},$$

oppure

$$\frac{\#\{m \colon x_m = x\}}{M} \cdot \frac{\#\{m \colon y_m = y\}}{M} \;\approx\; \frac{\#\{m \colon x_m = x, y_m = y\}}{M},$$

e nel limite $M \to \infty$

$$p_x q_y \;=\; P(\{(x, y)\}) \;=:\; p_{(x,y)}.$$

Poiché $p_{(x,y)} \le q_y$, la formula precedente vale anche nel caso $q_y = 0$.

Come vediamo la probabilità del risultato (x, y), composto dai risultati x e y di due esperimenti che non hanno niente a che fare l'uno con l'altro, è il prodotto delle probabilità p_x è q_y dei risultati individuali.

$$\boxed{p_{(x,y)} \;=\; p_x q_y} \tag{2.2.1}$$

Questo risultato è facilmente generalizzabile alla descrizione simultanea di n esperimenti che non s'influenzano, come ci serve nell'Esempio 2.2.6.

2.2.2 Teorema. *Siano* $(\Omega_1, \mathcal{P}(\Omega_1), P)$ *e* $(\Omega_2, \mathcal{P}(\Omega_2), Q)$ *spazi di probabilità discreti, con* $p_{\omega_1} := P(\{\omega_1\})$ *e* $q_{\omega_2} := Q(\{\omega_2\})$. *Allora i numeri* $p_{(\omega_1,\omega_2)} := p_{\omega_1} q_{\omega_2} \geq 0$ *soddisfanno*

$$\sum_{(\omega_1,\omega_2)\in\Omega_1\times\Omega_2} p_{(\omega_1.\omega_2)} \;=\; 1$$

e $P(C) := \sum_{(\omega_1,\omega_2)\in C} p_{(\omega_1,\omega_2)}$ *è l'unica probabilità booleana su* $\mathcal{P}(\Omega_1\times\Omega_2)$ *che soddisfa* $P(A\times B) = P(A)Q(B)$ *per ogni* $A \subset \Omega_1, B \subset \Omega_2$.

Dimostrazione. Notiamo che $(\{\omega_1\} \times B)_{\omega_1\in A}$ è una partizione di $A \times B$. Secondo il Corollario 1.3.14 vale, quindi,

$$\begin{aligned} P(A\times B) \;&=\; \sum_{(\omega_1,\omega_2)\in A\times B} p_{(\omega_1,\omega_2)} \;=\; \sum_{\omega_1\in A}\Big(\sum_{\omega_2\in B} p_{\omega_1} q_{\omega_2}\Big) \\ &=\; \sum_{\omega_1\in A}\Big(p_{\omega_1}\Big(\sum_{\omega_2\in B} q_{\omega_2}\Big)\Big) \;=\; \Big(\sum_{\omega_1\in A} p_{\omega_1}\Big)\Big(\sum_{\omega_2\in B} q_{\omega_2}\Big) \;=\; P(A)Q(B). \end{aligned}$$

Questo vale, in particolare, per $A = \Omega_1$ e $B = \Omega_2$, di cui segue che i $p_{(\omega_1,\omega_2)}$ definiscono un'unica probabilità booleana con $P(\{(\omega_1, \omega_2)\}) = p_{(\omega_1,\omega_2)} = p_{\omega_1} q_{\omega_2} = P(\{\omega_1\})Q(\{\omega_2\})$. ∎

2.2.3 Osservazione. Il teorema precedente vale anche per il prodotto di un numero finito di spazi di probabilità discreti $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$ $(i = 1, \ldots, n)$. Infatti, se poniamo $p^i_{\omega_i} := P_i(\{\omega_i\})$, $\Omega := \Omega_1 \times \ldots_n$, e $p_\omega = p_{(\omega_1,\ldots,\omega_n)} := p^1_{\omega_1} \ldots p^n_{\omega_n}$, allora $P(A) = \sum_{\omega\in\Omega} p_\omega$ definisce l'unica probabilità booleana su Ω che soddisfa $P(A_1 \times \ldots \times A_n) = P_1(A_1) \ldots P_n(A_n)$.

Una dimostrazione per induzione applicherebbe il Teorema 2.2.2 al calcolo della probabilità su $\Omega = \Omega_1 \times \ldots \times \Omega_{n-1} \times \Omega_n = (\Omega_1 \times \ldots \times \Omega_{n-1}) \times \Omega_n$ come prodotto di quelle su $\Omega_1 \times \ldots \times \Omega_{n-1}$ e su Ω_n.

2.2.4 Osservazione. Notiamo che $A\times B$ è nient'altro che l'evento "il primo esperimento verifica l'evento A e il secondo esperimento verifica l'evento B". In termini "rilassati" si potrebbe dire che si trattassi dell'evento $A \cap B$. Però A è un evento di Ω_1 e B è un evento di Ω_2. Non possono essere intersecati prima di poterli considerare come eventi di $\Omega_1 \times \Omega_2$. Notiamo, però, che

$$A \times B \;=\; (A \times \Omega_2) \cap (\Omega_1 \times B).$$

(**Esercizio:** Verificarlo.) L'evento $A \times \Omega_2$ di $\Omega_1 \times \Omega_2$ consiste di tutte le coppie (ω_1, ω_2) tali che $\omega_1 \in A$, in altre parole, tali che il primo esperimento verifichi l'evento A. Il prodotto $A \times \Omega_2$ dell'evento A di Ω_1 con l'intero spazio campionario del secondo esperimento è la descrizione opportuna dell'evento A che riguarda solo il primo esperimento come evento dell'esperimento congiunto. Altrettanto, $\Omega_1 \times B$ è la descrizione opportuna dell'evento B che riguarda solo il secondo esperimento come evento dell'esperimento congiunto. E l'intersezione $A \times B$ è la descrizione dell'evento che accadano A e B allo stesso tempo.

Questa possibilità di "immergere" un evento A di Ω_1 in uno spazio prodotto ci sarà molto utile. Verifichiamo che rispetti le leggi dell'algebra booleana.

2.2.5 Proposizione. *Siano Ω_1 e Ω_2 insiemi non vuoti. Allora la funzione $\varphi\colon \mathcal{P}(\Omega_1) \to \mathcal{P}(\Omega_1 \times \Omega_2)$ definita come $\varphi(A) = A \times \Omega_2$ è un omomorfismo di algebre booleane (si veda la Definizione B.2.1), cioè, per ogni $A, B \subset \Omega_1$ vale*

$$\varphi(A \cap B) \;=\; \varphi(A) \cap \varphi(B), \qquad \varphi(A \cup B) \;=\; \varphi(A) \cup \varphi(B), \qquad \varphi(A^{\complement}) \;=\; \varphi(A)^{\complement}.$$

Inoltre, l'omomorfismo φ è iniettivo, ossia, un'immersione.

Dimostrazione. Lasciando le altre parti come **esercizio**, dimostriamo solo $\varphi(A^{\complement}) = \varphi(A)^{\complement}$.

$$\begin{aligned}\varphi(A)^{\complement} \;=\; (A \times \Omega_2)^{\complement} \;=\; \{(\omega_1, \omega_2)\colon \omega_1 \in A, \omega_2 \in \Omega_2\}^{\complement} \;=\; \{(\omega_1, \omega_2)\colon \omega_1 \notin A \ \text{ o } \ \omega_2 \notin \Omega_2\} \\ =\; \{(\omega_1, \omega_2)\colon \omega_1 \notin A\} \;=\; A^{\complement} \times \Omega_2 \;=\; \varphi(A^{\complement}).\end{aligned}$$ ■

2.2.6 Esempio. Adesso vogliamo discutere la moneta astratta reale. Cioè, gli spazi campionari per un, rispettivamente, per n lanci sono sempre $\Omega = \{0, 1\}$, rispettivamente, Ω^n. Ma non supponiamo più, che la moneta sia equilibrata. Piuttosto, supponiamo che la probabilità di avere in un lancio il risultato 1 (successo) sia $p \in (0, 1)$ mentre la probabilità di avere il risultato 0 (insuccesso) sia $q := 1 - p \in (0, 1)$. (La convenzione $p \neq 0, 1$, in gran parte, non è necessaria. Ma è standard in molti libri ed, oltre ad evitare certe banalità, serve soprattutto per evitare che in alcune formule appaia un quoziente per 0.)

Probabilizziamo $\Omega = \{0, 1\}$ secondo il Teorema 1.3.9 sugli spazi di probabilità discreti con $p_0 = q$ e $p_1 = p$. Adesso ci tocca a probabilizzare anche lo spazio finito Ω^n. Dobbiamo, quindi, indicare le probabilità di ogni risultato $\omega = (\omega_1, \ldots, \omega_n)$. Secondo la discussione precedente, la probabilità di questo risultato è il prodotto $p_{(\omega_1,\ldots,\omega_n)} = p_{\omega_1} \ldots p_{\omega_n}$ delle probabilità in ogni tentativo. Osserviamo che $p_{\omega_\ell} = p$ se $\omega_\ell = 1$ e $p_{\omega_\ell} = q$ se $\omega_\ell = 0$. Questo lo possiamo anche

scrivere in una forma più compatta come $p_{\omega_\ell} = p^{\omega_\ell} q^{1-\omega_\ell}$. (**Esercizio:** Verficarlo!) Ne segue

$$p_{(\omega_1,\ldots,\omega_n)} = p_{\omega_1} \cdots p_{\omega_n} = p^{\omega_1} q^{1-\omega_1} \cdots p^{\omega_n} q^{1-\omega_n}$$
$$= p^{\omega_1+\ldots+\omega_n} q^{(1-\omega_1)+\ldots+(1-\omega_n)} = p^{\omega_1+\ldots+\omega_n} q^{n-(\omega_1+\ldots+\omega_n)}$$

Vediamo che questa probabilità dipende solo dal numero $k := \omega_1 + \ldots + \omega_n$ di successi in n lanci: $p_{(\omega_1,\ldots,\omega_n)} = p^k q^{n-k}$.

Come nell'Esempio 2.1.12 ci possiamo chiedere della probabilità di vincere esattamente k volte in questi n tentativi. La probabilità di una specifica n–upla $(\omega_1, \ldots, \omega_n)$, che contiene k volte "1" e $n - k$ volte "0", abbiamo appena stabilita come $p^k q^{n-k}$. Ma veramente, per avere k successi, una n–upla con k successi vale l'altra. Dobbiamo quindi calcolare il numero di n–uple favorevoli. Come nell'Esempio 2.1.12 si tratta del problema di scegliere i k posti fra n su cui appare "1". Questo numero è sempre $\binom{n}{k}$. Solo che adesso non lo dobbiamo dividere per 2^n, il numero totale di n–uple possibili, ma moltiplicare con $p^k q^{n-k}$, la probabilità di una n–upla particolare con k successi. Infine,

$$\boxed{P(\text{“}k \text{ successi in } n \text{ tentativi”}) = \binom{n}{k} p^k q^{n-k} =: B_{n,p}(k)\,.}$$

Chiameremo ***legge binomiale*** di parametri $n \in \mathbb{N}_0$, $p \in (0,1)$ la probabilità su $\mathcal{P}(\{0,1,\ldots,n\})$ che assegna a $\{k\}$ la probabilità $B_{n,p}(k)$.

Nuovamente, troviamo ancora un caso particolare del Teorema A.1.14 (formula del binomio), più generale del Corollario 2.1.14, ma non ancora in piena generalità.

2.2.7 Corollario. *Per $p + q = 1, p \geq 0, q \geq 0$ vale* $1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$.

2.2.8 Nota matematica.

Per arrivare alla forma generale della formula del binomio $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$ tramite questo ragionamento probabilistico, non manca molto. Se x e y hanno lo stesso segno, allora possiamo definire $p = \frac{x}{x+y} \geq 0$ e $q = \frac{y}{x+y} \geq 0$ con $p + q = 1$ ed applicare il Corollario 2.2.7. Troviamo

$$1 = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = \frac{1}{(x+y)^n} \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Stranamente, il caso dove x e y hanno il segno opposto, crea problemi. Si potrebbe ragionare che, per una y fissata, $(x+y)^n$ come funzione di x è un polinomio $a_0 + a_1 x + \ldots + a_n x^n$ in x e che i coefficienti a_k sono già determinati conoscendo il polinomio solo per $x \geq 0$ o $x \leq 0$.

Fine della nota.

Se n è grande, le probabilità $B_{n,p}(k)$ della legge binomiale, numericamente, sono difficile da calcolare. Questo è dovuto non solo ai coefficienti binomiali (che sono facile da calcolare quando k o $n-k$ è piccolo) ma soprattutto al calcolo di $p^k q^{n-k}$ dove almeno uno degli esponenti k e $n-k$ è grande. Per questo esistono diverse approssimazioni della legge binomiale per grandi n.

Però, mandando semplicemente $n \to \infty$, troviamo

$$0 \;\le\; B_{n,p}(k) \;=\; \binom{n}{k} p^k q^{n-k} \;=\; \frac{p^k}{k!q^k} n(n-1)\dots(n-k+1)q^n \;\le\; \frac{p^k}{k!q^k} n^k q^n \;\to\; 0$$

per ogni k.

Il fattore $\frac{p^k}{k!q^k}$ non dipende da n. Quindi, basta di dimostrare che $n^k q^n \to 0$ per $|q| < 1$. Per $k = 0$, non c'è nulla da dimostrare. ($q^n \to 0$.) Per $k \ge 1$ l'affermazione è un fatto noto dall'analisi. Si può dimostrare applicando la seguente regola.

2.2.9 Regola di l'Hospital. *Siano $f\colon x \mapsto f(x)$ e $g\colon x \mapsto g(x)$ funzioni derivabili in un intorno di x_0 (con la possibile eccezione di x_0 stesso) che soddisfano $g(x) \neq 0$ per $x \neq x_0$ e $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = 0$ oppure $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = \infty$. Allora*

$$\lim_{x\to x_0} \frac{f(x)}{g(x)} \;=\; \lim_{x\to x_0} \frac{f'(x)}{g'(x)}$$

includendo il fatto che l'esistenza di uno dei limiti implica l'esistenza dell'altro. x_0 può essere anche ∞.

Per applicare questa regola sostituiamo n con x. Otteniamo $n^k q^n = x^k q^x = \frac{x^k}{q^{-x}}$. La derivata di x^k è kx^{k-1}. La derivata di q^{-x} è $-q^{-x}\ln q$. Osserviamo che $\lim_{x\to\infty} x^k = \lim_{x\to\infty} q^{-x} = \infty$. Quindi

$$\lim_{x\to\infty} \frac{x^k}{q^{-x}} \;=\; \lim_{x\to\infty} \frac{kx^{k-1}}{q^{-x}\cdot(-\ln q)} \;=\; \dots \;=\; \lim_{x\to\infty} \frac{k(k-1)\dots 1}{q^{-x}\cdot(-\ln q)^k} \;\to\; 0.$$

Infine, $\lim_{n\to\infty} B_{n,p}(k) = 0$ per ogni $k \in \mathbb{N}$.

Questo diventa plausibile, se disegniamo il grafico della legge binomiale per alcuni n. Si vede, che con n aumentando, il grafico diventa sempre più largo, ma anche sempre più piatto. Il "baricentro" della distribuzione (che conosceremo nella Sezione 4.4 come *attesa*) si sposta sempre più verso valori grandi, cosicché per un valore fissato di k la probabilità diventa sempre più piccola.

Un modo per evitare questo fenomeno, è diminuire p allo stesso modo come n diventa grande. Più formalmente, per ogni n scegliamo il parametro p_n della legge binomiale tale che il prodotto $p_n n$ assume un valore costante $\lambda > 0$.

2.2.10 Approssimazione di Poisson. *Sia $\lambda > 0$ e per ogni $n > \lambda$ poniamo $p_n = \frac{\lambda}{n}$. Allora*

$$\lim_{n\to\infty} B_{n,p_n}(k) \;=\; e^{-\lambda}\frac{\lambda^k}{k!} \;=:\; Poi_\lambda(k)$$

*per ogni $k \in \mathbb{N}_0$. La probabilità booleana su $\mathcal{P}(\mathbb{N}_0)$ che soddisfa $P(\{k\}) = Poi_\lambda(k)$ si chiama la **legge di Poisson** di parametro $\lambda > 0$.*

La verifica che i numeri $Poi_\lambda(k)$ definiscano una probabilità secondo il Teorema 1.3.9 sugli spazi di probabilità discreti, la lasciamo come **esercizio**.

Dimostrazione del Teorema 2.2.10. Inseriamo la forma concreta di $p_n = \frac{\lambda}{n}$ e $q_n = 1 - \frac{\lambda}{n}$:

$$\begin{aligned} B_{n,p_n}(k) \;&=\; \binom{n}{k} p_n^k (1-p_n)^k \;=\; \binom{n}{k}\Big(\frac{\lambda}{n}\Big)^k\Big(1-\frac{\lambda}{n}\Big)^{n-k} \\ &=\; \frac{\lambda^k}{k!} n(n-1)\dots(n-k+1)\Big(1-\frac{\lambda}{n}\Big)^n\Big(\frac{1}{n(1-\frac{\lambda}{n})}\Big)^k \\ &=\; \frac{\lambda^k}{k!}\Big(1-\frac{\lambda}{n}\Big)^n \frac{n}{n-\lambda}\cdot\frac{n-1}{n-\lambda}\cdot\ldots\cdot\frac{n-k+1}{n-\lambda}. \end{aligned}$$

Il fattore $\frac{\lambda^k}{k!}$ non dipende da n. Il fattore $\left(1-\frac{\lambda}{n}\right)^n$, secondo l'analisi, converge verso $e^{-\lambda}$. Ognuno dei k fattori rimasti converge verso 1, quindi, lo fa anche il loro prodotto. (Per questo è importante che il numero k di fattori è costante e non dipende da n.) ■

2.2.11 Esercizio. La probabilità che una persona scelta a caso soffra di una certa malattia sia $p = \frac{1}{20} = 5\%$. Scegliamo $n = 100$ volte una persona a caso. Calcolare le probabilità $B_{100,\frac{1}{20}}(k)$ per i valori $k = 0, 1, 2, 3, 4, 5$ (o almeno per $k = 5$):

1. secondo la formula esatta per $B_{n,p}(k)$.
2. secondo l'approssimazione di Poisson. (Qual è λ?)

Qual è la probabilità di "pescare"

3. il primo malato al quinto tentativo?
4. il secondo malato al decimo tentativo?

Ripetere l'esercizio per $n = 1000$, $p = \frac{1}{200} = 5‰$.

La legge di Poisson non è più concentrata su un insieme finito, ma abbiamo sempre a che fare con uno spazio di probabilità discreto. Nella prossima sezione lasciamo addirittura gli spazi di probabilità discreti.

2.3 Una costruzione di uno spazio di probabilità non discreto

Torniamo al lancio della moneta. Una domanda tipica al lancio della moneta è quale sia il momento del primo successo. Quante volte, in un esperimento concreto, dobbiamo lanciare la moneta finché non esca "1".

Dobbiamo affrontare un problema cruciale: Non possiamo sapere in anticipo il tempo da aspettare! In un numero finito n di tentativi è benissimo possibile di avere n volte 0. (Ciò accade con la probabilità $q^n > 0$.) Dobbiamo modellare l'esperimento in tal modo che sin dall'inizio è possibile di lanciare un numero arbitrario di volte. Uno spazio campionario opportuno è lo spazio

$$\Omega^\infty := \{\omega = (\omega_1, \omega_2, \ldots) \colon \omega_1, \omega_2, \ldots \in \Omega\}$$

di tutte le successioni di elementi di $\Omega = \{0, 1\}$, dove ω_i è il risultato del i–esimo lancio. Adesso ci tocca a probabilizzare questo spazio campionario. Se procedessimo come nel caso di n lanci, una singola successione ω dovrebbe avere come probabilità un prodotto infinito di numeri p e q, entrambi strettamente più piccoli di 1. Al meno uno di questi due numeri appare un numero infinito di volte. Un tale prodotto è 0. Vediamo subito, che non possiamo comporre le probabilità di eventi complessi dalle probabilità degli eventi elementari, perché questi sarebbero tutti 0.

Dobbiamo trovare altri eventi (diversi da eventi elementari) ai quali ci riesce di attribuire delle probabilità in modo ragionevole. Tali eventi potrebbero essere questi, di cui verifica dipende solo da un numero finito n di lanci, anche se n può variare per eventi diversi. L'evento "primo successo al k–esimo tentativo" è di questo tipo, perché alla domanda si può rispondere non appena il numero n di tentativi raggiunga k. (O c'è il primo successo in questo momento, o non lo è. Per stabilirlo basta di conoscere la successione $(\omega_1, \omega_2, \ldots)$ fino al punto k. Solo che, se non è avvenuto almeno un successo nei primi k tentativi, non sapremo nulla su quanto dobbiamo ancora lanciare.)

Come probabilità di un tale evento porremo semplicemente quella che conosciamo già dalla discussione del modello con n lanci. Precisando questa idea, vogliamo che gli eventi A_n di Ω^n sono anche eventi di Ω^∞, e che come eventi di Ω^∞ hanno la stessa probabilità che hanno in Ω^n.

A questo punto, se supponiamo che questo procedimento funzioni, siamo già in grado di calcolare la probabilità di avere il primo successo al k–esimo tentativo. Lo facciamo subito, rimandando la dimostrazione, che il procedimento funzioni, ad una discussione in caratteri piccoli.

2.3.1 Teorema. *In un lancio continuato di una moneta con probabilità $p \in (0, 1)$ di "testa", la probabilità di lanciare "testa" per la prima volta nel k–esimo lancio ($k \in \mathbb{N}$) è*

$$p(1-p)^{k-1} =: geom_p(k).$$

*La probabilità su $\mathcal{P}(\mathbb{N})$ che soddisfa $P(\{k\}) = geom_p(k)$ si chiama la **legge geometrica** di parametro $p \in (0, 1)$.*

La verifica che i numeri $geom_p(k)$ definiscano una probabilità secondo il Teorema 1.3.9 sugli

spazi di probabilità discreti, la lasciamo come **esercizio**.

Dimostrazione del Teorema 2.3.1. L'appartenenza di una successione $(\omega_1, \omega_2, \ldots)$ all'evento "k è il momento del primo successo" si decide entro k lanci, cioè, guardando solo la k–upla $(\omega_1, \ldots, \omega_k)$. La probabilità di questo evento va, quindi, calcolata come quella dell'evento

$$\text{“}k-1 \textit{ zero} \text{ nei } k-1 \text{ primi lanci e } \textit{uno} \text{ nel } k\text{–esimo lancio”} \;=\; \{(\underbrace{0,\ldots,0}_{k-1 \text{ volte}},1)\}$$

di Ω_k. La probabilità di tale evento elementare con esattamente un successo in k tentativi abbiamo calcolato come $p^1 q^{k-1}$. ■

Ripetiamo subito tale ragionamento per la domanda più generale del momento del n–esimo successo.

2.3.2 Teorema. *In un lancio continuato di una moneta con probabilità $p \in (0,1)$ di "testa", la probabilità di lanciare "testa" per l'n–esima volta nel k–esimo lancio ($k \in \{n, n+1, \ldots\}$) è*

$$\binom{k-1}{n-1} p^n (1-p)^{k-n} \;=:\; Pas_{n,p}(k).$$

*La probabilità su $\mathcal{P}(\{n, n+1, \ldots\})$ che soddisfa $P(\{k\}) = Pas_{n,p}(k)$ si chiama la **legge di Pascal** di parametri $n \in \mathbb{N}$, $p \in (0,1)$. Per $n = 1$ ritroviamo la legge geometrica.*

Dimostrazione. La verifica che i numeri $Pas_{n,p}(k)$ definiscano una probabilità secondo il Teorema 1.3.9 sugli spazi di probabilità discreti, la dobbiamo rimandare fino alla Sezione 4.6 sulle funzioni generatrici.

Come sopra, l'appartenenza di una successione $(\omega_1, \omega_2, \ldots)$ all'evento "k è il momento del n–esimo successo" si decide entro k lanci, cioè, guardando solo la k–upla $(\omega_1, \ldots, \omega_k)$. La probabilità di questo evento va, quindi, calcolata come quella dell'evento

$$\begin{aligned}&\text{“}k-n \textit{ zero} \text{ e } n-1 \textit{ uno} \text{ nei } k-1 \text{ primi lanci e } \textit{uno} \text{ nel } k\text{–esimo lancio”}\\ &\qquad = \{(\omega_1, \ldots, \omega_{k-1}, 1) \colon \omega_1 + \ldots + \omega_{k-1} = n-1\}\end{aligned}$$

di Ω_k. La probabilità di ogni elemento di questo evento con esattamente n successi in k tentativi abbiamo calcolato come $p^n q^{k-n}$. Il numero di elementi è uguale al numero $\binom{k-1}{n-1}$ di possibilità di scegliere gli $n-1$ posti fra $k-1$ tentativi dove appare *uno*. (L'ultimo posto è fissato di essere k). ■

Per giungere al termine di questa sezione, adesso ci occupiamo dell'esistenza di un'algebra booleana su Ω^∞ che contiene tutti gli eventi che corrispondo a domande che riguardano solo

un numero finito n di elementi di una successione $\omega = (\omega_1, \omega_2, \ldots, \omega_n, \omega_{n+1}, \ldots)$. Il numero n può variare dipendente dall'evento, ma vogliamo che per ogni evento A di quest'algebra booleana esista un n tale che appartenenza ad A si decide guardando solo i risultati dei primi n lanci $(\omega_1, \ldots, \omega_n)$. Quindi una successione $\omega = (\omega_1, \ldots \omega_n, \omega_{n+1}, \omega_{n+2}, \ldots)$ è di A se e solo anche la successione $\omega' = (\omega_1, \ldots \omega_n, \omega'_{n+1}, \omega'_{n+2}, \ldots)$ è di A, qualunque siano i risultati $\omega'_{n+1}, \omega'_{n+2}, \ldots$. Non è difficile da vedere che A debba avere la forma

$$A \;=\; A_n \times \Omega \times \Omega \times \ldots \;\subset\; \Omega^\infty \;=\; \Omega \times \Omega \times \ldots \;=\; \Omega^n \times \Omega \times \Omega \times \ldots$$

per un evento (unico; s.v. la Proposizione 2.2.5) $A_n \in \Omega_n$. La Proposizione 2.2.5 ci dice anche che

$$\mathcal{A}_n \;:=\; \{A_n \times \Omega \times \Omega \times \ldots : A_n \in \mathcal{P}(\Omega_n)\}$$

è un'algebra booleana su Ω^∞ isomorfa a $\mathcal{P}(\Omega_n)$.

Chiaramente vogliamo assegnare all'evento $A_n \times \Omega \times \Omega \times \ldots$ di Ω^∞ la medesima probabilità dell'evento A_n di Ω^n. Il seguente teorema ci dice che ciò funziona.

2.3.3 Teorema. *La famiglia*

$$\mathcal{A}(\Omega) \;:=\; \bigcup_{n\in\mathbb{N}} \{A_n \times \Omega \times \Omega \times \ldots : A_n \subset \Omega^n\} \;=\; \bigcup_{n\in\mathbb{N}} \mathcal{A}_n$$

di sottoinsiemi di Ω^∞ è un'algebra booleana su Ω^∞. La funzione P_∞ che manda $A_n \times \Omega \times \Omega \times \ldots \in \mathcal{A}(\Omega)$ a $P_n(A_n)$ (la probabilità dell'evento $A_n \in \Omega^n$), è ben definita e definisce un probabilità booleana $P_\infty \colon \mathcal{A}(\Omega) \to \mathbb{R}$ su $\mathcal{A}(\Omega)$. Chiameremo lo spazio di probabilità booleano $(\Omega^\infty, \mathcal{A}(\Omega), P_\infty)$ lo ***schema successo-insuccesso*** *o* ***lo schema di Bernoulli****.*

Dimostrazione. Osserviamo che la successione di sottoinsiemi $\mathcal{A}_n$ di $\mathcal{P}(\Omega^\infty)$ è crescente:

$$n \le m \implies \mathcal{A}_n \subset \mathcal{A}_m.$$

Infatti, se per $A_n \in \Omega^n$ poniamo $A_m := A_n \times \Omega^{m-n} \in \Omega^m$, si vede che $A_n \times \Omega \times \Omega \times \ldots = A_m \times \Omega \times \Omega \times \ldots$. Poi sappiamo già che per ogni $n \in \mathbb{N}$, la famiglia $\mathcal{A}_n$ è un'algebra booleana di sottoinsiemi di Ω^∞. Ne segue, che anche $\mathcal{A}(\Omega)$ è un'algebra booleana:

$$A, B \in \mathcal{A}(\Omega) \implies \exists n, m \colon A \in \mathcal{A}_n, B \in \mathcal{A}_m \implies A \in \mathcal{A}_k, B \in \mathcal{A}_k, \quad \text{dove} \quad k = \max(n, m).$$

Allora, anche $A \cap B, A \cup B, A^\complement \in \mathcal{A}_k \subset \mathcal{A}(\Omega)$.

Sia $n \le m$ e $A \in \mathcal{A}_n \subset \mathcal{A}_m$. Dobbiamo dimostrare che la definizione di $P_\infty(A)$ non dipende da dalla nostra scelta di considerare A come elemento di $\mathcal{A}_n$ o di $\mathcal{A}_m$. Ricordiamoci che l'elemento $A_n \in \Omega^n$ che permette di scrivere A come $A_n \times \Omega \times \Omega \times \ldots$ è unico. Lo stesso vale per $A_m \in \Omega^m$ tale che $A = A_m \times \Omega \times \Omega \times \ldots$. Sappiamo già che $A_m = A_n \times \Omega^{m-n}$ è una scelta possibile, quindi, è l'unica scelta. Dall'Osservazione 2.2.4 segue

$$P_m(A_m) \;=\; P_m(A_n \times \Omega^{m-n}) \;=\; P_n(A_n) P_{m-n}(\Omega^{m-n}) \;=\; P_n(A_n).$$

P_∞ è chiaramente positiva e normalizzata. Come sopra, due eventi di $\mathcal{A}(\Omega)$ possono sempre essere considerati elementi dello stesso $\mathcal{A}_k$ (k abbastanza grande). Poiché P_k è additiva, lo è anche P_∞. ∎

Qual è la differenza fra il Teorema 2.3.1 (o il suo compagno più generale, il Teorema 2.3.2) con e senza essere a conoscenza del Teorema 2.3.3? Le dimostrazioni dei Teoremi 2.3.1 e 2.3.2 l'abbiamo fatte come se fossimo all'interno dello spazio di probabilità discreto Ω^k. Per poter calcolare la probabilità del primo (o dell'n–esimo) successo al k–esimo lancio questo è perfettamente sufficiente. Il problema è che per ogni k per cui poniamo la domanda, abbiamo modellato l'esperimento su uno spazio di probabilità diverso. Solo il Teorema 2.3.3 ci permette rispondere alla domanda per tutti i k su un medesimo spazio di probabilità Ω^∞. In particolare, solo Ω^∞ ci permette la modellazione dell'esperimento che risponde alla domanda "quando arriva il primo (l'n–esimo) successo?" con tutte le risposte $k \in \mathbb{N}$ (rispettivamente, $k = n, n+1, \ldots$) possibili.

Il prezzo da pagare è che Ω^∞ non è più uno spazio di probabilità discreto. Infatti, abbiamo già notato che i risultati individuali $(\omega_1, \omega_2, \ldots)$ hanno tutti la probabilità 0.

2.3.4 Nota. Parlando onestamente, se attribuiamo agli eventi elementari di Ω^∞ la probabilità 0, abbiamo commesso un errore formale, perché gli eventi elementari $\{(\omega_1, \omega_2, \ldots)\}$ non fanno parte di $\mathcal{A}(\Omega)$. (Per sapere che si verifichi un evento elementare, bisogna pure conoscere l'intera successione di lanci. Questo è in piena contraddizione con la descrizione degli elementi di $\mathcal{A}(\Omega)$ come eventi di cui realizzazione può essere deciso sempre dopo un numero finito di lanci, benché tale numero dipende dall'evento.) Aggiungendo all'algebra booleana anche gli eventi elementari senza ulteriori precauzioni, violiamo la chiusura dell'algebra sotto il complementare. È, però, possibile rimediare:

Definizione. Una famiglia $\mathcal{N}$ di sottoinsiemi di $\Omega \neq \emptyset$ si dice un ***ideale*** dell'algebra booleana $\mathcal{P}(\Omega)$ se soddisfa

$$N, M \in \mathcal{N} \implies N \cup M \in \mathcal{N}, \qquad \text{e} \qquad S \subset N \in \mathcal{N} \implies S \in \mathcal{N}.$$

Osservazione. Si nota che la seconda proprietà, per definizione, equivale $S \cap N \in \mathcal{N}$ per ogni $S \in \mathcal{P}(\Omega)$ e ogni $N \in \mathcal{N}$, e che grazie a $N \vartriangle M \subset N \cup M \in \mathcal{N}$ anche $N \vartriangle M$ è di $\mathcal{N}$. Quindi, $\mathcal{N}$ è chiusa sotto l'addizione $\vartriangle$, ed il prodotto $\cap$ di un elemento S arbitrario di $\mathcal{P}(\Omega)$ e di un elemento n di $\mathcal{N}$ è sempre di $\mathcal{N}$. Questo è che, nell'algebra, si chiama un *ideale* nell'anello $(\mathcal{P}(\Omega), \vartriangle, \cap)$; si vedano le dispense [Ske04]. Lasciamo come **esercizio** dimostrare l'affermazione opposta: Se $N \cap S \in \mathcal{N}$ e $N \vartriangle M \in \mathcal{N}$ per ogni $N, M \in \mathcal{N}$ e ogni $S \in \mathcal{P}(\Omega)$, allora $\mathcal{N}$ è un ideale di $\mathcal{P}(\Omega)$ nel senso della definizione precedente.

Esempio. L'insieme $\mathcal{N}_0 := \{N \subset \Omega \colon \exists A_1, A_2, \ldots ; N \subset A_i \in \mathcal{A}; P(A_i) \to 0\}$ è un ideale di $\mathcal{P}(\Omega)$. (**Esercizio:** Verificare che $\mathcal{N}_0$ è un ideale.) Nel caso dello spazio Ω^∞ per il lancio continuato, $\mathcal{N}_0$ contiene (fra l'altro) tutti i sottoinsiemi finiti di Ω^∞.

Teorema. *Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità booleano, e sia $\mathcal{N}$ un ideale di $\mathcal{P}(\Omega)$. Allora*

$$\beta(\mathcal{A}, \mathcal{N}) := \left\{ (A \cap N_1^{\complement}) \cup N_2 \;\middle|\; A \in \mathcal{A}, N_i \in \mathcal{N}, N_1 \cap N_2 = \emptyset \right\}$$

è un'algebra booleana su Ω, infatti, l'algebra booleana più piccola che contiene $\mathcal{A}$ e $\mathcal{N}$.

Dimostrazione. Notiamo che $N_1 \cap N_2 = \emptyset$ equivale $N_1 \subset N_2^{\complement}$ equivale $N_2 \subset N_1^{\complement}$. Ne segue che $(A \cap N_1^{\complement}) \cup N_2 = (A \cup N_2) \cap N_1^{\complement}$ e, di conseguenza, $((A \cap N_1^{\complement}) \cup N_2)^{\complement} = (A^{\complement} \cup N_1) \cap N_2^{\complement}$ è di $\beta(\mathcal{A}, \mathcal{N})$. Calcoliamo

$$\begin{aligned}&((A \cap N_1^{\complement}) \cup N_2) \cap ((B \cap M_1^{\complement}) \cup M_2)\\ &\quad = \Big(\underbrace{(A \cap B)}_{=:C} \cap \big(\underbrace{N_1 \cup M_1}_{=:L_1}\big)^{\complement}\Big) \cup \Big(\underbrace{(A \cap N_1^{\complement} \cap M_2) \cup (N_2 \cap B \cap M_1^{\complement}) \cup (N_2 \cap M_2)}_{=:L_2}\Big),\end{aligned}$$

dove $C = A \cap B$ è di $\mathcal{A}$, dove $L_1 = N_1 \cup M_1$ e $L_2 \subset N_2 \cup M_2$) sono tutti e due di $\mathcal{N}$, e dove $L_1 \cap L_2 = \emptyset$. Quindi, la famiglia $\beta(\mathcal{A}, \mathcal{N})$, essendo chiusa sotto complementari ed intersezioni, è un'algebra booleana su Ω e non ci può essere un'algebra booleana più piccola che contiene sia $\mathcal{A}$ che $\mathcal{N}$. ■

Lemma.

1. *$A, B \in \mathcal{A}$, allora $|P(A) - P(B)| \leq P(A \vartriangle B)$.*
2. *Per qualsiasi $A, B, C \in \mathcal{P}(\Omega)$ vale $(C \vartriangle A) \vartriangle (C \vartriangle B) = (A \vartriangle B)$.*
3. *$\beta(\mathcal{A}, \mathcal{N}) \ni C = (A \cap N_1^{\complement}) \cup N_2 = (B \cap M_1^{\complement}) \cup M_2)$, allora $A \vartriangle B \in \mathcal{N}$.*
4. *$(A \cap N_1^{\complement}) \cup N_2 = \emptyset$, allora $A \in \mathcal{N}$.*
5. *$((A \cap N_1^{\complement}) \cup N_2) \cap (B \cap M_1^{\complement}) \cup M_2) = \emptyset$, allora $A \cap B \in \mathcal{N}$.*

Dimostrazione.

1. Calcoliamo

$$\begin{aligned}|P(A) - P(B)| &= \big|(P(A \cap B) + P(A \cap B^{\complement})) - (P(A \cap B) + P(A^{\complement} \cap B))\big|\\ &= |P(A \cap B^{\complement}) - P(A^{\complement} \cap B)| \leq P(A \cap B^{\complement}) + P(A^{\complement} \cap B) = P(A \vartriangle B).\end{aligned}$$

2. Ricordiamoci che $\vartriangle$ è un operazione associativa (e, ovviamente, commutativa). Inoltre $C \vartriangle C = (C \cap C^{\complement}) \cup (C^{\complement} \cap C) = \emptyset$ e $\emptyset \vartriangle C = (\emptyset \cap C^{\complement}) \cup (\emptyset^{\complement} \cap C) = C$. Quindi,

$$(C \vartriangle A) \vartriangle (C \vartriangle B) = C \vartriangle C \vartriangle A \vartriangle B = \emptyset \vartriangle A \vartriangle B = A \vartriangle B.$$

3. Per $A, B, C, N_1, N_2, M_1, M_2$ come indicati, dimostriamo solo che $C \vartriangle A \in \mathcal{N}$, perché con (2) ne segne che anche $A \vartriangle B = (C \vartriangle A) \vartriangle (C \vartriangle B) \subset (C \vartriangle A) \cup (C \vartriangle B)$ è di $\mathcal{N}$. Per dimostrare $C \vartriangle A = (C \cap A^{\complement}) \cup (C^{\complement} \cap A) \in \mathcal{N}$, basta osservare che

$$C \cap A^{\complement} = ((A \cap N_1^{\complement}) \cup N_2) \cap A^{\complement} = (A \cap N_1^{\complement} \cap A^{\complement}) \cup (N_2 \cap A^{\complement}) = N_2 \cap A^{\complement} \in \mathcal{N},$$

perché $N_2 \in \mathcal{N}$, e che

$$C^{\complement} \cap A = ((A \cap N_1^{\complement}) \cup N_2)^{\complement} \cap A = (A^{\complement} \cup N_2) \cap N_1^{\complement} \cap A = N_2 \cap (N_1^{\complement} \cap A) \in \mathcal{N},$$

perché $N_2 \in \mathcal{N}$.

4. Se $(A \cap N_1^{\complement}) \cup N_2 = \emptyset$, allora $(A \cap N_1^{\complement}) \cup N_2 \supset A \cap N_1^{\complement} = \emptyset$. Quindi,

$$A = A \cap (N_1 \cup N_1^{\complement}) = (A \cap N_1) \cup (A \cap N_1^{\complement}) = (A \cap N_1) \cup \emptyset = A \cap N_1 \in \mathcal{N},$$

perché $N_1 \in \mathcal{N}$.

5. Se $C \supset C'$ e $D \supset D'$, allora $C \cap D \supset C' \cap D'$. (**Esercizio:** Dimostrarlo!) Quindi,

$$\emptyset = ((A \cap N_1^{\complement}) \cup N_2) \cap (B \cap M_1^{\complement}) \cup M_2) \supset (A \cap N_1^{\complement}) \cap (B \cap M_1^{\complement}) = (A \cap B) \cap (N_1 \cup M_1)^{\complement},$$

e di conseguenza $(A \cap B) \cap (N_1 \cup M_1)^{\complement} = \emptyset$. Poiché $N_1 \cup M_1$ è di $\mathcal{N}$, applicando il Numero (4) troviamo $A \cap B \in \mathcal{N}$. ■

L'idea è, di definire una probabilità booleana Q su $\beta(\mathcal{A}, \mathcal{N})$ in modo tale che gli elementi di $\mathcal{N}$ diventino eventi impossibili, i loro cpmplementi, quindi, eventi sicuri. In altre parole, vogliamo porre $Q((A \cap N_1^{\complement}) \cup N_2) := P(A)$. Poiché questo sia una definizione valida, occorre è basta che $Q(C)$ non dipende dal modo in ciui scriviamo $C \in \beta(\mathcal{A}, \mathcal{N})$ come $(A \cap N_1^{\complement}) \cup N_2$. Ma una volta stabilito questi, ne segue subito che Q estende P (infatti, per $C \in \mathcal{A}$ abbiamo $C = (C \cap \emptyset^{\complement}) \cup \emptyset$ con $\emptyset \in \mathcal{N}$, quindi $Q(C) = P(C)$), che $Q(C) \geq 0$, e che $Q(\Omega) = P(\Omega) = 1$. E se gli elementi $(A \cap N_1^{\complement}) \cup N_2$ e $(B \cap M_1^{\complement}) \cup M_2)$ di cN sono disgiunti, allora secondo il Numero (5) del lemma $A \cap B$ è un elemento di $\mathcal{N}$ e, quindi, $Q(A \cap B) = Q((\emptyset \cap \emptyset^{\complement}) \cup (A \cap B)) = P(\emptyset) = 0$. Quindi, se Q esiste, allore è una probabilità booleana su $\beta(\mathcal{A}, \mathcal{N})$ che estende P.

Vediamo alcuni casi dove riusciamo dimostrarlo.

Corollario 1. *Se $\mathcal{A} \cap \mathcal{N} = \{\emptyset\}$, allora la funzione $Q((A \cap N_1^{\complement}) \cup N_2) := P(A)$ definisce una probabilità booleana Q su $\beta(\mathcal{A}, \mathcal{N})$ che estende P.*

Dimostrazione. Rimane da dimostrare che la definizione sia valida, ossia, che uguaglianza di $(A \cap N_1^{\complement}) \cup N_2$ e$(B \cap M_1^{\complement}) \cup M_2)$ implichi che $P(A) = P(B)$. Poiché, A e B sono di $\mathcal{A}$, anche $A \vartriangle B$ è di $\mathcal{A}$. Dall'altra parte sappiamo dal Numero (3) del lemma che $A \vartriangle B$ è di $\mathcal{N}$. Per ipotesi, l'unico insieme che soddisfa tutti e due le condizioni è $\emptyset$. Quindi, usando anche il Numero (1) del lemma, troviamo

$$|P(A) - P(B)| \leq P(A \vartriangle B) = P(\emptyset) = 0,$$

ossia, $P(A) = P(B)$. ■

Corollario 2. *Se per ogni $N \in \mathcal{N}$ esiste una successione $A_1, A_2, \ldots \supset N$ di eventi in $\mathcal{A}$ tale che $\lim_{n\to\infty} P(A_n) = 0$, allora $Q((A \cap N_1^{\complement}) \cup N_2) := P(A)$ definisce l'unica probabilità booleana su $\beta(\mathcal{A}, \mathcal{N})$ che estende P. (Si noti che in questo caso non abbiamo richiesto $\mathcal{A} \cap \mathcal{N} = \emptyset$.)*

Dimostrazione. Come nella dimsotrazione precedente, abbiamo $A \vartriangle B \in \mathcal{N}$. Secondo l'ipotesi esistono eventi $C_1, C_2, \ldots$ tutti contenenti $A \vartriangle B$ tali che $P(C_n) \to 0$. Quindi,

$$|P(A) - P(B)| \leq P(A \vartriangle B) \leq P(C_n) \to 0$$

non può essere che 0. ■

Per concludere la Nota 2.3.4, l'esempio ed il Corollario 2 della discussione precedente ci dicono che possiamo aggiungere all'algebra booleana di eventi $\mathcal{A}$ di qualsiasi spazio di probabilità booleano tutti gli eventi $N \subset \mathcal{N}_0$ (ossia, quelli che possono essere "rinchiusi" in eventi A_n di $\mathcal{A}$ con probabilià sempre più piccole $P(A_n) \to 0$) dandoli la probabilità $Q(N) = 0$, e ovviamente anche gli eventi dell'algebra booleana $\beta(\mathcal{A}, \mathcal{N}_0)$ generata da loro, definendo così un'unica probabilità booleana su tutta quest algebra.

Uno spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$ si dice ***completo***, se $(\Omega, \beta(\mathcal{A}, \mathcal{N}_0), Q)$ è uguale a $(\Omega, \mathcal{A}, P)$. Non è difficile convincersi, che iterando la procedura applicandola di nuovo ad $(\Omega, \beta(\mathcal{A}, \mathcal{N}_0), Q)$, non si aggiungo nuovi eventi. (**Esercizio!**) Quindi, $(\Omega, \beta(\mathcal{A}, \mathcal{N}_0), Q)$ è sempre completo. Si chiama il ***completamento*** di $(\Omega, \mathcal{A}, P)$.

2.3.5 Nota. Nell lancio continuato della moneta, un elemento $\omega = (\omega_1, \omega_2, \ldots) \in \Omega$ lo possiamo interpretare come frazione binaria $0, \omega_1\omega_2 \ldots \in [0, 1]$. In questa notazione, l'evento $\{(\omega_1, \ldots, \omega_n)\} \times \Omega^\infty$ di $\mathcal{A}(\Omega)$ corrisponde al sottointervallo $[0, \omega_1 \ldots \omega_n, 0\,,\ \omega_1 \ldots \omega_n + \frac{1}{2^n}]$ di $[0, 1]$. Se la moneta è equilibrata (ossia, se $p = q = \frac{1}{2}$) la probabilità dell'evento di $\mathcal{A}(\Omega)$ che corrisponde tale intervallo, è proprio $\frac{1}{2^n}$, ossia, la lunghezza dell'intervallo. Siamo vicini alla situazione del *teorema di Vitali* e degli altri risultati discussi nelle note conclusive del primo capitolo. Dobbiamo notare, però, che la corrispondenza far elementi di Ω^∞ e elementi dell'intervallo $[0, 1]$ non è biunivoca. (Infatti, un numero binario con periodo $1\ldots$ corrisponde ad una frazione con espansione finita.) Ma è possibile assicurare in modo consistente, che "i problemi hanno probabilità 0".

2.3.6 Nota. Abbiamo visto diversi teoremi che garantiscono l'esistenza di probabilità prodotti su spazi prodotti. C'è il Teorema 2.2.2 sul prodotto di due spazi di probabilità discreti, e la sua generalizzazione ad un prodotto di un numero finito di fattori nell'Osservazione 2.2.3. Poi abbiamo visto il Teorema 2.3.3 sulla probabilizzazione di uno spazio prodotto infinito, ma molto particolare. Infatti, si trattava dello spazio prodotto più semplice in cui tutti i fattori hanno solo due elementi (un fattore che ha un solo elemento è banale) e le probabilità di queste due elementi sono le stesse in ogni fattore. Però la dimostrazione del Teorema 2.3.3 generalizza facilmente (ed è scritta tale che la generalizzazione si veda facilmente) al caso di un prodotto di una successione $(\Omega_n, \mathcal{P}(\Omega_n), P_n)$ $(n \in \mathbb{N})$ di spazi di probabilità discreti. Non è neanche troppo difficile fare lo stesso per famiglie arbitrarie di spazi di probabilità discreti. Una volta stabilito un prodotto di due (o di un numero finito) di spazi di probabilità booleani arbitrari $(\Omega_i, \mathcal{A}_i, P_i)$, anche i prodotti di famiglie arbitrarie si possono costruire allo stesso modo come sopra. I metodi per prodotti finiti, li trattiamo nell'Appendice C. Torniamo alla formulazione del risultato per un prodotto finito nel Teorema 3.3.1.

Capitolo 3

Indipendenza e probabilità condizionata

Che cosa potrebbe significare la frase "l'evento A è indipendente dall'evento B"? Ricordandoci che l'unica cosa che possiamo attribuire all'evento A è la probabilità che si verifichi, arriviamo alla frase "la probabilità dell'evento A *non dipende* da informazioni che, eventualmente, acquistiamo sull'evento B". E l'unica informazione su un evento B, che possiamo acquisire, è se si sia verificato o meno. Quindi, A sarebbe indipendente da B se la probabilità di A non cambia, quando acquistiamo l'informazione "si è verificato B" (togliendo così dallo spazio campionario tutti i casi che non verifichino B; si paragoni con la discussione nel Punto 2.2.1 sulle frequenze relative).

Questo ci lascia con il problema di indicare che cosa sarebbe la probabilità di A dato l'informazione che si è verificato B, una probabilità che vogliamo indicare con il simbolo $P(A|B)$. Solo quando abbiamo stabilito il significato di una tale *probabilità condizionata* nella Sezione 3.1, possiamo proseguire e dire che A sia *indipendente* da B, se $P(A|B) = P(A)$. Veramente, come spesso nell'assiomatizzazione, nella Sezione 3.2 sceglieremo una definizione di indipendenza, che permette di dedurre tale uguaglianza ma che risulta più flessibile nelle applicazioni.

3.1 Probabilità condizionata

Per dare un significato al simbolo $P(A|B)$, cioè alla probabilità di A *condizionata* dall'informazione che si è verificato B, ci lasciamo guidare dalla parola "probabilità", richiedendo che, soprattutto, la funzione $P(\bullet|B)\colon A \mapsto P(A|B)$ sia, a tutti gli effetti, una probabilità booleana sull'algebra booleana $\mathcal{A}$ di eventi dello spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$.

Per evitare problemi, chiediamo che $P(B)$ non è 0. (Non è molto logico, considerare il caso che si sia verificato un evento B che si verifichi con probabilità 0.) Se $P(\bullet|B)$ è una probabilità

su $\mathcal{A}$, allora vale anche la formula della probabilità totale (Teorema 1.2.11)

$$P(A|B) \;=\; P(A \cap B|B) + P(A \cap B^{\complement}|B).$$

Abbiamo, quindi, la suddivisione in due casi: Un caso di un evento $A \cap B$ contenuto in B, e l'altro caso di un evento $A \cap B^{\complement}$ disgiunto da B.

Guardiamo prima il secondo caso, cioè supponiamo che A sia un evento disgiunto da B. L'informazione che si sia verificato l'evento B, esclude che si sia verificato anche l'evento A. Una probabilità prendendo in considerazione tale informazione non può fare altro che assegnare ad A la probabilità condizionata 0. Quindi, per una probabilità condizionata sensata dobbiamo chiedere che $A \cap B = \emptyset \Longrightarrow P(A|B) = 0$.

Per eventi contenuti in B è meno ovvio che dobbiamo fare. Però, in una serie di verifiche per gli eventi $A_1 \subset B$ e $A_2 \subset B$ i numeri assoluti N_1 e N_2 dei tentativi nei quali si verifichino gli eventi A_1 e A_2, rispettivamente, non cambiano se prendiamo in considerazione solo i tentativi nei quali si è verificato anche l'evento B, perché in ogni tentativo in cui si è verificato $A_1 \subset B$ o $A_2 \subset B$ sappiamo che si è verificato anche B. Quindi, l'informazione che si sia verificato B non cambia in nessun modo il rapporto $\frac{N_1}{N_2} \approx \frac{P(A_1)}{P(A_2)}$. Perché le probabilità condizionate riflettano questo fatto, chiediamo $\frac{P(A_1)}{P(A_2)} = \frac{P(A_1|B)}{P(A_2|B)}$ qualunque siano gli eventi A_1 e A_2 contenuti in B e con $P(A_2) \neq 0$. Soprattutto, se scegliamo per A_1 un evento qualsiasi $A \subset B$ contenuto in B, e per A_2 lo stesso evento B, allora troviamo

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

dove per forza $P(B|B)$, cioè la probabilità che si verifichi B dato l'informazione che si è verificato B, la dobbiamo mettere uguale ad 1.

Intanto, per un evento qualsiasi A di $\mathcal{A}$, se applichiamo le due argomentazioni precedenti all'evento $A \cap B$ contenuto in B, rispettivamente, all'evento $A \cap B^{\complement}$ disgiunto da B, troviamo

$$P(A|B) \;=\; P(A \cap B|B) + P(A \cap B^{\complement}|B) \;=\; \frac{P(A \cap B)}{P(B)} + 0 \;=\; \frac{P(A \cap B)}{P(B)}.$$

Sarà questa la nostra definizione della probabilità condizionata.

3.1.1 Definizione. Sia B un evento di $\mathcal{A}$ con la probabilità $\boxed{P(B) \neq 0 \text{ (!!!)}}$. Allora per ogni evento $A \in \mathcal{A}$ la ***probabilità condizionata di A dato B*** è definita

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

Abbiamo motivato la definizione chiedendo, fra altre proprietà naturali, che la funzione $P(\bullet|B)$ definita su $\mathcal{A}$ sia una probabilità booleana. Adesso ci tocca a dimostrare che sia veramente così:

3.1.2 Proposizione. *Per ogni $B \in \mathcal{A}$, $P(B) \neq 0$ la funzione $P(\bullet|B)\colon A \mapsto P(A|B)$ è una probabilità booleana su $\mathcal{A}$.*

Dimostrazione. Dobbiamo verificare i tre assiomi. Chiaramente $P(A|B) = \frac{P(A\cap B)}{P(B)}$ è un numero positivo, dato che si tratta del quoziente di un numero positivo $P(A \cap B) \geq 0$ e di un numero strettamente positivo $P(B) > 0$. (Soprattutto $P(\bullet|B)$ è una funzione a valori in $\mathbb{R}$, un fatto di cui verifica avevamo trascurato finora.) Ovviamente $P(\Omega|B) = \frac{P(B)}{P(B)} = 1$. Siano A_1 e A_2 eventi disgiunti. Allora anche gli eventi $A_1 \cap B$ e $A_1 \cap B$ sono disgiunti. (Perché?) Troviamo

$$P(A_1 \cup A_2|B) = \frac{P\Big((A_1 \cup A_2) \cap B\Big)}{P(B)} = \frac{P\Big((A_1 \cap B) \cup (A_2 \cap B)\Big)}{P(B)}$$
$$= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P(A_1|B) + P(A_2|B).$$

Quindi, $P(\bullet|B)$ è additiva. ■

Convenzione: Se $P(B) = 0$ poniamo, per definizione, $P(A|B) := 0$ per ogni $A \in \mathcal{A}$. Si tratta di pura comodità, perché con questa convenzione alcune formule funzioneranno senza che dovessimo escludere tanti casi particolari. Chiaramente, se $P(B) = 0$, la funzione $P(\bullet|B)$ non è una probabilità ($P(\Omega|B) = 0$) e in questo caso il simbolo non può essere chiamato *probabilità condizionata* per nessun motivo.

N.B.: Abbiamo dimostrato che la probabilità condizionata $P(\bullet|B)\colon A \mapsto P(A|B)$ è una probabilità. Questo ***non è*** assolutamente valido come funzione dell'altra variabile B. La funzione $P(A|\bullet)\colon B \mapsto P(A|B)$ non è ***quasi mai*** una probabilità.

3.1.3 Esercizio. Che significa $P(A|\Omega) = 1$ per $P(A)$? Per un A che soddisfi tale condizione, quale valore risulterebbe per $P(A|B)$ per ogni B con $P(B) \neq 0$? Verificare che nello spazio di probabilità booleano $(\mathbb{N}, \mathcal{A}, P)$ degli Esempi 1.1.15 e 1.2.4 la funzione $B \mapsto P(\Omega|B)$ definisce una probabilità booleana su $\mathcal{A}$. Esiste uno spazio di probabilità booleano $(\Omega, \mathcal{P}(\Omega), P)$, tale che $B \mapsto P(\Omega|B)$ definisce una probabilità booleana su $\mathcal{P}(\Omega)$?

3.1.4 Esercizio. Usufruire della Proposizione 3.1.2 e dell'Osservazione 1.2.5(3) per dare un'altra dimostrazione di $A \subset B \Longrightarrow P(A) \leq P(B)$ (Corollario 1.2.8) nel caso $P(B) > 0$.

3.1.5 Esempio. Torniamo all'Esempio 2.1.9 con lo spazio di probabilità elementare

$$\Omega_2 = \{(k_1, k_2) \in \{1, \ldots, N\}^2 \colon k_1 \neq k_2\}$$

che descrive l'estrazione di due palline, una dopo l'altra, dall'urna con n palline bianche ed $N - n$ palline rosse, ed i suoi due eventi

$$A_1 = \{(k_1, k_2) \in \Omega_2 \colon k_1 \in \{1, \ldots, n\}\} \qquad A_2 = \{(k_1, k_2) \in \Omega_2 \colon k_2 \in \{1, \ldots, n\}\}$$

indicando che la prima, rispettivamente, la seconda delle palline estratte risulta bianca. Con le probabilità $P(A_1 \cap A_2) = \frac{n(n-1)}{N(N-1)} = \frac{n}{N} \cdot \frac{n-1}{N-1}$, $P(A_1 \cap A_2^{\complement}) = \frac{n(N-n)}{N(N-1)} = \frac{n}{N} \cdot \frac{N-n}{N-1}$ e $P(A_1) = \frac{n}{N}$, troviamo

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{n-1}{N-1} \qquad P(A_2^{\complement}|A_1) = \frac{P(A_1 \cap A_2^{\complement})}{P(A_1)} = \frac{N-n}{N-1}.$$

(Notiamo che $P(A_2|A_1) + P(A_2^{\complement}|A_1) = 1$ come si deve.) In ogni caso, la probabilità condizionata che otteniamo è uguale alla probabilità che otterremmo se avessimo, sin dall'inizio, estratto una pallina da un'urna con $N-1$ palline, $n-1$ di esse bianche, le altre $N-n$ rosse. In altre parole, è come se lo spazio di probabilità fosse ridotto a quello che rimane dopo l'estrazione della prima pallina. Questo comportamento è tipico per gli spazi di probabilità elementari.

Sia $(\Omega, \mathcal{P}(\Omega), P)$ uno spazio di probabilità elementare e sia $B \subset \Omega$ un suo evento con $P(B) \neq 0$ (ossia, $B \neq \emptyset$). Allora

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

È come se avessimo calcolato la probabilità del sottoinsieme $A \cap B$ (cioè la parte di A compatibile con la condizione B) nello spazio di probabilità elementare B.

Continuiamo l'esempio con l'urna. Abbiamo già notato una certa simmetria rispetto allo scambio delle coordinate degli elementi di Ω_2. Per esempio, $P(A_1) = P(A_2)$. Ne segue,

$$P(A_1|A_2) = \frac{P(A_2|A_1)P(A_1)}{P(A_2)} = P(A_2|A_1).$$

Come possiamo interpretare questa probabilità condizionata? Estraiamo la prima pallina, ma non prendiamo nota del risultato. (All'urna adesso manca una pallina, ma non ne sappiamo il colore.) Poi estraiamo la seconda pallina che risulta bianca, cioè si è verificato l'evento A_2. Poi ci possiamo chiedere, prendendo in considerazione che la seconda pallina è bianca, con che probabilità anche la prima pallina era bianca, cioè, se adesso decidessimo di guardare anche la prima pallina, qual è la probabilità condizionata dell'evento A_1 (la prima palline risulta bianca) dato A_2 (la seconda era bianca).

Similmente, troviamo

$$P(A_1|A_2^{\complement}) = \frac{P(A_2^{\complement}|A_1)P(A_1)}{P(A_2^{\complement})} = \frac{(1 - P(A_2|A_1))P(A_1)}{P(A_2^{\complement})} = \frac{(1 - \frac{n-1}{N-1}) \cdot \frac{n}{N}}{\frac{N-n}{N}} = \frac{\frac{N-n}{N-1} \cdot \frac{n}{N}}{\frac{N-n}{N}} = \frac{n}{N-1}.$$

Anche questo è uguale a $P(A_2|A_1^{\complement})$.

In quest'ultima parte dell'Esempio 3.1.5 abbiamo usufruito della probabilità condizionata $P(A|B)$ come

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

per esprimere la probabilità condizionata $P(B|A)$. Questo, assieme alla probabilità totale ha delle conseguenze molto importanti.

3.1.6 Teorema. *Sia $B_1, \ldots, B_m \in \mathcal{A}$ una partizione di Ω. Allora valgono:*

1. **(Seconda forma della formula della probabilità totale.)**

$$\begin{aligned} P(A) &= P(A \cap B_1) + \ldots + P(A \cap B_m) \\ &= P(A|B_1)P(B_1) + \ldots + P(A|B_m)P(B_m). \end{aligned}$$

2. **(Formula di Bayes.)** Sia $P(A) \neq 0$. Allora

$$\begin{aligned} P(B_k|A) &= \frac{P(A|B_k)P(B_k)}{P(A)} \\ &= \frac{P(A|B_k)P(B_k)}{P(A|B_1)P(B_1) + \ldots + P(A|B_m)P(B_m)} \end{aligned}$$

per ogni $k = 1, \ldots, m$.

Nel caso particolare di una partizione in due eventi $B_1 = B$ e $B_2 = B^{\complement}$ abbiamo

1. $P(A) = P(A|B)P(B) + P(A|B^{\complement})P(B^{\complement})$,
2. $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)} = \dfrac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^{\complement})P(B^{\complement})}$.

Le formule che abbiamo scritto forniscono subito la dimostrazione, se ci ricordiamo del fatto che, per convenzione, nei casi dove alcune delle $P(B_\ell)$ dovessero risultare 0, allora anche $P(A \cap B_\ell) = 0 = P(A|B_\ell)P(B_\ell)$.

3.1.7 Nota. Riteniamo notevole il fatto che questa versione della probabilità totale e la corrispondente formula di Bayes valgono per una qualsiasi partizione $(B_s)_{s\in S} \subset \mathcal{A}$ che soddisfa le ipotesi del Teorema 1.3.101. In particolare, se lo spazio di probabilità è discreto, allora le formule valgono per tutte le partizioni $(B_s)_{s\in S}$.

Anticipiamo che esempi tipici per partizioni, sono partizioni del tipo $(B_x)_{x\in\mathbb{R}}$, dove X è una variabile aleatoria e B_x è l'evento che X assuma il valore $x \in \mathbb{R}$. Soprattutto quando X è una variabile aleatoria *discreta*, la corrispondente partizione soddisfa le ipotesi del Teorema 1.3.101.

Le variabili aleatorie discrete le discuteremo nel Capitolo 4, e ci ricorderemo in alcune occasioni di questa nostra osservazione.

N.B.: In applicazioni tipici del Teorema 3.1.6, come illustrato nel seguente esempio, sia calcola prima $P(A)$ secondo la prima parte (normalmente solo per la partizione in due eventi B e $B^{\complement}$) e poi la probabilità condizionata $P(B|A)$ secondo la seconda parte. Non è necessario per nessun motivo di ricalcolare $P(A)$ già calcolata nella prima parte. Invece, calcoliamo molto meglio secondo la formula $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$. (Notiamo che anche il numeratore $P(A|B)P(B)$ era già stato calcolato come risultato parziale della prima parte.

3.1.8 Esempio. (Tipico per l'esame.) In una classe di 15 ragazze e 10 ragazzi, un terzo delle ragazze e la metà dei ragazzi non hanno fatto i compiti.

Qual è la probabilità che una persona scelta a caso non abbia fatto i compiti? Ci sono due caratteristiche delle persone, la prima "niente compiti" e la seconda "ragazza", con le loro rispettivi complementari. La probabilità di "niente compiti" è richiesta, mentre la probabilità di "ragazza" c'è data come $\frac{15}{15+10} = \frac{3}{5}$. Guardando le formule del Teorema 3.1.6, pare opportuno di annotare l'evento con la probabilità ricercata "niente compiti' 'con A e quello con la probabilità data "ragazza" con B. Poi ci sono dati i rapporti $\frac{1}{3}$ e $\frac{1}{2}$. Che sono questi rapporti in termini probabilistici? Se scelgo una persona a caso fra tutte le ragazze, con probabilità $\frac{1}{3}$ la persona scelta non ha fatto i compiti. Quindi, la probabilità di "niente compiti" data l'informazione "ragazza" è $\frac{1}{3}$. Allora si tratta della probabilità condizionata $P(A|B)$. Similmente, la probabilità di "niente compiti" dato l'informazione "ragazzo" è $P(A|B^{\complement}) = \frac{1}{2}$. Una volta individuate queste 3 probabilità, è facile calcolare

$$P(A) \;=\; P(A|B)P(B) + P(A|B^{\complement})P(B^{\complement}) \;=\; \frac{1}{3}\cdot\frac{3}{5} + \frac{1}{2}\cdot\frac{2}{5} \;=\; \frac{2}{5}.$$

(Evidentemente, si può arrivare alla stessa conclusione determinando il numero delle persone che non hanno fatto i compiti come $\frac{1}{3}$ delle 15 ragazze più $\frac{1}{2}$ dei 10 ragazzi, cioè $\frac{1}{3}\cdot 15 + \frac{1}{2}\cdot 10 = 10$ e dividerlo per il numero totale 25. Ma questo funziona solo in spazi di probabilità elementari e non centra con lo schema generale che doveva illustrare l'esempio.)

Qual è la probabilità che una persona che non ha fatto i compiti sia una ragazza? Questa è, chiaramente, la probabilità condizionata $P(B|A)$. Avendo già calcolato $P(A) = \frac{2}{5}$ e $P(A|B)P(B) = \frac{1}{3}\cdot\frac{3}{5} = \frac{1}{5}$, troviamo subito

$$P(B|A) \;=\; \frac{\frac{1}{5}}{\frac{2}{5}} \;=\; \frac{1}{2}.$$

Normalmente, dopo aver formalizzato il testo in modo opportuno in formule, ci sono due eventi A e B con i loro eventi complementari, la probabilità $P(B)$, oppure $P(B^{\complement})$ di cui segue $P(B) = 1 - P(B^{\complement})$, e le probabilità condizionate $P(A|B)$ e $P(A|B^{\complement})$, oppure le stesse espressioni per $A^{\complement}$. Ricordiamoci che $P(A|B)$ e $P(A^{\complement}|B)$ sono collegate tramite $P(A|B) + P(A^{\complement}|B) = 1$, e lo stesso sostituendo B con $B^{\complement}$. Al contrario, non c'è assolutamente nessun legame fra $P(A|B)$ e $P(A|B^{\complement})$. Esse possono essere scelte liberamente fra 0 e 1 da chi concepisce il problema.
Attenzione: $P(A|B)$ e $P(A \cap B)$ di solito non sono uguali!

3.1.9 Esercizio. Ci sono due software ALPHA e BETA per proteggere il computer dai virus. 60% delle persone che usano un portatile comprano il prodotto ALPHA (gli altri BETA), mentre 80% delle persone che usano il PC a casa comprano il prodotto BETA (gli altri ALPHA). Due terzi di un negozio per computer usano il portatile (gli altri il PC a casa).

1. Qual è la probabilità che un cliente del negozio compri il prodotto ALPHA?
2. Qual è la probabilità che un cliente che ha comprato il software ALPHA possieda un portatile?

3.1.10 Esempio. (Non tipico.) L'Esercizio 1.2.16 lo discutiamo sotto il punto di vista della ricetta precedente. Soprattutto ci interessa vedere perché non centra proprio nello schema. Sappiamo che d'avanti l'aula con 150 studenti dei quali vogliamo scegliere uno a caso, ci sono 70 ombrelli. Quindi la probabilità che la persona scelta ha portato con se un ombrello è data come $\frac{70}{150} = \frac{7}{15}$. Nello schema questo evento "ombrello" con la probabilità conosciuta dovrebbe essere annotato con la lettera B, mentre l'esercizio ci chiedeva di calcolare la probabilità dell'evento "ragazza" che, seguendo lo schema, dovremmo annotare con la lettera A. Però ci sono date la probabilità condizionata $P(B|A) = 60\% = \frac{3}{5}$ che una ragazza porti l'ombrello e la probabilità condizionata $P(B^{\complement}|A^{\complement}) = 80\% = \frac{4}{5}$ che un ragazzo porti l'impermeabile. Queste, secondo lo schema, sono le probabilità condizionate "sbagliate", e la soluzione del problema richiede una strategia originale. Infatti, l'unica equazione (che conosciamo) che collega queste due probabilità condizionate, è

$$P(B) = P(B|A)P(A) + P(B|A^{\complement})P(A^{\complement}).$$

Sappiamo che $P(A^{\complement}) = 1 - P(A)$ e $P(B|A^{\complement}) = 1 - P(B^{\complement}|A^{\complement})$. Allora

$$\frac{7}{15} = P(B) = P(B|A)P(A) + (1 - P(A))(1 - P(B^{\complement}|A^{\complement})) = \frac{3}{5}P(A) + \frac{1}{5} - P(A)\frac{1}{5} = \frac{2}{5}P(A) + \frac{1}{5}.$$

Ne segue $P(A) = \frac{5}{2} \cdot \frac{7-3}{15} = \frac{2}{3}$.

3.1.11 Nota. Nell'estrazione di due palline, una dopo l'altra, dall nostra urna (Esempio 3.1.5) abbiamo calcolato delle probabilita condizionate come $P(A_2|A_1)$ secondo la definizione come $\frac{P(A_2\cap A_1)}{P(A_1)}$ supponendo che Ω_2 sia uno spazio di probabilità elementare. Quasi più importante è la seguente osservazione: Se estraiamo dall'urna una pallina delle N, diciamo quella con il numero k_1, allora ci rimane un'urna con $N-1$ palline. Quante di queste siano bianche dipende pure da k_1, ma una volte estratta la pallina con il numero k_1, per l'estrazione della seconda pallina ci troviamo nuovamente di fronte di un'esperimento tipo urna e sappiamo come probabilizzarlo. Le probabilità che otteniamo sono le probabilità condizionate per eventi riguardanti la seconda pallina, dato l'evento B_{k_1} :="la prima pallina e quella con il numero k_1". Allora dato l'evento B_{k_1}, se $k_2 \neq k_1$, l'evento C_{k_2} :="la seconda pallina e quella con il numero k_2" ha la probabilità condizionata

$$P(C_{k_2}|B_{k_1}) \;=\; \frac{1}{N-1},$$

di cui segee che $P(\{(k_1,k_2)\}) \;=\; P(B_{k_1}\cap C_{k_2}) \;=\; P(B_{k_1})P(C_{k_2}|B_{k_1}) \;=\; \frac{1}{N}\frac{1}{N-1} \;=\; \frac{1}{N(N-1)}$, indipendentemente da $\{(k_1,k_2)\}\in\Omega_2$. Riconfermiamo che Ω_2 è uno spazio di probabilità elementare.

Questo modo di probabilizzare spazi prodotti $\Omega_1\times\Omega_2$ dando le probabilità dei eventi elementari di Ω_1 e poi dando delle *probabilità di transizione* da un certo punto di Ω_1 ad un punto di Ω_2 è un metodo che funziona sempre (salvo che le probabilità condizionate prescelte siano veramenre probabilità). È molto importante per definire *catene di Markov* o poi *processi di Markov*. Le loro probabilità si esprimono quasi completamente in termini di probabilità condizionate. Non abbiamo tempo di approfondire questo aspetto importantissimo delle probabilità condizionate.

3.2 Indipendenza

Ci siamo resi conto che per stabilire quando un evento A sia indipendente dalle informazioni che conosciamo sul conto dell'evento B, era necessario stabilire prima che fosse la probabilità $P(A|B)$ di A dato l'informazione "si è verificato B". Nel caso che A non dipende da quest'informazione, le due probabilità $P(A)$ e $P(A|B)$ dovrebbero risultare uguali. Per un'evento B con $P(B)\neq 0$ risulta la condizione

$$P(A) \;=\; P(A|B) \;=\; \frac{P(A\cap B)}{P(B)} \qquad \text{ossia} \qquad P(A\cap B) \;=\; P(A)\cdot P(B).$$

La seconda equazione ha senso anche se $P(B)=0$. Infatti, la definizione sarà questa.

3.2.1 Definizione. ***L'evento*** A ***è indipendente dall'evento*** B, se

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

N.B.: Non è richiesto né $P(B)\neq 0$ né $P(A)\neq 0$.
La definizione è simmetrica, cioè A è indipendente da B se e solo se B è indipendente da A. Perciò, nel seguito diciamo anche ***gli eventi*** A ***e*** B ***sono indipendenti***.

3.2.2 Esempio. Vogliamo decidere sull'indipendenza di alcuni eventi del lancio di due dadi. Lo spazio di probabilità elementare è $\Omega = \{1,\dots,6\} \times \{1,\dots,6\}$.

Consideriamo gli eventi

$$\begin{aligned} A &= \text{“primo dado pari”} = \{(n,m)\colon n \in \{2,4,6\}, m \in \{1,\dots,6\}\} = \{2,4,6\} \times \{1,\dots,6\}, \\ B &= \text{“secondo dado 5”} = \{(n,m)\colon n \in \{1,\dots,6\}, m \in \{5\}\} = \{1,\dots,6\} \times \{5\}. \end{aligned}$$

Allora

$$A \cap B = \{2,4,6\} \times \{5\}.$$

Troviamo $\#\Omega = 6 \cdot 6 = 36$,

$$\#A = 3 \cdot 6 = 18, \qquad \#B = 6 \cdot 1 = 6, \qquad \#(A \cap B) = 3 \cdot 1 = 3,$$

e, quindi,

$$\begin{aligned} P(A) = \frac{3 \cdot 6}{6 \cdot 6} = \frac{1}{2}, \qquad P(B) = \frac{6 \cdot 1}{6 \cdot 6} = \frac{1}{6}, \qquad P(A \cap B) &= \frac{3 \cdot 1}{6 \cdot 6} = \frac{1}{12} \\ &= \frac{1}{2} \cdot \frac{1}{6} = P(A) \cdot P(B). \end{aligned}$$

Allora, A e B sono indipendenti.

Tale comportamento è tipico per una coppia di eventi di un prodotto cartesiano

$$(\Omega_1 \times \Omega_2, \mathcal{P}(\Omega_1 \times \Omega_2), P)$$

di due spazi di probabilità elementari $(\Omega_1, \mathcal{P}(\Omega_1), P_1)$ e $(\Omega_2, \mathcal{P}(\Omega_2), P_2)$, dove il primo evento riguarda solo il risultato ω_1 della coppia (ω_1, ω_2) e il secondo evento riguarda solo il risultato ω_2. Infatti, siano $A_1 \subset \Omega_1$ un evento di Ω_1 e $A_2 \subset \Omega_2$ un evento di Ω_2. Allora l'evento “$\omega_1 \in A_1$” di $\Omega_1 \times \Omega_2$ è nient'altro che $A_1 \times \Omega_2 \subset \Omega_1 \times \Omega_2$, mentre l'evento “$\omega_2 \in A_2$” di $\Omega_1 \times \Omega_2$ è proprio $\Omega_1 \times A_2 \subset \Omega_1 \times \Omega_2$. L'intersezione delle due “$\omega_1 \in A_1$ e $\omega_2 \in A_2$” è

$$(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2) = A_1 \times A_2.$$

(Si veda l'Osservazione 2.2.4.) Quindi,

$$\begin{aligned} P(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) = P(A_1 \times A_2) &= \frac{\#(A_1 \times A_2)}{\#(\Omega_1 \times \Omega_2)} = \frac{\#A_1 \cdot \#A_2}{\#\Omega_1 \cdot \#\Omega_2} \\ &= \frac{\#A_1}{\#\Omega_1} \cdot \frac{\#A_2}{\#\Omega_2} = P_1(A_1) \cdot P_2(A_2) = P(A_1 \times \Omega_2) \cdot P(\Omega_1 \times A_2), \end{aligned}$$

perché

$$P_1(A_1) \;=\; \frac{\#A_1}{\#\Omega_1} \;=\; \frac{\#A_1}{\#\Omega_1}\cdot\frac{\#\Omega_2}{\#\Omega_2} \;=\; \frac{\#(A_1\times\Omega_2)}{\#(\Omega_1\times\Omega_2)} \;=\; P(A_1\times\Omega_2)$$

e

$$P_2(A_2) \;=\; \frac{\#A_2}{\#\Omega_2} \;=\; \frac{\#\Omega_1}{\#\Omega_1}\cdot\frac{\#A_2}{\#\Omega_2} \;=\; \frac{\#(\Omega_1\times A_2)}{\#(\Omega_1\times\Omega_2)} \;=\; P(\Omega_1\times A_2).$$

In altre parole, $A_1\times\Omega_2$ e $\Omega_1\times A_2$ sono indipendenti. Nella Sezione 3.3 vedremo che il prodotto cartesiano di spazi di probabilità booleane arbitrari può sempre essere probabilizzato tale che tutte le coppie di eventi $A_1\times\Omega_2$ e $\Omega_1\times A_2$ sono indipendenti.

Per essere indipendenti gli eventi di uno spazio prodotto non debbano necessariamente avere la forma $A = A_1\times\Omega_2$ e $B = \Omega_1\times A_1$. Torniamo al lancio dei due dadi.

3.2.3 Esempio. L'evento A sia quello dell'Esempio 3.2.2, cioè "primo dado pari", mentre adesso consideriamo l'evento

$$B' \;=\; \text{“somma pari”} \;=\; \Big(\{2,4,6\}\times\{2,4,6\}\Big)\cup\Big(\{1,3,5\}\times\{1,3,5\}\Big)$$

nella decomposizione in un'unione di due prodotti disgiunti. Quindi,

$$A\cap B' \;=\; \Big(\{2,4,6\}\times\{2,4,6\}\Big)\cup\emptyset \;=\; \{2,4,6\}\times\{2,4,6\}.$$

Per le probabilità troviamo

$$P(B') \;=\; \frac{3}{6}\cdot\frac{3}{6}+\frac{3}{6}\cdot\frac{3}{6} \;=\; \frac{1}{2} \qquad\text{e}\qquad P(A\cap B') \;=\; \frac{3}{6}\cdot\frac{3}{6} \;=\; \frac{1}{4} \;=\; P(A)\cdot P(B').$$

Quindi, A e B' sono indipendenti. Più generalmente, se $A' = A_1\times\{1,\dots,6\}$ è un evento che riguarda solo il primo dado, vediamo che

$$A'\cap B' \;=\; \Big((A_1\cap\{2,4,6\})\times\{2,4,6\}\Big)\cup\Big((A_1\cap\{1,3,5\})\times\{1,3,5\}\Big).$$

Per le probabilità troviamo

$$P(A') \;=\; \frac{\#A_1}{6}\cdot\frac{6}{6} \;=\; \frac{\#A_1}{6}$$

e

$$\begin{aligned} P(A'\cap B') \;&=\; \frac{\#(A_1\cap\{2,4,6\})}{6}\cdot\frac{3}{6}+\frac{\#(A_1\cap\{1,3,5\})}{6}\cdot\frac{3}{6} \\ &=\; \frac{\#(A_1\cap\{2,4,6\})+\#(A_1\cap\{1,3,5\})}{6}\cdot\frac{1}{2} \;=\; \frac{\#A_1}{6}\cdot\frac{1}{2} \;=\; P(A')\cdot P(B'). \end{aligned}$$

Quindi, B' è indipendente da un qualsiasi evento A' che riguarda solo il primo (o, per simmetria, solo il secondo) dado. (Questo è anche comprensibile: Qualunque sia l'evento A_1, per ogni $\omega_1 \in A_1$ c'è sempre una metà dei risultati ω_2 che fa la somma $\omega_1 + \omega_2$ pari e l'atra metà che fa la somma dispari.)

Se poniamo

$$A'' \;=\; \text{“somma} < 4\text{”} \;=\; \{(1,1),(1,2),(2,1)\},$$

troviamo

$$A'' \cap B' \;=\; \{(1,1)\}.$$

Le probabilità sono

$$P(A'') \;=\; \frac{3}{36} \;=\; \frac{1}{12} \qquad \text{e} \qquad P(A'' \cap B') \;=\; \frac{1}{36},$$

però

$$P(A'') \cdot P(B') \;=\; \frac{1}{12} \cdot \frac{1}{2} \;=\; \frac{1}{24} \;\neq\; \frac{1}{36}.$$

Quindi, A'' e B' non sono indipendenti.

Per non dimenticarlo mai, fissiamo una volta per tutte lo schema generale per controllare se **due** eventi A e B siano indipendenti o no: Calcolare l'intersezione $A \cap B$. Calcolare le **tre** probabilità $P(A)$, $P(B)$ e $P(A \cap B)$. Controllare se valga $P(A)\cdot P(B) = P(A\cap B)$ ($\Longrightarrow$ A e B sono indipendenti) o se valga $P(A)\cdot P(B) \neq P(A\cap B)$ ($\Longrightarrow$ A e B non sono indipendenti).

3.2.4 Esercizio. Nell'Esempio 3.1.10 degli ombrelli, gli eventi A e B sono o non sono indipendenti?

3.2.5 Esercizio. Sia $\Omega = \{(n,m)\colon n,m \in \mathbb{N}; 1 \leq n \leq N; 1 \leq m \leq M\} = \{1,\dots,N\} \times \{1,\dots,M\}$ dove $N, M \in \mathbb{N}$ uno spazio di probabilità elementare. Per $1 \leq k \leq N$ e $1 \leq \ell \leq M$ si annoti $A = \{1,\dots,k\} \times \{1,\dots,M\}$ e $B = \{1,\dots,N\} \times \{1,\dots,\ell\}$. Gli eventi A e B sono indipendenti? (Fare un disegno! Si potrebbe fare un simile disegno anche nell'esercizio precedente?)

1. Siano $N = 7$ e $M = 8$. Sia C l'evento che la somma $n + m$ sia pari. Calcolare la probabilità di C. Per $\ell = 5$ calcolare $P(C|B)$ e $P(B|C)$.
2. Ripetere la parte 1 con i valori $N = 6$ e $M = 7$ (la stessa $\ell = 5$) .

Nota: in ogni caso fare un disegno!

3.2.6 Osservazione. Raccogliamo alcune proprietà generali dell'indipendenza.

1. $P(A) = 0 \Longrightarrow A$ è indipendente da un qualsiasi evento B.

   Infatti, abbiamo $0 = P(A) \geq P(A \cap B) \geq 0$, allora $0 = P(A \cap B) = P(A) \cdot P(B)$.

2. A e B indipendenti $\Longleftrightarrow$ $A^{\complement}$ e B indipendenti.

   Infatti, valga $P(A \cap B) = P(A) \cdot P(B)$. Allora

   $$P(A^{\complement} \cap B) \;=\; P(B) - P(A \cap B) \;=\; P(B) - P(A) \cdot P(B) \;=\; (1 - P(A)) \cdot P(B) \;=\; P(A^{\complement}) \cdot P(B).$$

   Scambiando A con $A^{\complement}$ avremo l'atra direzione.

3. $P(A) = 1 \Longrightarrow A$ è indipendente da un qualsiasi evento B.

   Infatti, $P(A^{\complement}) = 0$. Così l'affermazione segue dal Punto (1) applicato ad $A^{\complement}$ e Punto (2).

Ovviamente, scambiando B con A o con $A^{\complement}$, Punto (2) ci da A e B indipendenti $\Longleftrightarrow$ $A^{\complement}$ e B indipendenti $\Longleftrightarrow$ A e $B^{\complement}$ indipendenti $\Longleftrightarrow$ $A^{\complement}$ e $B^{\complement}$ indipendenti.

N.B.: Punti (1) e (3) ci dicano che, soprattutto, $\emptyset$ ed Ω sono indipendenti da un qualsiasi evento B. In un elenco di tutti gli eventi indipendenti da B non devono mai essere dimenticati!

Possiamo raccogliere tutte le informazioni dell'Osservazione 3.2.6 in una sola proposizione. È questa la forma che ci permette di generalizzare (in modo utile) la definizione d'indipendenza di due eventi all'indipendenza di più di due eventi.

3.2.7 Proposizione. *Gli eventi A e B sono indipendenti, se e solo se*

$$P(A^{\#} \cap B^{\#}) \;=\; P(A^{\#}) \cdot P(B^{\#})$$

per ogni scelta di eventi $A^{\#} \in \beta(A) = \{\emptyset, A, A^{\complement}, \Omega\}$ e $B^{\#} \in \beta(B) = \{\emptyset, B, B^{\complement}, \Omega\}$.

3.2.8 Definizione. ***Gli eventi*** $A_1, \ldots, A_m$ ***sono indipendenti***, se

$$P(A_1^{\#} \cap \ldots \cap A_m^{\#}) \;=\; P(A_1^{\#}) \cdot \ldots \cdot P(A_m^{\#})$$

per ogni scelta di eventi $A_1^{\#} \in \beta(A_1), \ldots, A_m^{\#} \in \beta(A_m)$.

3.2.9 Nota. In genere, si definisce ***indipendente*** una famiglia $\mathcal{F} \subset \mathcal{A}$ arbitraria di eventi, se $P(A_1 \cap \ldots \cap A_m) = P(A_1) \cdot \ldots \cdot P(A_m)$ vale per ogni sottofamiglia $\{A_1, \ldots, A_m\} \subset \mathcal{F}, m \in \mathbb{N}, \#\{A_1, \ldots, A_m\} = m$ finita di $\mathcal{F}$. Per una famiglia $\mathcal{F}$ finita, tale proprietà equivale la Definizione 3.2.8.

3.2.10 Esercizio. Sia $\Omega = \{r, b, v, rbv\}$ uno spazio di probabilità con quattro elementi e la distribuzione uniforme. (Ci possiamo pensare del lancio di un tetraedro, con i quattro lati dipinti uno in rosso, uno in blu, uno in verde e l'ultimo in tutti e tre i colori.) Consideriamo i tre eventi

$$A_r \;:=\; \{r, rbv\}, \qquad A_b \;:=\; \{b, rbv\}, \qquad A_v \;:=\; \{v, rbv\}.$$

Dimostrare che i tre eventi A_r, A_b, A_v sono due a due indipendenti ma non sono indipendenti.

3.2.11 Esercizio. 1. Sia A un evento indipendente da se stesso. Che si può dire di $P(A)$?

2. Usare la prima parte per costruire uno spazio di probabilità elementare e tre suoi eventi A, B, C, tali $P(A \cap B \cap C) = P(A)P(B)P(C)$ ma che nessuna delle coppie A, B, A, C e B, C sia indipendente.

Vediamo più esempi di eventi indipendenti che emergono da variabili aleatorie indipendenti nella Sezione 4.3.

3.3 Il prodotto di spazi di probabilità booleani

Il seguente teorema risolve il problema come modellare uno spazio comune per esperimenti diversi il tal modo che gli eventi che riguardano gli esperimenti individuali diventino indipendenti. Così otteniamo anche facilmente esempi per un numero arbitrario di eventi indipendenti.

3.3.1 Teorema. *Siano* $(\Omega_i, \mathcal{A}_i, P_i)$ $(i = 1, \ldots, m)$ *spazi di probabilità booleani. Allora*

$$\mathcal{A} := \left\{ \bigcup_{i=1}^{k} A_1^i \times \ldots \times A_m^i : \quad A_j^i \in \mathcal{A}_j \ (j = 1, \ldots, m; i = 1, \ldots, k) \right\}$$

è un'algebra booleana su $\Omega := \Omega_1 \times \ldots \times \Omega_m$*, l algebra booleana più piccola, che contiene tutti gli eventi* $A_1 \times \ldots \times A_m$ $(A_j \in \mathcal{A}_j; j = 1, \ldots, m)$*. Su* $\mathcal{A}$ *esiste un'unica probabilità booleana* $P \colon \mathcal{A} \to \mathbb{R}$*, tale che*

$$P(A_1 \times \ldots \times A_m) \ = \ P_1(A_1) \cdot \ldots \cdot P_m(A_m)$$

per ogni $A_1 \in \mathcal{A}_1, \ldots, A_m \in \mathcal{A}_m$*. Chiameremo la terna* $(\Omega, \mathcal{A}, P)$ *il* ***prodotto*** *degli spazi di probabilità booleani* $(\Omega_i, \mathcal{A}_i, P_i)$*.*

Con un certo sforzo, sarebbe possibile discutere il caso del prodotto di due spazi di probabilità booleane e poi procedere per induzione. Visto, però, che il teorema è un caso particolare della Proposizione C.3.1 sugli prodotti di semialgebre booleane e le loro probabilità, non discuteremo tale dimostrazione qua ma la rimandiamo tutto all'Appendice C.

Le probabilità dell'intersezione degli eventi "$\omega_j \in A_j^{\#}$" $= \Omega_1 \times \ldots \times A_j^{\#} \times \ldots \times \Omega_m$ $(j = 1, \ldots, m)$ fattorizzano, qualunque siano le scelte $A_j^{\#} \in \beta(A_j)$. Quindi gli eventi "$\omega_j \in A_j$" $(j = 1, \ldots, m)$ sono indipendenti.

3.3.2 Esempio. Gli eventi $\{0, 1\} \times \ldots \times \{0, 1\} \times A_i \times \{0, 1\} \times \ldots \times \{0, 1\}$ di $\{0, 1\}^n$ nel lancio della moneta per n volte, sono indipendenti. Infatti, qualunque siano gli eventi $A_i \subset \{0, 1\}$ la

probabilità dell'evento $A_1 \times \ldots \times A_n$ è

$$P(A_1 \times \ldots \times A_n) = \sum_{(\omega_1,\ldots,\omega_m)\in A_1\times\ldots\times A_n} p^{\omega_1}q^{1-\omega_1} \cdot \ldots \cdot p^{\omega_n}q^{1-\omega_n}$$

$$= \sum_{\omega_1\in A_1} \ldots \sum_{\omega_n\in A_n} p^{\omega_1}q^{1-\omega_1} \cdot \ldots \cdot p^{\omega_n}q^{1-\omega_n} = \sum_{\omega_1\in A_1} p^{\omega_1}q^{1-\omega_1} \ldots \sum_{\omega_n\in A_n} p^{\omega_n}q^{1-\omega_n}$$

$$= \left(\sum_{\omega_1\in A_1} p^{\omega_1}q^{1-\omega_1}\right) \cdot \ldots \cdot \left(\sum_{\omega_n\in A_n} p^{\omega_n}q^{1-\omega_n}\right) = P(A_1) \cdot \ldots \cdot P(A_n).$$

Quell'esempio è un caso particolare dell'Osservazione 2.2.3, la quale, a suo turno, è un caso particolare del Teorema 3.3.1. Veramente, l'esempio generalizza al lancio continuato modellato su Ω^∞ secondo il Teorema 2.3.3: Tutti i lanci sono indipendenti secondo la definizione nella Nota 3.2.9. Ci saranno più esempi concreti nella Sezione 4.3 sull'indipendenza di variabili aleatorie.

3.3.3 Esempio. Nel lancio continuato di una moneta modellato secondo il Teorema 2.3.3, sia $B = B_n \times \Omega^\infty$ con $B_n \subset \Omega^n$ un evento che riguarda solo i primi n lanci. Ci interessa il numero k di lanci che, dopo n lanci, dobbiamo ancora aspettare al prossimo successo. In altre parole ci interessa l'evento

$$A := \Omega^n \times \{(0,\ldots,0,1) \in \Omega^k\} \times \Omega^\infty.$$

Secondo il Teorema 2.3.3, per calcolare delle probabilità basta limitarsi su gli eventi di Ω^{n+k}. Secondo l'Esempio 3.3.2, troviamo

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P_{n+k}\Big(B_n \times \{(0,\ldots,0,1) \in \Omega^k\}\Big)}{P_{n+k}(B_n \times \Omega^k)}$$

$$= \frac{P_n(B_n)P_k\Big(\{(0,\ldots,0,1) \in \Omega^k\}\Big)}{P_n(B_n)P_k(\Omega^k)} = P_k(\{(0,\ldots,0,1)\}) = geom_p(k).$$

Interpretazione: Avendo lanciato già n volte, il numero di volte ancora da lanciare fino al prossimo successo non dipende dalla *preistoria* avvenuta nei primi n lanci. Per questo comportamento, la legge geometrica si chiama anche la legge (discreta) ***senza memoria***.

Capitolo 4

Variabili aleatorie discrete

Un compito del calcolo delle probabilità è fare predizioni, come indicare un *valore atteso* di un esperimento. Di un campione statistico possiamo calcolare la *media statistica* e, intuitivamente, ci aspettiamo che per campioni sempre più grandi la media si avvicini sempre di più al valore atteso. Questa nostra aspettativa è esattamente la stessa che ci fa aspettare che le frequenze relative dovrebbero avvicinarsi sempre più alla probabilità. Un *valore* atteso è soprattutto un dato numerico. Non ha senso chiedere quale sia la media di "testa" e "croce" per il lancio di una moneta. Però se vinciamo per ogni "testa" un euro e per ogni "croce" niente, la media su tutti i tentativi ci indica un *valore medio* delle vincite: È come se avessimo vinto in ogni tentativo questo valore medio.

Quello che abbiamo fatto non è altro che trasformare i risultati ω di Ω, con cui in genere non è possibile calcolare, in valori numerici, con cui è possibile calcolare. (Ricordiamoci quanto ci è stato utile il significato numerico di "1" (per "testa") e "0" (per "croce") quando abbiamo calcolato la probabilità di avere k "teste" in n lanci negli Esempi 2.1.12 e 2.2.6!) In altre parole, per fare dei calcoli ci tocca a trasformare i risultati ω di Ω in valori numerici. In altre parole, dobbiamo considerare delle funzioni X da Ω in $\mathbb{R}$.

Questo capitolo è dedicato alle variabili aleatorie *discrete*, una sottoclasse di tutte le variabili aleatorie particolarmente buona con un minimo di difficoltà tecniche da risolvere. Prima di passare alle variabili discrete vorremmo, però, discutere un po' alcuni aspetti delle variabili aleatorie generali come fanno apparizione in molti testi di probabilità. La ragione è che tutti i concetti e quasi tutti i teoremi che discuteremo in questo capitolo mantengono validità per i casi più generali. Per esempio, nel Capitolo 5 ci limitiamo a indicare tali proprietà delle variabili aleatorie discrete che valgono anche per le variabili aleatorie *continui*. È possibile saltare il resto di questa parte introduttiva, e continuare subito con la Sezione 4.1. È, però, opportuno tornarci prima della lettura del Capitolo 5.

Sia $X\colon \Omega \to \mathbb{R}$ una funzione. Di tale funzione non possiamo dire più del valore $X(\omega)$ che assume (dipendente dal risultato $\omega \in \Omega$ che si verifica). Se il verificarsi di un certo evento $A \subset \Omega$ può essere deciso guardando solo il valore di X, allora possiamo indicare un sottoinsieme B di $\mathbb{R}$, che contiene esattamente questi valori $x \in \mathbb{R}$ per i quali si verifica l'evento A di Ω. Conseguentemente, $B^{\complement}$ contiene i valori $x \in \mathbb{R}$ per i quali l'evento A non si verifica. Il fatto che la domanda A (cioè, "$\omega \in A$, si o no?") può essere decisa guardando solo $X(\omega)$ limita le possibili apparenze per un tale evento. Infatti, se per un certo $\omega \in \Omega$ il valore $x = X(\omega)$ ci dice $\omega \in A$ (ossia, $x \in B$), allora anche un qualsiasi ω' per il quale $X(\omega') = x$ deve appartenere ad A. Noi, per decidere, possiamo solo riferirci al valore x. Non sappiamo se questo valore $x \in B$ è apparso perché il risultato era ω o era ω' o forse era un qualsiasi altro ω'' che soddisfi $X(\omega'') = x \in B$. Con questo ragionamento vediamo che, necessariamente, A ha la forma

$$A \;=\; \{\omega \in \Omega\colon X(\omega) \in B\} \;=:\; X^{-1}(B) \quad (\text{la } \textbf{\textit{controimmagine}} \text{ di } B \text{ sotto } X).$$

In altre parole, gli eventi che *riguardano* i valori di X sono le controimmagini $A = X^{-1}(B) \subset \Omega$ di sottoinsiemi $B \subset \mathbb{R}$. Annotiamo un tale evento anche con $A =$"$X \in B$".

Una probabilità $P(A)$ per un evento $A \subset \Omega$ dello spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$ è definita solo se A appartiene all'algebra booleana $\mathcal{A}$. Quindi, se vogliamo calcolare la probabilità $P(X \in B)$ dell'evento "$X \in B$", è necessario che "$X \in B$"$= X^{-1}(B)$ sia un elemento di $\mathcal{A}$. Chiaramente, se $\mathcal{A} = \mathcal{P}(\Omega)$ non risulta nessuna restrizione. Ma con l'esempio del lancio continuato della moneta (Teorema 2.3.3) conosciamo già un esempio dove non siamo riusciti a definire una probabilità su tutto $\mathcal{P}(\Omega)$.

La condizione $X^{-1}(B) \in \mathcal{A}$ può essere letta come condizione per la funzione X (se B è dato) o come condizione per B (se X è data). In prassi, non è opportuno di porre condizioni ai sottoinsiemi B di $\mathbb{R}$ che dipendono dalla funzione X in considerazione. Sarebbe troppo scomodo dover aggiustare un'algebra booleana di eventi B per i valori di X per ogni scelta specifica di X. Dall'altra parte sappiamo, però, che in genere non ci sarà possibile ammettere per B proprio tutti gli sottoinsiemi di $\mathbb{R}$. Nel Capitolo 5 arriviamo alla conclusione che una variabile aleatoria X "buona" dovrebbe assegnare una probabilità almeno all'evento "$X \in [a, b]$" per ogni intervallo. Tale condizione, infatti, rimane valida per una qualsiasi variabile aleatoria a valori in $\mathbb{R}$ in un qualsiasi testo di probabilità.

La definizione delle variabili aleatorie discrete, che forma la base di partenza in questo capitolo, ci assicura la possibilità di calcolare in modo consistente una probabilità $P(X \in B)$ per un qualsiasi sottoinsieme $B \subset \mathbb{R}$, addirittura anche se "$X \in B$" dovesse risultare non essere elemento di $\mathcal{A}$. In questo capitolo permettiamoci questa libertà senza preoccuparci troppo della soluzione del problema formale e marginale che, comunque, può sempre essere risolto; si veda la Nota 4.1.5.

4.1 Variabili aleatorie discrete e le loro leggi

4.1.1 Definizione. Una ***variabile aleatoria (reale) discreta*** su uno spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$ è una funzione $X\colon \Omega \to \mathbb{R}$ tale che

$$\text{“}X = x\text{”} \;:=\; X^{-1}(\{x\}) \;\in\; \mathcal{A} \quad \text{per ogni } x \in \mathbb{R}, \qquad \text{e} \qquad \sum_{x\in\mathbb{R}} p(x) \;=\; 1,$$

dove $p(x) := P(X = x)$. Chiameremo la funzione $p\colon x \mapsto p(x)$ la ***densità (discreta)*** di X. La probabilità booleana P_X su $\mathcal{P}(\mathbb{R})$ definita secondo il Teorema 1.3.9 come

$$P_X(B) \;=\; \sum_{x\in B} p(x).$$

chiameremo la ***legge*** di X.

N.B.: La proprietà di essere discreta è una proprietà della variabile, non dello spazio di probabilità booleano su cui è definita. Infatti, $(\Omega, \mathcal{A}, P)$ era assunto arbitrario.

N.B.: Più della legge non possiamo dire di una variabile aleatoria. Ω sparisce dalla circolazione. Se due variabili aleatorie (discrete) X e Y hanno ***la stessa legge***, lo annotiamo $\boxed{X \sim Y}$. (Non significa che debbano essere uguali, visto che possono essere definite su spazi di probabilità diversi. Anche se sono definite sullo stesso spazio, non coincidono necessariamente; si veda l'Esempio 4.2.5.)

N.B.: Per essere una variabile aleatoria vera e propria, come fanno apparizione in ogni testo, X debba soddisfare anche “$X \in [a, b]$”$\in \mathcal{A}$ per ogni $a \le b$. Mentre quest'ipotesi non basta, per esempio, per dimostrare la Proposizione 4.2.7 per funzioni φ arbitrari, vedremo nell'Osservazione 4.1.5 che per una variabile aleatoria discreta X è sempre possibile fare $\mathcal{A}$ più grande in modo consistente tale che contenga tutti gli eventi “$X \in B$” per qualsiasi $B \subset \mathbb{R}$.

4.1.2 Osservazione. Per ogni funzione $X\colon \Omega \to \mathbb{R}$ poniamo B_x :=“$X = x$”. Ovviamente (**esercizio!**) la famiglia $(B_x)_{x\in\mathbb{R}}$ è una partizione di Ω. Per definizione, X è una variabile aleatoria discreta se e solo se la partizione $(B_x)_{x\in\mathbb{R}}$ soddisfa le ipotesi della Parte 1 del Teorema 1.3.10.

4.1.3 Corollario. *Ogni funzione $X\colon \Omega \to \mathbb{R}$ su uno spazio di probabilità discreto $(\Omega, \mathcal{P}(\Omega), P)$ è una variabile aleatoria discreta.*

Dimostrazione. Secondo la Parte 2 del Teorema 1.3.10, ogni partizione di uno spazio di probabilità discreto soddisfa la Parte 1 . ■

La legge P_X di una variabile aleatoria discreta ci permette calcolare $P(X \in B)$ salvo che “$X \in B$“ sia di $\mathcal{A}$.

4.1.4 Corollario. *Sia $B \subset \mathbb{R}$ tale che "$X \in B$"$\in \mathcal{A}$. Allora $P(X \in B) = P_X(B)$.*

Dimostrazione. L'Equazione (1.3.4) ci dice per l'evento $A =$"$X \in B$" di $\mathcal{A}$, che

$$P(X \in B) \;=\; \sum_{x\in\mathbb{R}} P(\text{"}X \in B\text{"} \cap \text{"}X = x\text{"}) \;=\; \sum_{x\in B} P(X = x) \;=\; P_X(B). \blacksquare$$

4.1.5 Osservazione. Ricordiamoci che la legge P_X di una variabile aleatoria discreta X ci permette di calcolare $P_X(B)$ pero **ogni** sottoinsieme B di $\mathbb{R}$. Il Corollario 4.1.4 ci dice che se "$X \in B$"$\subset \Omega$ dovesse essere un evento proprio di $\mathcal{A} \subset \mathcal{P}(\Omega)$, allora la sua probabilità $P(X \in B)$ coincide con quella $P_X(B)$ calcolata secondo la legge di X. Vogliamo estendere $\mathcal{A}$ e P tale che l'estensione contenga tutti gli eventi "$X \in B$" e li dia le probabilità $P_X(B)$.

Infatti, l'estensione $\mathcal{B}$ di $\mathcal{A}$ rispetto alla partizione $(B_x)_{x\in\mathbb{R}}$ secondo il Teorema 1.3.15, contiene tutti gli eventi "$X \in B$" (perché, "$X \in B$"$\cap B_x$ è $B_x \in \mathcal{A}$ se $x \in B$ ed è $\emptyset \in \mathcal{A}$ se $x \notin B$) e la probabilità $Q(X \in B)$ di tale evento coincide per definizione con $P_X(B)$.

Dopo questo risultato, nel seguito supponiamo sempre che $\mathcal{A} \ni$"$X \in B$" per ogni $B \subset \mathbb{R}$.

4.1.6 Esempio. Abbiamo discusso spesso la somma di due dadi (l'esercizio dell'introduzione continuato negli Esercizi 1.1.1 e 2.1.1). Nella solita descrizione sullo spazio di probabilità elementare $\Omega = \{1, \dots, 6\}^2$, la somma la otteniamo come funzione

$$X(\omega) \;=\; \omega_1 + \omega_2$$

di $\omega = (\omega_1, \omega_2) \in \Omega$. Supponiamo di avere due giocatori di cui il primo lancia il primo dado ed il secondo lancia il secondo. Se ci interessiamo del risultato del giocatore numero i ($i = 1, 2$), l'otteniamo come valore della funzione

$$X_i(\omega) \;=\; \omega_i$$

di ω. Vediamo che la somma la possiamo esprimere usando X_1 e X_2 come $X = X_1 + X_2$. Supponiamo che i due si giocano soldi. Chi ha il numero più grande dell'avversario, riceve da lui un euro. Con parità, nessuno paga nulla. La vincita Y_1 del primo giocatore è

$$Y_1(\omega) \;=\; \begin{cases} 1 & \omega_1 > \omega_2, \\ 0 & \omega_1 = \omega_2, \\ -1 & \omega_1 < \omega_2. \end{cases}$$

La vincita Y_2 del secondo giocatore è la perdita del primo: $Y_2 = -Y_1$. Adesso cambiamo le regole: Ogni giocatore riceve la differenza fra suo dado e quello del suo avversario. La vincita del primo giocatore secondo le nuove regole è

$$Z_1(\omega) \;=\; \omega_1 - \omega_2 \qquad \text{ossia} \qquad Z_1 \;=\; X_1 - X_2.$$

Nuovamente, la vincita del secondo giocatore è $Z_2 = X_2 - X_1 = -Z_1$.

4.1.7 Esempio. Per ogni probabilità P su $\mathcal{P}(\mathbb{R})$ tale che $(\mathbb{R}, \mathcal{P}(\mathbb{R}), P)$ diventi uno spazio di probabilità discreto, l'identità $X = \mathsf{id}_{\mathbb{R}} : x \mapsto x$ è una variabile aleatoria discreta su $\mathbb{R}$ con la legge $P_X = P$. Tutte le probabilità che abbiamo chiamate leggi nel Capitolo 2 sono, infatti, esempi di leggi di variabili aleatorie discrete.

1. La variabile aleatoria X che indica il numero k di successi in n ripetizioni di un esperimento bernoulliano ha la legge binomiale $B_{n,p}$ di parametri $n \in \mathbb{N}_0$, $p \in (0,1)$ su $\{0, \ldots, n\}$ con $P(X = k) = p(k) = B_{n,p}(k)$. In particolare, la variabile di un solo esperimento bernoulliano che assume il valore 1 nel caso di successo e il valore 0 nel caso di insuccesso ha la legge di Bernoulli $B_{1,p}$ di parametro $p \in (0,1)$ su $\{0,1\}$.

2. La variabile aleatoria X che indica il momento k del n–esimo successo nel lancio continuato di una moneta reale ha la legge di Pascal $Pas_{n,p}$ di parametri $n \in \mathbb{N}, p \in (0,1)$ su $\{n, n+1, \ldots\}$ con $p(k) = Pas_{n,p}(k)$. In particolare, la variabile che indica il momento del primo successo, ha la legge geometrica $geom_p = Pas_{1,p}$ di parametro $p \in (0,1)$ su $\{1, 2, \ldots\}$.

3. La legge di Poisson Poi_λ di parametro $\lambda > 0$ è la legge di una variabile aleatoria su $\mathbb{N}_0 = \{0, 1, \ldots\}$. Per esempio, possiamo porre $X = \mathsf{id}\colon \mathbb{N}_0 \colon \mathbb{N}_0 \subset \mathbb{R}$ dove $\Omega = \mathbb{N}_0$ come discusso nel Teorema 2.2.10.

4. Similmente, la legge ipergeometrica $H_{N,n,k}$, come ci siamo convinti senza dare troppi dettagli, è concentrata su un sottoinsieme finito di $\mathbb{N}_0$ e definisce la legge di una variabile aleatoria discreta.

4.1.8 Esercizio. Indicare nell'Esempio 2.1.3 una variabile aleatoria concreta (cioè una funzione da Ω in $\mathbb{R}$) che ha una legge ipergeometrica.

4.1.9 Esempio. Continuiamo l'Esempio 4.1.6: Secondo l'Esercizio 2.1.1 sappiamo

$$P(X = 2) = P(\text{“somma è 2”}) = \frac{1}{36}, \qquad P(X = 3) = P(\text{“somma è 3”}) = \frac{2}{36} = \frac{1}{18},$$
$$P(X = 8) = P(\text{“somma è 8”}) = \frac{5}{36}, \qquad P(\text{“}X \text{ è pari”}) = P(\text{“somma è pari”}) = \frac{1}{2},$$
$$P(X < 4) = P(\text{“somma} < 4\text{”}) = \frac{3}{36} = \frac{1}{12}.$$

Per le variabili aleatorie Y_i troviamo

$$P(Y_1 = 0) = P(Y_2 = 0) = P(\{(1,1), \ldots, (6,6)\}) = \frac{6}{36} = \frac{1}{6}$$

e, per simmetria, $P(Y_1 = 1) = P(Y_2 = 1)$

$$P(Y_{1/2} = 1) \;=\; P(Y_{2/1} = -1) \;=\; \frac{\frac{36-6}{2}}{36} \;=\; \frac{5}{12}.$$

4.1.10 Esercizio. Determinare le leggi delle variabili aleatorie Y_1 e Y_2 dell'Esempio 4.1.6.

4.1.11 Esempio. Per ogni $A \in \mathcal{A}$ la funzione indicatrice $X := \chi_A \colon \Omega \to \{0, 1\} \subset \mathbb{R}$ è una variabile aleatoria discreta con $p(0) = P(X = 0) = P(A^{\complement}) =: q$ e $p(1) = P(X = 1) = P(A) =: p$. Quindi X ha la legge di Bernoulli con la densità $p(x) = B_{1,p}(x)$.

4.2 Leggi congiunti

Se abbiamo due variabili aleatorie X e Y con le loro leggi, non sappiamo ancora tutto di loro. La conoscenza delle due leggi individuali, in genere, non ci permette ancora di calcolare una probabilità come $P(X = x, Y = y)$. (Si veda l'Esempio 4.2.5. Nella Sezione 4.3 impareremo che le leggi individuali (ossia *marginali*) determinano tutto (ossia la legge *congiunta*), se le variabili aleatorie sono *indipendenti*.)

4.2.1 Definizione. Un'm–upla $X = (X_1, \ldots, X_m)$ di variabili aleatorie si chiama ***vettore aleatorio discreto*** ossia ***variabile aleatoria discreta a valori in*** $\mathbb{R}^m$ ossia ***variabile aleatoria discreta*** m***–dimensionale***, se ogni X_i è una variabile aleatoria discreta. Per ogni $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$ sia

$$p(x) \;=\; p(x_1, \ldots, x_m) \;:=\; P(X = x) \;=\; P(X_1 = x_1, \ldots, X_m = x_m).$$

Chiameremo la funzione $p \colon p \to p(x)$ la ***densità congiunta (discreta)*** di X.
N.B.: Le nozioni "vettore aleatorio" o "variabile aleatoria a valori in $\mathbb{R}^m$" si spiegano da sole se notiamo che $X \colon \Omega \mapsto (X_1(\omega), \ldots, X_m(\omega)) \in \mathbb{R}^m$ definisce una funzione a valori in $\mathbb{R}^m$, ossia in uno spazio vettoriale.

Ogni evento "$X_i = x_i$" ($i = 1, \ldots, m$) è di $\mathcal{A}$, di cui segue che anche l'evento ("$X = x$") = ("$X_1 = x_1$")$\cap \ldots \cap$("$X_m = x_m$") è di $\mathcal{A}$ e, quindi, $p(x)$ è ben definito. Dal Teorema 1.3.10 ed il fatto che X_k è una variabile aleatoria discreta, segue che

$$\sum_{x_k \in \mathbb{R}} P(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, X_k = x_k) \;=\; P(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}).$$

Applicazione multipla di questa formula ci da:

4.2.2 Proposizione.
$$\sum_{x \in \mathbb{R}^m} p(x) \;=\; 1.$$

4.2.3 Definizione. Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio discreto. Chiameremo la ***legge congiunta discreta*** di X la probabilità booleana P_X su $\mathcal{P}(\mathbb{R}^m)$ definita secondo il Teorema 1.3.9 come

$$P_X(B) \;=\; \sum_{x \in B} p(x).$$

La ***legge marginale*** della componente X_i di X con la sua ***densità marginale*** rispetto ad X sono semplicemente legge P_{X_i} e densità $p_i(x_i)$ della variabile aleatoria discreta X_i. Più generale, per $1 \le k < m$ sia $f\colon \{1, \dots, k\} \to \{1, \dots, m\}$ una funzione iniettiva (ossia, una disposizione; si veda l'Appendice A.1). Poniamo $Y = (Y_1, \dots, Y_k)$ con $Y_\ell := X_{f(\ell)}$. (Parlando in modo più rilassato, Y è una sotto–k–upla dell'm–upla X.) La ***legge marginale*** e la ***densità marginale*** di Y rispetto ad X sono semplicemente legge P_Y e densità $q(y)$ del vettore aleatorio discreto Y.

4.2.4 Esempio. La densità marginale di X_1 rispetto ad X è

$$p_1(x_1) \;=\; P(X_1 = x_1) \;=\; \sum_{x_2 \in \mathbb{R}} \dots \sum_{x_m \in \mathbb{R}} p(x_1, \dots, x_m).$$

Notiamo che le sommatorie vanno solo sulle variabili $x_2, \dots, x_m$, mentre l'argomento x_1 è fissato. Analogamente, se $Y = (X_1, \dots, X_k)$, troviamo

$$q(y_1, \dots, y_k) \;=\; P(X_1 = y_1, \dots, X_k = y_k) \;=\; \sum_{x_{k+1} \in \mathbb{R}} \dots \sum_{x_m \in \mathbb{R}} p(y_1, \dots, y_k, x_{k+1}, \dots, x_m).$$

4.2.5 Esempio. Le leggi marginali non determinano la legge congiunta. Sia $\Omega = \{0,1\} \times \{0,1\} = \{(0,0), (0,1), (1,0), (1,1)\}$ uno spazio di probabilità elementare. Siano $X_i(\omega_1, \omega_2) = \omega_i$ due variabili aleatorie. Sia $X = (X_1, X_2)$. Allora

$$p(x_1, x_2) \;=\; \begin{cases} \frac{1}{4} & (x_1, x_2) \in \Omega \subset \mathbb{R}^2, \\ 0 & \text{altrimenti.} \end{cases}$$

Verifichiamo facilmente (**esercizio!**) che $p_i(x_i) = \frac{1}{2}$ se $x_i \in \{0,1\}$ e 0 altrimenti. In particolare, $X_1 \sim X_2$.

Adesso sia $Y = (X_1, X_1)$. Chiaramente $Y_1 = Y_2 = X_1$ hanno tutte la stessa medesima legge. Però la densità congiunta di Y è

$$q(y_1, y_2) \;=\; \begin{cases} \frac{1}{2} & y_1 = y_2 \in \{0,1\}, \\ 0 & \text{altrimenti.} \end{cases}$$

Ovviamente p e q sono densità diverse su $\mathbb{R}^2$. Infatti, vedremo nella prossima sezione che le due variabili aleatorie della coppia X sono *indipendenti* mentre nella coppia Y l'una delle variabili aleatorie determina completamente l'altra.

4.2.6 Osservazione. Continuiamo usare la notazione B_x ="$X = x$" dell'Osservazione 4.1.2, solo che adesso x è di $\mathbb{R}_m$. Vale l'analogo (con lo stesso ragionamento) del Corollario 4.1.3: Una qualsiasi funzione $X = (X_1, \dots, X_m)\colon \Omega \to \mathbb{R}^m$ su uno spazio di probabilità discreto è un vettore aleatorio discreto.

Secondo la Proposizione 4.2.2, anche la partizione $(B_x)_{x\in\mathbb{R}^m}$ soddisfa le ipotesi della Teorema 1.3.10(1). Vale, quindi, l'affermazione analoga del Corollario 4.1.4: Se per un certo $B \subset \mathbb{R}^m$ vale "$X \in B$"$\in \mathcal{A}$, allora $P(X \in B) = P_X(B)$.

Come nell'Osservazione 4.1.5, possiamo estendere l'algebra booleana $\mathcal{A}$ e la probabilità, tale che contenga anche gli eventi "$X \in B$" per ogni $B \subset \mathbb{R}^m$ e che, quindi, $P(X \in B) = P_X(B)$ per ogni $B \subset \mathbb{R}^m$.

D'ora in poi supponiamo sempre che (dopo, ove necessario, aver applicato l'Osservazione 4.2.6) che per un vettore aleatorio discreto $X = (X_1, \dots, X_m)$ valga "$X_i \in B_i$"$\in \mathcal{A}$ per ogni $B_i \subset \mathbb{R}$ $(i = 1, \dots, m)$.

La seguente proposizione è assai importante. Ci permette calcolare le probabilità di eventi che riguardano una funzione $Z = \varphi \circ X$ del vettore aleatorio X in termini che coinvolgono solo la legge congiunta di X. Non occorre, quindi, conoscere la legge di Z stessa. Anche l'affermazione che qualsiasi funzione di un vettore aleatorio discreto è di nuovo una variabile aleatoria discreta, è un risultato forte.

4.2.7 Proposizione. *Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio discreto (tale che "$X \in B$" per ogni $B \subset \mathbb{R}^m$) e sia $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ una funzione. Allora anche la funzione $Z = \varphi \circ X$ definita come*

$$Z(\omega) \;:=\; \varphi(X_1(\omega), \dots, X_m(\omega))$$

è una variabile aleatoria discreta con densità discreta $q(z) = P(X \in B_z) = P_X(B_z)$, dove definiamo

$$B_z \;:=\; \{x \in \mathbb{R}^m\colon \varphi(x_1, \dots, x_m) = z\}.$$

Dimostrazione. Osserviamo che $(B_z)_{z\in\mathbb{R}}$ è una partizione dello spazio di probabilità discreto $(\mathbb{R}^m, \mathcal{P}(\mathbb{R}^m), P_X)$. Allora, secondo il Teorema 1.3.10(1), $\sum_{z\in\mathbb{R}} P(X \in B_z) = \sum_{z\in\mathbb{R}} P_X(B_z) = 1$. Evidentemente "$Z = z$" e "$X \in B_z$" sono due modi equivalenti di descrivere lo stesso evento di $\mathcal{A}$. Quindi, $\sum_{z\in\mathbb{R}} q(z) = 1$ dove $q(z) := P(Z = z) = P(X \in B_z)$. ∎

4.3 Indipendenza di variabili aleatorie discrete

È naturale dire che due variabili aleatorie X e Y sono indipendenti se gli eventi "$X \in B$" che riguardano solo la prima sono indipendenti dagli eventi "$Y \in C$" che riguardano solo la seconda. Notiamo che se $A :=$"$X \in B$" è un evento di $\mathcal{A}$ che riguarda X allora anche $A^{\complement} =$"$X \in B^{\complement}$" (**esercizio:** dimostrarlo!) è un evento di $\mathcal{A}$ che riguarda X. Ovviamente, $\emptyset =$"$X \in \emptyset$" e $\Omega =$"$X \in \mathbb{R}$" sono eventi di $\mathcal{A}$ che riguardano X. Grazie a questa preparazione possiamo dare subito la definizione generale.

4.3.1 Definizione. Sia $X = (X_1, \ldots, X_m)$ un vettore aleatorio discreto. Allora le variabili aleatorie $X_1, \ldots, X_m$ sono ***indipendenti*** se

$$P(X_1 \in B_1, \ldots, X_m \in B_m) \;=\; P(X_1 \in B_1) \cdot \ldots \cdot P(X_m \in B_m)$$

per ogni $B_1, \ldots, B_m \subset \mathbb{R}$.

Secondo la preparazione, questo significa che gli eventi "$X_1 \in B_1$", ..., "$X_m \in B_m$" sono indipendenti per ogni scelta $B_1, \ldots, B_m \subset \mathbb{R}$.

4.3.2 Osservazione. Notiamo che

$$\begin{aligned} P(X_1 \in B_1, \ldots, X_m \in B_m) \;&=\; P(X \in B_1 \times \ldots \times B_m) \;=\; P_X(B_1 \times \ldots \times B_m) \\ &=\; P(X_1 \in B_1) \cdot \ldots \cdot P(X_m \in B_m) \;=\; P_{X_1}(B_1) \cdot \ldots \cdot P_{X_m}(B_m). \end{aligned}$$

In altre parole, $(\mathbb{R}^m, \mathcal{P}(\mathbb{R}^m), P_X)$ è probabilizzato come il prodotto degli spazi di probabilità discreti $(\mathbb{R}, \mathcal{P}(\mathbb{R}), P_{X_i})$ secondo l'Osservazione 2.2.3.

4.3.3 Esempio. Ricordiamoci dall'Esempio 4.1.11 che gli eventi $A_1, \ldots, A_m$ di $\mathcal{A}$ possono essere descritti in modo equivalente come variabili aleatorie $X_i := \chi_{A_i}$ a valori in $\{0, 1\} \subset \mathbb{R}$. Troviamo

$$\text{"}X_i \in B_i\text{"} \;=\; \left.\begin{cases} \emptyset & 1 \notin B_i, 0 \notin B_i \\ A_i & 1 \in B_i, 0 \notin B_i \\ A_i^{\complement} & 1 \notin B_i, 0 \in B_i \\ \Omega & 1 \in B_i, 0 \in B_i \end{cases}\right\} \;=:\; A_i^{\#} \;\in\; \beta(A_i),$$

cosicché

$$P(X_1 \in B_1, \ldots, X_m \in B_m) \;=\; P(A_1^{\#} \cap \ldots \cap A_m^{\#}).$$

Allora, le variabili aleatorie $X_1, \ldots, X_m$ sono indipendenti, se e solo se gli eventi $A_1, \ldots, A_m$ sono indipendenti.

4.3.4 Proposizione. *Sia $X = (X_1, \ldots, X_m)$ un vettore aleatorio discreto. Allora le variabili aleatorie $X_1, \ldots, X_m$ sono indipendenti se e solo se*

$$p(x) \;=\; p_1(x_1) \cdot \ldots \cdot p_m(x_m).$$

Dimostrazione. Lo dimostriamo solo per $m = 2$. (La dimostrazione generale richiederebbe solo più indici.) Siano X e Y variabili aleatorie discrete con densità marginali $p(x)$ e $q(y)$.

"$\Longrightarrow$". Se X e Y sono indipendenti, allora troviamo per la densità congiunta

$$p(x,y) \;=\; P(X = x, Y = y) \;=\; P(\text{“}X = x\text{”} \cap \text{“}Y = y\text{”}) \;=\; P(X = x) \cdot P(Y = y) \;=\; p(x) \cdot q(y).$$

"$\Longleftarrow$". Se $p(x,y) = p(x) \cdot q(y)$, allora per ogni $B, C \subset \mathbb{R}$ vale

$$\begin{aligned} P(X \in B, Y \in C) \;&=\; \sum_{(x,y) \in B \times C} p(x,y) \;=\; \sum_{x \in B} \left(\sum_{y \in C} p(x) q(y) \right) \\ &=\; \sum_{x \in B} \left(p(x) \sum_{y \in C} q(y) \right) \;=\; \left(\sum_{x \in B} p(x) \right) \cdot \left(\sum_{y \in C} q(y) \right) \;=\; P(X \in B) \cdot P(Y \in C). \blacksquare \end{aligned}$$

4.3.5 Esempio. Torniamo alla moneta reale lanciata n volte dell'Esempio 2.2.6. L'abbiamo descritta sullo spazio campionario $\Omega^n = \{0,1\}^n = \{\omega = (\omega_1, \ldots, \omega_n) \colon \omega_i = 0, 1\}$ probabilizzato proprio come prodotto

$$p_{(\omega_1, \ldots, \omega_n)} \;=\; p^{\omega_1} q^{1-\omega_1} \cdot \ldots \cdot p^{\omega_n} q^{1-\omega_n} \;=\; p_1(\omega_1) \cdot \ldots \cdot p_n(\omega_n),$$

dove $p_i(\omega_i) = P(X_i = \omega_i)$ è proprio la densità della variabile aleatoria discreta X_i che da l'i–esima coordinata ω_i. Quindi, le $X_1, \ldots, X_n$ sono indipendenti.

4.3.6 Esempio. Consideriamo il lancio continuato della moneta descritto su $\Omega^\infty = \{0,1\}^\infty = \{\omega = (\omega_1, \omega_2, \ldots) \colon \omega_i = 0, 1\}$ e probabilizzato secondo il Teorema 2.3.3. Per $n = 1, 2, \ldots$ definiamo le variabili aleatorie

$$X_n(\omega) \;:=\; \text{“numero del tentativo nella successione } \omega \text{ con l'}n\text{–esimo successo”}.$$

Sappiamo che X_n è distribuita secondo la legge di Pascal $P(X_n = k) = Pas_{n,p}(k) = \binom{k-1}{n-1} p^n q^{k-n}$ per $k = n, n+1, \ldots$. Poniamo

$$Y_n \;=\; X_{n+1} - X_n,$$

ossia il numero di tentativi che dobbiamo aspettare dopo l'n–esimo successo finché non arrivi il prossimo.

Attenzione! Nonostante Y_n sia una funzione dei variabili discreti X_n e X_{n+1}, l'evento "$Y_n = k$" non è un evento di $\mathcal{A}(\Omega)$. (Infatti, $\mathcal{A}(\Omega)$ contiene tutte l domande binarie a cui è possibile rispondere dopo un numero finito di lanci. Però "$Y_n = k$" è l'unione (disgiunta) di tutti gli eventi "$X_n = \ell$"$\cap$"$X_{n+1} = \ell + k$" ($\ell \in \mathbb{N}$). Evidentemente, non è possibile verificare tutti questi eventi dopo un numero finito di lanci.) Come dicevamo, la Proposizione 4.2.7 vale solo sotto l'ipotesi che l'algebra booleana contenga tutti gli eventi "$(X_n, X_{n+1}) \in B$" ($B \subset \mathbb{R}^2$). Possiamo però definire per ogni $n \in \mathbb{N}$ l'algebra booleana $\mathcal{B}_n$ ottenuto dal vettore aleatorio discreto $(X_1, \ldots, X_n)$ aggiungendo, come nell'Osservazione 4.2.6, ad $\mathcal{A}(\Omega)$ gli eventi "$(X_1, \ldots, X_n) \in B_n$" ($B_n \subset \mathbb{R}^n$) e annotando con Q_n l'estensione di P su $\mathcal{B}_n$. Con un procedimento simile alla dimostrazione del Teorema 2.3.3, si arriva addirittura all'algebra booleana $\mathcal{B} := \bigcup_{n\in\mathbb{N}} \mathcal{B}_n$ con la probabilità Q. (Chi vuole, può ancora aggiungere tutti gli eventi elementari secondo il teorema nella Nota 2.3.4.) Questa algebra booleana contiene tutti gli eventi "$Y_n = k$" ($n, k \in \mathbb{N}$).

Intuitivamente, ci aspettiamo che in un certo momento il tempo d'attesa fino al prossimo successo non dipenda dal numero di successi già avvenuti. Quindi Y_n dovrebbe risultare indipendente da X_n. Inoltre, il tempo d'attesa per un solo successo ha la legge geometrica. Quindi lo dovrebbe valere anche per Y_n. Vogliamo verificare questi due sospetti.

Bisogna calcolare la densità congiunta di (X_n, Y_n) e controllare se fattorizzi. Abbiamo $X_n = k$ e $Y_n = \ell$, se e solo se $X_n = k$ e $X_{n+1} = k + \ell$. Allora

$$P(X_n = k, Y_n = \ell) \;=\; P(X_n = k, X_{n+1} = k + \ell) \;=\; \binom{k-1}{n-1} p^{n+1} q^{k+\ell-n-1}$$

perché dobbiamo distribuire $n + 1$ volte il numero "1" su $k + \ell$ posti (spiegando il fattore $p^{n+1}q^{k+\ell-n-1}$ che indica la probabilità di avere un certo risultato ($\omega_1, \ldots, \omega_{k+\ell}$ con $n + 1$ successi), in tal modo che l'ultimo e il k–esimo sono "1" mentre i posti strettamente compresi fra k e $k + \ell$ sono "0" (spiegando il fattore combinatorio $\binom{k-1}{n-1}$ per distribuire i $n - 1$ successi rimasti su $k - 1$ posti ancora disponibili). Questa probabilità fattorizza come

$$\binom{k-1}{n-1} p^{n+1} q^{k+\ell-n-1} \;=\; \left(\binom{k-1}{n-1} p^n q^{k-n}\right) \cdot (pq^{\ell-1}) \;=\; Pas_{n,p}(k) \cdot geom_p(\ell).$$

Troviamo confermate le nostre congetture.

4.3.7 Proposizione. *Siano $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ variabili aleatorie discrete indipendenti. Siano $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ e $\psi\colon \mathbb{R}^n \to \mathbb{R}$ funzioni. Allora anche le variabili aleatorie $Z_\varphi = \varphi \circ X$ e $Z_\psi = \psi \circ Y$ definite come*

$$\begin{aligned} Z_\varphi(\omega) &:= \varphi(X_1(\omega), \ldots, X_m(\omega)) \\ Z_\psi(\omega) &:= \psi(Y_1(\omega), \ldots, Y_n(\omega)) \end{aligned}$$

sono indipendenti.

Dimostrazione. Per calcolare $P(Z_\varphi = a, Z_\psi = b)$ usufruiremo di un'altra formula molto utile per il calcolo con funzioni indicatrici: Per ogni scelta di eventi $A, B \in \Omega$ vale

$$\chi_{A\cap B}(\omega) \;=\; \chi_A(\omega)\cdots\chi_B(\omega).$$

(**Esercizio:** Dimostrarlo!) Quindi,

$$\begin{aligned}
&P(Z_\varphi = a, Z_\psi = b)\\
&= \sum_{(x_1,\ldots,x_m,y_1,\ldots,y_n)\in\mathbb{R}^{m+n}} \chi_{Z_\varphi=a,Z_\psi=b}(x_1,\ldots,x_m,y_1,\ldots,y_n)p_1(x_1)\cdot\ldots\cdot p_m(x_m)\cdot q_1(y_1)\cdot\ldots\cdot q_n(y_n)\\
&= \sum_{x\in\mathbb{R}^m}\sum_{y\in\mathbb{R}^n} \chi_{Z_\varphi=a}(x_1,\ldots,x_m,y_1,\ldots,y_n)p_1(x_1)\cdot\ldots\cdot p_m(x_m)\\
&\qquad\qquad \cdot\chi_{Z_\psi=b}(x_1,\ldots,x_m,y_1,\ldots,y_n)q_1(y_1)\cdot\ldots\cdot q_n(y_n)\\
&= \left(\sum_{x\in\mathbb{R}^m}\chi_{\Phi=a}(x_1,\ldots,x_m)p_1(x_1)\cdot\ldots\cdot p_m(x_m)\right)\cdot\left(\sum_{y\in\mathbb{R}^n}\chi_{\Psi=b}(y_1,\ldots,y_n)q_1(y_1)\cdot\ldots\cdot q_n(y_n)\right),\\
&= P(Z_\varphi = a)\cdot P(Z_\psi = b),
\end{aligned}$$

dove negli ultimi due passi abbiamo preso in considerazione che $\chi_{Z_\varphi=a}$ su $\mathbb{R}^{m+n}$ non dipende dagli argomenti $(y_1,\ldots,y_n)\in\mathbb{R}^n$ e come funzione degli $(x_1,\ldots,x_m)\in\mathbb{R}^m$ coincide con la funzione indicatrice del medesimo evento "$Z_\varphi = a$" per la legge marginale P_X e, analogamente, per P_Y nel secondo fattore. ■

4.3.8 Esempio. Nel lancio della moneta $n+m$ volte le variabili $X_1,\ldots,X_{n+m}$ (con X_i indicando il risultato del i–esimo lancio) sono indipendenti. Allora con $\varphi(x_1,\ldots,x_m) := x_1+\ldots+x_m$ e $\psi(x_{m+1},\ldots,x_{m+n}) = x_{m+1}+\ldots+x_{m+n}$ segue che le variabili aleatorie $X_1+\ldots+X_m$ e $X_{m+1}+\ldots+X_{m+n}$ sono indipendenti.

4.3.9 Esercizio. Lanciamo in continuazione una moneta ed un dado. Qual è la probabilità che la moneta dia testa prima che il dado dia 6?

Suggerimenti: Pare naturale di modellare l'esperimento sullo spazio campionario

$$\Big(\{0,1\}\times\{1,\ldots,6\}\Big)^\infty,$$

cioè con l'idea in mente che lanciamo ogni volta la coppia "(moneta,dado)" e lo ripetiamo ad infinito. Però possiamo anche pensare alla descrizione su

$$\Omega \;=\; \{0,1\}^\infty\times\{1,\ldots,6\}^\infty,$$

dove due persone eseguono due esperimenti, l'una il lancio della moneta e l'altra il lancio del dado. Matematicamente, le due descrizioni sono equivalenti. Però la prima mette l'accento sull'indipendenza di un tentativo da ogni altro, mentre la seconda mette l'accento sull'indipendenza di tutte le domande che poniamo alla moneta da tutte le domande che poniamo al dado. Alla moneta possiamo porre la domanda quando arriva il primo successo e definire così una variabile aleatoria

$$X \;=\; \text{“momento della prima testa”}.$$

Similmente, al dado possiamo porre la domanda quando arriva il primo successo e definire così una variabile aleatoria

$$Y \;=\; \text{“momento del primo 6”}.$$

X e Y si riferiscono a fattori diversi di Ω e sono chiaramente indipendenti, tutte e due distribuite secondo una legge geometrica (di parametri diversi). L'evento che ci interessa è “la moneta da testa prima che il dado da 6”, ossia, “$X < Y$”. La probabilità è quindi la sommatoria su tutte le probabilità $p(k,\ell) := P(X = k, Y = \ell) = P(X = k) \cdot P(Y = \ell)$ dove $k < \ell$. Bisogna solo calcolarla.

4.4 Attese di variabili aleatorie discrete

Siano $x_1, \ldots, x_M \in \mathbb{R}$ dati numerici (per esempio, i risultati di M misurazioni di una variabile aleatoria). La statistica descrittiva definisce la ***media*** di M dati come

$$\bar{x} \;:=\; \frac{1}{M}\sum_{i=1}^{M} x_i \;=\; \frac{x_1 + \ldots + x_M}{M}.$$

Con il numero $\#\{k\colon x_k = x\}$ di apparizioni del valore $x \in \mathbb{R}$ nel campione $x_1, \ldots, x_M$ e la frequenza relativa $h_x^M = \frac{\#\{k\colon x_k = x\}}{M}$ di x, possiamo riscrivere la media come

$$\bar{x} \;=\; \frac{x_1 + \ldots + x_M}{M} \;=\; \sum_{x\in\mathbb{R}} \frac{x \cdot \#\{k\colon x_k = x\}}{M} \;=\; \sum_{x\in\mathbb{R}} x \cdot h_x^M.$$

Se ci ricordiamo della distribuzione empirica (Esempio 1.2.3), vediamo che $p(x) = h_x^M$ è la densità della variabile aleatoria discreta $X = \mathsf{id}_R$ (come nell'Esempio 4.1.7) su $\mathbb{R}$, e la media diventa $\sum_{x\in\mathbb{R}} x \cdot p(x)$. In questa forma, la media (di dati concreti!) ci fa la guida per la definizione dell'attesa (teorica!) di una variabile aleatoria.

4.4.1 Definizione. Sia $X\colon \Omega \to \mathbb{R}$ una variabile aleatoria discreta. Allora, l'***attesa*** di X è definita come

$$\mathbb{E}X \;:=\; \sum_{x\in\mathbb{R}} x \cdot p(x),$$

se la sommatoria converge.

N.B.: Se la sommatoria non converge, allora $\sum_{x\in\mathbb{R}} |x|\, p(x) = \infty$. (Si veda il Lemma 1.3.8.) Se l'attesa di X esiste, diciamo anche X ha ***l'attesa finita***.

Sia esistenza che valore di $\mathbb{E}X$ dipendono solo dalla legge, ossia, dalla densità $p(x)$ di X. Però secondo la definizione, finora bisognerebbe conoscere la densità di ogni variabile aleatoria di cui volessimo calcolare l'attesa. Per fortuna esiste un teorema che spesso ci facilita la vita.

4.4.2 Teorema. *Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio discreto con densità $p(x)$ e sia $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ una funzione. Allora la variabile aleatoria $Z = \varphi \circ X$ come nella Proposizione 4.2.7 ha l'attesa finita*

$$\mathbb{E}Z \;=\; \sum_{x\in\mathbb{R}^m} \varphi(x)\cdot p(x)$$

se e solo se la sommatoria converge.

Dimostrazione. Poiché $x \in B_z$, se e solo $\varphi(x) = z$, dalla Proposizione 4.2.7 segue che

$$\sum_{z\in\mathbb{R}} z p_Z(z) \;=\; \sum_{z\in\mathbb{R}} z \sum_{x\in B_z} p(x) \;=\; \sum_{z\in\mathbb{R}} \sum_{x\in B_z} z p(x) \;=\; \sum_{z\in\mathbb{R}} \sum_{x\in B_z} \varphi(x) p(x) \;=\; \sum_{x\in\mathbb{R}^m} \varphi(x) p(x),$$

dove nell'ultimo passo abbiamo applicato il Corollario 1.3.14 (che include anche l'affermazione sull'esitenza dell'attesa). ■

4.4.3 Esempio. Notiamo che per una variabile aleatoria discreta che con probabilità 1 assume solo valori positivi ($P(X \geq 0) = 1$) ha senso di parlare di $\mathbb{E}X = \sum_{x\geq 0} x \cdot p(x) = \infty$ se l'attesa non dovesse essere finita. In particolare, X ha l'attesa finita, se e solo se $\mathbb{E}\,|X| \neq \infty$.

4.4.4 Osservazione. Sia $X(\omega) = 1$ una variabile aleatoria costante. Allora X è discreta con $p(1) = 1$ e $p(x) = 0$ se $x \neq 1$ e con l'attesa finita

$$\mathbb{E}1 \;=\; \mathbb{E}\,|1| \;=\; \sum_{x\in\mathbb{R}} x \cdot p(x) \;=\; 1 \cdot 1 \;=\; 1.$$

4.4.5 Corollario. *Siano X, Y variabili aleatorie discrete con l'attesa finita e sia $c \in \mathbb{R}$. Allora esistono anche $\mathbb{E}(cX)$ e $\mathbb{E}(X + Y)$ e valgono*

$$\mathbb{E}(cX) \;=\; c\mathbb{E}X \qquad e \qquad \mathbb{E}(X + Y) \;=\; \mathbb{E}X + \mathbb{E}Y.$$

In particolare, $\mathbb{E}c = \mathbb{E}(c \cdot 1) = c \cdot \mathbb{E}1 = c$.

DIMOSTRAZIONE. 1.) Poniamo $\varphi(x) = cx$. Allora

$$\mathbb{E}(cX) = \sum_{x\in\mathbb{R}} \varphi(x)p(x) = \sum_{x\in\mathbb{R}} cxp(x) = \sum_{x\in\mathbb{R}} c\cdot|x|\,p(x) = c\sum_{x\in\mathbb{R}} xp(x) = c\mathbb{E}X.$$

2.) Sia $\varphi(x,y) = x+y$. Allora

$$\mathbb{E}(X+Y) = \sum_{(x,y)\in\mathbb{R}^2} \varphi(x,y)p(x,y) = \sum_{(x,y)\in\mathbb{R}^2} (x+y)p(x,y)$$

$$\sum_{(x,y)\in\mathbb{R}^2} xp(x,y) + \sum_{(x,y)\in\mathbb{R}^2} yp(x,y) = \sum_{x\in\mathbb{R}} xp(x) + \sum_{y\in\mathbb{R}} yq(y) = \mathbb{E}X + \mathbb{E}Y,$$

dove nell'ultimo passo abbiamo applicato l'Esempio 4.2.4 e di nuovo il Corollario 1.3.14. ■

Se definiamo $\varphi_X(x,y) = x$, allora possiamo X considerare come variabile aleatoria individuale oppure come funzione $\varphi_X \circ (X,Y)$ della coppia (X,Y). Notiamo quanto è stato importante il passo dove abbiamo scambiato il calcolo di $\mathbb{E}X$ con X come funzione della coppia (X,Y) (cioè, con la densità congiunta) con il calcolo con X come variabile aleatoria individuale (cioè, con la densità marginale di X).

4.4.6 Corollario. *Siano X, Y variabili aleatorie discrete con l'attesa finita. Se X e Y sono* ***indipendenti****, allora esiste anche $\mathbb{E}(X\cdot Y)$ e vale*

$$\mathbb{E}(X\cdot Y) = \mathbb{E}X\cdot\mathbb{E}Y.$$

DIMOSTRAZIONE. Sia $\varphi(x,y) = xy$. Allora

$$\mathbb{E}(X\cdot Y) = \sum_{(x,y)\in\mathbb{R}^2} \varphi(x,y)p(x,y) = \sum_{(x,y)\in\mathbb{R}^2} xyp(x)q(y)$$

$$= \left(\sum_{x\in\mathbb{R}} x\cdot p(x,y)\right)\cdot\left(\sum_{y\in\mathbb{R}} y\cdot q(y)\right) = \mathbb{E}X\cdot\mathbb{E}Y$$

(di nuovo applicando l'Esempio 4.2.4 ed il Corollario 1.3.14). ■

La seguente proposizione è importante quanto appare innocua: È importantissima!

4.4.7 Proposizione. *Siano X, Y variabili aleatorie discrete con l'attesa finita. Allora:*

1. *$X \geq Y$ (cioè $X(\omega) \geq Y(\omega) \forall \omega \in \Omega$) $\Longrightarrow \mathbb{E}X \geq \mathbb{E}Y$.*
2. *$|\mathbb{E}X| \leq \mathbb{E}|X|$.*

Dimostrazione. 1.) Poniamo $Z := X - Y$. Allora $P(Z \geq 0) = 1$. Quindi,

$$\mathbb{E}X - \mathbb{E}Y \;=\; \mathbb{E}(X - Y) \;=\; \mathbb{E}Z \;=\; \sum_{z \geq 0} zP(Z = z) \;\geq\; 0,$$

cosicché $\mathbb{E}X \geq \mathbb{E}Y$.

2.) $\pm X \leq |X|$, quindi $\pm\mathbb{E}X \leq \mathbb{E}\,|X|$. ■

La dimostrazione della prima parte funziona, senza cambiare una parola, anche se solo $P(X \geq Y) = 1$. (Che cosa fanno le funzioni X e Y su sottoinsiemi di Ω con probabilità 0 non centra con il calcolo dell'attesa.) Il seguente corollario dimostra che una variabile aleatoria non può fare a meno che avere un'atesa compresa fra il valore minimo e il valore massimo che assume su un evento sicuro di Ω.

4.4.8 Corollario. *Siano $a \leq b$ tali che $P(a \leq X \leq b) = 1$. Allora anche $a \leq \mathbb{E}X \leq b$.*

4.4.9 Esempio. 1. $X = \chi_A$. Allora

$$\mathbb{E}X \;=\; \sum_{x=0,1} xP(\chi_A = x) \;=\; 0 \cdot P(\chi_A = 0) + 1 \cdot P(\chi_A = 1)$$

$$= \; 0 \cdot P(A^{\complement}) + 1 \cdot P(A) \;=\; P(A).$$

2. La legge di Bernoulli $X \sim B_{1,p}$, cioè $P(X = 1) = p$ e $P(X = 0) = q$. Allora

$$\mathbb{E}X \;=\; 0 \cdot q + 1 \cdot p \;=\; p.$$

Se $A =$"$X = 1$", ritroviamo la parte precedente con $p = P(A)$.

3. La legge binomiale $X \sim B_{n,p}$. Osserviamo che $X = X_1 + \ldots + X_n$ è la somma di n variabili bernoulliane X_i. Quindi $\mathbb{E}X = \mathbb{E}X_1 + \ldots + \mathbb{E}X_n = np$.

Possiamo anche considerare $Y = X_1 \cdot \ldots \cdot X_n$. Visto che gli $X_1, \ldots, X_n$ sono addirittura indipendenti, concludiamo $\mathbb{E}Y = p^n$. Questo segue anche dal fatto che Y è la funzione indicatrice dell'evento elementare "$X_1 = \ldots = X_n = 1$" che ha la probabilità p^n.

4.5 Varianza, covarianza e la legge dei grandi numeri

Per il resto del capitolo $X, Y, \ldots$ sono variabili aleatorie discrete.

4.5.1 Definizione. Sia X una variabile aleatoria discreta e sia $k \in \mathbb{N}_0$. Diciamo X ha il ***k–esimo momento*** $\mathbb{E}X^k$ finito se $\mathbb{E}\,|X|^k < \infty$.

Chiaramente $\mathbb{E}X^0 = \mathbb{E}1 = 1$ esiste sempre e $\mathbb{E}X^1 = \mathbb{E}X$ è l'attesa se esiste.

4.5.2 Proposizione. *Se $\mathbb{E}\,|X|^n < \infty$ per un $n \in \mathbb{N}_0$, allora $\mathbb{E}\,|X|^k < \infty$ per ogni $k \le n$.*

Dimostrazione. Osserviamo che

$$|x| \le 1 \Longrightarrow |x|^k \le 1,$$
$$|x| \ge 1 \Longrightarrow |x|^k \le |x|^n .$$

Allora vale

$$|x|^k \ \le \ 1 + |x|^n$$

su tutta la retta. Quindi $\mathbb{E}\,|X|^k \le \mathbb{E}(1 + |X|^n) = \mathbb{E}1 + \mathbb{E}\,|X|^n < \infty$. ■

4.5.3 Proposizione. *Se $\mathbb{E}\,|X|^n < \infty$ e $\mathbb{E}\,|Y|^n < \infty$, allora anche $\mathbb{E}\,|X + Y|^n < \infty$.*

Dimostrazione. Abbiamo

$$\begin{aligned} |x + y|^n \ \le \ (|x| + |y|)^n \ \le \ (2\max(|x|\,,|y|))^n \ &= \ 2^n(\max(|x|\,,|y|))^n \\ &= \ 2^n \max(|x|^n\,,|y|^n) \ \le \ 2^n(|x|^n + |y|^n), \end{aligned}$$

dove nel passo dalla prima alla seconda riga abbiamo usato $|x| \le |y| \Longleftrightarrow |x|^n \le |y|^n$. Quindi, $\mathbb{E}\,|X + Y|^n \le 2^n(\mathbb{E}\,|X|^n + \mathbb{E}\,|Y|^n) < \infty$. ■

4.5.4 Definizione. La ***varianza*** di una variabile aleatoria X la definiamo

$$\mathbb{V}X \ := \ \mathbb{E}(X - \mathbb{E}X)^2,$$

se il secondo momento $\mathbb{E}X^2$ (e, quindi, anche il primo $\mathbb{E}X$) esiste, e $\mathbb{V}X = \infty$ altrimenti.

4.5.5 Osservazione. $(X - \mathbb{E}X)^2$ è una funzione positiva. Qindi, $\boxed{\mathbb{V}X \ge 0}$.

La varianza è invariata sotto traslazioni, cioè $\boxed{\text{se } Y = X + a, \text{ allora } \mathbb{V}Y = \mathbb{V}X}$. Infatti,

$$\mathbb{V}Y \ = \ \mathbb{E}((X - a) - (\mathbb{E}X - a))^2 \ = \ \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{V}X.$$

N.B.: È spesso opportuno calcolare la varianza secondo la formula

$$\begin{aligned} \mathbb{V}X \ = \ \mathbb{E}(X - \mathbb{E}X)^2 \ &= \ \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) \\ &= \ \mathbb{E}X^2 - 2\mathbb{E}(X\mathbb{E}X) + (\mathbb{E}X)^2 \ = \ \mathbb{E}X^2 - (\mathbb{E}X)^2. \end{aligned}$$

4.5.6 Corollario. $\boxed{\mathbb{V}(cX) = \mathbb{E}(cX)^2 - (\mathbb{E}(cX))^2 = c^2\mathbb{E}X^2 - (c\mathbb{E}X)^2 = c^2\mathbb{V}X.}$

4.5.7 Osservazione. La varianza di X è la media della *deviazione quadratica* $(X - \mathbb{E}X)^2$ di X dalla sua attesa e, quindi, una misura quanto i valori di X siano sparpagliati in giro ad $\mathbb{E}X$. Ci sarebbero altre possibilità come la media di $|X - \mathbb{E}X|$ oppure di $|X - \mathbb{E}X|^\alpha$ ($\alpha > 0$). Ma per prima cosa, il modulo con la sua definizione a pezzi è scomodo. È poi la deviazione quadratica si gode della seguente proprietà particolare: Osserviamo che $\mathbb{V}X < \infty$, se e solo se $\mathbb{E}(X - a)^2 = \mathbb{E}X^2 - 2a\mathbb{E}X + a^2 < \infty$. Se calcoliamo la derivata per a di $\mathbb{E}(X - a)^2$, troviamo

$$\frac{d}{da}\mathbb{E}(X - a)^2 \;=\; -2\mathbb{E}X + 2a.$$

Quindi la parabola $\mathbb{E}(X - a)^2$ assume un valore minimo per $a = \mathbb{E}X$.

Osserviamo, inoltre, che $0 = \mathbb{V}X = \sum_{x\in\mathbb{R}}(x - \mathbb{E}X)^2 p(x)$ implica che per ogni $x \in \mathbb{R}$ vale $x - \mathbb{E}X = 0$ o $p(x) = 0$. Quindi, l'unico valore per x, dove $p(x)$ possa essere diverso da 0, è $x = \mathbb{E}X$. Da $1 = \sum_{x\in\mathbb{R}} p(x) = p(\mathbb{E}X)$ concludiamo che $P(X = \mathbb{E}X) = 1$, ossia, X (con probabilità 1) è la funzione costante $\mathbb{E}X$.

Una delle disuguaglianze più fondamentali, che ci aiuta di dimostrare la legge dei grandi numeri (come promesso nell'introduzione), è la seguente:

4.5.8 Disuguaglianza di Cebicev. *Per ogni variabile aleatoria X e per ogni $\varepsilon > 0$ vale*

$$P(|X - \mathbb{E}X| \geq \varepsilon) \quad\leq\quad \frac{\mathbb{V}X}{\varepsilon^2}.$$

Dimostrazione. Ricordiamoci che la probabilità di un evento è l'attesa della sua funzione indicatrice. Quindi $P(|X - \mathbb{E}X| \geq \varepsilon) = \mathbb{E}\chi_{\text{“}|X - \mathbb{E}X| \geq \varepsilon\text{”}}$. La dimostrazione consiste in nient'altro che trovare la parabola $c(X - a)^2$ più bassa che risulta sempre più grande di $\chi_{\text{“}|X - \mathbb{E}X| \geq \varepsilon\text{”}}$. Il risultato (unico) di questo tentativo è $\frac{(X-\mathbb{E}X)^2}{\varepsilon^2}$.

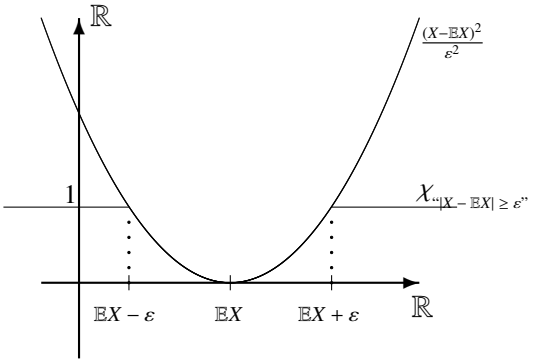


Figura 4.1: $\frac{(X-\mathbb{E}X)^2}{\varepsilon^2} \geq \chi_{\text{“}|X - \mathbb{E}X| \geq \varepsilon\text{”}}$.

Quindi,

$$P(|X - \mathbb{E}X| \geq \varepsilon) \;=\; \mathbb{E}\chi_{\text{“}|X - \mathbb{E}X| \geq \varepsilon\text{”}} \;\leq\; \mathbb{E}\Big(\frac{(X - \mathbb{E}X)^2}{\varepsilon^2}\Big) \;=\; \frac{\mathbb{V}X}{\varepsilon^2}. \blacksquare$$

La stima in questa dimostrazione, e conseguentemente anche la disuguaglianza di Cebicev è piuttosto brutta. Soprattutto, se deviazioni di X dalla media $\mathbb{E}X$ appaiono con probabilità non piccole, la parabola è decisamente troppo alta per dare una limitazione soddisfacente. Inoltre, se vogliamo fare ε sempre più piccolo senza variare anche X, il quoziente $\frac{\mathbb{V}X}{\varepsilon^2}$ supera prima o poi il valore 1 e la disuguaglianza non ci da più nessun'informazione. Nelle applicazioni, se vogliamo fare ε sempre più piccolo, dobbiamo anche variare X in dipendenza da ε. Un modo opportuno è la dimostrazione della legge dei grandi numeri che vediamo fra poco.

Prima ci serve un altro calcolo di varianze. Vogliamo calcolare $\mathbb{V}(X+Y)$. Rimordiamoci che se le varianze di X e di Y sono finite, allora secondo la Proposizione 4.5.3 lo è anche la varianza della somma. Calcoliamo

$$\begin{aligned}\mathbb{V}(X+Y) \;&=\; \mathbb{E}(X+Y)^2 - (\mathbb{E}(X+Y))^2 \;=\; \mathbb{E}X^2 + 2\mathbb{E}(XY) + \mathbb{E}Y^2 - (\mathbb{E}X)^2 - 2\mathbb{E}X\mathbb{E}Y - (\mathbb{E}Y)^2 \\ &=\; \mathbb{V}X + \mathbb{V}Y + 2(\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y) \;=\; \mathbb{V}X + \mathbb{V}Y + 2\,\mathsf{cov}(X,Y),\end{aligned}$$

dove abbiamo definito la ***covarianza*** $\mathsf{cov}(X,Y) := \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$.

4.5.9 Nota. Normalmente si definisce $\mathsf{cov}(X,Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$ in analogia con la varianza. In questa forma ci si vede subito che la covarianza è una funzione bilineare di suoi argomenti X e Y. È facile verificare come per la varianza, che le due definizioni coincidono.

4.5.10 Corollario. *Se X e Y sono* ***non correlate*** *(cioè, se $\mathsf{cov}(X,Y) = 0$), allora vale*

$$\mathbb{V}(X+Y) \;=\; \mathbb{V}X + \mathbb{V}Y.$$

Attenzione! $\mathbb{V}(X+X) = \mathbb{V}(2X) = 4\mathbb{V}X \neq 2\mathbb{V}X$ se $\mathbb{V}X \neq 0$. Infatti, $\mathsf{cov}(X,X) = \mathbb{V}X$, cioè X e X sono correlatissime.

Se X e Y sono indipendenti, allora $\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y = 0$. Quindi variabili aleatorie indipendenti sono sempre non correlate. L'affermazione opposta è sbagliata.

4.5.11 Esercizio. Sia $\Omega = \{-1, 0, 1\}$ uno spazio di probabilità elementare, sia $X(\omega) = \omega$ e $Y = |X|$. Verificare che X e Y sono non correlate ma non sono indipendenti.

4.5.12 Osservazione. Siano $X_1, \ldots, X_m$ ***due a due non correlate*** (cioè, $\mathsf{cov}(X_i, X_j) = 0$ se $i \neq j$). Allora abbiamo

$$\mathsf{cov}(X_1 + \ldots + X_{k-1}, X_k) \;=\; \mathsf{cov}(X_1, X_k) + \ldots + \mathsf{cov}(X_{k-1}, X_k) \;=\; 0$$

per ogni $k > 1$. Quindi

$$\mathbb{V}(X_1 + \ldots + X_{m-1} + X_m) \;=\; \mathbb{V}(X_1 + \ldots + X_{m-1}) + \mathbb{V}(X_m) \;=\; \ldots \;=\; \mathbb{V}(X_1) + \ldots + \mathbb{V}(X_m).$$

Con questo, finalmente, siamo pronti:

4.5.13 La legge dei grandi numeri. (Versione debole.) *Sia X una variabile aleatoria con varianza (e quindi con attesa) finita. Siano $X_1, X_2, \ldots$ variabili aleatorie due a due non correlate con la medesima legge $X \sim X_1 \sim X_2 \sim \ldots$ di X. Allora*

$$\lim_{n\to\infty} P\left(\left| \frac{X_1 + \ldots + X_n}{n} - \mathbb{E}X \right| \geq \varepsilon \right) \;=\; 0$$

per ogni $\varepsilon > 0$.

Dimostrazione. Notiamo che $\mathbb{E}\left(\frac{X_1+\ldots+X_n}{n}\right) = \frac{\mathbb{E}X_1+\ldots+\mathbb{E}X_n}{n} = \frac{n\mathbb{E}X}{n} = \mathbb{E}X$. Quindi, secondo la disuguaglianza di Cebicev abbiamo

$$P\left(\left| \frac{X_1 + \ldots + X_n}{n} - \mathbb{E}X \right| \geq \varepsilon \right) \;\leq\; \frac{\mathbb{V}\left(\frac{X_1+\ldots+X_n}{n}\right)}{\varepsilon^2} \;=\; \frac{\mathbb{V}X_1 + \ldots + \mathbb{V}X_n}{n^2\varepsilon^2} \;=\; \frac{n\mathbb{V}X}{n^2\varepsilon^2} \;\longrightarrow\; 0$$

per $n \to \infty$. ■

4.5.14 Nota. La legge dei grandi numeri debole ci dice che in una successione $x_1, x_2, \ldots$ di verificazioni indipendenti della stessa quantità numerica aleatoria, per ogni $\varepsilon > 0$ la probabilità di avere una deviazione della media empirica dall'attesa teorica più grande di ε, diminuisce sempre di più se aumentiamo il numero n di valori di cui calcoliamo la media. ***Non*** ci dice che la media stessa convergerebbe verso l'attesa! Questa è l'affermazione della legge dei grandi numeri ***forte***. Più precisamente, la versione forte afferma che l'insieme A di tutte le successioni $x_1, x_2, \ldots$ per i quali le medie non convergono verso l'attesa, ha la probabilità $P(A) = 0$. (La dimostrazione richiede una preparazione matematica più elevata. Per iniziare, se prendiamo nota del fatto, che per controllare l'affermazione "le medie convergono verso l'attesa" è necessario conoscere tutta la successione. Non si tratta, quindi, di un evento contenuto nella nostra algebra booleana per i modelli con ripetizioni ad infinito.) Tutte e due le leggi, sia la versione debole che la versione forte, non dicono assolutamente nulla sulla velocità della convergenza. Questa dipende dal modello concreto e fa parte dell'inferenza statistica.

4.6 La funzione generatrice (di Laplace)

In questa sezione impariamo un metodo efficace di calcolare attese e varianze con l'aiuto della *funzione generatrice*. Questo metodo è, però, ristretto alle sole variabili aleatorie a valori in $\mathbb{N}_0$. Un attrezzo similmente potente, che funziona per tutte le probabilità su $\mathbb{R}$, sarebbe la sua *funzione caratteristica*. Non la discuteremo perché richiederebbe l'uso dei numeri complessi. Comunque, quasi tutti gli esempi concreti di leggi che abbiamo discusso finora sono leggi di variabili aleatorie a valori in $\mathbb{N}_0$.

4.6.1 Definizione. Sia X una variabile aleatoria a valori in $\mathbb{N}_0$. Con ciò intendiamo che X è discreta con la probabilità $P(X \in \mathbb{N}_0) = 1$. Tale probabilità è determinata dalla sua densità discreta $p(n) = P(X = n)$. La ***funzione generatrice*** ψ (oppure ψ_X, se vogliamo indicare la variabile) di X è definita come

$$\psi(z) \;:=\; \sum_{n=0}^{\infty} z^n p(n) \;=\; \mathbb{E}(z^X)$$

dove esiste.

La serie converge (assolutamente) almeno per ogni $z \in [-1, 1]$, perché in questo caso

$$\sum_{n=0}^{\infty} |z^n p(n)| \;\le\; \sum_{n=0}^{\infty} p(n) \;=\; 1.$$

Attenzione: Non tutte le serie di potenza $\sum_{n=0}^{\infty} z^n a_n$ sono funzioni generatrici di variabili aleatorie. Infatti, occorre e basta che $a_n \ge 0$ e $\sum_{n=0}^{\infty} a_n = 1$.

Secondo i ben noti teoremi dell'analisi sulle serie di potenza, le derivate $\frac{d^n\psi}{dz^n}$ esistono per ogni n almeno nell'interno $|z| < 1$ dell'intervallo $[0, 1]$. In particolare, troviamo $\frac{d^n\psi}{dz^n}(0) = n!p(n)$, cioè, ψ determina univocamente la legge di X.

Al limite $z \uparrow 1$ dell'intervallo vale $\frac{d^n\psi}{dz^n}(1) = \lim_{z\uparrow 1} \frac{d^n\psi}{dz^n}(z)$, se tale limite esiste. Se non esiste il limite, allora neanche la derivata esiste. (Per la dimostrazione è cruciale che i $p(n)$ sono tutti positivi. Non vale per $z \downarrow -1$.)

Come detto, la funzione generatrice serve per calcolare l'attesa $\mathbb{E}X = \sum_{n=0}^{\infty} np(n)$. L'idea è che $\frac{d(z^n)}{dz} = nz^{n-1}$ è, quindi, $\frac{d(z^n)}{dz}\big|_{z=1} = n \cdot 1^{n-1} = n$ se inseriamo $z = 1$. Allora se esiste $\psi'(1)$ troviamo

$$\psi'(1) \;=\; \frac{d}{dz}\Big(\sum_{n=0}^{\infty} z^n p(n)\Big)\Big|_{z=1} \;=\; \sum_{n=0}^{\infty} nz^{n-1} p(n)\Big|_{z=1} \;=\; \sum_{n=0}^{\infty} np(n). \;=\; \mathbb{E}X.$$

4.6.2 Proposizione. *L'attesa di X esiste se e solo $\psi'(1)$ esiste, ed in tal caso vale*

$$\mathbb{E}X \;=\; \psi'(1).$$

Un modo più formale di scriverlo sarebbe $\frac{d\mathbb{E}(z^X)}{dz}\Big|_{z=1} = \mathbb{E}(Xz^{X-1})\Big|_{z=1} = \mathbb{E}X$. Tentiamo lo stesso per la varianza di X. Calcoliamo

$$\psi''(1) \;=\; \frac{d^2\mathbb{E}(z^X)}{dz^2}\Big|_{z=1} \;=\; \mathbb{E}(X(X-1)z^{X-2})\Big|_{z=1} \;=\; \mathbb{E}X(X-1) \;=\; \mathbb{E}X^2 - \mathbb{E}X.$$

Questo non è ancora il secondo momento. Ma conosciamo già il primo momento $\mathbb{E}X = \psi'(1)$. Allora:

4.6.3 Proposizione. *La varianza di X esiste se e solo $\psi''(1)$ esiste, ed in tal caso vale*

$$\mathbb{V}X \;=\; \psi''(1) + \psi'(1) - \psi'(1)^2.$$

Aggiungiamo ancora un risultato che ci permette di calcolare la funzione generatrice della somma $X + Y$ di variabili aleatorie indipendenti se conosciamo le loro funzioni generatrici.

4.6.4 Proposizione. *Siano X e Y variabili aleatorie* ***indipendenti*** *a valori in $\mathbb{N}_0$. Allora anche $X + Y$ è a valori in $\mathbb{N}_0$ e vale*

$$\psi_{X+Y} \;=\; \psi_X \cdot \psi_Y.$$

Dimostrazione. Infatti, se X e Y sono indipendenti allora lo sono anche le funzioni z^X e z^Y. Quindi,

$$\psi_{X+Y}(z) \;=\; \mathbb{E}(z^{X+Y}) \;=\; \mathbb{E}(z^X \cdot z^Y) \;=\; \mathbb{E}(z^X) \cdot \mathbb{E}(z^Y) \;=\; \psi_X(z) \cdot \psi_Y(z). \blacksquare$$

4.6.5 Esempio. 1. Legge di Bernoulli $X \sim B_{1,p}$. Allora $\psi(z) = \mathbb{E}(z^X) = z^0 \cdot q + z^1 \cdot p = zp + q$. Quindi,

$$\begin{aligned} \psi'(z) \;&=\; p \quad &\Longrightarrow \quad \psi'(1) \;&=\; p \quad &\Longrightarrow \quad \mathbb{E}X \;&=\; p, \\ \psi''(z) \;&=\; 0 \quad &\Longrightarrow \quad \psi''(1) \;&=\; 0 \quad &\Longrightarrow \quad \mathbb{V}X \;&=\; 0 + p - p^2 \;=\; pq. \end{aligned}$$

2. Legge binomiale $X \sim B_{n,p}$. Allora

$$\psi(z) \;=\; \sum_{k=0}^{n} z^k \binom{n}{k} p^k q^{n-k} \;=\; \sum_{k=0}^{n} \binom{n}{k} (zp)^k q^{n-k} \;=\; (zp + p)^n.$$

Notiamo che, come si deve, la funzione generatrice di X è l'n–esima potenza della funzione generatrice della variabile bernoulliana, perché e la somma di n tali variabili indipendenti. Da quest'osservazione segue anche che se $Y \sim B_{m,p}$ indipendente da X, allora $X + Y \sim B_{n+m,p}$.

Continuiamo con X:

$$\begin{aligned} \psi'(z) &= np(pz+q)^{n-1} \qquad \Longrightarrow \qquad \psi'(1) = np(p+q)^{n-1} = np, \\ \psi''(z) &= n(n-1)p^2(p+q)^{n-2} \qquad \Longrightarrow \qquad \psi''(1) = n(n-1)p^2. \end{aligned}$$

Quindi,

$$\mathbb{E}X = np, \qquad \mathbb{V}X = n(n-1)p^2 + np - n^2p^2 = n(p-p^2) = npq.$$

Confermiamo non solo che l'attesa è n volte l'attesa della variabile bernoulliana, ma anche le varianze delle n variabili bernoulliane indipendenti si sommano.

3. Legge geometrica $X \sim geom_p$. Allora $\psi(z) = \sum_{k=1}^{\infty} z^k pq^{k-1} = zp\sum_{k=0}^{\infty}(zq)^k = \frac{zp}{1-zq}$. Troviamo

$$\begin{aligned} \psi'(z) &= \frac{p(1-zq) - zp(-q)}{(1-zq)^2} = \frac{p}{(1-zq)^2} \qquad \Longrightarrow \qquad \psi'(1) = \frac{p}{(1-q)^2} = \frac{1}{p}, \\ \psi''(z) &= \frac{(-2)p(-q)}{(1-zq)^3} = \frac{2pq}{(1-zq)^3}. \qquad \Longrightarrow \qquad \psi''(1) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}, \end{aligned}$$

Quindi

$$\mathbb{E}X = \frac{1}{p}, \qquad \mathbb{V}X = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}.$$

4. Legge di Pascal $X \sim Pas_{n,p}$. Quello che abbiamo detto per la transizione dalla legge di Bernoulli alla legge binomiale (somma di n copie indipendenti) vale anche per la transizione dalla legge geometrica alla legge di Pascal, se prendiamo in considerazione l'Esempio 4.3.6, che ci permette dimostrazioni con l'induzione per i seguenti fatti:

$$\psi(z) = \Big(\frac{zp}{1-zq}\Big)^n, \qquad \mathbb{E}X = \frac{n}{p}, \qquad \mathbb{V}X = \frac{nq}{p^2}.$$

Ci impegniamo solo di calcolare la funzione generatrice direttamente dalla definizione.

$$\begin{aligned} \psi(z) &= \sum_{k=n}^{\infty} z^k \binom{k-1}{n-1} p^n q^{k-n} = \sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (zq)^{k-n} z^n \\ &= \sum_{k=n}^{\infty} \binom{k-1}{n-1} (1-zq)^n (zq)^{k-n} z^n \frac{p^n}{(1-zq)^n} = \Big(\frac{zp}{1-zq}\Big)^n \sum_{k=n}^{\infty} \binom{k-1}{n-1} (1-zq)^n (zq)^{k-n}. \end{aligned}$$

Poiché la legge di Pascal $P_{n,1-zq}$ è normalizzata, sappiamo che

$$\sum_{k=n}^{\infty} \binom{k-1}{n-1} (1-zq)^n (zq)^{k-n} \;=\; 1.$$

Quindi, $\psi(z) = (\frac{zp}{1-zq})^n$.

5. Legge di Poisson $X \sim Poi_\lambda$. Troviamo $\psi(z) = \sum_{k=0}^{\infty} z^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(z\lambda)^k}{k!} = e^{-\lambda} e^{z\lambda} = e^{(z-1)\lambda}$. Quindi,

$$\begin{aligned} \psi'(z) \;&=\; \lambda\psi(z) \qquad \Longrightarrow \qquad \psi'(1) \;=\; \lambda \qquad \Longrightarrow \qquad \mathbb{E}X \;=\; \lambda, \\ \psi''(z) \;&=\; \lambda^2\psi(z) \qquad \Longrightarrow \qquad \psi''(1) \;=\; \lambda^2 \qquad \Longrightarrow \qquad \mathbb{V}X \;=\; \lambda^2 + \lambda - \lambda^2 \;=\; \lambda. \end{aligned}$$

4.6.6 Esercizio. Calcolare l'attesa e la varianza per le cinque leggi su $\mathbb{N}_0$ dell'esempio precedente direttamente dalla definizione

$$\mathbb{E}X \;=\; \sum_{k=0}^{\infty} kp(k) \qquad \mathbb{E}X^2 \;=\; \sum_{k=0}^{\infty} k^2 p(k) \qquad \mathbb{V}X \;=\; \mathbb{E}X^2 - (\mathbb{E}X)^2,$$

cioè senza far riferimento alla funzione generatrice, e confermare così i risultati dell'esempio ottenuti con l'uso della funzione generatrice.

4.6.7 Esercizio. Finora abbiamo accettato senza controllare che la legge di Pascal è normalizzata come si deve. (Se non abbiamo sbagliato nel probabilizzare lo spazio dello schema di Bernoulli, non dobbiamo dubitare. Ma chi lo sa?) Dare una dimostrazione diretta (qualsiasi) che

$$\sum_{k=n}^{\infty} \binom{k-1}{n-1} p^n (1-p)^{k-n} \;=\; 1.$$

4.6.8 Esercizio. Dato due variabili aleatorie indipendenti $X \sim Poi_\lambda$ e $Y \sim Poi_\mu$, qual è la legge di $X + Y$?

4.7 Un esempio esteso

Torniamo al problema della somma di due dadi (l'esercizio dell'introduzione con i suoi successori 1.1.1, 2.1.1 e l'Esempio 4.1.6), modellato sullo spazio campionario $\Omega_c = \{2, \ldots, 12\}$ che contiene i possibili valori della somma. Supponiamo ci siamo resi conto che le probabilità $p_k = P(\text{“somma= } k\text{”})$ devono essere *proporzionali* al numero $f(k) = 6 - |7 - k|$ delle coppie ordinate che realizzino il valore k della somma. (**Esercizio:** Verificare che questa formula dia il numero giusto.) Ma non ci è venuto in mente che, considerando lo spazio Ω_a delle coppie, potremmo ridurre il problema ad uno spazio di probabilità elementare.

A questo punto, ci tocca a risolvere il seguente problema (esame!):

Data una funzione $f\colon \mathbb{R} \to \mathbb{R}_+$ diversa da zero su un sottoinsieme numerabile Ω di $\mathbb{R}$ in $\mathbb{R}_+$, trovare una costante $c > 0$ tale che la funzione

$$p(x) := cf(x) = \begin{cases} cf(x) & x \in \Omega, \\ 0 & \text{altrimenti.} \end{cases}$$

diventi la densità discreta di una variabile aleatoria X.

Per la soluzione bisogna trovare c tale che $\sum_{x\in\mathbb{R}} p(x) = \sum_{x\in\Omega} p(x) = 1$. Il problema ha la soluzione $c = \frac{1}{\sum_{x\in\Omega} f(x)}$ se $\sum_{x\in\Omega} f(x) < \infty$. Altrimenti non ha soluzione. Soprattutto quando Ω è un insieme finito, il problema ha sempre una soluzione.

Nel caso nostro troviamo $f(2) + \ldots + f(12) = 1 + 2 + 3 + 4 + 5 + 6 + 5 + 4 + 3 + 2 + 1 = 36$ (confermando che il numero totale delle coppie ordinate è 36), e quindi $p(x) = \frac{6-|7-x|}{36}$ ($x = 2, \ldots, 12$) e 0 altrimenti.

Una volta individuata la legge della variabile aleatoria X ="somma dei due dadi", ci possiamo chiedere (esame!) dell'attesa di X e della varianza di X. L'attesa era definita $\mathbb{E}X = \sum_{x\in\Omega} x \cdot p(x) = \sum_{x\in\Omega} x \cdot cf(x)$.

Nel nostro caso troviamo

$$\mathbb{E}X = \frac{2\cdot1+3\cdot2+4\cdot3+5\cdot4+6\cdot5+7\cdot6+8\cdot5+9\cdot4+10\cdot3+11\cdot2+12\cdot1}{36} = \frac{(2+12)\cdot1+(3+11)\cdot2+(4+10)\cdot3+(5+9)\cdot4+(6+8)\cdot5+7\cdot6}{36}$$
$$= \frac{14\cdot(1+2+3+4+5)+7\cdot6}{36} = \frac{210+42}{36} = \frac{252}{36} = \frac{42}{6} = 7.$$

È un risultato facilmente intuibile. Infatti, il risultato più probabile è 7 e le deviazioni da questo valore centrale appaiono in tutte e due le direzioni con probabilità uguali. Un simile calcolo per il secondo momento $\mathbb{E}X^2 = \sum_{x\in\Omega} x^2 \cdot p(x) = \sum_{x\in\Omega} x^2 \cdot cf(x)$ da $\mathbb{E}X^2 = \frac{329}{6}$ di cui risulta

$$\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{329}{6} - 7^2 = \frac{329 - 49\cdot 6}{6} = \frac{35}{6}.$$

Avremmo potuto facilitarci la vita enormemente, prendendo un po' in considerazione più della struttura concreta del problema. Infatti, sullo spazio di probabilità elementare di tutte le coppie $\Omega_a = \{1, \ldots, 6\} \times \{1, \ldots, 6\}$, la somma X è proprio la somma $X = X_1 + X_2$ dove $X_i(\omega_1, \omega_2) = \omega_i$ sono i risultati dei dadi individuali. Quindi, non solo $\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2$, dove l'attesa del lancio di una dado è $\mathbb{E}X_i = \frac{1+\ldots+6}{6} = \frac{21}{6} = \frac{7}{2}$. Ma anche, per l'indipendenza delle variabili X_1 e X_2, abbiamo $\mathbb{V}X = \mathbb{V}X_1 + \mathbb{V}X_2$. E $\mathbb{V}X$ è molto più facile da calcolare. Troviamo $\mathbb{E}X_i^2 = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$ di cui risulta $\mathbb{V}X_i = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$, confermando così $\mathbb{V}X = 2\mathbb{V}X_i = \frac{35}{6}$.

Una terza possibilità è usare la funzione generatrice. Troviamo $\psi_i(z) := \psi_{X_i} := \frac{z+z^2+\ldots+z^6}{6} = \frac{z-z^7}{6(1-z)}$. Dalla Proposizione 4.6.4 segue immediatamente che $\psi(z) := \psi_X(z) = (\psi_{X_i}(z))^2 = (\frac{z-z^7}{6(1-z)})^2$, un risultato difficilmente trovabile direttamente dalla definizione di $\psi(z)$ per X.

4.7.1 Esercizio. Calcolare l'attesa e la varianza di X_i e di X con l'aiuto della funzione generatrice.

Capitolo 5

Variabili aleatorie continue

Non sempre possiamo limitarci al caso di una variabile aleatoria che assume (con probabilità 1) solo valori in un sottoinsieme numerabile di $\mathbb{R}$. Basta pensare all'Esempio I.4, in cui, soprattutto per ragioni pratici, abbiamo dovuto ammettere tutta la semiretta positiva come possibili valore. Uno dei problemi più vecchi (risolto molto prima che Kolmogorov enunciasse i suoi assiomi della teoria della probabilità) è quello di specificare che potrebbe essere l'*equidistribuzione* su un intervallo limitato $(a, b]$. Per avere un'idea di un esperimento concreto dove appare questo problema, guardiamo la seguente figura.

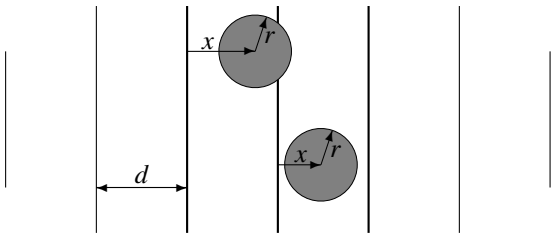


Figura 5.1: Palla di neve di raggio r lanciata "a caso" su una griglia di larghezza d.

Una palla di neve di raggio r la lanciamo "a caso" verso una griglia con la distanza d fra le sbarre. Vogliamo sapere, per esempio, se la palla si rompe, perché tocca una sbarra, oppure no. Dire "a caso" significa proprio che intendiamo non favorire nessuna posizione del centro della pallina rispetto alla griglia. Ci conviene essere un po' più specifico. Per prima cosa, idealizzando un po', la griglia la ci immaginiamo estesa ad infinito. Notiamo che la y–coordinata non centra per il problema se la palla tocchi la griglia, o meno. E per quanto riguarda l'x–coordinata notiamo, che l'intero problema è periodico con periodo d. Quindi, quello che ci interessa veramente è la distanza del centro della palla alla prossima sbarra, diciamo, alla sua sinistra, ossia un valore x dell'intervallo $(0, d]$. Chiaramente l'evento A ="la palla non si rompe" corrisponde al sottoin-

sieme $(r, d - r)$ di $(0, d]$. (Se ci dovesse capitare $d - r < r$, ossia $d < 2r$, cioè il diametro $2r$ della palla è più grande della distanza fra le sbarre, allora la palla semplicemente non ci passa e $A = \emptyset$. Questo caso banale l'escludiamo.)

Ci sono alcune cose da notare. Per primo, se tutti i punti $x \in (0, d]$ hanno la stessa probabilità p, allora p deve essere 0. (Altrimenti potremmo costruire addirittura sottoinsiemi B finiti di $(0, d]$, diciamo con n punti, tale che $P(B) = np$ fosse più grande di 1.) Per secondo, chiunque sarebbe d'accordo che, se raddoppiassimo la lunghezza del sottointervallo $(a, b]$, allora anche la probabilità di centrare l'intervallo lungo il doppio si raddoppierebbe. Continuando, giungiamo alla conclusione che la probabilità di trovare x nell'intervallo $(a, b]$ dovrebbe essere proporzionale alla sua lunghezza $b - a$. Soprattutto, la probabilità dell'intervallo intero $(0, d]$ dovrebbe essere uno. Chiediamo, quindi,

$$P((a, b]) = \frac{b - a}{d} \qquad (*)$$

per ogni intervallo $(a, b] \subset (0, d]$.

Dagli spazi di probabilità discreti (o le variabili aleatorie discrete) eravamo abituati, di poter attribuire senza problemi delle probabilità ad un qualsiasi sottoinsieme A di Ω (o di $\mathbb{R}$). Fu proprio un italiano, Vitali, a dimostrare che questo, per il problema appena discusso, non è più possibile: Non esiste nessuna probabilità (booleana) definita su $\mathcal{P}((0, d])$ che soddisfa $(*)$. Quindi, se ci teniamo ad avere tali distribuzioni sugli intervalli (e, infatti, ci teniamo moltissimo!), allora dobbiamo pur dire addio all'idea di poter sempre definire delle probabilità per tutti i sottoinsiemi di uno spazio campionario. Quali siano le algebre booleane più opportune e come possiamo definire su loro delle probabilità, lo discuteremo in questo capitolo almeno per $\Omega = \mathbb{R}$, ossia, per le leggi di variabili aleatorie. Non nascondiamo che la maggior parte delle difficoltà matematiche da risolvere, pur essendo sempre elementari, le rimandiamo all'Appendice C.

5.1 Probabilità su $\mathbb{R}$

Vogliamo avere a disposizione variabili aleatorie reali, cioè funzioni $X\colon \Omega \to \mathbb{R}$, che non più (come le variabili aleatorie discrete) assumono con probabilità 1 valori in un sottoinsieme numerabile. Vogliamo che fra gli insiemi ammissibili $B \subset \mathbb{R}$ (termine matematico: ***misurabili***) per l'evento "$X \in B$" ci siano almeno gli intervalli. Ci tocca, quindi, trovare due cose: L'algebra booleana su $\mathbb{R}$ generata dagli intervalli. E stabilire quando una funzione che assegna ad ogni intervallo un numero positivo definisce un'unica probabilità booleana su tale algebra booleana.

È opportuno procedere in due passi e guardare prima solo gli intervalli semiaperti (per convenzione semiaperti a sinistra). Un intervallo ***semiaperto a sinistra*** è un intervallo della forma $(a, b] = \{x \in \mathbb{R}\colon a < x \leq b\}$ o della forma $(a, \infty) = \{x \in \mathbb{R}\colon a < x\}$, dove $-\infty \leq a \leq b < \infty$.

(Notiamo che, secondo questa definizione, gli intervalli $(-\infty, b]$, (a, ∞) e $(a, a] = \emptyset$ sono tutti semiaperti a sinistra.) Annotiamo con $\mathcal{H}_]$ l'insieme di tutti questi intervalli semiaperti a sinistra. Non è difficile dimostrare che l'algebra booleana su $\mathbb{R}$ più piccola che contiene $\mathcal{H}_]$, è

$$\mathcal{B}_] := \beta(\mathcal{H}_]) := \{I_1 \cup \ldots \cup I_n \colon n \in \mathbb{N}, I_i \in \mathcal{H}_], i \neq j \Rightarrow I_i \cap I_j \emptyset\},$$

l'insieme di tutte le unioni di un numero finito di elementi due a due disgiunti di $\mathcal{H}_]$.

5.1.1 Nota matematica.

$\mathcal{H}_]$ è una semialgebra booleana secondo la Definizione C.1.1. (Infatti, $\emptyset \in \mathcal{H}_]$. Per ogni $I, I' \in \mathcal{H}_]$ anche $I \cap I' \in \mathcal{H}_]$. E per ogni $(a, b] \in \mathcal{H}_]$ il complementare $(a, b]^{\complement} = (-\infty, a] \cup (b, \infty)$ è l'unione di un numero finito di elementi due a due disgiunti di $\mathcal{H}_]$.) È facile convincersi, che l'insieme $\mathcal{B}_]$ è invariata sotto intersezioni. Per definizione di semialgebra booleana, il complementare di $I \in \mathcal{H}_]$ è un elemento di $\mathcal{B}_]$. Poi il complementare di $I_1 \cup \ldots \cup I_n$, come intersezione di elementi di $\mathcal{B}_]$, è di $\mathcal{B}_]$. Quindi, $\mathcal{B}_]$, essendo chiusa sotto intersezioni e complementari, è un'algebra booleana. Si veda, comunque, la Proposizione C.1.2 e, soprattutto, la sua dimostrazione.

Fine della nota.

Supponiamo che P sia una probabilità booleana definita su $\mathcal{B}_]$. Allora, se definiamo la ***funzione di ripartizione*** $F = F^{\leq} \colon \mathbb{R} \to \mathbb{R}$ come $F^{\leq}(x) := P((-\infty, x])$, possiamo calcolare

$$P((a, b]) = P(\,(-\infty, b] \backslash (-\infty, a]\,) = P((-\infty, b]) - P((-\infty, a]) = F^{\leq}(b) - F^{\leq}(a),$$

perché $(-\infty, b] = (-\infty, a] \cup (a, b]$ e $(-\infty, a] \cap (a, b] = \emptyset$.

Notiamo che la funzione $F^{\leq}$ determina la probabilità su tutto $\mathcal{B}_]$. Infatti, abbiamo visto che $F^{\leq}$ determina P su $\mathcal{H}_]$. Poi, per un elemento arbitrario $I_1 \cup \ldots \cup I_n$ di $\mathcal{B}_]$, visto che gli intervalli I_i sono due a due disgiunti, troviamo $P(I_1 \cup \ldots \cup I_n) = P(I_1) + \ldots + (P(I_n)$.

Inoltre, $F^{\leq}$ soddisfa $0 \leq F^{\leq}(x) \leq 1$ per ogni $x \in \mathbb{R}$ e $x \leq y \Rightarrow F^{\leq}(x) \leq F^{\leq}(y)$, cioè $F^{\leq}$ è ***crescente***. Sommiamo:

> Per ogni probabilità booleana P su $\mathcal{B}_]$ c'è una funzione $F^{\leq}$ crescente a valori in $[0, 1]$, la funzione di ripartizione, tale che $F^{\leq}$ determina P univocamente secondo $P((a, b]) = F^{\leq}(b) - F^{\leq}(a)$ per ogni $\infty < a < b < \infty$.

Una domanda naturale è, se ogni tale funzione crescente a valori in $[0, 1]$ è la funzione di ripartizione di una probabilità P su $\mathcal{B}_]$. La risposta a questa domanda è affermativa:

> Ogni funzione $F^{\leq}$ crescente a valori in $[0, 1]$ determina un'unica probabilità $P \colon \mathcal{B}_] \to \mathbb{R}$, di cui è la funzione di ripartizione.

5.1.2 Nota matematica.

Per dimostrarlo, occorre dimostrare che la funzione $P\colon (a,b] \mapsto F^{\leq}(b) - F^{\leq}(a)$ definita su $\mathcal{H}_{]}$, è additiva su $\mathcal{H}_{]}$ secondo la Definizione C.1.4. Con ciò intendiamo, che qualunque siano gli intervalli $I_1, \ldots, I_n \in \mathcal{H}_{]}$ semiaperti a sinistra e due a due disgiunti, se (ma solo se!) la loro unione è nuovamente un intervallo $I \in \mathcal{H}_{]}$, allora

$$P(I_1 \cup \ldots \cup I_n) \;=\; P(I_1) + \ldots + P(I_n).$$

Quindi, l'additività non si chiede per tutte le unioni disgiunte di intervalli, ma solo dove ha senso, cioè, dove l'unione è nuovamente un intervallo. Per le altre unioni disgiunte non ci sono condizioni! La Proposizione C.1.5 ci dice, che questo già basta.

Fine della nota.

Vediamo che è molto facile indicare una probabilità sull'agebra booleana $\mathcal{B}_{]}$: Basta indicare una funzione crescente $F^{\leq}\colon \mathbb{R} \to [0,1]$. Però, c'è un prezzo da pagare per questa semplicità. L'algebra booleana $\mathcal{H}_{]}$ non ancora contiene tutti gli intervalli. Infatti, non contiene gli eventi elementari $\{a\} = [a,a]$. Ci disturba anche una certa asimmetria. Che cosa rende più interessanti gli intervalli semiaperti a sinistra? (Infatti, avremmo benissimo potuto partire con la semialgebra booleana $\mathcal{H}_{[}$ degli intervalli semiaperti a destra. Tutto funziona allo stesso modo, tranne il fatto, che $F^{\leq}(x)$ dovrebbe essere sostituita con la funzione crescente $F^{<}(x) := P((-\infty, x))$, oppure con funzione decrescene $F^{\geq}(x) := P([x,\infty)) = 1 - F^{<}(x)$.)

Gli eventi elementari li possiamo facilmente aggiungere. Infatti, anche tutti gli intervalli aperti, semiaperti a sinistra o a destra, e chiusi, finiti ed infiniti formano una semialgebra booleana che annotiamo con $\mathcal{H}$. Ma già lo sforzo di descrivere tutti questi intervalli, ci indica che le notazioni diventano più complicate. Secondo la stessa Proposizione C.1.2, l'algebra booleana generata da $\mathcal{H}$ è

$$\mathcal{B} \;:=\; \beta(\mathcal{H}) \;:=\; \{I_1 \cup \ldots \cup I_n \colon n \in \mathbb{N}, I_i \in \mathcal{H}, i \neq j \Rightarrow I_i \cap I_j = \emptyset\}.$$

Anche qua, se P è una probabilità booleana, allora $F(x) = P((-\infty, x])$ definisce una funzione crescente a valori in $[0,1]$, tale che $P((a,b]) = F(b) - F(a)$. Però non sappiamo che succede con gli altri intervalli. Visto che $[a,b]$ è l'unione disgiunta di $\{a\} = [a,a]$ e $(a,b]$, se consociamo F, allora l'unica incertezza che rimane , sono proprio le probabilità degli eventi elementari.

Una discussione dettagliata di che cosa possa accadere nel caso generale la rimandiamo all'Appendice C.2. Per la comprensione completa occorre la conoscenza di alcuni delle proprietà analitiche eccezionali di cui le funzioni monotone si godono. A noi, in questo capitolo interessa solo un caso particolare: Supponiamo che la funzione crescente F sia continua. (Di questo fatto le variabili aleatorie *continue* prenderanno il nome.)

Supponiamo che F sia una funzione crescente continua a valori in $[0, 1]$. Questa continuità include i punti $-\infty$ ed ∞, cioè supponiamo che $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$. Osserviamo che per ogni $a < b < c$ vale $P((a, c)) \geq P((a, b])$, perché $(a, c) \supset (a, b]$. Allora

$$\begin{aligned} 0 \;\leq\; P(\{c\}) \;&=\; P((a,c]\backslash(a,c)) \;=\; P((a,c]) - P((a,c)) \\ &\leq\; P((a,c]) - P((a,b]) \;=\; P((b,c]) \;=\; F(c) - F(b). \end{aligned}$$

Visto che F l'abbiamo assunta continua e che $0 \leq P(\{c\}) \leq F(c) - F(b)$ vale proprio per ogni $b < c$, allora nel limite $b \uparrow c$ non rimane altro che $P(\{c\}) = 0$ per ogni c. L'incertezza che possa essere la probabilità di un evento elementare, svanisce nel nulla!

Nel ragionamento precedente abbiamo usato solo la ***continuità a sinistra*** di F. Ma questo basta già per avere un'unica probabilità:

5.1.3 Teorema. *Sia $F : \mathbb{R} \to \mathbb{R}$ continua che soddisfa:*

1. *F è crescente.*
2. *$\lim_{x\to-\infty} F(x) = 0$.*
3. *$\lim_{x\to\infty} F(x) = 1$.*

Allora esiste un'unica probabilità booleana su $\mathcal{B}$ che soddisfa

$$P([a,b]) \;=\; P((a,b]) \;=\; P([a,b)) \;=\; P((a,b)) \;=\; F(b) - F(a) \tag{5.1.1}$$

per ogni $a < b$.

Per quanto riguarda la dimostrazione, è simile a quella per $\mathcal{B}_]$; si veda la Nota 5.1.2. Dobbiamo solo dimostrare che la probabilità definita sulla semialgebra booleana $\mathcal{H}$ è additiva su $\mathcal{H}$. Poiché, secondo la discussione precedente, i punti individuali $\{a\}$ "non hanno peso" (nel senso della (5.1.1)), la dimostrazione dell'additività per $\mathcal{H}$ è più o meno la stessa come quella dell'additività per $\mathcal{H}_]$. Ci sono solo un po' più di difficoltà formali da affrontare.

5.1.4 Nota matematica.

Attenzione! Abbiamo concluso dalla continuità a sinistra di F che la probabilità di un evento elementare $\{a\}$ è 0. Non vale però l'affermazione opposta. Una probabilità booleana che assegna ad ogni evento elementare la probabilità 0 non debba avere necessariamente una funzione di ripartizione continua. Basta guardare la probabilità su $\mathcal{H}$ che assegna ad ogni intervallo della forma (a, b) con $a \leq 0$ e $b > 0$ la probabilità 1, e a tutti gli altri intervalli la probabilità zero. Quindi, un intervallo ha probabilità 1 già se "tocca lo zero" dall'alto, anche se 0 non è necessariamente contenuto. La probabilità di ogni evento elementare è

0. Comunque la funzione di ripartizione $F(x)$ è 0 per $x \leq 0$ e 1 per $x > 0$, quindi, non è continua.

Notiamo che questo comportamento è in contrasto con la teoria standard basata su σ–algebre e probabilità σ–additive. Infatti, la σ–additività porta con se che la funzione di ripartizione è continua a destra, e la funzione di ripartizione che abbiamo appena costruita non lo è.

Notiamo, inoltre, che neanche continuità a sinistra è necessaria per $P(\{a\}) = 0$. Scambiamo nell'esempio appena discusso i segni delle disuguaglianze in $P((a,b)) = 1$ se $a < 0$ e $b \geq 0$, e 0 in tutti gli altri casi. Allora $F(x) = 0$ per $x < 0$ e 1 per $x \geq 0$. Adesso F è continua a destra, ma non a sinistra. Comunque, rimane $P(\{a\}) = 0$ per ogni a.

Fine della nota.

Come abbiamo visto, una qualsiasi funzione F, che soddisfa le ipostesi del Teorema 5.1.3, definisce una probabilità booleana so $\mathcal{B}$. Un modo opportuno di procurarcene una, è il seguente:

Sia $p\colon \mathbb{R} \to \mathbb{R}$ una funzione *integrabile* (in un senso opportuno; si veda la seguente nota matematica), che soddisfa le seguenti condizioni:

1. p è ***positiva***, cioè $p(x) \geq 0$ per ogni $x \in \mathbb{R}$.
2. $\int p(x)\,dx := \int_{-\infty}^{\infty} p(x)\,dx = 1$.

Allora la funzione

$$F(x) := \int_{-\infty}^{x} p(y)\,dy$$

soddisfa le ipotesi del Teorema 5.1.3 e, quindi, definisce una probabilità su $\mathcal{B}$ tale che

$$P([a,b]) = P((a,b]) = P([a,b)) = P((a,b)) = F(b) - F(a) = \int_a^b p(x)\,dx \tag{5.1.2}$$

per ogni $a \leq b$. Chiamiamo la funzione p la ***densità*** della probabilità P.

5.1.5 Nota matematica.

Il significato del termine ***integrabile*** appena usato, dipende dalla *teoria d'integrazione* che abbiamo a disposizione (ossia, dalla nostra preparazione matematica). La teoria più generale conosciuta, è l'integrale di Lebesgue. Gli assiomi della probabilità dati da Kolmogorov si riferiscono a questa teoria. Se usiamo l'integrabilità nel senso di Lebesgue, il collegamento fra funzioni di ripartizioni continui F e la loro rappresentazione come integrale $F(x) = \int_{-\infty}^{x} p(y)\,dy$ per una densità opportuna diventa reciproco. (Il teorema di Radon-Nikodym della teoria dell'integrale di Lebesgue afferma proprio l'esistenza di una tale densità sotto un'ipotesi più forte di continuità che una funzione monotona, comunque, soddisfa non appena sia continua.)

Noi, dai corsi di Matematica I e II, abbiamo a disposizione l'integrale di Riemann, proprio ed improprio. (L'integrale proprio è quello $\int_a^b p(x)\,dx$ su un intervallo limitato. Mentre se, nel caso improprio, dove almeno uno dei limiti è infinito, intendiamo il limite dell'integrale proprio mandando a a $-\infty$ e/o b ad ∞. Per la nostra fortuna, l'esistenza dell'integrale di Riemann improprio su tutta la retta per una funzione **positiva**(!) implica l'integrabilità nel senso dell'integrale di Lebesgue.) Una classe di funzioni per noi assolutamente sufficienti, fra cui scegliere delle densità, sono le funzioni continui o almeno continui ***a pezzi***. Con ciò intendiamo che possiamo suddividere la retta in un numero finito di intervalli, tali che su ogni intervallo la funzione è continua.

Fine della nota.

Spesso una densità ci è data a meno di una costante moltiplicativa. Ripetiamo quello che abbiamo già fatto per le leggi discrete (esame!):

Data una funzione $f\colon \mathbb{R} \to \mathbb{R}_+$ dobbiamo trovare una costante $c > 0$ tale che la funzione

$$p(x) \;:=\; cf(x)$$

diventi la densità di una variabile aleatoria X.

Bisogna, quindi trovare c, tale che $\int p(x)\,dx = 1$. Se esiste l'integrale $\int f(x)\,dx < \infty$, allora il problema ha la soluzione $c = \frac{1}{\int f(x)\,dx<\infty}$. Si vedano gli esempi della Sezione 5.5.

5.1.6 Esempio. Torniamo al problema discusso nell'introduzione di questo capitolo. Sia $[a, b]$ un intervallo limitato di lunghezza $b - a$ diversa da zero. Vogliamo che la probabilità $P([a', b'])$ di ogni intervallo $[a', b'] \subset [a, b]$ sia proporzionale alla sua lunghezza $b' - a'$. Osserviamo che

$$\int_{a'}^{b'} dx \;=\; b' - a'.$$

Allora ci possiamo provare con la densità p come multiplo opportuno cf della funzione seguente:

$$f(x) \;=\; \begin{cases} 1 & a \le x \le b, \\ 0 & \text{altrimenti.} \end{cases}$$

La costante c la dobbiamo scegliere tale che $P(\mathbb{R})$ diventi 1. Calcoliamo

$$c\int f(x)\,dx \;=\; c\left(\int_{-\infty}^{a} 0\cdot dx + \int_a^b 1\cdot dx + \int_b^{\infty} 0\cdot dx\right) \;=\; c\left(0 + [x]_a^b + 0\right) \;=\; c(b-a),$$

di cui segue $c = \frac{1}{b-a}$. Chiameremo la legge continua con la densità $\frac{1}{b-a}$ nell'intervallo $[a, b]$ e 0 altrove, la ***distribuzione uniforme*** sull'intervallo $[a, b]$. Scriviamo $X \sim uni_{[a,b]}$ se la variabile aleatoria ha la legge continua con la densità $uni_{a,b}(x) := cf(x)$.

5.1.7 Esempio. Sia $f(x) = e^{-\lambda x}$ per $x \geq 0$ e $f(x) = 0$ altrove, dove $\lambda > 0$ è un parametro. La costante c la dobbiamo scegliere tale che $P(\mathbb{R})$ con la densità $p = cf$ diventi 1. Allora

$$c \int f(x)\,dx \;=\; c \int_0^\infty e^{-\lambda x}\,dx \;=\; c\Big[\frac{e^{-\lambda x}}{-\lambda}\Big]_0^\infty \;=\; c\Big[\frac{0-1}{-\lambda}\Big] \;=\; \frac{c}{\lambda},$$

di cui segue $c = \lambda$. Questa legge la chiamiamo la ***legge esponenziale*** di parametro λ. Scriviamo $X \sim exp_\lambda$ se la variabile aleatoria ha la legge continua con la densità $exp_\lambda(x) := cf(x)$.

Si vede subito che la funzione di ripartizione di $X \sim exp_\lambda$ è $F(x) = \int_{-\infty}^x exp_\lambda(y)\,dy = \int_0^x \lambda e^{-\lambda y}\,dy = 1 - e^{-\lambda x}$ per $x \geq 0$ e $F(x) = 0$ per $x \leq 0$. Per $s, t \geq 0$ troviamo

$$\begin{aligned} P(X \in [s, s+t] | X \geq s) \;&=\; \frac{P(\text{“}X \in [s, s+t]\text{”} \cap \text{“}X \geq s\text{”})}{P(X \geq s)} \;=\; \frac{P(X \in [s, s+t])}{P(X \geq s)} \\ &=\; \frac{F(s+t) - F(s)}{1 - F(s)} \;=\; \frac{e^{-\lambda s} - e^{-\lambda(s+t)}}{e^{-\lambda s}} \;=\; 1 - e^{-\lambda s} \;=\; F(t). \end{aligned}$$

La legge esponenziale è un modello per descrivere fenomeni di *decadenza* (per esempio, il tempo fino alla decadenza di un atomo radioattivo, o il tempo fino al guasto di un'apparecchio. Come la legge geometrica, anche la legge esponenziale, una volta arrivato al punto $X = s$, non ricorda la preistoria come ci sia arrivato, e la probabilità che la decadenza arrivi in un momento del intervallo successivo ad s di lunghezza t, è la stessa come se iniziassimo a $s = 0$.

5.2 Leggi e attese di variabili aleatorie continue

In questa sezione definiamo in piena generalità che cosa sia una *variabile aleatoria* a valori in $\mathbb{R}$. Tale definizione è motivata dal desiderio di potere attribuire alla variabile una legge, ossia, una probabilità booleana su $\mathcal{B}$. Poi useremo le nostre conoscenze sulla funzione di ripartizione per definire che cosa siano una variabile aleatoria *continua* e una variabile aleatoria continua *con densità*. Poi discutiamo le loro attese.

5.2.1 Definizione. Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità booleano. Una funzione $X\colon \Omega \to \mathbb{R}$ è una ***varabile aleatoria (reale)*** se per ogni intervallo I di $\mathcal{H}$ l'evento "$X \in I$" è in $\mathcal{A}$.

5.2.2 Nota matematica.

> Questa definizione compatibile con quella solita in altri testi di probabilità. Solo che la nostra nozione di spazio do probabilità booleano è più ampia di quella di spazio di probabilità.

Fine della nota.

Notiamo che, se X è una variabile aleatoria, allora la funzione $P_X(B) := P(X \in B)$ per ogni $B \in \mathcal{B}$ è una probabilità booleana su $\mathcal{B}$. (**Esercizio:** Verificarlo!)

5.2.3 Definizione. Chiameremo P_X la ***legge*** di X.

5.2.4 Nota matematica.

Anche questa definizione è, *cum grano salis*, compatibile con quella solita. Una variabile aleatoria definita su uno spazio di probabilità definisce una probabilità. Però tale probabilità non è definita solo su l'algebra booleana $\mathcal{B}$ ma su una σ–algebra decisamente più grande di $\mathcal{B}$. (Si chiama la ***σ–algebra boreliana*** di $\mathbb{R}$ e, infatti, si tratta della σ–algebra più piccola che contiene $\mathcal{B}$ (o $\mathcal{H}$ o $\mathcal{H}_]$ o $\mathcal{H}_[$ o $\mathcal{B}_]$, proprio non importa). Il fatto che lo spazio su cui la variabile aleatoria è definita sia uno spazio di probabilità, garantisce che la probabilità su $\mathcal{B}$ estende sul dominio più grande. Ma questo è teoria della misura, che non trattiamo.

Fine della nota.

Infine, arriviamo alle variabili aleatorie continue, il soggetto che ci interessa.

5.2.5 Definizione. Diciamo un variabile aleatoria X su Ω è ***continua***, se ha la ***funzione di ripartizione*** $F(x) := P(X \leq x)$ continua. Diciamo che X è continua ***con densità***, se la legge P_X ha una densità $p\colon \mathbb{R} \to \mathbb{R}_+$, tale che P_X è determinata dalla (5.1.2).

È importante stabilire se una variabile aleatoria X sia continua o abbia addirittura una densità. Per definizione, identificare una funzione $p\colon \mathbb{R} \to \mathbb{R}_+$ come densità di X, significa verificare che $P(X \in [a,b]) = \int_a^b p(x)\,dx$. È sufficiente verificare $F_X(x) = \int_{-\infty}^x p(u)\,du$. Per trovare candidati possiamo, quindi, calcolare la derivata di F_X. Saremo confrontati con problemi di questo tipo in tutto ill resto del capitolo. La seguente proposizione ne da un esempio.

5.2.6 Proposizione. *Sia X una variabile aleatoria e poniamo $Y := \sigma X + \mu$ ($\sigma > 0, \mu \in \mathbb{R}$).*

1. *Anche Y è una variabile aleatoria.*
2. *Se X è continua, lo è anche Y.*
3. *Se X ha la densità, lo ha anche Y. Infatti, una possibile scelta è la funzione*

$$q(y) := \tfrac{1}{a} p(\tfrac{y-b}{a})$$

Dimostrazione. Dall'elementare equivalenza

$$Y \in [a,b] \iff \sigma X + \mu \in [a,b] \iff X \in [\tfrac{a-\mu}{\sigma}, \tfrac{b-\mu}{\sigma}]$$

segue, che ogni evento "$Y \in [a,b]$" è di $\mathcal{A}$ e che con F_X anche la funzione $F_Y(y) = F_X(\frac{y-\mu}{\sigma})$ è continua. Infine, se F_X ha la densità p, allora

$$F_Y(y) \;=\; F_X(\frac{y-\mu}{\sigma}) \;=\; \int_{-\infty}^{\frac{y-\mu}{\sigma}} p(x)\,dx \;=\; \int_{-\infty}^{y} p(\tfrac{u-\mu}{\sigma})\,du \;=\; \int_{-\infty}^{y} q(u)\,du,$$

che verifica q come possibile densità. ■

5.2.7 Esercizio. Sia X una variabile aleatoria. Dimostrare che:

1. Anche $|X|$ è una variabile aleatoria.
2. Se X è continua, lo è anche $|X|$.
3. Se X ha la densità, lo ha anche $|X|$. (Indicare una densità!)

5.2.8 Esercizio. Formulare e dimostrare una versione della Proposizione 5.2.6 per $\sigma < 0$.

5.2.9 Esercizio. Sia X una variabile aleatoria continua con densità p. Dimostrare che anche $Y := X^2$ è una variabile aleatoria continua, che ha la densità

$$q(y) \;=\; \begin{cases} \frac{p(\sqrt{y})+p(-\sqrt{y})}{2\sqrt{y}} & y > 0, \\ 0 & y \le 0. \end{cases}$$

Passiamo all'attesa di una variabile aleatoria. Questa parte fino al Teorema 5.2.11 serve per motivare la definizione dell'attesa di una variabile aleatoria generale (si veda il Teorema 5.2.10) per poi derivarne la forma nel caso di una variabile continua con densità. La difficoltà matematica, benché sempre elementare e a disposizione delle nostre conoscenze, e chi non è interessato a questa motivazione, può anche saltarla e andare subito alla Definizione 5.2.12.

L'attesa di una variabile aleatoria discreta X con densità p era definita come

$$\mathbb{E}X \;=\; \sum_{x\in\mathbb{R}} x \cdot p(x)$$

salvo che esista, cioè, salvo che $\sum_{x\in\mathbb{R}} |x| \cdot p(x) < \infty$. Sia $(I_n)_{n\in\mathbb{N}}$ una partizione di $\mathbb{R}$ in intervalli $I_n \in \mathcal{H}_1$ non vuoti, e tali che $\sup_{n\in\mathbb{N}} \lambda(I_n) = L < \infty$, dove con $\lambda(I)$ annotiamo la lunghezza di I. Chiamiamo L l'***ampiezza*** della partizione. Per ogni $n \in \mathbb{N}$ scegliamo un $x_n \in I_n$. Ogni x si trova in uno e uno solo degli intervalli I_n. Se definiamo y_x di essere quello degli x_n tale che x ed y_x si trovino nello stesso intervallo. Vediamo che

$$\Big|\sum_{x\in\mathbb{R}} x\cdot p(x) - \sum_{x\in\mathbb{R}} y_x \cdot p(x)\Big| \;=\; \Big|\sum_{x\in\mathbb{R}} (x-y_x)\cdot p(x)\Big| \;\le\; \sum_{x\in\mathbb{R}} |x-y_x|\cdot p(x) \;\le\; \sum_{x\in\mathbb{R}} L\cdot p(x) \;=\; L.$$

Quindi, se facciamo la lunghezza massima L degli intervalli sempre più piccola e per qualsia scelta di punti x_n in questi intervalli, le sommatorie $\sum_{x\in\mathbb{R}} y_x \cdot p(x)$ convergono verso l'attesa $\mathbb{E}X$. Poiché gli I_n formano una partizione, troviamo

$$\sum_{x\in\mathbb{R}} y_x \cdot p(x) \;=\; \sum_{n\in\mathbb{N}}\sum_{x\in I_n} y_x \cdot p(x) \;=\; \sum_{n\in\mathbb{N}}\sum_{x\in I_n} x_n \cdot p(x) \;=\; \sum_{n\in\mathbb{N}} x_n \sum_{x\in I_n} p(x) \;=\; \sum_{n\in\mathbb{N}} x_n \cdot P(X\in I_n).$$

Per i due calcoli precedenti era essenziale che la legge della variabile aleatorira discreta rende $\mathbb{R}$ uno spazio di probabilità discreto, e che a causa di questo ogni partizione $(I_n)_{n\in\mathbb{N}}$ soddisfa $\sum_{n\in\mathbb{N}} P(I_n) = 1$. La cosa bella è, che per calcolare l'ultima sommatoria a destra e, poi, mandare L a 0, questa proprietà è l'unica necessaria. Chiamiamo una variabile aleatoria ***partizionabile*** se soddisfa $\sum_{n\in\mathbb{N}} P(I_n) = 1$ per ogni partizione di $\mathbb{R}$ in intervalli di ampiezza finita.

5.2.10 Teorema. *Sia X una variabile aleatoria partizionabile su uno spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$. Allora vale una delle due possibilità:*

1. *Per una partizione in intervalli $I_n \in \mathcal{H}_1$ di ampiezza finita e una scelta di punti $x_n \in I_n$ vale*
$$\sum_{n\in\mathbb{N}} |x_n| \cdot P(X\in I_n) \;=\; \infty.$$
In tal caso questa uguaglianza vale per tutte tali partizioni e scelte di x_n. Scriviamo $\mathbb{E}\,|X| = \infty$.

2. *Per una partizione in intervalli $I_n \in \mathcal{H}_1$ di ampiezza finita e una scelta di punti $x_n \in I_n$ vale*
$$\sum_{n\in\mathbb{N}} |x_n| \cdot P(X\in I_n) \;<\; \infty.$$
In tal caso questa disuguaglianza vale per tutte tali partizioni e scelte di x_n. Inoltre, qualsiasi sia la scelta di $0 \neq L_m \to 0$, per $m \to \infty$ le sommatorie $\sum_{n\in\mathbb{N}} |x_n| \cdot P(X \in I_n)$ convergono verso lo stesso limite che annotiamo con $\mathbb{E}\,|X|$.

*Nel secondo caso convergono anche tutte le sommatorie $\sum_{n\in\mathbb{N}} x_n \cdot P(X \in I_n)$ convergono verso un solo limite che annotiamo con $\mathbb{E}X$ e che chiamiamo **attesa** di X.*

Bozza di una dimostrazione. Non è troppo difficile di capire (anche se richiede molto allenamento matematico) che se per una partizione la sommatoria e finita, allora la differenza di quella con una la sommatoria per una qualsiasi altra partizione non è più grande somma delle ampiezze delle due partizioni. In particolare, anche la seconda sommatoria è finita. (Procedendo per *contrapposizione*, questo dimostra subito (1).) Poi scegliendo in (2) una successione di partizioni con gli L_m come indicati, si vede che le sommatorie formano una *successione di Cauchy* che, per la completezza di $\mathbb{R}$, ha un limite. Scegliendo una altra successione e prendendo in considerazione che le differenze convergono verso 0, vediamo che i limite devono coincidere. ■

Dalla discussione precede il teorema, segue che una variabile aleatoria discreta è sempre partizionabile, e che l'attesa secondo la Definizione 4.4.1 coincide con l'attesa secondo il Teorema 5.2.10, esistenza inclusa.

Passiamo alle varibili aleatorie continue con densità.

5.2.11 Teorema. *Sia X una variabile aleatoria continua su uno spazio di probabilità booleano con densità p. Allora X è partizionabile ed $\mathbb{E}X$ esiste se e solo se $\int |x| \cdot p(x)\,dx < \infty$, ed in tal caso*

$$\mathbb{E}X \;=\; \int x \cdot p(x)\,dx.$$

La dimostrazione dipende (come la classe delle funzioni a cui p possa appartenere) dalla teoria dell'integrazione che abbiamo a disposizione. (Se $I = (a, b]$ allora $P(X \in I) = \int_a^b p(x)\,dx$. Secondo una versione opportuna del *teorema della media*, in I esiste un y tale che $\int_a^b x \cdot p(x)\,dx = y \int_a^b p(x)\,dx$. Usando questo fatto, la dimostrazione riesce simile a quella del Teorema 5.2.10.) L'omettiamo. Invece **ignoriamo** il Teorema 5.2.11, e prendiamo il suo contenuto come definizione dell'attesa di una variabile aleatoria continua.

5.2.12 Definizione. Sia X una variabile aleatoria continua su uno spazio di probabilità booleano con densità p. Se $\int |x| \cdot p(x)\,dx < \infty$, definiamo l'***attesa*** di X

$$\mathbb{E}X \;:=\; \int x \cdot p(x)\,dx$$

e diciamo X ***l'attesa finita***. Altrimenti l'attesa di X non è definita.

5.2.13 Esercizio. Sia X una variabile aleatoria continua con densità p. Usare il risultato dell'Esercizio 5.2.9 per dimostrare che $\mathbb{E}X^2 = \int x^2 \cdot p(x)\,dx$.

L'esercizio precedente, ci permette di calcolare subito le varianze di alcune variabili aleatorie continue con densità come $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$, anche se la giustificazione viene solo dopo.

5.2.14 Esempio. Torniamo all'Esempio 5.1.6 della variabile X distribuita secondo la distribuzione uniforme $uni_{[a,b]}$ sull'intervallo finito $[a, b]$. Abbiamo capito una volta per tutte, che gli integrali si svolgono solo sull'intervallo dove $p(x)$ è diverso da 0. Troviamo

$$\mathbb{E}X \;=\; \int xp(x)\,dx \;=\; \frac{1}{b-a}\int_a^b x\,dx \;=\; \frac{1}{b-a}\Big[\frac{x^2}{2}\Big]_a^b \;=\; \frac{b^2-a^2}{2(b-a)} \;=\; \frac{b+a}{2}.$$

Vediamo con soddisfazione (perché con soddisfazione?) che l'attesa è la media, il baricentro, dell'intervallo. Per la varianza calcoliamo prima il secondo momento. Troviamo

$$\mathbb{E}X^2 \;=\; \int x^2 p(x)\,dx \;=\; \frac{1}{b-a}\int_a^b x^2\,dx \;=\; \frac{1}{b-a}\Big[\frac{x^3}{3}\Big]_a^b \;=\; \frac{b^3-a^3}{3(b-a)} \;=\; \frac{b^2+ba+a^2}{3}.$$

La varianza risulta

$$\mathbb{V}X \;=\; \mathbb{E}X^2 - (\mathbb{E}X)^2 \;=\; \frac{b^2+ba+a^2}{3} - \frac{b^2+2ab+a^2}{4} \;=\; \frac{b^2-2ab+a^2}{12} \;=\; \frac{(b-a)^2}{12}.$$

La varianza dipende solo dalla lunghezza dell'intervallo, non dalla sua posizione. Troviamo riconfermato che la varianza è invariata sotto traslazioni.

5.2.15 Esempio. Torniamo all'Esempio 5.1.7 della variabile X distribuita secondo la la legge esponenziale exp_λ di parametro λ. Gli integrali $\int_0^\infty xe^{-\lambda x}\,dx$ e $\int_0^\infty x^2e^{-\lambda x}\,dx$ per calcolare l'attesa ed il secondo momento possono essere risolti con integrazioni per parti. Noi useremo un trucco che si serve della derivata di un integrale per un parametro. Infatti, vale

$$\frac{d}{d\lambda}\int_0^\infty e^{-\lambda x}\,dx \;=\; -\int_0^\infty xe^{-\lambda x}\,dx \qquad \text{e} \qquad \frac{d^2}{d\lambda^2}\int_0^\infty e^{-\lambda x}\,dx \;=\; \int x^2e^{-\lambda x}\,dx.$$

Il valore $\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}$ l'avevamo già calcolato. Quindi,

$$\int_0^\infty xe^{-\lambda x}\,dx \;=\; -\frac{d}{d\lambda}\int_0^\infty e^{-\lambda x}\,dx \;=\; -\frac{d}{d\lambda}\left(\frac{1}{\lambda}\right) \;=\; \frac{1}{\lambda^2}$$

e

$$\int_0^\infty x^2e^{-\lambda x}\,dx \;=\; \frac{d^2}{d\lambda^2}\int_0^\infty e^{-\lambda x}\,dx \;=\; \frac{d^2}{d\lambda^2}\left(\frac{1}{\lambda}\right) \;=\; \frac{2}{\lambda^3}.$$

Allora

$$\mathbb{E}X \;=\; \lambda\int_0^\infty xe^{-\lambda x}\,dx \;=\; \frac{1}{\lambda} \qquad \text{e} \qquad \mathbb{E}X^2 \;=\; \lambda\int_0^\infty x^2e^{-\lambda x}\,dx \;=\; \frac{2}{\lambda^2}.$$

Allora la varianza è $\mathbb{V}X = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

5.2.16 Esercizio. Sia X una variabile aleatoria continua con densità p, e poniamo $Y := \sigma X + \mu$. Allora per **ogni** $\sigma \in \mathbb{R}$ e $\mu \in \mathbb{R}$ vale $\mathbb{E}Y = \sigma\mathbb{E}X + \mu$ e $\mathbb{V}Y = \sigma^2\mathbb{V}X$. (Prestare particolare attenzione ai casi $\sigma < 0$ e $\sigma = 0$! È sempre vero che Y sia una varibile aleatoria continua?)

5.3 Vettori aleatori continui e le loro funzioni

Un ***vettore aleatorio*** su uno spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$ è semplicemente un'm–upla $X = (X_1, \ldots, X_m)$ di variabili aleatorie su Ω secondo la Definizione 5.2.1. Come li, anche la nostra definizione di vettore aleatorio su uno spazio di probabilità booleano è compatibile con la versione per spazi di probabilità. Visto che ogni $X_i \in I_i$ è un evento di $\mathcal{A}$, lo è anche la loro intersezione. Annotiamo con $\mathcal{B}^m := \beta(\mathcal{H}^m)$ l'algebra booleana su $\mathbb{R}^m$ generata dalla famiglia

$$\mathcal{H}^m \;:=\; \{I_1 \times \ldots \times I_m \colon I_i \in \mathcal{H}\}$$

di tutti gli ***intervalli m–dimensionali***. Allora $P_X(B) := P(X \in B)$ definisce una probabilità booleana su $\mathcal{B}^m$ (**esercizio:** dimostrarlo!), che chiamiamo la ***legge congiunta*** di X. Infine, notiamo che la famiglia $\mathcal{H}^m$ è una semialgebra booleana su $\mathbb{R}^m$. (Si veda l'Appendice C.3 per dettagli.) Quindi, una probabilità su $\mathcal{B}^m$ è già determinata dai valori che assume su $\mathcal{H}^m$. (Proposizione C.1.5.)

5.3.1 Esempio. Sia $(\mathbb{R}^m, \mathcal{B}^m, P)$ uno spazio di probabilità booleano. Se definiamo le funzioni $X_i(x_1, \ldots, x_m) := x_i$ da $\mathbb{R}^m$ in $\mathbb{R}$, allora l'm–upla $X = (X_1, \ldots, X_m)$ è un vettore aleatorio che ha la legge congiunta $P_X = P$. Notiamo che X come funzione a valori in $\mathbb{R}^m$ non è che l'identità $\mathsf{id}_{\mathbb{R}^m}$ su $\mathbb{R}^m$.

Anche se, in principio, sarebbe possibile di generalizzare la discussione con funzioni di ripartizione, tale discussione per più variabili diventa molto scomoda. Non definiremo vettori aleatori continui generali, ma solo quelli con densità, e lo facciamo in analogia con la definizione per una sola variabile.

5.3.2 Definizione. Un ***vettore aleatorio continuo con densità*** è un vettore aleatorio $X = (X_1, \ldots, X_m)$ tale che esiste una funzione integrabile $p \colon \mathbb{R}_m \to \mathbb{R}_+$ che soddisfa

$$P(X_1 \in I_1, \ldots, X_m \in I_m) \;=\; \int_{I_1 \times \ldots \times I_m} p(x)\, dx$$

per ogni scelta di $I_i \in \mathcal{H}$. Chiamiamo p una ***densità*** di X. Diciamo la legge P_X è ***continua con densità*** p.

Ci sono diverse cose da notare. Innanzitutto, l'integrale è un integrale sul dominio m–dimensionale $I_1 \times \ldots \times I_m$ di $\mathbb{R}^m$. Per chi non lo conosce, tale integrale può essere calcolato in termini di integrali un-dimensionali iterati, secondo la formula

$$\int_{I_1 \times \ldots \times I_m} p(x)\, dx \;=\; \int_{I_m} \left(\ldots \left(\int_{I_2} \left(\int_{I_1} p(x_1, \ldots, x_m)\, dx_1 \right) dx_2 \right) \ldots \right) dx_m.$$

È anche un teorema che il valore non dipende dall'ordine in cui eseguiamo gli integrali unidimensionali. È anche chiaro che la densità p la possiamo variare su un sottoinsieme di cui funzione indicatrice ha l'integrale 0. La densità è, quindi, non unica. Non è difficile vedere che

$$P_X(B) \;=\; \int_B p(x)\,dx \;:=\; \int_{\mathbb{R}^m} \chi_B(x)\cdot p(x)\,dx.$$

(Segue subito dal fatto che l'indicatrice è "additiva" sotto unioni disgiunti.) In particolare,

$$\int_{\mathbb{R}^m} p(x)\,dx \;=\; P_X(\mathbb{R}^m) \;=\; P(X\in\mathbb{R}^m) \;=\; P(\Omega) \;=\; 1.$$

Molto più interessante è il fatto che, di nuovo grazie alla "additività" delle indicatrici, vale anche l'affermazione opposta:

5.3.3 Teorema. *Sia $p\colon \mathbb{R}^m \to \mathbb{R}_+$ una funzione integrabile tale che $\int_{R^m} p(x)\,dx = 1$. Allora*

$$P(B) \;=\; \int_B p(x)\,dx$$

definisce una probabilità booleana su $\mathcal{B}^m$.

Quasi tutti i risultati del Capitolo 4 sulle variabili aleatorie discrete valgono anche per le variabili ed i vettori aleatori continui con densità, ed anche più generali. Per esempio, nella Definizione 4.3.1 di ***indipendenza*** per le variabili discrete, scriviamo semplicemente variabili arbitrari ma i sottoinsiemi B_i di $\mathbb{R}$ gli dobbiamo limitare ad essere elementi di $\mathcal{H}$. Con tale definizione la Proposizione 4.3.4 sulla fattorizzazione delle densità rimane valida nel senso seguente:

5.3.4 Proposizione. *Sia $X = (X_1 \ldots, X_m)$ vettore aleatorio continuo con densità p e annotiamo con*

$$p_i(x_i) \;:=\; \int_{I_1\times\ldots\times I_{i-1}\times I_{i+1}\times\ldots\times I_m} p(x_1,\ldots,x_m)\,d(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m)$$

*la **densità marginale** di X_i. Allora vale:*

1. *X_i è una variabile aleatoria continua con densità p_i. (Lo stesso vale per densità marginali a più variabili da definire in un senso opportuno.)*
2. *X è indipendente se e solo se la funzione*

$$(x_1,\ldots,x_m) \;\longmapsto\; p_1(x_1)\cdot\ldots\cdot p_m(x_m)$$

è una densità per X.

DIMOSTRAZIONE. Basta verificare alla lettera le definizioni. ■

Fra tutte le affermazioni che valgono con piccole modifiche o proprio senza cambiare una virgola, un'eccezione drammatica è la Proposizione 4.2.7. Non è più valido che ogni funzione di un vettore aleatorio continuo con densità sia una variabile aleatoria continua con densità. Per prima cosa, per una funzione qualsiasi $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ non è neanche detto che la funzione $Z := \varphi \circ X$ sia una variabile aleatoria.

5.3.5 Nota matematica.

Per le variabili aleatorie su spazi di probabilità, la teoria della misura ci dice che φ deve essere ***misurabile***, ossia, per ogni $B \in \mathcal{H}$ la controimmagine $\varphi^{-1}(B)$ deve essere un'evento della σ–algebra generata da $\mathcal{H}^m$. A noi, invece, servirebbe che $\varphi^{-1}(B)$ sia un elemento della famiglia più piccola $\mathcal{B}^m$. Una buona volta le nostre condizioni sono più ristrette che quelle della teoria generale. Se dovessimo essere costretti a considerare solo queste variabili aleatorie, allora questo si che significherebbe una limitazione severa.

Fine della nota.

Per fortuna, abbiamo altri metodi a disposizione per rimediare. Per prima cosa osserviamo che un integrale della forma come nel Teorema 5.3.3 può essere calcolato per sottoinsiemi ben più generali che i soli elementi di $\mathcal{B}^m$.

5.3.6 Teorema. *Sia $p\colon \mathbb{R}^m \to \mathbb{R}_+$ una funzione integrabile tale che $\int_{R^m} p(x)\,dx = 1$. Allora vale:*

1. *La famiglia*
$$\mathcal{B}_p := \{B \subset \mathbb{R}^m\colon \chi_B \cdot p \text{ è “integrabile”}\}$$
è un'algebra booleana su $\mathbb{R}^m$ (contenente $\mathcal{B}^m$).

2. *La funzione $P_p\colon \mathcal{B}_p \to \mathbb{R}$ definita come*
$$P_p(B) := \int_B p(x)\,dx$$
è una probabilità booleana (estendo P_X se p è densita di X).

5.3.7 Nota matematica.

Il termine “integrabile”, e quindi anche la definizione di $\mathcal{B}_p$, dipende dalla teoria d'integrazione a nostra disposizione. Però in qualsiasi teoria d'integrazione la somma ed il prodotto di funzioni integrabili sono integrabili. Questo vale soprattutto per $\chi_{B\cap B'} = \chi_B \cdot \chi_{B'}$ e per $\chi_{B^\complement} = 1 - \chi_B$ visto che le funzioni costanti sono integrabili. Con ciò, $\mathcal{B}_p$ è un'algebra booleana. Che P_p definisca una probabilità booleana, segue come sempre.

Fine della nota.

5.3.8 Definizione. Diciamo che $(\Omega, \mathcal{A}, P)$ è ***completo rispetto al vettore aleatorio continuo*** X ***con densità*** p se "$X \in B$"$\in \mathcal{A}$ per ogni $B \in \mathcal{B}_p$.

5.3.9 Nota matematica.

Si può dimostrare che ogni Ω possiede un ***completamento rispetto a*** p passando semplicemente da $\mathcal{A}$ a $\beta(\mathcal{A}, \{\text{"}X \in B\text{"} : B \in \mathcal{B}_p\})$. Data una famiglia di vettori aleatori con densità, si può addirittura aumentare $\mathcal{A}$ in tal modo che sia completa rispetto ad ogni densità dei vettori della famiglia. La dimostrazione dipende moltissimo dall teoria d'integrazione e dal fatto che la legge di un vettore aleatorio continuo con densità è, effettivamente, σ–additiva.

Un modo efficiente di ottenere uno spazio di probabilità booleano completo rispeto ad X, che evita tutti i problemi, è sostituire $\mathcal{A}$ con l'agebra booleana $X^{-1}(\mathcal{B}_p)$ che contiene tutti gli eventi "$X \in B$" per $B \in \mathcal{B}_p$.

Fine della nota.

5.3.10 Teorema. *Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio continuo con densità p sullo spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$, e sia $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ una funzione tale che $\varphi^{-1}(I) \in \mathcal{B}_p$ per ogni $I \in \mathcal{H}$. In altre parole, sia φ è una variabile aleatoria su $(\mathbb{R}_m, \mathcal{B}_p, P_p)$. Allora*

$$\mathbb{E}\varphi \;=\; \int \varphi(x) \cdot p(x)\, dx,$$

esistenza reciproca inclusa.
Sia, inoltre, Ω completo rispetto a p. Allora $Z = \varphi \circ X$ è una variabile aleatoria con l'attesa $\mathbb{E}Z = \mathbb{E}\varphi$.

5.3.11 Definizione. Sia $X = (X_1, \dots, X_m)$ un vettore aleatorio continuo con densità p sullo spazio di probabilità booleano $(\Omega, \mathcal{A}, P)$. Chiameremo $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ ***ammissibile rispetto a*** p se è una variabile aleatoria su $(\mathbb{R}_m, \mathcal{B}_p, P_p)$. Chiameremo una funzione $Z\colon \Omega \to \mathbb{R}$ una ***variabile aleatoria ammissibile rispetto ad*** X se esiste una funzione $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ ammissibile rispetto a p tale che $Z = \varphi \circ X$.

Chiameremo una funzione $Z = (Z_1, \dots, Z_k)\colon \Omega \to \mathbb{R}^k$ un ***vettore aleatorio ammissibile rispetto ad*** X se ogni Z_i è una variabile aleatoria ammissibile rispetto ad X.

Notiamo che qualsiasi teoria d'integrazione garantisce che l'insieme di tutte la variabili aleatorie ammissibili rispetto ad X è invariato sotto somma e prodotto di funzioni $\Omega \to \mathbb{R}$. Con Z, anche $|Z|$ è ammissibile. È facile vedere che per ogni funzione continua $\psi\colon \mathbb{R} \to \mathbb{R}$ che possiede una partizione di $\mathbb{R}$ in un numero finito di intervalli I_n, tale che tutte le restrizioni $\psi \restriction I_n$ sono iniettive, con Z anche la funzione $\psi \circ Z$ è una variabile aleatoria ammissibile. Si può anche dimostrare facilmente che ogni variabile aleatoria discreta è ammissibile rispetto ad

una variabile aleatoria $X \sim uni_{[0,1]}$. Ovviamente rimane valida la Proposizione 4.3.7: Funzioni di blocchi indipendenti di vettori indipendenti sono indipendenti.

Vediamo che adesso copriamo un'insieme (infatti, un'algebra) di variabili aleatorie abbastanza grande, che ci permette di lavorare senza preoccupazioni. Vediamo anche che, grazie al Teorema 5.3.10, il calcolo di attese di variabili aleatorie Z come integrale non dipende dal modo in cui abbiamo espresso Z come $\varphi \circ X$. Possiamo trasportare subito la definizione dei momenti.

5.3.12 Definizione. Sia Z una variabile aleatoria ammissibile rispetto ad X, e sia $Z = \varphi \circ X$ una sua realizzazione. Allora diciamo Z ha l'n–esimo momento finito, se $\mathbb{E}\,|X|^n = \int |\varphi(x)|^n \cdot p(x)\,dx < \infty$. In tal caso esiste anche $\mathbb{E}X^n = \int \varphi(x)^n \cdot p(x)\,dx$ e lo chiamiamo l'***n–esimo momento*** di Z. Come sempre, il primo momento $\mathbb{E}X$ lo chiamiamo l'***attesa*** di X e definiamo la ***varianza*** come $\mathbb{V}X := \mathbb{E}(X - \mathbb{E}X)$.

Senza modifiche rimangono validi:

1. I Corollari 4.4.5 e 4.4.6 con le regole del calcolo.

2. La Proposizione 4.4.7 ed il suo Corollario 4.4.8.

3. Le Proposizioni 4.5.2 e 4.5.3 sui momenti.

4. I risultati sulla varianza dell'Osservazione 4.5.5 fra l'altro la formula $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2$, il Corollario 4.5.6 e, soprattutto, la disuguaglianza di Cebicev 4.5.8.

5. La definizione di covarianza con il Corollario 4.5.10 con la conseguenza fondamentale, la legge dei grandi numeri 4.5.13.

5.4 La legge normale ed il teorema limite centrale

Come abbiamo ottenuto la legge di Poisson dalla legge binomiale tramite l'approssimazione di Poisson, otterremo la legge normale tramite un'altra approssimazione della legge binomiale. l'approssimazione di De Moivre e Laplace. La ***legge normale*** N_{μ,σ^2} di parametri $\mu \in \mathbb{R}, \sigma^2 > 0$ è una legge continua che ha la densità

$$n_{\mu,\sigma^2}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

In particolare, la legge $N_{0,1}$ si chiama la ***legge normale standardizzata***. Tralasciamo per il momento la dimostrazione che $n_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ sia la densità di una probabilità su $\mathcal{B} \subset \mathcal{P}(\mathbb{R})$.

Però, tenendolo per scontato, dalla Proposizione 5.2.6 si evince che se $X \sim N_{0,1}$ è una variabile aleatoria normale standardizzata, allora $\sigma X + \mu \sim N_{\mu,\sigma^2}$. Grazie a risultati come quello dell'Esercizio 5.2.16, basta studiare la legge normal standardizzata $N_{0,1}$.

Per ogni $n \in \mathbb{N}_0$, la funzione $x^n e^{-\frac{x^2}{2}}$ è limitata. (Infatti, per induzione dalla *regola di L'Hospital* segue che il limite per $x \to \pm\infty$ è addirittura 0.) Visto che $x^{n+2}e^{-\frac{x^2}{2}}$ è limitata, diciamo da una costante $C > 0$, ne segue che $x^n e^{-\frac{x^2}{2}}$ è limitata da $\frac{C}{x^2}$. La maggiorante $\frac{C}{x^2}$ è integrabile a $\pm\infty$, quindi la legge normale possiede momenti di ogni ordine n.

I momenti dispari di $X \sim N_{0,1}$, in particolare l'attesa, sono facile da calcolare, visto che la densità è ***simmetrica***: $n_{0,1}(-x) = n_{0,1}(x)$. Infatti, per qualsiasi funzione integrabile f ***antisimmetrica***, cioè, $f(-x) = -f(x)$, vale

$$\begin{aligned}\int f(x)\,dx \;&=\; \int_{-\infty}^{0} f(x)\,dx + \int_0^{\infty} f(x)\,dx \\ &= -\int_{\infty}^{0} f(-x)\,dx + \int_0^{\infty} f(x)\,dx \;=\; \int_0^{\infty} (f(-x)+f(x))\,dx \;=\; 0,\end{aligned}$$

quindi $\mathbb{E}X^{2n-1} = \int x^{2n-1} n_{0,1}(x)\,dx = 0$. In particolare, $\mathbb{E}X = 0$.

La normalizzazione $\frac{1}{\sqrt{2\pi}}$ della funzione $e^{-\frac{x^2}{2}}$ non è troppo facile da calcolare. (Non è possibile con i mezzi del calcolo di una variabile reale.) Ma una volta nota (come supponiamo noi), un trucco simile a quello che abbiamo usato per la legge esponenziale (o anche con la funzione generatrice di Laplace) ci permette di calcolare anche gli altri momenti pari $\mathbb{E}X^{2n}$. Infatti, per $\alpha > 0$ calcoliamo

$$g(\alpha) \;:=\; \int e^{-\frac{\alpha x^2}{2}}\,dx \;=\; \int e^{-\frac{\alpha x^2}{2}}\,\frac{d(\sqrt{\alpha}x)}{\sqrt{\alpha}} \;=\; \frac{1}{\sqrt{\alpha}}\int e^{-\frac{y^2}{2}}\,dy \;=\; \frac{\sqrt{2\pi}}{\sqrt{\alpha}}.$$

L'n–esima derivata di g, da una parte ci da $\frac{1}{(-2)^n}\int x^{2n} e^{-\frac{\alpha x^2}{2}}\,dx$. Dall'altra parte ci da

$$\frac{d^n g}{dx^n} \;=\; \frac{2n-1}{-2}\cdot\ldots\cdot\frac{1}{-2}\cdot\frac{\sqrt{2\pi}}{\alpha^{\frac{2n+1}{2}}}.$$

Quindi, mettendo $\alpha = 1$, troviamo

$$\mathbb{E}X^{2n} \;=\; \frac{(-2)^n}{\sqrt{2\pi}}\frac{d^n g}{dx^n}(1) \;=\; (2n-1)\cdot\ldots\cdot 1.$$

In particolare, $\mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1 - 0^2 = 1$. Applicando il risultato dell'Esercizio 5.2.16, ne segue la legge normale N_{μ,σ^2} ha attesa $\mathbb{E}(\sigma X + \mu) = \mathbb{E}X + \mu = \mu$ e $\mathbb{V}(\sigma X + \mu) = \sigma^2 \mathbb{V}X = \sigma^2$.

La funzione di ripartizione della legge normale standardizzata $N_{0,1}$ si chiamal a ***funzione di Gauß dell'errore*** è va annotata con

$$\Phi(x) \;:=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{y^2}{2}},dy.$$

Non è fra le funzione che conosciamo. I valori $\Phi(x)$ si trovano in tabelle. (Tipicamente, per x fra 0 e 3 in passi da un centesimo. Per $x >$, il valore $\Phi(x)$ è già più grande di $0,99$ e le variazioni che ci sono ancora, servono proprio solo nelle applicazioni numeriche più sofisticate. Grazie alla simmetria di $n_{0,1}(x)$, i valori di Φ per argomenti negativi possono essere calcolate come $\Phi(-x) = P(X \leq -x) = P(X \geq x) = 1 - P(X < x) = 1 - \Phi(x)$ da quelli positivi.) Oggi, la funzione di Gauß è dispopnibili con software come SPSS o Matematica.

L'approssimazione locale di De Moivre-Laplace locale permette di approssimare $B_{n,p}(k)$ con l'aiuto della legge normale uniformenente se k è in un certo intervallo (che dipende da n). Va usata per dimostrare l'approssimazione globale, che è nient'altro che un caso particolare del *teorema limite centrale* che discuteremo dopo.

5.4.1 Teorema. *(Approssimazione locale di De Moivre-Laplace.) Siano $p \in (0,1)$, $q = 1 - p$ e $x_n(k) = \frac{k-np}{\sqrt{npq}}$. Allora per ogni $\varepsilon > 0, A > 0$ esiste un N tale che*

$$\left| \frac{\sqrt{npq} B_{n,p}(k)}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_n(k)^2}{2}}} - 1 \right| \;\leq\; \varepsilon$$

per ogni $n \geq N$ e ogni k tale che $|x_n(k)| \leq A$, cioè $k \in [np - \sqrt{npq}A\,,\, np + \sqrt{npq}A]$.

(Approssimazione globale di De Moivre-Laplace.) Sia $X_1 \sim X_2 \sim \ldots \sim B_{1,p}$ una successione di variabili aleatori indipendenti di legge bernoulliana $B_{1,p}$. In altre parole, sia $S_n = X_1 + \ldots + X_n \sim B_{n,p}$. Poniamo $S_n^* = \frac{S_n - np}{\sqrt{npq}}$ (cosicché $\mathbb{E}S_n^* = 0$ e $\mathbb{V}S_n^* = 1$). Allora

$$\lim_{n\to\infty} P(S_n^* \in [a,b]) \;=\; \Phi(b) - \Phi(a)$$

(uniformemente) per ogni $a \leq b$.

L'idea dell'approssimazione della somma di una successione di variabili indipendenti con attesa 0 (perciò *centrali*) e varianza 1 normalizzata in modo opportuno con la legge normale (come per le variabili bernoulliane nel paragrafo precedente) funziona in piena generalità. "Normalizzata in modo opportuno" significa tale che il risultato è non banale. Se normalizziamo con $\frac{1}{n}$, la legge dei grandi numeri ci dice che la convergenza è verso la costante 0. Se non normalizziamo la varianza n diventa sempre più grande, quindi non converge. La scelta "giusta" è $\frac{1}{\sqrt{n}}$. (Una dimostrazione che consiste solo nel "contare" sotto l'ipotesi che esistano tutti i momenti, motivata dalla *probabilità quantistica*, ci farebbe vedere anche il perché.)

5.4.2 Teorema. *(Teorema limite centrale.) Sia $X_1 \sim X_2 \sim \ldots$ una successione di variabili aleatori indipendenti della stessa legge con $\mathbb{E}X_i = 0$ e $\mathbb{V}X_i = 1$. Allora*

$$\lim_{n\to\infty} P\Big(\tfrac{X_1+\ldots+X_n}{\sqrt{n}} \in [a,b]\Big) \;=\; \Phi(b) - \Phi(a)$$

per ogni $a \leq b$.

L'importanza del teorema limite centrale consiste nel fatto che spesso un fenomeno aleatorio può essere considerato il risultato di molti effetti piccoli indipendenti. Perciò possiamo supporre che, approssimativamente, il fenomeno ha una legge normale. Un modello matematico di un'evoluzione di tali fenomeni è il *moto browniano*.

La dimostrazione del teorema limite centrale si serve della funzione caratteristica. Mentre la funzione generatrice è riservata alle variabili a valori in $\mathbb{N}_0$, che sono discrete, la ***funzione caratteristica*** di una probabilità è un metodo generale. Per le variabili con densità si tratta della *trasformata di Fourier*. Non è neanche difficile di dimostrare che le funzioni caratteristiche delle somme $\frac{X_1+\ldots+X_n}{\sqrt{n}}$ convergono verso la funzione caratterstica della legge normale $N(0,1)$. La dimostrazione che questo basti per la convergenza delle probabilitè su gli intervalli (*convergenza in legge*) richiede, però, una bella porzione di conoscenze dell'analisi che noi non abbiamo.

Un'altro metodo di dimostrazione, nell caso che tutte le X_i posseggano momenti di ogni ordine, è bastato sulla *convergenza dei momenti*. Di nuovo, non è difficile, benché fastidioso, di dimostrare tale convergenza, ma la dimostrazione che questi basti per convergenza in legge va oltre le nostre possibilità.

5.5 Convoluzioni ed altri altri esempi con densità

Il prossimo esempio discute un impostazione tipica di problemi intorno a leggi continue con densità. *Determinare la costante* è un compito importante. Se, poi, derviamo nuove densità da quelli che conosciamo (per esemprio tramite *convoulzioni*), dove sapiamo che il risultato è per forza una densità, questo ci fornisce molte formule fra diverse costanti di normalizzazione.

5.5.1 Esempio. Sia

$$f(x) \;=\; \begin{cases} x & 0 \le x \le 1, \\ 0 & \text{altrimenti.} \end{cases}$$

La costante c la dobbiamo scegliere tale che $P(\mathbb{R})$ con la densità $p = cf$ diventi 1. Calcoliamo

$$c\int f(x)\,dx \;=\; c\int_0^1 x\,dx \;=\; c\Big[\frac{x^2}{2}\Big]_0^1 \;=\; c\frac{1^2-0^2}{2} \;=\; \frac{c}{2},$$

di cui segue $c = 2$. Per l'attesa troviamo

$$\mathbb{E}X \;=\; \int xp(x)\,dx \;=\; 2\int_0^1 x^2\,dx \;=\; 2\Big[\frac{x^3}{3}\Big]_0^1 \;=\; 2\frac{1^2-0^2}{3} \;=\; \frac{2}{3}.$$

Per il secondo momento troviamo

$$\mathbb{E}X^2 \;=\; \int x^2p(x)\,dx \;=\; 2\int_0^1 x^3\,dx \;=\; 2\Big[\frac{x^4}{4}\Big]_0^1 \;=\; 2\frac{1^3-0^3}{4} \;=\; \frac{1}{2}.$$

La varianza risulta

$$\mathbb{V}X \;=\; \mathbb{E}X^2 - (\mathbb{E}X)^2 \;=\; \frac{1}{2} - \frac{4}{9} \;=\; \frac{9-8}{18} \;=\; \frac{1}{18}.$$

5.5.2 Convoluzione di densità. Siano X e Y variabili aleatorie indipendenti continue con densità p e q. Ci possiamo chiedere della legge della variabile $Z = X + Y$.

Proposizione. *Z è una variabile continua con densità data come prodotto di* ***convoluzione***

$$p * q(z) \;:=\; \int p(u)q(z-u)\,du.$$

Dimostrazione. Ricordiamoci che, grazie all'indipendenza una possibile densità del vettore aleatorio continuo (X, Y) è $(x, y) \mapsto p(x)q(y)$. La probabilità che Z assuma un valore in $[a, b]$ è, quindi, data da

$$\begin{aligned} P(Z \in [a,b]) \;&=\; \int dx \int dy\, \chi_{[a,b]}(x+y)p(x)q(y) \;=\; \int dx \int dz\, \chi_{[a,b]}(z)p(x)q(z-x) \\ &=\; \int dx \int_a^b dz\, p(x)q(z-x) \;=\; \int_a^b dz \int dx\, p(x)q(z-x). \end{aligned}$$

Questo identifica $z \mapsto \int dx\, p(x)q(z-x)$ come possibile densità di Z. ∎

Questa proposizione include l'affermazione la convoluzione $p * q$ di due densità è una densità. Dal fatto che $Z = X + Y = Y + X$ segue anche $p * q = q * p$. (**Esercizio:** Verificarlo direttamente dalla definizione.)

5.5.3 Legge gamma $\Gamma_{\lambda,r}$ di parametri $\lambda > 0, r > 0$. Siano $X \sim Y \sim exp_\lambda$ variabili aleatorie indipendenti. Allora la legge di $Z = X + Y$ è

$$\begin{aligned} exp_\lambda * exp_\lambda(z) \;&=\; \int exp_\lambda(u) \cdot exp_\lambda(z-u)\,du \;\overset{!}{=}\; \int_0^z \lambda e^{-\lambda u} \cdot \lambda e^{-\lambda(z-u)}\,du \\ &=\; \lambda^2 e^{-\lambda z} \int_0^z du \;=\; \lambda^2 z e^{-\lambda z}, \end{aligned}$$

dove $z \geq 0$. (Questo spiega i limiti dell'integrale nel passo $\overset{!}{=}$.) Quindi, modulo la costante di normalizzazione, è la prima potenza di z per la funzione esponenziale $e^{-\lambda z}$.

Anche se $X \sim N_{0,1}$, secondo l'Esercizio 5.2.9, troviamo per la densità di $Y = X^2$

$$y \;\longmapsto\; \frac{n_{0,1}(\sqrt{y}) + n_{0,1}(-\sqrt{y})}{2\sqrt{y}} \;=\; \frac{2e^{-\frac{\sqrt{y}^2}{2}}}{2\sqrt{2\pi}\sqrt{y}} \;=\; \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}$$

per $y \geq 0$, una potenza di y per una funzione esponenziale module la costante di normalizzazione.

Questo ci motiva di studiare tutte leggi con delle densità proporzionali ad una delle funzioni $x^{r-1}e^{-\lambda x}$ epr $x \geq 0$. Perché x^{r-1} e non x^r? Va bene, una delle ragioni è che tutto vada bene per $r > 0$, e tale scelta, x^{r-1}, lo so soddisfa. Un'altra ragione, come vediamo sotto, è l'ottimo comportamento del parametro sotto convoluzioni. Il parametro λ, chiaramente, deve essere strettamente più grande di 0 se la funzione deve essere integrabile. Definiamo come la ***legge gamma*** $\Gamma_{\lambda,r}$ di parametri $\lambda > 0, r > 0$ la legge con densità

$$\gamma_{\lambda,r}(x) \ := \ \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x \geq 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Infatti, l'integrale su questa funzione esiste per ogni scelta di $r > 0$, e $\Gamma(r)$ bisogna determinarla per ottenere la normalizzazione giusta. Dalla Proposizione 5.2.6, si nota subito che se $X \sim \Gamma_{1,r}$, allora $Y = \frac{X}{\lambda} \sim \Gamma_{\lambda,r}$ e che, infatti, la costante

$$\Gamma(r) \ := \ \int_0^\infty x^{r-1} e^{-x} \, dx$$

non dipende da λ. Si tratta della famosa ***funzione gamma***,da cui la legge gamma prende il nome, e che provvede un'*interpolazione* del fattoriale. Infatti, è un **esercizio** in *integrazione per parti* di dimostrare (**farlo!**) che

$$\Gamma(r+1) \ = \ r\Gamma(r),$$

di cui si evinca in particolare che $n! = \Gamma(n+1)$.

Siano $X \sim \Gamma_{1,r}$ e $Y \sim \Gamma_{1,s}$ indipendenti. Allora la legge di $Z = X + Y$ è

$$\gamma_{1,r} * \gamma_{1,s}(z) \ = \ \frac{1}{\Gamma(r)} \cdot \frac{1}{\Gamma(s)} \int_0^z u^{r-1} e^{-u} (z-u)^{s-1} e^{-(z-u)} \, du \ = \ \frac{e^{-z}}{\Gamma(r)\Gamma(s)} \int_0^1 (zv)^{r-1} (z - zv)^{s-1} z \, dv$$

$$= \ \frac{z^{r+s-1} e^{-z}}{\Gamma(r)\Gamma(s)} \int_0^1 v^{r-1} (1-v)^{s-1} \, dv \ = \ \frac{B(r,s)}{\Gamma(r)\Gamma(s)} z^{r+s-1} e^{-z},$$

dove abbiamo definito la ***funzione beta*** $B(r,s) := \int_0^1 v^{r-1}(1-v)^{s-1} \, dv$ per ogni $r, s, > 0$. Vediamo che $\gamma_{1,r} * \gamma_{1,s}$ è proporzionale a $\gamma_{1,r+s}$. Da questo si evincono due conseguenze. La prima: Visto che $\gamma_{1,r} * \gamma_{1,s}$ è una densità non puè fare a meno di essere proprio uguale a $\gamma_{1,r+s}$. La seconda: Per forza la costante di normalizzazione deve soddisfare $\frac{B(r,s)}{\Gamma(r)\Gamma(s)} = \frac{1}{\Gamma(r+s)}$ o

$$B(r,s) \ = \ \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)},$$

una formula difficilmente intuibile. Lasciamo come **esercizio(!)**, di usare il risultato $\gamma_{1,r} * \gamma_{1,s} = \gamma_{1,r+s}$ e la Proposizione 5.2.6 per dimostrare $\gamma_{\lambda,r} * \gamma_{\lambda,s} = \gamma_{\lambda,r+s}$ per ogni $\lambda > 0$.

Una formula altrettano bella, però un po' difficile da dimostrare, è $N_{\mu,\sigma^2} * N_{\nu,\tau^2} = N_{\mu+\nu,\sigma^2+\tau^2}$. Nell'Esercizio 4.6.8 abbiamo visto che le leggi di Poisson sono un'esempio di una familgia di leggi discreti che si compoirtano bene sotto somma di variabili indipendenti.

5.5.4 Legge beta $B_{r,s}$ di parametri $r, s > 0$. Per definizione della funzione beta, la funzione

$$\beta_{r,s}(x) := \begin{cases} \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1} & x \in (0,1), \\ 0 & \text{altrimenti,} \end{cases}$$

è la densità di una probabilità che chiamiamo la ***legge beta*** di parametri $r, s > 0$. Ne torniamo in un attimo.

5.5.5 Legge del χ^2. Una delle ragioni di introdurre le leggi gamma, era la convoluzione di due leggi esponenziali. Infatti, le leggi gamma includono la legge esponenziale come $exp_\lambda = \Gamma_{\lambda,1}$. Quindi troviamo $\underbrace{exp_\lambda * \ldots * exp_\lambda}_{n \text{ fattori}} = \Gamma_{\lambda,n}$, confermando il cso $n = 2$ che avevamo già calcolato.

Abbiamo anche visto che se $X \sim N_{0,1}$, allora $X^2 \sim \gamma_{\frac{1}{2},\frac{1}{2}} =: \chi^2$, che si chiama anche la ***legge del*** χ^2. Ne segue che, se $X_1 \sim \ldots \sim X_n \sim N_{0,1}$ sono indipendenti allora troviamo $X_1^2 + \ldots + X_n^2 \sim \Gamma_{\frac{1}{2},\frac{n}{2}} =: \chi^2_n$. Chiamiamo χ^2_n la ***legge del*** χ^2 ***di*** n ***gradi di libertà***. Le leggi del χ^2 fanno importantissima apparizione nel seguente teorema.

5.5.6 Teorema. *(Student alias W.S. Gosset.) Siano $X_1 \sim \ldots \sim X_n \sim N_{\mu,\sigma^2}$ indipendenti. Poniamo $M := \frac{X_1+\ldots+X_n}{n}$ e $V^* := \frac{(X_1-M)^2+\ldots+(X_n-M)^2}{n-1}$. Allora:*

1. *M e V^* sono indipendenti.*
2. *$M \sim N_{\mu,\frac{\sigma^2}{n}}$ e $\frac{n-1}{\sigma^2} V^* \sim \chi^2_{n-1}$.*
3. *$\frac{\sqrt{n}(M-\mu)}{\sqrt{V^*}} \sim t_{n-1}$ dove la **legge di Student** t_n su $\mathbb{R}$ ha la densità*

$$\tau_n(x) = \frac{1}{B(\frac{1}{2},\frac{n}{2})\sqrt{n}(1+\frac{x^2}{n})^{\frac{n+1}{2}}}.$$

Questo teorema è cruciale nella statistica per la costruzione di *intervalli di fiducia*.

5.5.7 Calcolo di momenti per densità con il fattore x^n. Siano X_n distribuite secondo delle densità $p_n(x) = c_n x^n f(x)$, dove f è una funzione integrabile $\mathbb{R} \to \mathbb{R}_+$ e c_n sono costanti di normalizzazione. Quindi,

$$x p_n(x) = c_n x^{n+1} f(x) = \frac{c_n}{c_{n+1}} p_{n+1}(x),$$

di cui

$$\mathbb{E}X_n^{m+1} \;=\; \int x^{m+1} p_n(x)\,dx \;=\; \frac{c_n}{c_{n+1}} \int x^m p_{n+1}(x)\,dx \;=\; \frac{c_n}{c_{n+1}} \mathbb{E}X_{n+1}^m.$$

Di conseguenza

$$\mathbb{E}X_n^m \;=\; \frac{c_n}{c_{n+1}} \mathbb{E}X_{n+1}^{m-1} \;=\; \ldots \;=\; \frac{c_n}{c_{n+1}} \cdot \frac{c_{n+1}}{c_{n+2}} \cdot \ldots \cdot \frac{c_{n+m-1}}{c_{n+m}} \mathbb{E}X_{n+m}^0 \;=\; \frac{c_n}{c_{n+m}},$$

in particolare, $\mathbb{E}X_n = \frac{c_n}{c_{n+1}}$,$\mathbb{E}X_n^2 = \frac{c_n}{c_{n+2}}$, quindi, $\mathbb{V}X_n = \frac{c_n}{c_{n+2}} - \frac{c_n^2}{c_{n+1}^2}$.

Le leggi gamma $X_n \sim \Gamma_{\lambda,r+n}$ sono di questo tipo, con $f(x) = x^{r-1}e^{-\lambda x}$ $(x > 0)$ e $c_n = \frac{\lambda^{r+n}}{\Gamma(r+n)}$. Lo stesso vale per le leggi beta $X_n \sim B_{r+n,s}$ con $f(x) = xr - 1(1-x)^{s-1}$ $(0 < x < 1)$ e $c_n = \frac{1}{B(r+n,s)}$. Giocando un po' con le ricursioni

$$\Gamma(r+1) \;=\; r\Gamma(r), \qquad\qquad \frac{B(r+1,s)}{B(r,s)} \;=\; \frac{\Gamma(r+1)\Gamma(s)}{\Gamma(r+s+1)} \cdot \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} \;=\; \frac{r}{r+s},$$

troviamo i seguenti risultati:

$$\mathbb{E}\Gamma_{\lambda,r} \;=\; \frac{r}{\lambda}, \qquad \mathbb{V}\Gamma_{\lambda,r} \;=\; \frac{r}{\lambda^2}, \qquad \mathbb{E}B_{r,s} \;=\; \frac{r}{r+s}, \qquad \mathbb{V}B_{r,s} \;=\; \frac{rs}{(r+s)^2(r+s+1)}.$$

Lasciamo i dettagli come **esercizio!**

5.5.8 Esempio. Abbiamo già discusso la distribuzione uniforme su un intervallo finito. In genere, se A è un sottoinsieme di $\mathbb{R}$ tale che l'indicatrice χ_A è integrabile, allora possiamo definire la ***distribuzione uniforme*** uni_A ***su*** A come probabilità su $\mathcal{B}$ con densità

$$uni_A(x) \;:=\; \frac{\chi_A(x)}{\int \chi_A(y)\,dy}.$$

Allora se $B \subset A$, troviamo che $P(\bullet|B) \sim uni_B$.

5.5.9 Esercizio. Calcolare $uni_{[0,1]} * uni_{[-1,0]}$. Quali sono attesa e varianza di una variabile con questa densità?

5.5.10 Tempo d'attesa in una coda: La legge esponenziale a la legge di Poisson. La variabile aleatoria X che indica il tempo necessario a servire un cliente ad uno sportello (banca, poste, cassa del supermercato ...), sia distribuito secondo la legge esponenziale di parametro λ. (Oppure, possiamo considerare il tempo di funzionamento di una lampadina che, quando si rompe, va sostituito con la prossima. Oppure, possiamo considerare il tempo di decadenza di un atomo radioattivo.) Per $t > 0$, ci interessa la variabile aleatoria Y_t che indica quanti clienti vanno serviti nel periodo di tempo $[0, t]$. (Oppure, vogliamo sapere, quante lampadine servono

nell'intervallo $[0, 1]$ per tenere in funzione l'illuminazione. Oppure, ci chiediamo quanti atomi decadono nell'intervallo $[0, t]$)

Sia X_n la variabile aleatoria che indica il tempo che passa fra inizio e termine del servizio all'n–esimo cliente (oppure, il tempo che passa fra l'istallazione dell'n–esima lampadina fino alla sua rottura, oppure, il tempo che passa fra la decadenza dell'n–esimo atomo e quella dell'$(n + 1)$–esimo). L'evento "$Y_t = k$" che nel periodo $[0, t]$ vanno serviti esattamente k clienti (oppure, ecc.), significa che la somma $Z_k := X_1 + \ldots + X_k$ dei tempi $X_1, \ldots, X_k$ consumati dai primi k clienti, non deve superare t (altrimenti, il numero dei clienti serviti sarebbe più piccolo di k), mentre quello $Z_{k+1} = Z_k + X_{k+1}$ consumato dai primi $k+1$ clienti deve superare t (altrimenti, sarebbero stati serviti almeno $k + 1$ clienti). Ci interessa, quindi, la probabilità dell'evento

$$Z_k \;\leq\; t \;\leq\; Z_k + X_{k+1}.$$

Notiamo che il tempo X_{n+1} che passa dal momento Z_n fino al momento Z_{n+1} è indipendente dal valore attuale di Z_n. Ne seguono due conclusioni. La prima: Z_k è la somma di k variabili esponenziali di parametro λ e, quindi, $Z_k \sim \gamma_{\lambda,k}$. La seconda: La legge congiunta di Z_k e X_{k+1} è $p(z_k, x) = \gamma_{\lambda,k}(z_k) \cdot exp_\lambda(x)$. Troviamo, quindi,

$$\begin{aligned} P(Y_t = k) \;&=\; P(Z_k \leq t \leq Z_k + X_{k+1}) \;=\; \int_0^t dz_k \int_{t-z_k}^{\infty} dx\, \gamma_{\lambda,k}(z_k) \cdot exp_\lambda(x) \\ &=\; \int_0^t dz_k \, \frac{\lambda^k}{\Gamma(k)} z_k^{k-1} e^{-\lambda z_k} \int_{t-z_k}^{\infty} dx\, \lambda e^{-\lambda x} \;=\; \frac{\lambda^k}{(k-1)!} \int_0^t dz_k \, z_k^{k-1} e^{-\lambda z_k} e^{-\lambda(t-z_k)} \\ &=\; \frac{\lambda^k e^{-\lambda t}}{(k-1)!} \int_0^t dz_k \, z_k^{k-1} \;=\; \frac{\lambda^k e^{-\lambda t}}{(k-1)!} \, \frac{t^k}{k} \;=\; \frac{(\lambda t)^k e^{-\lambda t}}{k!} \;=\; Poi_{\lambda t}(k). \end{aligned}$$

Appendice A

Alcuni rinfrescamenti

In questo appendice ripetiamo alcuni risultati noti dai corsi di Matematica I e II. L'Appendice A.1 si occupa dei risultati basilari del calcolo combinatorio. Ci servono durante il testo per determinare il numero di elementi di eventi. Dopo aver ricordato le proprietà basilari dei limiti nell'Appendice A.2, l'Appendice A.3 si occupa dei risultati basilari sulle serie assolutamente convergenti. Ci servono per poter probabilizzare gli spazi di probabilità discreti e per calcolare delle attese di variabili aleatorie discrete. Soprattutto è importante che le sommatorie (o serie) che calcoliamo non dipendono dal modo in cui lo facciamo.

Per tutte e due le appendici è indispensabile avere presente delle idee chiarissime su che cosa sia una funzione ***biettiva***, ossia, equivalentemente, una funzione ***invertibile***. Non possiamo ripetere anche questo. Riferiamo il lettore che si sente (ancora!) insicuro di questo concetto *sine qua non*, ai corsi di Matematica I e II; si vedano le dispense [Ske04] del corso di algebra.

Alla base della matematica c'è l'idea di contare degli *oggetti*. Il risultato sono gli ***assiomi di Peano*** dei numeri naturali; si veda [Ske04]. Un proverbio, non solo fra matematici, dice:

I numeri naturali vengono da Dio; tutto il resto è opera umana.

L'esistenza dei numeri naturali costituisce l'***assioma fondamentale della matematica***. Strettamente collegato con la definizione dei numeri naturali è la tecnica della ***dimostrazione per induzione***, che supponiamo nota; si veda [Ske04]. Questa tecnica fa apparizione in quasi tutte le dimostrazioni del calcolo combinatorio, ed in tante altre dimostrazioni di queste dispense.

Contare il numero di oggetti, significa determinare la ***cardinalità*** $\#\Omega$ dell'insieme Ω che li contiene. La cardinalità, invece, è una nozione della matematica ben precisa, che non discutiamo in piena generalità; si veda nuovamente [Ske04]. A noi, del concetto generale di cardinalità, ci interessano i seguenti fatti:

1. L'insieme vuoto ha la cardinalità $\#\emptyset = 0$.

2. Un insieme Ω ha la cardinalità $\#\Omega = n \in \mathbb{N}$ (oppure, l'insieme Ω ha n elementi), se esiste una biezione $f\colon \{1,\dots,n\} \to \Omega$. Un insieme Ω è ***finito***, se $\#\Omega = n$ per un $n \in \mathbb{N}$. Se Ω non è finito, allora è ***infinito***.

   Per la definizione del ***segmento*** $\{1,\dots,n\} := \{k \in \mathbb{N}\colon\ 1 \le k \le n\}$ usando l'ordinamento di $\mathbb{N}$, ci riferiamo a [Ske04].

   Con la notazione $\Omega = \{\omega_1,\dots,\omega_n\}$ per un insieme di cardinalità $\#\Omega = n$ intendiamo, che abbiamo scelta una biezione $f\colon \{1,\dots,n\} \to \Omega$ e abbiamo elencato gli elementi di Ω come $\omega_k := f(k)$ $(k = 1,\dots,n)$.

3. Un insieme Ω ha la cardinalità $\#\Omega = \aleph_0$ (cioè, aleph zero), se esiste una biezione $f\colon \mathbb{N} \to \Omega$. In particolare, $\#\mathbb{N} = \aleph_0$ (perché?). Un insieme Ω è ***numerabile***, se $\#\Omega = n$ per un $n \in \mathbb{N}$ o se $\#\Omega = \aleph_0$. Se Ω non è numerabile, scriviamo $\#\Omega > \aleph_0$.

   Con la notazione $\Omega = \{\omega_1, \omega_2, \dots\}$ per un insieme numerabile infinito intendiamo, che abbiamo scelta una biezione $f\colon \mathbb{N} \to \Omega$ e abbiamo elencato gli elementi di Ω come $\omega_k := f(k)$ $(k = 1, 2, \dots)$.

Attenzione: Alcuni autori escludono il caso degli insiemi finiti, quindi, per loro un insieme è numerabile, solo se ha la cardinalità $\aleph_0$. Noi ***non*** adottiamo questo punto di vista.

Per quanto riguarda la cardinalità, un insieme con cardinalità n vale l'altro. (Se $\#\Omega = n = \#\Omega'$ alloro esistono biezioni $f\colon \{1,\dots,n\} \to \Omega$ e $f'\colon \{1,\dots,n\} \to \Omega'$. Quindi $f' \circ f^{-1}$ è una biezione da Ω su Ω'.) Quindi, se ci interessano solo questioni di cardinalità e se Ω ha la cardinalità $\#\Omega = n$, allora possiamo supporre anche subito che $\Omega = \{1,\dots,n\}$.

A.1 Elementi del calcolo combinatorio

In parole povere, il ***calcolo combinatorio*** tratta il problema di determinare la cardinalità di insiemi finiti. I problemi, quasi sempre, si possono interpretare come trovare un numero di possibilità eseguire delle scelte seguendo certe regole. Per esempio il ***prodotto cartesiano*** (o semplicemente ***prodotto***) di due insiemi A e B è definito come

$$A \times B \ := \ \{(a,b)\colon\ a \in A, b \in B\}.$$

Evidentemente ad ogni scelta di un elemento a di A ed un elemento di b corrisponde esattamente una coppia (a,b). Allora il numero di elementi di $A \times B$ è il numero di possibilità di scegliere

un elemento di A ***ed*** un elemento di B. Più generalmente, gli elementi $(a_1, \dots, a_M)$ del ***prodotto cartesiano*** (o semplicemente ***prodotto***)

$$\bigtimes_{m=1}^{M} A_m \;=\; A_1 \times \ldots \times A_M \;:=\; \{(a_1, \dots, a_M)\colon a_1 \in A_1, \dots a_M \in A_M\}$$

corrispondono esattamente alle possibili scelte di esattamente un elemento a_m di ogni insieme A_m $(m = 1, \dots, M)$.

A.1.1 Proposizione. *Se A e B sono insiemi con cardinalità $\#A = n$ e $\#B = m$ allora la cardinalità del prodotto $A \times B$ è $\#(A \times B) = nm$.*

Intuitivamente, se organizziamo gli elementi di $A \times B$ in uno schema rettangolare (cioè, una tabella), l'affermazione è quasi immediata. Questo ragionamento, anche se molto utile per trovare la soluzione, non è ancora una dimostrazione a tutti gli effetti. Come minimo bisognerebbe, che indicassimo una funzione biettiva da $\{1, 2, \dots, nm-1, nm\}$ su $\{1, \dots, n\} \times \{1, \dots, m\}$. Tale funzione potrebbe essere definita come

$$f(k) \;:=\; (\ell(k), k - \ell(k)m)$$

dove $\ell(k)$ è tale che $\ell(k)m < k \leq \ell(k)m$. Non è proprio facilissimo (richiede basi della teoria della divisibilità), convincersi che tale funzione sia biettiva.

Prepariamo la dimostrazione con un lemma altrettanto chiaro all'intuito.

A.1.2 Lemma. *Siano X e Y insiemi finiti disgiunti. Allora vale $\#(X \cup Y) = \#X + \#Y$.*

Dimostrazione. Siano $f\colon \{1, \dots, n\} \to X$ e $g\colon \{1, \dots, m\} \to Y$ biezioni. Allora anche la funzione $h\colon \{1, \dots, n+m\} \to X \cup Y$ definita come

$$h(k) \;=\; \begin{cases} f(k) & k \leq n, \\ g(k-n) & k > n, \end{cases}$$

è un biezione. (**Esercizio:** Verificarlo!). ∎

Dimostrazione della Proposizione A.1.1. Procediamo con l'induzione per il numero n di elementi di A. Se A è vuoto, cioè se $n = 0$, allora anche $A \times B$ 'vuoto e, quindi, $nm = 0 = \#(A \times B)$. Per la conclusione $n \Rightarrow n+1$ supponiamo che l'affermazione valga per $\#A = n$. Sia $\widetilde{A}$ un insieme con $\#\widetilde{A} = n + 1$. Scegliamo un $a \in \widetilde{A}$ e poniamo $A := \widetilde{A} \backslash \{a\}$. Poiché A e $\{a\}$ sono disgiunti e $A \cup \{a\} = \widetilde{A}$, il lemma ci dice che $\#A + \#\{a\} = n + 1$, oppure $\#A = n$. Poi

$$\widetilde{A} \times B \;=\; (A \cup \{a\}) \times B \;=\; (A \times B) \cup (\{a\} \times B) \qquad \text{e} \qquad (A \times B) \cap (\{a\} \times B) \;=\; \emptyset.$$

(**Esercizio:** Verificarlo!) Dal lemma segue che

$$\#(\widetilde{A} \times B) \;=\; \#(A \times B) + \#(\{a\} \times B) \;=\; nm + m \;=\; (n+1)m.$$

Quindi, se l'affermazione vale per $n = 0$, e se vale per n, vale anche per $n+1$. Con questo concludiamo la dimostrazione per induzione. ■

A.1.3 Corollario. $\#(A_1 \times \ldots \times A_M) \;=\; \#A_1 \cdot \ldots \cdot \#A_M.$

Dimostrazione. Abbiamo

$$\begin{aligned}\#(A_1 \times \ldots \times A_M) \;&=\; \#\big((A_1 \times \ldots \times A_{M-1}) \times A_M\big) \;=\; \#(A_1 \times \ldots \times A_{M-1}) \cdot \#A_M \\ &=\; \#(A_1 \times \ldots \times A_{M-2}) \cdot \#A_{M-1} \cdot \#A_M \;=\; \quad \ldots \quad \;=\; \#A_1 \cdot \ldots \cdot \#A_M. \blacksquare\end{aligned}$$

A.1.4 Nota. Anche qui, formalmente ci servirebbe un'induzione. La dimostrazione per induzione l'abbiamo indicata con i punti "…". L'abbiamo fatto molte volte nelle dimostrazioni dei Capitoli 1–5. Ossia, ovunque appaiano i punti "…" per indicare una successione di equazioni (o conclusioni), il lettore, se vuole una dimostrazione completa, si dovrebbe sentire invitato a mettere in pratica una dimostrazione per induzione.

Il matematico pignolo potrebbe anche protestare perché non abbiamo dimostrato che la cardinalità dei due insiemi $A_1 \times \ldots \times A_M$ e $(A_1 \times \ldots \times A_{M-1}) \times A_M$ siano uguali. (Infatti, per definizione gli insiemi sono diversi. Però $(a_1, \ldots, a_{M-1}, a_M) \mapsto ((a_1, \ldots, a_{M-1}),\, a_M)$, ovviamente, definisce una biezione.)

D'ora in poi non ci occupiamo più troppo di questi aspetti formali. Dal punto di vista intuitivo i punti "…" sono più che sufficienti. Chi non crede la conclusione è sempre libero a riempire i buchi nelle dimostrazioni.

Torniamo ai prodotti cartesiani. Finora gli insiemi A e B (o $A_1, \ldots, A_M$) erano diversi. La coppia (a, b) non può essere confusa in nessun modo con la coppia (b, a) (che fa parte dell'insieme $B \times A$). Ci dovremo ricordare di questo fatto, quando ci capita il caso $B = A$. Gli elementi (a_1, a_2) del prodotto cartesiano $A \times A$, anche se a_1 e a_2 tutti e due sono dell'insieme A, si "ricordano" da quale fattore sono venuti: Le coppie (a_1, a_2) sono ***ordinate*** e $(a_1, a_2) = (a_2, a_1)$ vale se e solo se $a_1 = a_2$. Più generale, gli elementi $(a_1, \ldots, a_M)$ e $(b_1, \ldots, b_M)$ di $A^M := A \times \ldots \times A$ (M fattori) sono uguali se e solo se $a_m = b_m$ per ogni $m = 1, \ldots, M$.

A.1.5 Corollario. $\#(A^k) \;=\; \#(A^{k-1})\,\#A \;=\; \ldots \;=\; (\#A)^k.$

Possiamo interpretare A^k anche come l'insieme $\{f\colon \{1,\dots,k\} \to A\}$ di tutte le funzioni f da $\{1,\dots,k\}$ in A. Infatti, ad ogni k–upla $(a_1,\dots,a_k)$ corrisponde in modo biunivoco la funzione $f\colon \ell \mapsto a_\ell$. Per la cardinalità, la natura dell'insieme $\{1,\dots,k\}$ con centra. Lo possiamo sostituire con un insieme arbitrario Ω di cardinalità $\#\Omega = k$. Quindi,

$$\#\{f\colon \Omega \to A\} \;=\; \#A^{\#\Omega}.$$

Un caso particolare è $A = \{0,1\}$, cosicché $\#A = 2$. Le funzioni da Ω in $\{0,1\}$ sono in corrispondenza biunivoca con i sottoinsiemi B di Ω come le loro funzioni indicatrici; si veda l'Esercizio 1.1.16. Allora $\mathcal{P}(\Omega) = \{f\colon \Omega \to \{0,1\}\}$ e, conseguentemente,

$$\#\mathcal{P}(\Omega) \;=\; 2^{\#\Omega}.$$

Scambiamo un'ultima volta l'interpretazione della cardinalità $\#(A^k) = (\#A)^k$ di A^k, dicendo è il numero di possibilità per scegliere k volte dall'insieme $A = \{1,\dots,n\}$ con n elementi un elemento, dove gli elementi a_ℓ si ricordano del numero ℓ del tentativo e in ogni tentativo è disponibile l'intero insieme A.

A.1.6 Corollario. *Il numero di possibilità di scegliere k volte da n oggetti, prendendo nota dell'ordine con ripetizioni, è n^k.*

La formulazione un po' torturata c'indica che per il modo nel quale scegliere k volte sempre dallo stesso insieme ci sono parecchie possibilità. La situazione del corollario descrive la situazione dove prendiamo in considerazione l'ordine in cui i risultati escono uno dopo l'altro e dove, dopo una scelta, prima della prossima l'elemento scelto va reinserito nell'insieme (l'urna). Un risultato si può ripetere. Per questo, la situazione si descrive con "estrazione k di n $(= \#A)$ con ripetizioni". Ovviamente, ci siano anche estrazioni non ordinate e senza o con ripetizioni dei quali ci occupiamo adesso.

Siano $0 \le k \le n$ numeri in $\mathbb{N}_0$. Annotiamo con D_k^n l'insieme di tutte le funzioni iniettive da $\{1,\dots,k\}$ in $\{1,\dots,n\}$ (le ***disposizioni***). In altre parole, scegliamo k volte da n individui ricordandoci dell'ordine delle estrazioni ($f \in D_k^n$ associa ad ogni $\ell \in \{1,\dots,k\}$ l'elemento $f(\ell)$ estratto nell'ℓ–esimo tentativo) ma senza ripetizioni (altrimenti f non sarebbe iniettiva).

Ricordiamoci che $n! := n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1$ per $n \in \mathbb{N}$ e $0! = 1$. (Definizione ricorsiva: $0! = 1$ e $(n+1)! = (n+1)n!$; si veda [Ske04] per definizioni ricorsive.)

A.1.7 Proposizione. $$\#D_k^n = \frac{n!}{(n-k)!} =: n(n-1)\dots(n-k+1).$$

Dimostrazione. Intuitivamente abbiamo

$$\begin{aligned} \#D_k^n \;=\;& n \text{ possibilità di scegliere il primo elemento} \\ \times\;& (n-1) \text{ possibilità di scegliere il secondo elemento} \\ \times\;& \quad \dots \quad \times \\ \times\;& (n-k+1) \text{ possibilità di scegliere il } k\text{–esimo elemento.} \end{aligned}$$

Adesso formalizziamo questo intuito. Per prima cosa il numero di iniezioni da un insieme A di cardinalità $\#A = k$ in un insieme B di cardinalità $\#B = n$ è sempre $\#D_k^n$. (La natura concreta degli insiemi $\{1,\dots,k\}$ e $\{1,\dots,n\}$ non centra proprio per questo conteggio.) La formula è giusta per $n = 0$. (In questo anche $k = 0$. E ponendo in modo giusto la definizione del termine "funzione", c'è una sola funzione che associa ad ogni elemento dell'insieme vuoto un elemento dell'insieme vuoto, e questa funzione è iniettiva.) Supponiamo che la formula sia giusta per n (e, quindi, per ogni $k \in \{0,\dots,n\}$). Se $k = 0$, allora esiste una sola funzione da $\emptyset$ in $\{1,\dots,n+1\}$ e questa funzione è iniettiva, quindi $D_k^{n+1} = 1 = \frac{(n+1)!}{(n+1-0)!}$. Se $k \geq 1$ possiamo decomporre D_k^{n+1} come

$$D_k^{n+1} \;=\; \bigcup_{\ell=1}^{n+1} \{f \in D_k^{n+1} \colon f(1) = \ell\}$$

come unione di insiemi due a due disgiunti. La cardinalità di ogni $\{f \in D_k^{n+1} \colon f(k) = \ell\}$ è sempre $\#D_{k-1}^n$. (Si tratta del problema di determinare il numero di iniezioni da $\{1,\dots,k-1\}$ in $\{1,\dots,n+1\}\backslash\{\ell\}$.) Allora

$$\begin{aligned} \#D_k^{n+1} \;=\; \sum_{\ell=1}^{n+1} \#\{f \in D_k^{n+1} \colon f(1) = \ell\} \;=\; & \sum_{\ell=1}^{n+1} \#D_{k-1}^n \\ = \;& (n+1)\frac{n!}{(n-(k-1)+1)!} \;=\; \frac{(n+1)!}{((n+1)-k+1)!}. \end{aligned}$$

Quindi, la formula vale anche per $n+1$. ∎

A.1.8 Nota. In molti rispetti è preferibile usare la formula $n(n-1)\dots(n-k+1)$. Per prima cosa, è questa formula che esprime in modo immediato la deduzione che abbiamo fatto. Poi, soprattutto se n è grande e k piccolo, è meno complessa da calcolare. (Sono solo $k-1$ prodotti da calcolare, mentre la formula $\frac{n!}{(n-k)!}$ ne richiede il $2n-k-1$ più una divisione.) Un vantaggio della formula con i fattoriali è, che sono più chiaramente definite se uno degli argomenti è 0.

A.1.9 Corollario. *Il numero dell'insieme S_n delle **permutazioni** (cioè delle biezioni) dell'insieme $\{1,\dots,n\}$ (o di un qualsiasi insieme di cardinalità n) è $D_n^n = \frac{n!}{0!} = n!$.*

Siano sempre $0 \le k \le n$ numeri in $\mathbb{N}_0$. Annotiamo con C_k^n l'insieme di tutti i sottoinsiemi S di $\{1, \dots, n\}$ che hanno la cardinalità $\#S = k$ (le ***combinazioni***). In altre parole, scegliamo da n individui k diversi (cioè senza ripetizioni) ma senza guardare l'ordine.

A.1.10 Proposizione. $\#C_k^n = \dfrac{\#D_k^n}{k!} = \dfrac{n!}{k!(n-k)!} =: \binom{n}{k}$ *(**coefficiente binomiale**).*

A.1.11 Nota. Anche qui, soprattutto quando k è piccolo, per i calcoli concreti è sempre utile ricordarsi della formula

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{1 \cdot 2 \dots \cdot k}.$$

Notiamo che i coefficienti binomiali sono possiedono una certa simmetria:

$$\binom{n}{k} = \binom{n}{n-k}$$

Conviene anche ricordarsi sempre almeno di alcuni valori particolari:

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = \binom{n}{n-1} = n, \qquad \binom{n}{2} = \binom{n}{n-2} = \frac{n(n-1)}{2}.$$

Dimostrazione della proposizione. Intuitivamente, i k risultati li possiamo immaginare sempre estratti uno dopo l'altro anche se alla fine ci dimentichiamo dell'ordine. In altre parole ogni composizione si può considerare il risultato di una disposizione, dove però invece della k–upla $(f(1), \dots, f(k))$ ci interessiamo solo dell'***immagine*** $f(\{1, \dots, k\}) := \{f(1), \dots, f(k)\}$ di $\{1, \dots, k\}$ sotto f. (Poiché f è iniettiva, la cardinalità di quest'immagine è proprio k.) In altre parole, della disposizione f ci interessa solo la sua ***classe d'equivalenza*** $[f]_\sim := \{g \in D_k^n : g \sim f\}$ sotto la ***relazione d'equivalenza*** definita come $f \sim g \Leftrightarrow f(\{1, \dots, k\}) = g(\{1, \dots, k\})$; si veda [Ske04] per relazioni e classi d'equivalenza. Quindi, C_k^n ha la stessa cardinalità dell'insieme quoziente $D_k^n/\sim$ delle disposizioni rispetto alla relazione $\sim$. Il numero di elementi in una classe d'equivalenza $[f]_\sim$ è sempre $\#[f]_\sim = \#S_k = k!$, il numero di permutazioni degli risultati $f(1), \dots, f(k)$ ottenuti. Quindi, $\#C_k^n = \frac{\#D_k^n}{k!}$. Però, non è facile formalizzare quest'intuito. Ma una volta capita la formula da dimostrare, possiamo procedere per induzione. (Si veda la nota sotto.)

Per $n = 0$ siamo apposto. La formula valga, allora, per n. Se $k = n + 1$, allora c'è un solo sottoinsieme di $\{1, \dots, n+1\}$ con $n + 1$ elementi. Quindi $\#C_{n+1}^{n+1} = 1 = \binom{n+1}{n+1}$. Allora supponiamo $k < n + 1$, ossia $k \le n$. Decomponiamo C_k^{n+1} in

$$C_k^{n+1} = \{S \in C_k^{n+1} : S \ni n+1\} \cup \{S \in C_k^{n+1} : S \not\ni n+1\}$$

Per le cardinalità troviamo (perché?)

$$\#C_k^{n+1} = \#\{S \in C_k^{n+1} : S \ni n+1\} + \#\{S \in C_k^{n+1} : S \not\ni n+1\} = \#C_{k-1}^n + \#C_k^n,$$

e usando l'ipotesi per n arriviamo a

$$\#C_k^{n+1} = \#C_{k-1}^n + \#C_k^n = \frac{n!}{(k-1)!(n-(k-1)+1)!} + \frac{n!}{k!(n-k+1)!}$$
$$= n!\frac{k+n-k+1}{(k!((n+1)-k+1)!} = \binom{n+1}{k}.$$

Quindi, la formula vale anche per $n+1$. ■

A.1.12 Nota. Dimostrazioni per induzione richiedono di conoscere già il risultato da dimostrare. (Va usato nella conclusione $n \Rightarrow n+1$.) Per questo sono preferibili dimostrazioni *costruttive* dove il risultato va *costruito* durante la dimostrazione, cioè dove per proseguire con la dimostrazione non è necessario di conoscere il risultato sin dall'inizio. Spesso succede, come in questa dimostrazione, che un ragionamento ci permette di indovinare la soluzione, ma il ragionamento è difficile da trasformare in una dimostrazione rigorosa. In questi casi, come adesso, una volta individuato il risultato si può procedere per induzione.

Durante la dimostrazione abbiamo anche dimostrato la seguente formula ricorsiva (esprime i coefficienti per $n+1$ in termini dei coefficienti per n) per i coefficienti binomiali.

A.1.13 Corollario. $$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}.$$

La formula può essere illustrata graficamente nel cosiddetto ***triangolo di Pascal***:

$n=0$	$\binom{0}{0}$	1
$n=1$	$\binom{1}{0}\ \binom{1}{1}$	1 1
$n=2$	$\binom{2}{0}\ \binom{2}{1}\ \binom{2}{2}$	1 2 1
$n=3$	$\binom{3}{0}\ \binom{3}{1}\ \binom{3}{2}\ \binom{3}{3}$	1 3 3 1
$n=4$	$\binom{4}{0}\ \binom{4}{1}\ \binom{4}{2}\ \binom{4}{3}\ \binom{4}{4}$	1 4 6 4 1
$n=5$	$\binom{5}{0}\ \binom{5}{1}\ \binom{5}{2}\ \binom{5}{3}\ \binom{5}{4}\ \binom{5}{5}$	1 5 10 10 5 1
...	... ...	... ...

Aggiungendo tacitamente i coefficienti $\binom{n}{-1} = \binom{n}{n+1} = 0$ alla sinistra rispettivamente alla destra di ogni riga, ogni coefficiente è la somma degli due coefficienti più vicini della riga precedente.

Il Corollario fa anche apparizione naturale nella dimostrazione del seguente teorema, una delle formule più importanti, con tante applicazioni in quasi tutte le parti della matematica.

A.1.14 Formula del binomio. *Per ogni coppia x e y di elementi di un campo (o un anello commutativo unitario) e per ogni $n \in \mathbb{N}_0$ vale*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}. \tag{A.1.1}$$

DIMOSTRAZIONE. Infatti, calcolando i primi esempi non banali, la formula binomiale $(x+y)^2 = y^2 + 2xy + y^2$, e $(x+y)^3 = y^3 + 3xy^2 + 3x^2y + x^3$ ci potremmo insospettire della validità di una tale formula generale. (Sicuramente, storicamente è andato così.) Una volta avuto indovinata la formula, bisogna dimostrarla per induzione.

Abbiamo $(x+y)^0 = 1 = \sum_{k=0}^{0} \binom{0}{k} x^k y^{0-k}$. Quindi, la formula vale per $n = 0$.

Allora ipotetizziamo che la formula valga per un certo n. Dobbiamo concludere, usando quest'ipotesi, che la formula vale anche per $n+1$. Calcoliamo

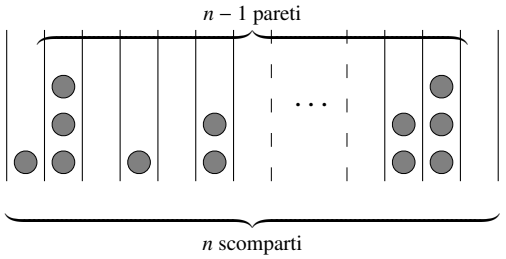
$$\begin{aligned}
(x+y)^{n+1} &\overset{(a)}{=} (x+y)(x+y)^n \overset{(b)}{=} (x+y)\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \\
&= x\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} + y\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}x^{k+1} y^{n-k} + \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k+1} \\
&\overset{(c)}{=} \sum_{k=1}^{n+1}\binom{n}{k-1}x^k y^{n-k+1} + \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k+1} \\
&= 1\cdot x^{n+1}y^0 + \sum_{k=1}^{n}\binom{n}{k-1}x^k y^{n-k+1} + 1\cdot x^0 y^{n+1} + \sum_{k=1}^{n}\binom{n}{k}x^k y^{n-k+1} \\
&= x^{n+1} + y^{n+1} + \sum_{k=1}^{n}\left(\binom{n}{k-1} + \binom{n}{k}\right)x^k y^{n-k+1} \\
&\overset{(d)}{=} x^{n+1} + y^{n+1} + \sum_{k=1}^{n}\binom{n+1}{k}x^k y^{n-k+1} = \sum_{k=0}^{n+1}\binom{n+1}{k}x^k y^{(n+1)-k}. \blacksquare
\end{aligned}$$

A.1.15 Nota matematica.

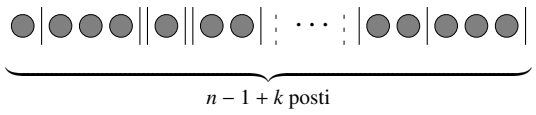
I singoli passi per la conclusione $n \Rightarrow n+1$ della dimostrazione precedente si motivano come segue. Bisogna pure iniziare da una parte dell'Equazione (A.1.1) per il caso $n+1$ per poi, utilizzando l'ipotesi per n, trasformarla nell'altra parte. Normalmente è opportuno iniziare con quella più semplice, in nostro caso con $(x+y)^{n+1}$. Poi bisogna scrivere questa parte in tal modo che faccia apparizione una parte dell'ipotesi, cioè dell'Equazione (A.1.1) per n. Questo lo facciamo nel Passo (a). Poi nel Passo (b) utilizziamo l'ipotesi. I passi seguenti fino a (c) sono normali elaborazioni del risultato dopo (b). Ad un certo punto bisogna non dimenticare dove vogliamo arrivare, cioè a $\sum_{k=0}^{n+1}\binom{n+1}{k}x^k y^{(n+1)-k}$, l'altra parte della (A.1.1) per $n+1$. Lo *shift* dell'indice della prima sommatoria nel Passo (c) è proprio diretto a questo scopo. Infine, nel Passo (d) stiamo proprio usando anche il Corollario A.1.13.

Fine della nota.

Adesso ci manca ancora il problema di scegliere k volte sempre da n individui (cioè con ripetizioni) ma senza guardare l'ordine del risultato. Questo problema equivale la distribuzione di k palline (tutte uguali) su n scomparti.



Ossia, il problema di distribuire le k palline fra le $n-1$ che risulta ad un totale di $n+k-1$ posti (da occupare o da una pallina o da una parete).



Quindi, il numero di tali scelte è

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

Il tentativo di ridurre il problema al risultato noto delle estrazioni k volte di n (cioè con ripetizioni) con riguardo dell'ordine con il numero n^k di possibilità, risulta difficile. Occorrerebbe calcolare quante k–uple $(n_1, \ldots, n_k)$ realizzino lo stesso risultato senza guardare l'ordine, e questo è decisamente più complicato. Soprattutto questo numero dipende dalla k–upla specifica. Se la k–upla contiene ℓ numeri diversi, diciamo $w_1, \ldots, w_\ell$, e se sono $h_{w_1}, \ldots, h_{w_\ell}$ le frequenze relative di questi risultati (i numeri di apparizioni nella k–upla), allora è noto che il numero delle k–uple che realizzino gli stessi dati $w_1, \ldots, w_\ell$ e $h_{w_1}, \ldots, h_{w_\ell}$ è il ***coefficiente multinomiale***

$$\binom{h_{w_1} + \ldots + h_{w_\ell}}{h_{w_1}\,,\ \ldots\ ,\ h_{w_\ell}} := \frac{n!}{h_{w_1}! \ldots h_{w_\ell}!}, \qquad \text{(dove notiamo, che } h_{w_1} + \ldots + h_{w_\ell} = n\text{)}.$$

Raccogliamo il numero delle possibilità per i diversi tipi di scelte di k di/da n in una tabella.

k di/da n	con ordine	senza ordine
con ripetezioni	n^k	$\binom{n+k-1}{n} = \binom{n+k-1}{k-1}$
senza ripetizioni	$\frac{n!}{(n-k)!} = n(n-1)\ldots(n-k+1)$	$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k(k-1)\ldots 1}$

A.2 Richiami ai limiti di numeri reali

Sia $R \subset \mathbb{R}$ un sottoinsieme dei numeri reali. Sia $a_1, a_2, \ldots$ una successione di numeri di R. Diciamo la successione ***converge*** in R verso il limite $a \in R$, in formula

$$\lim_{n\to\infty} a_n \;=\; a,$$

se per ogni $\varepsilon > 0$ esiste un N_ε (ovviamente, dipendente da ε), tale che

$$|a_n - a| \;\leq\; \varepsilon \qquad \text{per tutti gli} \qquad n \;\geq\; N_\varepsilon.$$

Tale limite è unico. (Infatti, supponiamo che $\lim_{n\to\infty} = a' \in R$. Se $a \neq a'$, allora $|a - a'| > 0$. Poniamo $\varepsilon = \frac{|a-a'|}{3}$ e scegliamo un N_ε. Allora per $n \geq N_\varepsilon$ varrebbe

$$|a - a'| \;=\; |(a - a_n) + (a_n - a')| \;\leq\; |a - a_n| + |a_n - a'| \;\leq\; \frac{2}{3}\,|a - a'|\,.$$

L'unica soluzione di questa disuguaglianza per il numero nonnegativo $|a - a'|$ è 0, contraddicendo l'ipotesi $a \neq a'$.)

Una successione $a_1, a_2, \ldots$ è ***limitata***, se esiste una costante M, tale che $|a_n| \leq M$ per ogni $n \in \mathbb{N}$. Ogni successione convergente è limitata. (Infatti, se scegliamo un N_ε opportuno, allora

$$|a_n| \;\leq\; |a_n - a| + |a| \;\leq\; \varepsilon + |a|$$

per ogni $n \geq N_\varepsilon$. Quindi, $M := \max\{\varepsilon + |a|\,, a_1, \ldots, a_{N_\varepsilon - 1}\}$ è una costante utile.)

Non ripetiamo tutte le regole per il calcolo con i limit. Menzioniamo solo che somma $a_n + b_n$ e prodotto $a_n b_n$ di successioni convergenti convergono verso la somma $a + b$ e il prodotto ab dei rispettivi limiti $a = \lim_{n\to\infty} a_n$ e $b = \lim_{n\to\infty} b_n$. Anche $a_n \geq b_n \Rightarrow a \geq b$ è un fatto rilevante.

Una successione convergente soddisfa il ***criterio di Cauchy***, cioè per ogni $\varepsilon > 0$ esiste un M_ε, tale che

$$|a_n - a_m| \;\leq\; \varepsilon \qquad \text{per tutti gli} \qquad n \;\geq\; M_\varepsilon.$$

(Infatti, $|a_n - a_m| \leq |a_n - a| + |a - a_m|$. Quindi, un M_ε opportuno è $M_\varepsilon = N_{\frac{\varepsilon}{2}}$.)

Una successione che soddisfa il criterio di Cauchy la chiameremo ***successione di Cauchy***. Il campo $\mathbb{R}$ può essere caratterizzato nel modo seguente: $\mathbb{R}$ è l'unica estensione (come campo) del campo dei numeri razionali $\mathbb{Q}$, che soddisfa:

- $\mathbb{R}$ è ***completo***, cioè ogni successione di Cauchy ha un limite in $\mathbb{R}$.
- $\mathbb{Q}$ è ***denso*** in $\mathbb{R}$, cioè ogni numero di $\mathbb{R}$ è il limite di una successione di numeri razionali.

In $\mathbb{R}$ il criterio di Cauchy equivale, quindi, la convergenza. Il vantaggio enorme del criterio di Cauchy consiste nel fatto, che per applicarlo non occorre conoscere il limite della successione. Il seguente esempio illustra la sua potenza.

A.2.1 Esempio. Una successione $a_1, a_2, \ldots$ è ***monotona***, se vale $a_{n+1} \geq a_n$ per ogni $n \in \mathbb{N}$ o se vale $a_{n+1} \leq a_n$ per ogni $n \in \mathbb{N}$. Sappiamo già che ogni successione convergente è limitata. Dimostriamo, usando il criterio di Cauchy, che una successione monotona e limitata è convergente.

Possiamo supporre che la successione $a_1, a_2, \ldots$ sia crescente. (Altrimenti, passiamo alla successione $-a_1, -a_2, \ldots$.) Supponiamo $a_1, a_2, \ldots$ non sia convergente. Allora non è una successione di Cauchy. Conseguentemente, esiste un $\varepsilon > 0$ e per ogni N esistono $n, m \geq N$ con $a_n - a_m > \varepsilon$. Ne segue (per definizione ricorsiva!) l'esistenza di numeri naturali

$$m_1 < n_1 \leq m_2 < n_2 \leq m_3 < n_3 \leq \ldots$$

tali che $a_{n_k} - a_{m_k} > \varepsilon$ per ogni $k \in \mathbb{N}$. Allora la sottosuccessione

$$\begin{aligned} a_{n_k} &= (a_{n_k} - a_{m_k}) + (a_{m_k} - a_{n_{k-1}}) + (a_{n_{k-1}} - a_{m_{k-1}}) + (a_{m_{k-1}} - a_{n_{k-2}}) + \ldots + (a_{n_1} - a_{m_1}) + a_{m_1} \\ &\leq \varepsilon + 0 + \varepsilon + 0 + \ldots + \varepsilon + a_{m_1} = k\varepsilon + a_{m_1} \qquad (k = 1, 2, \ldots) \end{aligned}$$

non è limitata. *A fortiori* non lo è l'intera successione.

Chiaramente se $a_n \leq M$, allora vale anche $a \leq M$. Ne segue che la costante più piccola che limita una successione crescente è proprio il suo limite.

A.2.2 Corollario. *Sia $a_1, a_2, \ldots$ una successione di numeri positivi. Allora*

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k$$

esiste, se e solo se esiste una costante M, tale che $\sum_{k=1}^{n} a_k \leq M$ per ogni $n \in \mathbb{N}$. Nel caso affermativo, la costante più piccola è il limite della serie.

Dimostrazione. La successione $b_n := \sum_{k=1}^{n} a_k$ è monotona. ∎

A.3 Serie assolutamente convergenti

Piano piano, siamo passati dalle successioni alle ***serie***, cioè alle successioni i cui elementi hanno la forma $b_n := \sum_{k=1}^{n} a_k$, dove a_n sono gli elementi di un'altra successione. Diciamo la serie

$\sum_{n=1}^{\infty} a_n$ è ***convergente*** al limite a se vale $\lim_{n\to\infty} \sum_{k=1}^{n} a_k = a$ e scriviamo

$$\sum_{n=1}^{\infty} a_n \;=\; a.$$

Se il limite non esiste, diciamo la serie $\sum_{n=1}^{\infty} a_n$ è ***divergente***. Diciamo una serie $\sum_{n=1}^{\infty} a_n$ è ***assolutamente convergente***, se la serie $\sum_{n=1}^{\infty} |a_n|$ è convergente. Una serie che è convergente ma non assolutamente convergente, è ***condizionalmente(?) convergente***.

Se $\sum_{n=1}^{\infty} a_n$ converge, allora $\lim_{n\to\infty} a_n = 0$. (Infatti, secondo il criterio di Cauchy esiste per ogni $\varepsilon > 0$ un N_ε tale che $\left|\sum_{k=m+1}^{n} a_k\right| \le \varepsilon$ per ogni $n > m \ge N_\varepsilon$. In particolare, per $n = m + 1$ troviamo $|a_n - 0| = |a_n| = \left|\sum_{k=m+1}^{n} a_k\right| \le \varepsilon$ per ogni $n > N_\varepsilon$.)

A.3.1 Il grande teorema sugli riordinamenti. *Siano $a_1, a_2, \ldots$ e $b_1, b_2, \ldots$ successioni tali che $|a_n| \le b_n$ e che $\sum_{n=1}^{\infty} b_n$ converge. (In particolare, con $b_n = |a_n|$ sia $\sum_{n=1}^{\infty} a_n$ assolutamente convergente.) Allora $\sum_{n=1}^{\infty} a_n$ converge, e il limite è indipendente da un qualsiasi **riordinamento** della successione $a_1, a_2, \ldots$. Con ciò intendiamo che*

$$\sum_{n=1}^{\infty} a_n \;=\; \sum_{n=1}^{\infty} a_{f(n)}$$

per ogni biezione $f\colon \mathbb{N} \to \mathbb{N}$.

Dimostrazione. Presentiamo una dimostrazione "classica" come si trova anche in testi di un primo semestre di analisi. Vorremmo, però, far presente che il teorema è incluso nel Lemma 1.3.8. Speriamo che l'eleganza e semplice naturalezza (paragonata alla dimostrazione che segue) del trattamento del Lemma 1.3.8 possa convincere il lettore, che le sommatorie $\sum_{\omega\in\Omega} a_\omega$ sono un mezzo più adeguato per lavorare efficacemente.

Per ogni $\varepsilon > 0$ esiste un N_ε, tale che

$$\Big|\sum_{k=m+1}^{n} a_k\Big| \;\le\; \sum_{k=m+1}^{n} |a_k| \;\le\; \sum_{k=m+1}^{n} b_k \;\le\; \varepsilon$$

per ogni $n > m \ge N_\varepsilon$. Allora $\sum_{n=1}^{\infty} a_n$ converge, diciamo ad un limite a.

Per ogni n definiamo con $\nu(n) := \max\{f^{-1}(1), \ldots, f^{-1}(n)\}$ il numero minimo, tale che

$$f(\{1, \ldots, \nu(n)\}) \;\supset\; \{1, \ldots, n\}.$$

Allora la sommatoria $\sum_{k=1}^{\nu(n)} a_{f(k)}$ contiene, fra l'altro, anche gli addendi $a_1, \ldots, a_n$. Se definiamo con $N(m) =: \max\{f(1), \ldots, f(m)\}$ il numero minimo, tale che

$$\{1, \ldots, N(m)\} \;\supset\; \{f(1), \ldots, f(m)\},$$

allora la sommatoria $\sum_{k=1}^{N(m)} a_k$ contiene, fra l'altro, anche gli addendi $a_{f(1)}, \dots, a_{f(m)}$. Ne segue per ogni $m \geq \nu(n)$ che

$$\left|\sum_{k=1}^{m} a_{f(k)} - \sum_{k=1}^{n} a_k\right| \;\leq\; \sum_{k=n+1}^{N(m)} a_k.$$

Per $n \geq N_{\frac{\varepsilon}{2}}$, questo termine è limitato da $\frac{\varepsilon}{2}$. Inoltre, esiste $M_{\frac{\varepsilon}{2}}$ tale che $\left|\sum_{k=1}^{n} a_k - a\right| \leq \frac{\varepsilon}{2}$ per ogni $n \geq M_{\frac{\varepsilon}{2}}$. Poniamo $K_\varepsilon := \max\{N_{\frac{\varepsilon}{2}}, M_{\frac{\varepsilon}{2}}\}$. Allora

$$\left|\sum_{k=1}^{m} a_{f(k)} - a\right| \;\leq\; \left|\sum_{k=1}^{m} a_{f(k)} - \sum_{k=1}^{K_\varepsilon} a_k\right| + \left|\sum_{k=1}^{K_\varepsilon} a_k - a\right| \;\leq\; \varepsilon$$

per ogni $m \geq \nu(K_\varepsilon)$. ∎

A.3.2 Nota. Il ***piccolo teorema sugli riordinamenti*** dice il seguente: Sia $\sum_{n=0}^{\infty} a_n$ una serie convergente ma non assolutamente convergente. Allora per ogni $a \in \mathbb{R}$ esiste una biezione $f_a \colon \mathbb{N} \to \mathbb{N}$, tale che

$$\sum_{n=1}^{\infty} a_{f_a(n)} \;=\; a.$$

In un senso opportuno questo vale anche per i valori $a = \pm\infty$.

Non ci occupiamo di una dimostrazione. (Basta definire un riordinamento secondo la regola: Se la somma parziale $\sum_{k=1}^{n} a_{\nu(k)}$ è più piccola di a, allora scegliere $\nu(n+1)$ is più piccolo numero non ancora scelto tale che $a_{\nu(n+1)} \geq 0$; altrimenti scegliere il numero tale che $a_{\nu(n+1)} < 0$. Poi bisogna convincersi (usando in modo essenziale l'ipotesi che la serie non converga assolutamente!) che questo definisca veramente un riordinamento, e (usando in modo essenziale l'ipotesi che la serie converga!) che il limite del riordinamento sia proprio a.) Ma la versione piccola del teorema ci dice chiaramente che la convergenza assoluta delle serie per calcolare l'attesa (cioè, l'esistenza di $\mathbb{E}\,|X|$) è un'ipotesi indispensabile se vogliamo che l'attesa non dipenda dal modo in cui abbiamo enumerato gli elementi dello spazio campionario.

Il teorema A.3.1 ci conferma che la nostra definizione delle sommatorie, che abbiamo usato nel testo, è consistente. Ricordiamoci che per un insieme finito R abbiamo definito

$$\sum_{r \in R} a_r \;:=\; \sum_{n=1}^{\#R} a_{f(n)},$$

dove $f \colon \{1, \dots, \#R\} \to R$ è una qualsiasi biezione. Teniamo per scontato, che tale definizione non dipenda dalla scelta di f. (**Esercizio:** Dimostrarlo con l'induzione. Attenzione, è difficile, ma solo per il problema di sviluppare la notazione giusta. Quindi non disperarsi troppo.) Notiamo che abbiamo usato questo fatto anche all'interno della dimostrazione del Teorema A.3.1. (Domanda: Dove l'abbiamo usato?)

A.4 Le funzioni monotone

A.5 Integrazione m–dimensionale

Appendice B

Algebra booleana astratta

La nostra lingua ci permette di fare affermazioni. Normalmente una tale affermazione può essere o *vero* o *falso*. È da escludere che sia vero e falso allo stesso tempo, oppure che non sia né vero né falso, per esempio perché non è sensata. Non facciamo nessuno sforzo di mettere su una base solida questi termini, che fanno parte della logica matematica. Chiameremo comunque una *proposizione* un'affermazione p che ha un ben definito *valore di verità* $V(p)$: o $V(p) = v$ (vero) o $V(p) = f$ (falso). Data una proposizione p possiamo definire un'altra proposizione $\neg p$, che ha il valore di verità $V(\neg p) = v$, se e solo se $V(p) = f$. Date due proposizioni p e q possiamo definire due altre proposizioni $p \wedge q$ e $p \vee q$, dove $V(p \wedge q) = v$ se e solo $V(p) = V(q) = v$ (logico "e"), e dove $V(p \vee q) = v$ se e solo $V(p) = v$ oppure $V(q) = v$ (logico "o").

In fondo, le proposizioni non sono nient'altro che indeterminate (o variabili) dell'insieme $\{v, f\}$. Le regole che le tre operazioni $\wedge, \vee, \neg$ soddisfano possiamo riconoscere come quelli di un'algebra booleana astratta con due elementi come definita nella Sezione B.1. Ogni espressione in termini di quest'algebra booleana può essere calcolata usando una *tabella della verità*. Se l'espressione coinvolge n indeterminate, allora la tabella ha 2^n righe, una per ogni scelta dei n valori delle n indeterminate in $\{v, f\}$. Questo significa che controllare equazioni fra tali espressioni è un lavoro, a volte fastidioso, ma sempre limitato.

Esercizio.

(a) Verificare con una tabella che $(p \wedge \neg q) \vee (q \wedge \neg p) = (p \vee p) \wedge \neg(p \wedge q)$. Tale operazione si chiama l'*o esclusivo* (o p o q ma non p e q) e va annotata come $p \veebar q$.

(b) Dimostrare che $(\{v, f\}, \veebar, \wedge)$ è un campo. (Suggerimento: A meno di isomorfismo esiste un solo campo con due elementi. Questi devono essere 0, l'elemento neutro dell'addizione, e $1 \neq 0$. Basta, quindi, verificare che le regole di calcolo per f, v sono quelli del campo $\{0, 1\} = \mathbb{Z}_2 = \mathbb{Z}/(2\mathbb{Z})$, perché di $\{0, 1\}$ sappiamo, nell'un modo o nell'altro, che soddisfa tutti gli assiomi del campo.)

La forza delle operazioni logiche consiste proprio nel fatto che sono così semplice. Qualunque siano le altre strutture a cui vanno sovrapposte operazioni logiche, alla fine il problema (o almeno la parte del problema che riguarda la logica) si riduce sempre a calcoli nel campo $\mathbb{Z}_2 = \{0, 1\} = \{f, v\}$. Un esempio sono le *espressioni logiche*. Un'espressione è una funzione $A\colon \Omega \to \{f, v\}$ definita su uno spazio campionario (cioè in insieme non vuoto) Ω che assegna ad ogni elemento $\omega \in \Omega$ un valore di verità $A(\omega) \in \{f, v\}$. Per esempio se A è un sottoinsieme di Ω, possiamo definire $A(\omega) = V(\omega \in A)$, cioè $A(\omega) = v$ se la proposizione $\omega \in A$ è vera e $A(\omega) = f$ se la proposizione $\omega \in A$ è falsa. In altre parole, se identifichiamo $\{f, v\}$ con $\{0, 1\}$, la funzione $A\colon \omega \mapsto A(\omega)$ non è nient'altro che la funzione indicatrice χ_A dell'evento A di Ω. Per espressioni A e B definiamo le espressioni

$$A \wedge B\colon \omega \longmapsto A(\omega) \wedge B(\omega), \qquad A \vee B\colon \omega \longmapsto A(\omega) \vee B(\omega), \qquad \neg A\colon \omega \longmapsto \neg A(\omega).$$

Esercizio. Ricordiamoci che nell'Esercizio 1.1.16 abbiamo stabilito una corrispondenza biunivoca fra funzioni a valori in $\{0, 1\}$, funzioni indicatrici χ_A ed eventi $A \subset \Omega$. Siano A, B eventi di Ω e annotiamo le espressioni corrispondenti come funzioni indicatrici $\chi_A\colon \omega \mapsto V(\omega \in A)$ e $\chi_B\colon \omega \mapsto V(\omega \in B)$. Verificare che

$$\chi_{A\cap B} = \chi_A \wedge \chi_B, \qquad \chi_{A\cup B} = \chi_A \vee \chi_B, \qquad \chi_{A^\complement} = \neg\chi_A.$$

Quindi le operazioni fra insiemi soddisfano le stesse leggi che soddisfano le espressioni o le proposizioni logiche. (La transizione $A \mapsto \chi_A$ dagli eventi di Ω alle funzioni indicatrici è quello che chiameremo un isomorfismo di algebre booleane.)

Sia $(R, +, \cdot)$ un anello. Ricordiamoci dall'algebra, che l'insieme delle funzioni

$$\mathcal{F}(\Omega, R) := \{f\colon \Omega \to R\}$$

da Ω in R con le operazioni $f + g\colon \omega \mapsto f(\omega) + g(\omega)$ e $f \cdot g\colon \omega \mapsto f(\omega) \cdot g(\omega)$ è un anello. (Verificarlo!) Quindi, le funzioni $\mathcal{F}(\Omega, \mathbb{Z}_2)$ da Ω nel campo $\mathbb{Z}_2 = \{0, 1\}$ formano un anello. Usare questo fatto per dimostrare che $(\mathcal{P}(\Omega), \triangle, \cap)$ è un anello. (In particolare, si vede che la differenza simmetrica $\triangle$ è un'operazione associativa. Veramente, $(\mathcal{P}(\Omega), \triangle, \cap)$ è un anello unitario, dove $A \cap A = A$, cioè dove tutti gli elementi sono idempotenti rispetto alla moltiplicazione. Si può dimostrare che ogni anello unitario con questa proprietà è un'algebra booleana. Inoltre, $\mathcal{F}(\Omega, \mathbb{Z}_2)$ è anche uno spazio vettoriale sul campo $\mathbb{Z}_2$, quindi, è un algebra su $\mathbb{Z}_2$. Questo spiega il nome *algebra* booleana.)

Come vediamo ci sono due strutture con operazioni che seguono le stesse regole. Vediamo anche sono intimamente correlate l'una con l'altra. Anche nell'interpretazione degli eventi come domande binarie (cioè, con le risposte possibili "si" o "no") abbiamo notato la stessa struttura (infatti quasi identica a quella delle espressioni logiche). Solo poi, mettendo queste domande in relazione con i risultati ω di uno spazio campionario Ω siamo arrivati ad algebre booleane di eventi di Ω. Ma non ci siamo mai stancati di mettere l'accento sul fatto che la scelta di Ω per la modellazione dell'esperimento che permette di rispondere a certe domande non è unica.

Nel seguito ci poniamo due problemi. Il primo: Dare una definizione assiomatica della struttura in questione. Come sempre questo include l'impegno di dare un numero di assiomi più piccolo possibile e di identificare il numero massimo delle proprietà come conseguenze degli assiomi. Il vantaggio è che per identificare la struttura, il numero di proprietà da controllare è minimo, e tutte le proprietà dimostrate in questo contesto generale valgono automaticamente in ogni realizzazione, senza che occorra di darne un'altra dimostrazione. Lo faremo nella Sezione B.1. Il secondo problema: Individuare quale sia la realizzazione "generica" della struttura astratta. Infatti, nelle Sezioni B.3 e B.4 dimostreremo il teorema di Stone. Tale teorema afferma che ogni algebra booleana è isomorfa ad una sottoalgebra di un $\mathcal{P}(\Omega)$ opportuno (Sezione B.4). Soprattutto le algebre booleane finite sono isomorfe ad un $\mathcal{P}(\Omega)$ opportuno (Sezione B.3). Quindi la scelta di algebre booleane su spazi campionari per modellare algebre booleane di domande binarie è giustificata dalla struttura e non significa nessuna restrizione della generalità.

B.1 Algebra booleana

B.1.1 Definizione. Un'***algebra booleana*** $(\mathcal{B}, \sqcap, \sqcup, {}^{-})$ è un insieme $\mathcal{B}$ non vuoto con due operazioni binarie $\sqcap, \sqcup\colon \mathcal{B} \times \mathcal{B} \to \mathcal{B}$ e un'operazione unaria ${}^{-}\colon \mathcal{B} \to \mathcal{B}$, che soddisfano le seguenti assiomi:

$(C_\sqcap)$	$x \sqcap y$	$= y \sqcap x$	$(C_\sqcup)$	$x \sqcup y$	$= y \sqcup x$
$(D_\sqcap)$	$x \sqcap (y \sqcup z)$	$= (x \sqcap y) \sqcup (x \sqcap z)$	$(D_\sqcup)$	$x \sqcup (y \sqcap z)$	$= (x \sqcup y) \sqcap (x \sqcup z)$
$(N_\sqcap)$	$x \sqcap (y \sqcup \overline{y})$	$= x$	$(N_\sqcup)$	$x \sqcup (y \sqcap \overline{y})$	$= x$

per ogni $x, y, z \in \mathcal{B}$. Gli assiomi (C), (D) e (N) si chiamano le ***leggi commutative***, le ***leggi distributive*** e l'***esistenza dell'elemento neutro***, rispettivamente, per le operazioni indicate.

B.1.2 Proposizione. *Esistono un unico* ***elemento neutro*** $\mathbf{1} \in \mathcal{B}$ *rispetto a* $\sqcap$ *(cioè,* $x \sqcap \mathbf{1} = x$ *per ogni* $x \in \mathcal{B}$*) ed un unico* ***elemento neutro*** $\mathbf{0} \in \mathcal{B}$ *rispetto a* $\sqcup$ *(cioè,* $x \sqcup \mathbf{0} = x$ *per ogni* $x \in \mathcal{B}$*). In particolare, vale* $y \sqcup \overline{y} = \mathbf{1}$ *e* $y \sqcap \overline{y} = \mathbf{0}$ *per ogni* $y \in \mathcal{B}$*.*

Dimostrazione. L'esistenza degli elementi neutri segue dagli assiomi (N). Rimane da dimostrare che siano unici. Se sapessimo già che $(\mathcal{B}, \sqcap)$ e $(\mathcal{B}, \sqcup)$ sono semigruppi ci potremmo riferire al risultato ben noto che un elemento neutro di un semigruppo, se esiste, è sempre unico. Ma la dimostrazione dell'associatività delle operazioni $\sqcap$ e $\sqcup$ viene molto più tardi. Dobbiamo rifare la dimostrazione dei semigruppi ed convincerci che, infatti, non abbiamo bisogno delle leggi associativi.

Siano, allora, $\mathbf{1}$ e $\mathbf{1}'$ due elementi neutri per l'operazione $\sqcap$. Usando la legge commutativa troviamo

$$\mathbf{1} \;=\; \mathbf{1} \sqcap \mathbf{1}' \;=\; \mathbf{1}' \sqcap \mathbf{1} \;=\; \mathbf{1}'.$$

Lasciamo la dimostrazione analoga per $\mathbf{0}$ come esercizio. ■

B.1.3 Nota. D'ora in poi non menzioniamo più l'uso della legge commutativa.

B.1.4 Proposizione. *Valgono*

$$(\mathrm{I}_\sqcap) \qquad x \sqcap x \;=\; x \qquad\qquad (\mathrm{I}_\sqcup) \qquad x \sqcup x \;=\; x$$

per ogni $x \in \mathcal{B}$. Chiamiamo (I) *le* ***leggi idempotenti****.*

Dimostrazione. $x \overset{\mathrm{N}_\sqcap}{=} x \sqcap (x \sqcup \overline{x}) \overset{\mathrm{D}_\sqcap}{=} (x \sqcap x) \sqcup (x \sqcap \overline{x}) \overset{\mathrm{N}_\sqcup}{=} x \sqcap x$. Lasciamo $\mathrm{I}_\sqcup$ come esercizio. ■

Con questo le leggi dove un lato dell'equazione si può trasformare facilmente nell'altro lato sono quasi esauriti. Non vogliamo dire che non sia più possibile. Infatti, è sempre possibile, ma non è più facile di metterlo in pratica. Come illustrazione invitiamo il lettore di dimostrare tutte le leggi che seguono ancora in tal modo. Il seguente lemma ci indica metodi per facilitare i calcoli. L'idea principale è il seguente calcolo che abbiamo fatto già una decina di volte per gli eventi di uno spazio campionario:

$$y \;=\; y \sqcap (x \sqcup \overline{x}) \;=\; (y \sqcap x) \sqcup (y \sqcap \overline{x}).$$

Per insiemi la formula corrisponde alla suddivisione di y nella parte appartenente ad x e nella parte appartenente ad $\overline{x}$. Per le espressioni logiche la formula corrisponde alla suddivisione della dimostrazione di un'affermazione y in due casi, uno dove vale l'affermazione x ed un altro dove x è falso. Cruciale è l'osservazione che conoscendo le parti $y \sqcap x$ e $y \sqcap \overline{x}$ di y, è possibile di ricostruire y univocamente. In altre parole, se un altro elemento z ha le stesse parti $z \sqcap x = y \sqcap x$ e $z \sqcap \overline{x} = y \sqcap \overline{x}$ come y allora z ed y devono essere uguali.

B.1.5 Lemma. *Siano $y, z \in \mathcal{B}$ ed esista un $x \in \mathcal{B}$ tale che vale una delle tre condizioni*

$$\Big[y \sqcap x = z \sqcap x \;\; e \;\; y \sqcap \overline{x} = z \sqcap \overline{x}\Big] \quad o \quad \Big[y \sqcup x = z \sqcup x \;\; e \;\; y \sqcup \overline{x} = z \sqcup \overline{x}\Big]$$

$$o \quad \Big[y \sqcap x = z \sqcap x \;\; e \;\; y \sqcup x = z \sqcup x\Big].$$

Allora $y = z$.

Dimostrazione. La prima conclusione segue come indicato d'avanti al lemma:

$$y \;=\; y \sqcap (x \sqcup \overline{x}) \;=\; (y \sqcap x) \sqcup (y \sqcap \overline{x}) \;=\; (z \sqcap x) \sqcup (z \sqcap \overline{x}) \;=\; z \sqcap (x \sqcup \overline{x}) \;=\; z.$$

La seconda conclusione può essere dimostrata ripetendo lo stesso calcolo scambiando ovunque $\sqcap \leftrightarrow \sqcup$. Visto però che dobbiamo anche dimostrare la terza conclusione, procediamo in modo diverso. Osserviamo che

$$(y \sqcup x) \sqcap \overline{x} \;=\; (y \sqcap \overline{x}) \sqcup (x \sqcap \overline{x}) \;=\; y \sqcap \overline{x}. \tag{B.1.1}$$

Quindi, "intersecando" l'equazione $y \sqcup x = z \sqcup x$ con $\overline{x}$, otteniamo $y \sqcap \overline{x} = z \sqcap \overline{x}$. In altre parole, la terza condizione si trasforma nella prima, di cui segue $y = z$. Se, "intersechiamo" tutte e due le equazioni della seconda condizione con $\overline{x}$ anche questa condizione si trasforma nella prima, di cui nuovamente segue $y = z$. ■

Un'illustrazione semplice è il seguente corollario. Il risultato $\overline{\overline{x}} = x$ ci dice, soprattutto, che l'operazione $\overline{\ }$ è autoinversa e, quindi, una biezione su $\mathcal{B}$. Notiamo che non abbiamo bisogno del corollario prima della dimostrazione delle *regole di De Morgan*.

B.1.6 Corollario. *Siano $x, y \in \mathcal{B}$ tali che*

$$(x \sqcup y) \sqcap z \;=\; z \;=\; (x \sqcap y) \sqcup z$$

per ogni $z \in \mathcal{B}$ (ossia, $x \sqcup y = \mathbf{1}$ e $x \sqcap y = \mathbf{0}$ per la Proposizione B.1.2). Allora $y = \overline{x}$ e, per simmetria, $x = \overline{y} = \overline{\overline{x}}$.

Dimostrazione. Scegliamo $z = \overline{x}$. Ne segue

$$\overline{x} \;=\; \begin{cases} (x \sqcup y) \sqcap \overline{x} \;=\; (x \sqcap \overline{x}) \sqcup (y \sqcap \overline{x}) \;=\; y \sqcap \overline{x}, \\ (x \sqcap y) \sqcup \overline{x} \;=\; (x \sqcup \overline{x}) \sqcap (y \sqcup \overline{x}) \;=\; y \sqcup \overline{x}. \end{cases}$$

Dall'idempotenza segue $\overline{x} \sqcap \overline{x} = \overline{x} = \overline{x} \sqcup \overline{x}$. Quindi, secondo la terza condizione del Lemma B.1.5 segue $y = \overline{x}$. ■

B.1.7 Proposizione. *Valgono*

$$(\mathrm{P}_\sqcap) \qquad x \sqcap \mathbf{0} \;=\; \mathbf{0} \qquad\qquad (\mathrm{P}_\sqcup) \qquad x \sqcup \mathbf{1} \;=\; \mathbf{1}$$

per ogni $x \in \mathcal{B}$. Chiamiamo (I) *le* ***leggi dell'assorbimento***.

DIMOSTRAZIONE. Sappiamo che $\mathbf{0} \sqcup z = z$ per ogni z. Se troviamo almeno un elemento y tale che $(x \sqcap \mathbf{0}) \sqcup y = y$ e $(x \sqcap \mathbf{0}) \sqcup \overline{y} = \overline{y}$, dalla seconda condizione del Lemma B.1.5 seguirebbe $\mathbf{0} = x \sqcap \mathbf{0}$. Calcoliamo

$$(x \sqcap \mathbf{0}) \sqcup z = (x \sqcup z) \sqcap (\mathbf{0} \sqcup z) = (x \sqcup z) \sqcap z.$$

Inserendo $z = x$ troviamo

$$(x \sqcap \mathbf{0}) \sqcup x = (x \sqcup x) \sqcap x = x \sqcap x = x$$

e inserendo $z = \overline{x}$ troviamo

$$(x \sqcap \mathbf{0}) \sqcup \overline{x} = (x \sqcup \overline{x}) \sqcap \overline{x} = \overline{x}.$$

Come detto questo implica $x \sqcap \mathbf{0} = \mathbf{0}$. Lasciamo come esercizio l'altra formula. ■

B.1.8 Corollario. *$x \sqcap (x \sqcup y) = x = x \sqcap (x \sqcup y)$ per ogni $x, y \in \mathcal{B}$.*

DIMOSTRAZIONE. $x \sqcap (x \sqcup y) = (x \sqcup \mathbf{0}) \sqcap (x \sqcup y) = x \sqcup (\mathbf{0} \sqcap y) = x \sqcup \mathbf{0} = x$. ■

Infine, siamo in grado di dimostrare anche le leggi associative e le regole di De Morgan.

B.1.9 Proposizione. *Valgono*

$$(A_\sqcap) \quad (x \sqcap y) \sqcap z = x \sqcap (y \sqcap z) \qquad (A_\sqcup) \quad (x \sqcup y) \sqcup z = x \sqcup (y \sqcup z)$$

per ogni $x, y, z \in \mathcal{B}$. Chiamiamo (A) *le* ***leggi associative****.*

DIMOSTRAZIONE. "Intersechiamo" ($A_\sqcup$) con z. Troviamo

$$\begin{aligned} [(x \sqcup y) \sqcup z] \sqcap z &= z, \\ [x \sqcup (y \sqcup z)] \sqcap z &= (x \sqcup z) \sqcap \underbrace{[(y \sqcup z) \sqcap z]}_{=z} = z. \end{aligned}$$

Poi "intersechiamo" ($A_\sqcup$) con $\overline{z}$. Usando diverse volte l'Equazione B.1.1, troviamo

$$\begin{aligned} [(x \sqcup y) \sqcup z] \sqcap \overline{z} &= (x \sqcup y) \sqcap \overline{z} = (x \sqcap \overline{z}) \sqcup (y \sqcap \overline{z}), \\ [x \sqcup (y \sqcup z)] \sqcap \overline{z} &= (x \sqcap \overline{z}) \sqcup [(y \sqcup z) \sqcap \overline{z}] = (x \sqcap \overline{z}) \sqcup (y \sqcap \overline{z}). \end{aligned}$$

Secondo la prima condizione del Lemma B.1.5 i due lati di ($A_\sqcup$) coincidono. Lasciamo la dimostrazione analoga di ($A_\sqcap$) come esercizio. ■

B.1.10 Nota. D'ora in poi non scriviamo più le parentesi per indicare un ordine fra operazioni uguali, se non per indicare una scelta nostra di eseguire una certa operazione prima delle altre uguali. Per nessun motivo possiamo lasciare da parte le parentesi fra operazioni diversi come nelle leggi distributive. ***Non adottiamo*** nessuna convenzione che considerasse $\sqcap$ più vincolante di $\sqcup$ (o *vice versa*). Come abbiamo notato le due operazioni di un'algebra booleana giocano ruoli assolutamente simmetrici. Non pare opportuno applicare delle convenzioni che distruggano tale simmetria. (Ogni formula valida si trasforma in un'altra formula valida se eseguiamo lo scambio $\sqcap \leftrightarrow \sqcup$ in tutta la formula, evitando l'uso dei simboli $\mathbf{0}$ e $\mathbf{1}$ e non toccando i complementari. Questo non sarebbe più valido se considerassimo una delle operazioni più vincolante dell'altra.) Si guarda anche l'Esercizio B.2.7.

B.1.11 Proposizione. *Valgono*

$$(DM_{\sqcap}) \qquad \overline{x \sqcap y} \;=\; \overline{x} \sqcup \overline{y} \qquad\qquad (DM_{\sqcup}) \qquad \overline{x \sqcup y} \;=\; \overline{x} \sqcap \overline{y}$$

per ogni $x, y \in \mathcal{B}$. Chiamiamo (DM) *le* ***regole di De Morgan***.

Dimostrazione. Calcoliamo

$$\begin{aligned}(x \sqcap y) \sqcap (\overline{x} \sqcup \overline{y}) \;&=\; (x \sqcap y \sqcap \overline{x}) \sqcup (x \sqcap y \sqcap \overline{y}) \;=\; (y \sqcap \mathbf{0}) \sqcup (x \sqcap \mathbf{0}) \;=\; \mathbf{0} \sqcup \mathbf{0} \;=\; \mathbf{0},\\ (x \sqcap y) \sqcup (\overline{x} \sqcup \overline{y}) \;&=\; (x \sqcup \overline{x} \sqcup \overline{y}) \sqcap (y \sqcup \overline{x} \sqcup \overline{y}) \;=\; (\mathbf{1} \sqcup \overline{y}) \sqcap (\mathbf{1} \sqcup \overline{x}) \;=\; \mathbf{1} \sqcap \mathbf{1} \;=\; \mathbf{1}.\end{aligned}$$

Allora secondo il Corollario B.1.6 $\overline{x} \sqcup \overline{y}$ è il "complementare" di $x \sqcap y$., ossia, vale ($\mathrm{DM}_{\sqcap}$). Possiamo ripetere lo stesso calcolo per ($\mathrm{DM}_{\sqcup}$). (Esercizio!) Oppure, possiamo sostituire x con $\overline{x}$ e y con $\overline{y}$ di cui segue con le stesse due formule che $\overline{x} \sqcap \overline{y}$ è il "complementare" di $x \sqcup y$. ■

B.2 Omomorfismi ed isomorfismi di algebre booleane

Abbiamo detto nell'introduzione di voler dimostrare che ogni algebra booleana è isomorfa ad un algebra booleana su un insieme Ω. In questa sezione breve specifichiamo i termini usati in quest'affermazione.

B.2.1 Definizione. Siano $(\mathcal{B}^1, \sqcap^1, \sqcup^1, {}^{-1})$ e $(\mathcal{B}^2, \sqcap^2, \sqcup^2, {}^{-2})$ due algebre booleane. Una funzione $\varphi\colon \mathcal{B}^1 \to \mathcal{B}^2$ è un ***omomorfismo*** di algebre booleane se vale

$$\varphi(x \sqcap^1 y) \;=\; \varphi(x) \sqcap^2 \varphi(y) \qquad \varphi(x \sqcup^1 y) \;=\; \varphi(x) \sqcup^2 \varphi(y) \qquad \varphi(\overline{x}^1) \;=\; \overline{\varphi(x)}^2$$

per ogni $x, y \in \mathcal{B}^1$. Un'***immersione*** è un omomorfismo iniettivo. Un ***isomorfismo*** è un omomorfismo biettivo.

B.2.2 Esercizio. Dimostrare che basta controllare una delle due condizioni $\varphi(x \sqcap^1 y) = \varphi(x) \sqcap^2 \varphi(y)$ o $\varphi(x \sqcup^1 y) = \varphi(x) \sqcup^2 \varphi(y)$. (Suggerimento: Usare le regole di De Morgan e la terza condizione.)

B.2.3 Esercizio. L'***immagine*** $\varphi(\mathcal{B}^1) := \{\varphi(x) \colon x \in \mathcal{B}^1\}$ di un omomorfismo φ è una sottoalgebra booleana di $\mathcal{B}^2$. La ***controimmagine*** $\varphi^{-1}(\mathcal{B}^2) := \{x \in \mathcal{B}^1 \colon \varphi(x) \in \mathcal{B}^2\}$ di un omomorfismo φ è una sottoalgebra booleana di $\mathcal{B}^1$.

B.2.4 Corollario. *Un'immersione $\varphi \colon \mathcal{B}^1 \to \mathcal{B}^2$ stabilisce sempre un isomorfismo da $\mathcal{B}^1$ **su** l'immagine $\varphi(\mathcal{B}^2)$. Quindi, l'esistenza di un'immersione $\mathcal{B}^1 \to \mathcal{B}^2$ equivale l'affermazione che $\mathcal{B}^1$ è isomorfa ad una sottoalgebra booleana di $\mathcal{B}^2$.*

B.2.5 Esercizio. Dimostrare che ogni omomorfismo rispetta gli elementi neutri, ossia, $\varphi(\mathbf{0}^1) = \mathbf{0}^2$ e $\varphi(\mathbf{1}^1) = \mathbf{1}^2$.

B.2.6 Esercizio. Dimostrare che l'inversa di un isomorfismo è un isomorfismo.

B.2.7 Esercizio. Sia $(\mathcal{B}, \sqcap, \sqcup, \overline{\ })$ un'algebra booleana. Dimostrare che la funzione φ su $\mathcal{B}$ definita come $\varphi(x) = \overline{x}$ è un isomorfismo dall'algebra booleana $(\mathcal{B}, \sqcap, \sqcup, \overline{\ })$ sull'algebra booleana $(\mathcal{B}, \sqcup, \sqcap, \overline{\ })$ con le operazioni $\sqcap$ a $\sqcup$ scambiate.

Concludiamo con un criterio quando un un omomorfismo è iniettivo.

B.2.8 Proposizione. *Un omomorfismo $\varphi \colon \mathcal{B}^1 \to \mathcal{B}^2$ è un'immersione, se e solo se $\varphi(x) = \mathbf{0} \Longrightarrow x = \mathbf{0}$.*

Dimostrazione. Questo segue facilmente dal fatto che algebre booleane sono anelli rispetto all'addizione $\triangle$ ed il prodotto $\sqcap$ (come discusso nell'introduzione a questo capitolo), e dal fatto che φ è anche un omomorfismo di anelli. (Un omomorfismo di anelli è iniettivo, se e solo se il suo nucleo è banale, cioè, se $\ker\varphi := \{x \in \mathcal{B} \colon \varphi(x) = \mathbf{0}\} = \{\mathbf{0}\}$.) Diamo, comunque, una dimostrazione con mezzi dell'algebra booleana.

Chiaramente, se $\varphi(x) = \mathbf{0}$ e $x \neq \mathbf{0}$, allora φ non è iniettivo (perché, secondo l'Esercizio B.2.5 anche $\varphi(\mathbf{0}) = \mathbf{0}$). Supponiamo, allora che φ non sia iniettivo, cioè, che esistano $x \neq y$ tali che $\varphi(x) = \varphi(y)$. Un'applicazione del Lemma B.1.5 dimostra, che vale $\overline{x} \sqcap y \neq \mathbf{0}$ o $x \sqcap \overline{y} \neq \mathbf{0}$. (Altrimenti, se $\overline{x} \sqcap y = \mathbf{0} = x \sqcap \overline{y}$, allora $x \sqcup \overline{y} = \mathbf{1}$ e, secondo la terza parte del lemma, varrebbe $x = y$.) Però, da $\varphi(x) = \varphi(y)$ segue $\varphi(\overline{x} \sqcap y) = \overline{\varphi(x)} \sqcap \varphi(y) = \mathbf{0}$ e $\varphi(x \sqcap \overline{y}) = \varphi(x) \sqcap \varphi(\overline{y}) = \mathbf{0}$. In altre parole, almeno un elemento, $\overline{x} \sqcap y$ o $x \sqcap \overline{y}$, che è diverso da $\mathbf{0}$ va mandato da φ a $\mathbf{0}$. ∎

B.2.9 Nota matematica.

Nella sezione precedente ci siamo impegnati di definire il termine *algebra booleana* con un minimo di assiomi, partendo di cui tutte le altre proprietà possono essere dimostrate. Nella letteratura, purtroppo, l'algebra booleana spesso va introdotta come caso particolare di altre strutture algebriche. Abbiamo già discusso nell'esercizio dell'introduzione a questa appendice, che le algebre booleane sono esattamente gli anelli unitari che hanno solo elementi idempotenti. Spieghiamo brevemente un altro modo di introdurre le algebre booleane come *reticoli* particolari.

Su un'algebra booleana $\mathcal{B}$ possiamo definire una relazione $\sqsubset$ come $x \sqsubset y$ se $x \sqcap y = x$. Questa relazione è un ***ordinamento parziale***. Cioè $\sqsubset$ è ***riflessiva*** (per l'idempotenza $x \sqcap x = x$, allora $x \sqsubset x$), è ***transitiva*** (se $x \sqsubset y$ e $y \sqsubset z$, allora

$$x \sqcap z \;=\; (x \sqcap y) \sqcap z \;=\; x \sqcap (y \sqcap z) \;=\; x \sqcap y \;=\; x,$$

cosicché $x \sqsubset z$), ed è ***antisimmetrica*** (se $x \sqsubset y$ e $y \sqsubset x$, allora $x = x \sqcap y = y \sqcap x = y$).

Esercizio. In analogia definiamo $x \sqsupset y$ se $x \sqcup y = x$. Dimostrare che $x \sqsubset y$, se e solo se $y \sqsupset x$, se e solo se $\overline{y} \sqsubset \overline{x}$.

Però l'insieme parzialmente ordinato $(\mathcal{B}, \sqsubset)$ ha ancora più struttura. È un ***reticolo***. (Cioè, per ogni $x, y \in \mathcal{B}$ gli elementi $x \sqcap y$ e $x \sqcup y$ sono l'unico ***minimo*** e l'unico ***massimo***, rispettivamente, di x e y. Con $x \sqcap y$ è il minimo, s'intende che $x \sqcap y \sqsubset x$ e $x \sqcap y \sqsubset y$ e se c'é un altro elemento z che soddisfa $z \sqsubset x$ e $z \sqsubset y$, allora $z \sqsubset x \sqcap y$. Analogamente per il massimo.) Il reticolo $(\mathcal{B}, \sqsubset)$ con queste due operazioni $\sqcap$ e $\sqcup$ è ***distributivo*** nel senso che valgono le leggi distributive (D). Inoltre $\mathcal{B}$ ha un ***minimo assoluto*** $\mathbf{0}$ (cioè $\mathbf{0} \sqsubset x$ per ogni x) e un ***massimo assoluto*** $\mathbf{1}$ (cioè $x \sqsubset \mathbf{1}$ per ogni x). Inoltre il reticolo con $\mathbf{0}, \mathbf{1}$ è ***complementato*** nel senso che per ogni x esiste $\overline{x}$ tale che vale $x \sqcap \overline{x} = \mathbf{0}$ e $x \sqcup \overline{x} = \mathbf{1}$. Si può dimostrare che ogni reticolo distributivo complementato emerge in questa maniera da un'algebra booleana. Per questo un tale reticolo si chiama anche un ***reticolo booleano***. Vediamo che tutta la struttura di un'algebra booleana è codificata nel suo ordinamento parziale.

Fine della nota.

B.3 Il teorema di Stone per algebre booleane finite

Lo scopo di questa sezione è dimostrare che ogni algebra booleana $\mathcal{B}$ finita è isomorfa ad un'algebra booleana $\mathcal{P}(\Omega)$ di tutte le parti di un insieme opportuno Ω, ossia il *teorema di Stone* per algebre booleane finiti. A questo punto dobbiamo prima osservare che di Ω abbiamo chiesto che non sia vuoto. Di cui risulta che $\mathcal{P}(\Omega)$ contiene almeno i due elementi diversi $\emptyset$ e Ω. Però, anche l'insieme $\mathcal{P}(\emptyset)$ di tutte le parti dell'insieme vuoto, che contiene un solo elemento, l'insieme vuoto, è un'algebra booleana a tutti gli effetti. Notiamo che esiste un unico modo di definire la struttura di un'algebra booleana su un insieme $\mathcal{B}$ che contiene un solo elemento. Quindi, se ammettiamo anche il caso $\Omega = \emptyset$, il teorema di Stone vale se $\mathcal{B}$ ha un solo elemento.

(Non nascondiamo che $\Omega \neq \emptyset$ significa che lo zero $\mathbf{0}$ e l'uno $\mathbf{1}$ dell'*anello booleano* $\mathcal{P}(\Omega)$ sono diversi. Questa condizione fa parte della definizione di *anello unitario*. Comunque, in questa sezione è più opportuno di ammettere anche $\mathcal{P}(\emptyset)$ come algebra booleana su un insieme, l'insieme vuoto.)

Partendo da un'algebra booleana astratta, per dimostrare il teorema di Stone è chiaro che il compito consiste soprattutto nell'individuare l'insieme Ω che permette di identificare $\mathcal{B}$ come sottoinsieme do $\mathcal{P}(\Omega)$. Ci poniamo la domanda, se $\mathcal{B}$ fosse già un sottoinsieme di $\mathcal{P}(\Omega)$, oppure addirittura se $\mathcal{B} = \mathcal{P}(\Omega)$, come sarebbe possibile di individuare gli eventi elementari $\{\omega\}$ senza far riferimento in nessun modo agli elementi di Ω, ma servendosi esclusivamente di proprietà esprimibili in termini delle operazioni $\cap, \cup, ^{\complement}$. Sorprendentemente, per dimostrare la versione del teorema di Stone per le algebre booleane finite, non occorre prendere questa strada e ne torniamo solo nella sezione seguente per la discussione del caso generale.

Per adesso ci lasciamo guidare da un'altra nostra conoscenza della discussione insiemistica. Se B è un sottoinsieme di Ω, sappiamo che ogni altro sottoinsieme A di Ω può essere scritto in modo unico come unione $A = (A \cap B) \cup (A \cap B^{\complement})$ della parte di A che appartiene a B e la parte di A che appartiene al complementare di B. L'insieme delle parti di B, cioè tutte le insiemi che possiamo scrivere $A \cap B$ ($A \in \mathcal{P}(\Omega)$), è invariato sotto unioni e sotto intersezioni. Per trasformare questo insieme di tutte le parti di B in un'algebra basta sostituire il complementare $A^{\complement} = \Omega \backslash A$ rispetto ad Ω con il complementare $A^{\complement_B} = B \backslash A = B \cap A^{\complement}$ rispetto a B. Evidentemente, lo stesso si può anche fare anche con l'insieme di tutte le parti di $B^{\complement}$.

Le domande che ci poniamo sono le seguenti: È possibile di decomporre in questo modo ogni algebra booleana rispetto ad un elemento scelto? È possibile di ricostruire l'algebra booleana di partenza dalle sue parti? Tutte e due le risposte sono affermative. Iniziamo preparando la risposta alla seconda domanda in un contesto un po' più generale.

Siano $(\mathcal{B}^1, \sqcap^1, \sqcup^1, ^{-1})$ e $(\mathcal{B}^2, \sqcap^2, \sqcup^2, ^{-2})$ due algebre booleane. Sullo spazio prodotto $\mathcal{B}^1 \times \mathcal{B}^2$ definiamo operazioni $\sqcap, \sqcup, ^{-}$ come

$$(x_1, x_2) \sqcap (y_1, y_2) := (x_1 \sqcap^1 y_1, x_2 \sqcap^2 y_2), \qquad (x_1, x_2) \sqcup (y_1, y_2) := (x_1 \sqcup^1 y_1, x_2 \sqcup^2 y_2),$$
$$\overline{(x_1, x_2)} := (\overline{x_1}^1, \overline{x_2}^2).$$

B.3.1 Esercizio. Verificare che $(\mathcal{B}^1 \times \mathcal{B}^2, \sqcap, \sqcup, ^{-})$ è un'algebra booleana. (Possiamo essere molto grati che il numero di assiomi da verificare sia il più piccolo possibile.)

Chiameremo $(\mathcal{B}^1 \times \mathcal{B}^2, \sqcap, \sqcup, ^{-})$ il ***prodotto*** di $(\mathcal{B}^1, \sqcap^1, \sqcup^1, ^{-1})$ e $(\mathcal{B}^2, \sqcap^2, \sqcup^2, ^{-2})$.

B.3.2 Esempio. Sia $\Omega \supset \Omega_1$ e poniamo $\Omega_2 := \Omega_1^{\complement}$. Vogliamo dimostrare che la funzione

$$\varphi\colon (A_1, A_2) \longmapsto A_1 \cup A_2$$

($A_i \subset \Omega_i$) definisce un isomorfismo $\mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2) \to \mathcal{P}(\Omega)$ di algebre booleane. Visto che per ogni sottoinsieme $A \subset \Omega$ vale $A = (A \cap \Omega_1) \cup (A \cap \Omega_2)$ dove $(A \cap \Omega_2) \subset \Omega_i$ la funzione φ è surgettiva. Visto che $\varphi(A_1, A_2) = \varphi(A_1', A_2')$ implica che $A_i = \Omega_i \cap \varphi(A_1, A_2) = \Omega_i \cap \varphi(A_1', A_2') = A_i'$, la funzione φ è anche iniettiva. Ricordiamoci che come elemento di $\mathcal{P}(\Omega_i)$ il complementare di A_i è rispetto ad Ω_i, cioè $A_i^{\complement_i} = \Omega_i \cap A_i^{\complement}$. Quindi,

$$\begin{aligned}\varphi(A_1^{\complement_1}, A_2^{\complement_2}) &= A_1^{\complement_1} \cup A_2^{\complement_2} = (\Omega_1 \cap A_1^{\complement}) \cup (\Omega_2 \cap A_2^{\complement}) \\ &= \big[(\Omega_2 \cup A_1) \cap (\Omega_1 \cup A_2)\big]^{\complement} = \big[(\Omega_2 \cap \Omega_1) \cup (\Omega_2 \cap A_2) \cup (A_1 \cap \Omega_1) \cup (A_1 \cap A_2)\big]^{\complement} \\ &= (\emptyset \cup A_2 \cup A_1 \cup \emptyset)^{\complement} = (A_1 \cup A_2)^{\complement} = \varphi(A_1, A_2)^{\complement}.\end{aligned}$$

Secondo l'Esercizio B.2.2 basta controllare una delle operazione $\cap$ o $\cup$. Lasciamo come esercizio di verificare che $\varphi((A_1, A_2) \cup (B_1, B_2)) = \varphi(A_1, A_2) \cup \varphi(B_1, B_2)$.

B.3.3 Nota. È importante notare che $\mathcal{P}(\Omega_1) \times \mathcal{P}(\Omega_2)$ non ha niente a che fare con il prodotto di spazi di probabilità $\Omega_1 \times \Omega_2$ con la loro algebra booleana $\mathcal{A} \subset \mathcal{P}(\Omega_1 \times \Omega_2)$ come discusso nel Teorema 3.3.1. La prima è isomorfa a $\mathcal{P}(\Omega)$ dove $\#\Omega = \#\Omega_1 + \#\Omega_2$ mentre l'ultima consiste di sottoinsiemi di $\Omega_1 \times \Omega_2$ che ha la cardinalità $\#(\Omega_1 \times \Omega_2) = \#\Omega_1 \cdot \#\Omega_2$.

Possiamo considerare l'esempio precedente come la decomposizione dell'algebra booleana $\mathcal{P}(\Omega)$ rispetto all'elemento Ω_1 ed il suo complementare Ω_2. Adesso vogliamo fare lo stesso per un algebra booleana arbitraria.

B.3.4 Lemma. *Sia $(\mathcal{B}, \sqcap, \sqcup, {}^{-})$ un'algebra booleana ed x un suo elemento. Definiamo*

$$\mathcal{B}^x := \{y \sqcap x\colon y \in \mathcal{B}\}.$$

Se definiamo l'operazione ${}^{-x}$ su $\mathcal{B}^x$ come $\overline{y}^x := \overline{y} \sqcap x$, allora $(\mathcal{B}^x, \sqcap, \sqcup, {}^{-x})$ è un'algebra booleana.

La funzione $\varphi\colon \mathcal{B}^x \times \mathcal{B}^{\overline{x}} \to \mathcal{B}$ definita come $\varphi(y^x, y^{\overline{x}}) := y^x \sqcup y^{\overline{x}}$ è un isomorfismo di algebre booleane.

Dimostrazione. Le operazioni $\sqcap, \sqcup, {}^{-x}$ lasciano $\mathcal{B}^x$ invariato, e $\sqcap$ e $\sqcup$ ereditano da $\mathcal{B}$ tutte le proprietà che non coinvolgono l'operazione ${}^{-}$, cioè (C) e (D). Per dimostrare che $\mathcal{B}^x$ è un'algebra booleana, basta quindi controllare (N). Troviamo

$$z \sqcap (y \sqcup \overline{y}^x) = z \sqcap [y \sqcup (\overline{y} \sqcap x)] = (z \sqcap y) \sqcup (z \sqcap \overline{y} \sqcap x) = (z \sqcap y) \sqcup (z \sqcap \overline{y}) = z,$$

dove $z \sqcap \overline{y} \sqcap x = z \sqcap x \sqcap \overline{y} = z \sqcap \overline{y}$ perché $z \in \mathcal{B}^x$. Con lo stesso ragionamento vale

$$z \sqcup (y \sqcap \overline{y}^x) \;=\; z \sqcup (y \sqcap \overline{y} \sqcap x) \;=\; z \sqcup (y \sqcap \overline{y}) \;=\; z.$$

La verifica che φ definisce un isomorfismo funziona esattamente come nell'Esempio B.3.2. La lasciamo come esercizio. ■

B.3.5 Corollario. (Il teorema di Stone per algebre booleane finite.) *Sia* $(\mathcal{B}, \sqcap, \sqcup, \overline{\ })$ *un'algebra booleana finita. Allora* $\mathcal{B}$ *è isomorfa a* $\mathcal{P}(\Omega)$ *per un insieme* Ω *opportuno. In particolare, la cardinalità di* $\mathcal{B}$ *è* $\#\mathcal{B} = 2^n$ *dove* $n = \#\Omega$.

Dimostrazione. Sappiamo che l'affermazione è vera per $\#\mathcal{B} = 1$ (dove $\Omega = \emptyset$). Adesso supponiamo che l'affermazione valga per tutte le cardinalità meno o uguale di un certo $m \in \mathbb{N}$. Sia $\mathcal{B}$ un'algebra booleana con cardinalità $\#\mathcal{B} = m + 1$. Se $m + 1 = 2$ abbiamo finito. (A meno di isomorfismo c'é una sola algebra booleana con due elementi $\{\mathbf{0}, \mathbf{1}\}$ e questa è isomorfa a $\mathcal{P}(\Omega)$ dove $\Omega = \{\omega\}$ ha un solo elemento ω.) Se $m + 1 > 2$ allora esiste un $x \in \mathcal{B}$ tale che $\mathbf{0} \neq x \neq \mathbf{1}$. Ne segue che $\#\mathcal{B}^x \leq m$ e $\#\mathcal{B}^{\overline{x}} \leq m$, perché nessuno delle due contiene $\mathbf{1}$. Per ipotesi ci sono insiemi opportuni Ω^x e $\Omega^{\overline{x}}$ tali che $\mathcal{B}^x \cong \mathcal{P}(\Omega^x)$ e $\mathcal{B}^{\overline{x}} \cong \mathcal{P}(\Omega^{\overline{x}})$. Sia Ω l'unione disgiunta di Ω^x e $\Omega^{\overline{x}}$.[1] Mettendo insieme l'Esempio B.3.2 con il Lemma B.3.4, segue che

$$\mathcal{P}(\Omega) \;\cong\; \mathcal{P}(\Omega^x) \times \mathcal{P}(\Omega^{\overline{x}}) \;\cong\; \mathcal{B}^x \times \mathcal{B}^{\overline{x}} \;\cong\; \mathcal{B}.$$

Quindi l'affermazione vale per tutte le cardinalità finite. ■

B.4 Il lemma di Zorn ed il teorema di Stone generale

Lo scopo di questa sezione è dimostrare che ***ogni*** algebra booleana $\mathcal{B}$ è isomorfa ad un'algebra booleana di sottoinsiemi di un insieme opportuno Ω, ossia il *teorema di Stone*. Nella sezione precedente abbiamo applicato il *principio dell'induzione* e siamo arrivati al risultato che ogni algebra booleana finita è addirittura isomorfa ad un intero $\mathcal{P}(\Omega)$ opportuno. Ricordiamoci, però, dell'Esempio 1.1.15 di un'algebra booleana che non è isomorfa a nessun $\mathcal{P}(\Omega)$. Per il risultato generale una semplice induzione non basta più. Abbiamo bisogno del *lemma di Zorn*, ossia del *principio dell'induzione transfinita*.

[1] L'***unione disgiunta*** di due insiemi A e B e un insieme Ω che contiene A e B in tal modo che $A^{\complement} = B$. Si può ottenere, per esempio, scegliendo un oggetto ω_0 che non è né un elemento di A né un elemento di B. Poi identificando A con il sottoinsieme $A \times \{\omega_0\}$ di $(A \cup \{\omega_0\}) \times (B \cup \{\omega_0\})$ e B con il sottoinsieme $\{\omega_0\} \times B$, i due sono disgiunti. Quindi, la loro unione Ω si gode delle proprietà richieste.

Il *lemma di Zorn* è un'affermazione che equivale all'*assioma della scelta* (lo spazio prodotto $\bigtimes_{i\in I} A_i$ di una famiglia $(A_i)_{i\in I}$ di insiemi A_i non è vuoto) e serve per stabilire l'esistenza di oggetti matematici in tanti teoremi. Per esempio, l'esistenza di una base in ogni spazio vettoriale, o di una base ortonormale in ogni spazio di Hilbert sono delle applicazioni tipiche. Ma anche il *teorema di Vitali* (non esiste una probabilità P definita su tutti i sottoinsiemi di $[0, 1]$ tale che $P([a, b]) = b - a$ per ogni $[a, b] \subset [0, 1]$) si serve dell'assioma della scelta.

Prima di poter formulare il lemma di Zorn, ci servono alcuni nozioni.

B.4.1 Definizione. Sia $(A, \leq)$ un ordinamento parziale. Un elemento $a \in A$ è ***massimale***, se $b \in A, a \leq b$ implica $a = b$.

Sia B un sottoinsieme di A. Un elemento $a \in A$ è un ***maggiorante*** di B, se $b \leq a$ per ogni $b \in B$.

$(A, \leq)$ è un ordinamento parziale ***induttivo***, se ogni sottoinsieme totalmente ordinato B di A possiede un maggiorante a in A.

B.4.2 Lemma di Zorn. *Ogni ordinamento parziale induttivo possiede un elemento massimale.*

Dopo questi preparativi possiamo procedere con la dimostrazione del teorema di Stone. Sia, allora, $(\mathcal{B}, \sqcap, \sqcup, \bar{\ })$ un'algebra booleana con almeno due elementi. Per stabilire un isomorfismo con un'algebra booleana $\mathcal{A} \subset \mathcal{P}(\Omega)$ (cioè, un immersione $\mathcal{B} \to \mathcal{P}(\Omega)$; si veda il Corollario B.2.4) occorre, soprattutto, individuare un insieme Ω opportuno. Dobbiamo trovare un modo di *codificare* le proprietà che determinano il *singoletto* $\{\omega\}$ di $\mathcal{P}(\Omega)$ in termini che non si servono di espressioni come $\omega \in A$ o $A \subset \Omega$. Notiamo che il sottoinsieme $P_\omega := \{A \subset \Omega \colon \omega \in A\}$ di $\mathcal{P}(\Omega)$ si gode di due proprietà:

1. Una qualsiasi intersezione di un numero finito di elementi di $P_\omega \subset \mathcal{P}(\Omega)$ non è mai vuota, cioè
$$A_1 \cap \ldots \cap A_n \neq \emptyset$$
per ogni scelta di $n \in \mathbb{N}$ e di $A_1, \ldots, A_n \in P_\omega$. (Infatti, tale intersezione contiene almeno l'elemento ω. Veramente, questa proprietà vale per un'intersezione di elementi di P_ω arbitraria. Ma nell'algebra booleana astratta $\mathcal{B}$ sappiamo eseguire solo un numero finito di "intersezioni" $\sqcap$.)

2. Sia A un altro elemento di $\mathcal{P}(\Omega)$ tale che l'intersezione con un numero finito di elementi di P_ω non è mai vuota, cioè
$$A \cap A_1 \cap \ldots \cap A_n \neq \emptyset$$
per ogni scelta di $n \in \mathbb{N}$ e di $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$. Allora anche $A \in P_\omega$. (Infatti, $\{\omega\} \in P_\omega$ e $A \cap \{\omega\} \neq \emptyset$ implica $\omega \in A$ cosicché anche $A \in P_\omega$.)

Queste due proprietà abbiamo formulate in termini che usano esclusivamente la struttura dell'algebra booleana $(\mathcal{P}(\Omega), \cap, \cup, ^{\complement})$. Quindi, le possiamo formulare anche per un'algebra booleana $\mathcal{B}$ qualsiasi. Definiamo, allora, l'insieme

$$FI(\mathcal{B}) \;:=\; \Big\{P \subset \mathcal{B}\colon\; x_1 \sqcap \ldots \sqcap x_n \neq \mathbf{0}\ (n \in \mathbb{N}, x_i \in P)\Big\}$$

di tutti i sottoinsiemi P di $\mathcal{B}$ che soddisfano la prima proprietà. Diciamo gli elementi P di $FI(\mathcal{B})$ si godono la ***proprietà delle "intersezioni" finite*** FI (inglese: *finite intersection property*).

La seconda é una proprietà di massimalità: Se P si gode la FI e x è un altro elemento di $\mathcal{B}$ tale che

$$x \sqcap x_1 \sqcap \ldots \sqcap x_n \;\neq\; \mathbf{0}$$

per ogni $x_i \in P$, allora possiamo aggiungere x a P e il risultato $P \cup \{x\}$ continua godersi la FI. Se $P \cup \{x\} = P$ per ogni tale $x \in \mathcal{B}$, allora P è un elemento massimale di $FI(\mathcal{B})$.

B.4.3 Proposizione. *Ogni $P \in FI(\mathcal{B})$ è contenuto in un elemento massimale M_P di $FI(\mathcal{B})$ nel senso che $P \subset Q \in FI(\mathcal{B})$ implica $Q \subset M_P$.*

Dimostrazione. Sia $\mathcal{P}_P = \{Q \in FI(\mathcal{B})\colon\; Q \supset P\}$ ordinato con l'inclusione $\subset$. Sia $\mathcal{Q}$ un sottoinsieme di $\mathcal{P}_P$ totalmente ordinato. Allora l'insieme

$$P_{\mathcal{Q}} \;:=\; \bigcup_{Q \in \mathcal{Q}} Q$$

è un maggiorante di $\mathcal{Q}$ in $\mathcal{P}_P$. Infatti, $P \in \mathcal{Q}$, cosicché $P \subset P_{\mathcal{Q}}$. Inoltre, $P_{\mathcal{Q}}$ si gode la FI. (Siano $x_1, \ldots x_n$ di $P_{\mathcal{Q}}$. Per ogni x_i c'è un $Q_i \in \mathcal{Q}$ che contiene x_i. Degli elementi $Q_1, \ldots, Q_n$ dell'insieme totalmente ordinato $\mathcal{Q}$ c'è ne uno più grande Q_k che contiene tutti gli altri. Visto che Q_k si gode della FI e che tutti gli x_i sono di Q_k, ne segue che $x_1 \sqcap \ldots \sqcap x_n \neq \mathbf{0}$.) Allora $P_{\mathcal{Q}} \in \mathcal{P}_P$. Chiaramente $Q \subset P_{\mathcal{Q}}$ per ogni $Q \in \mathcal{Q}$, cosicché $P_{\mathcal{Q}}$ è un maggiorante di $\mathcal{Q}$ in $\mathcal{P}_P$. Dal lemma di Zorn segue che $\mathcal{P}_P$ contiene un elemento massimale M_P.

Allora se $P \subset Q \in FI(\mathcal{B})$, ne segue soprattutto $Q \in \mathcal{P}_P$. Quindi, per massimalità di M_P in $\mathcal{P}_P$ ne segue $Q \subset M_P$. ∎

L'ultimo paragrafo della dimostrazione ci dice anche:

B.4.4 Corollario. *Ogni M_P è un elemento massimale di $FI(\mathcal{B})$ nel senso che $M_P \subset Q \in FI(\mathcal{B})$ implica $Q \subset M_P$.*

Adesso, che abbiamo stabilita l'esistenza di elementi massimali, possiamo procedere alla dimostrazione del teorema di Stone.

B.4.5 Teorema di Stone. *Ogni algebra booleana* $(\mathcal{B}, \sqcap, \sqcup, \overline{\ })$ *è isomorfa ad un algebra booleana* $\mathcal{A} \subset \mathcal{P}(\Omega)$ *di sottoinsiemi di un insieme opportuno* Ω*.*

Dimostrazione. Visto che conosciamo già il caso $\#\mathcal{B} = 1$, possiamo supporre nel seguito che $\mathcal{B}$ abbia almeno due elementi diversi. Annotiamo con $\Omega := \{M \in FI(\mathcal{B})\colon M \subset Q \in FI(\mathcal{B}) \Rightarrow Q \subset M\}$ l'insieme di tutti gli elementi massimali di $FI(\mathcal{B})$. Osserviamo che Ω non è vuoto. (Il sottoinsieme $\{\mathbf{1}\}$ di $\mathcal{B}$ si gode la FI e, secondo la Proposizione B.4.3, è contenuto in un elemento massimale di $FI(\mathcal{B})$.) Per ogni $x \neq \mathbf{0}$ di $\mathcal{B}$ definiamo il sottoinsieme

$$\varphi(x) \ := \ \{M \in \Omega\colon x \in M\}$$

di Ω. Notiamo che $\{x\}$ si gode la FI, quindi, secondo la Proposizione B.4.3, $\varphi(x)$ non è vuoto. Definiamo $\varphi(\mathbf{0}) := \emptyset \subset \Omega$. Notiamo che $\varphi(x) = \emptyset \Rightarrow x = \mathbf{0}$.

Per ogni $M \in \Omega$ e ogni $x \in \mathcal{B}$ vale, $x \in M \Rightarrow \overline{x} \notin M$. (Infatti, se $\overline{x}$ fosse di M allora M non si godrebbe la FI, perché $x \sqcap \overline{x} = \mathbf{0}$.) Scambiando x con $\overline{x}$, ne segue $\overline{x} \in M \Rightarrow x \notin M$. Infine, $x \in M \Leftrightarrow \overline{x} \notin M$. Allora $\varphi(\overline{x}) = \varphi(x)^{\complement}$.

Poi se $x \in M$, allora anche $x \sqcup y \in M$. (Infatti, $\{x \sqcup y\} \cup M$ si gode la FI e ha x come elemento. Visto che M è massimale, ne segue $M = \{x \sqcup y\} \cup M \ni x \sqcup y$.) Quindi $x \in M$ o $y \in M$ implica $x \sqcup y \in M$, ossia $\varphi(x) \cup \varphi(y) \subset \varphi(x \sqcup y)$. *Vice versa* se $x \notin M$ e $y \notin M$, allora $\overline{x} \in M$ e $\overline{y} \in M$. Ne segue che $x \sqcup y \notin M$, perché altrimenti non si godrebbe della FI. ($\overline{x} \sqcap (x \sqcup y) \sqcap \overline{y} = \mathbf{0}$.) Allora $\varphi(x)^{\complement} \cap \varphi(y)^{\complement} \subset \varphi(x \sqcup y)^{\complement}$, ossia, $\varphi(x) \cup \varphi(y) \supset \varphi(x \sqcup y)$. Infine, $\varphi(x) \cup \varphi(y) = \varphi(x \sqcup y)$.

Sommiamo: φ è un omomorfismo che soddisfa $\varphi(x) = \emptyset \Rightarrow x = \mathbf{0}$. Secondo la Proposizione B.2.8, φ è iniettivo. Secondo il Corollario B.2.4, l'immagine $\mathcal{A} := \varphi(\mathcal{B})$ è una sottoalgebra booleana di $\mathcal{P}(\Omega)$ isomorfa a $\mathcal{B}$. ∎

B.4.6 Nota. Nonostante la teoria delle algebre booleane sia simmetrica sotto lo scambio $\sqcap \leftrightarrow \sqcup$ (si veda l'Esercizio B.2.7), notiamo che le dimostrazioni delle die versioni del teorema di Stone che abbiamo discusso, chiaramente, interrompono questa simmetria. La radice per questo fenomeno si nasconde dietro il fatto, che il teorema di Stone fù trovato studiando gli *anneli booleani* dove le due operazioni $\triangle$ e $\sqcap$ non sono simmetrici sin dall'inizio. Chiaramente, sarebbe possibile modellare una dimostrazione anche partendo da una *finite union property*. Il risultato sarebbe lo stesso come se avessimo fatto la costruzione per l'algebra booleana opposta secondo l'Esercizio B.2.7.

B.4.7 Esempio. Anche se l'algebra booleana è data come algebra di eventi di un insieme, non è detto che la costruzione dia in dietro tale insieme. Vediamo l'Esempio 1.1.15, cioè sia $\mathcal{B} = \{A \subset \mathbb{N}\colon \#A < \infty \text{ o } \#A^{\complement} < \infty\} \subset \mathcal{P}(\mathbb{N})$. Sia $M \subset \mathcal{B}$ un sottoinsieme massimale che si goda la FI. Ci

sono due casi: o M contiene un insieme $A \in \mathcal{B}$ finito o tutti gli $A \in M$ hanno il complementare finito.

Sia, allora, $A \in M$ con $\#A = n < \infty$. Allora esiste un $m \in A$, tale che $m \in B$ per ogni altro $B \in M$. (Altrimenti potremmo scegliere per ogni $m_1, \ldots, m_n$ di A un insieme $B_k \in \mathcal{B}$ tale che $m_k \notin B_k$. Ne seguirebbe che $A \cap B_1 \cap \ldots \cap B_n = \emptyset$ contraddicendo la FI.) Poi notiamo che tale m è unico. (Infatti, anche $\{\{m\}\} \cup M$ si gode della FI, e per massimalità di M, ne segue $\{m\} \in M$. Se m' fosse un altro elemento di A contenuto in ogni $B \in M$, allora anche $\{m'\} \in M$. Per la FI segue $\{m\} \cap \{m'\} \neq \emptyset$, ossia, $m = m'$.) Intanto: Se M contiene come elemento un insieme finito, allora esiste un unico $m \in \mathbb{N}$ tale che

$$M \;=\; M_m \;:=\; \{A \in \mathcal{B} \colon m \in A\}.$$

Fino a questo punto vediamo che l'Ω della dimostrazione del teorema di Stone contiene in modo naturale tutti gli elementi di $\mathbb{N}$. Però ce ne sono di più.

Infatti, supponiamo adesso che M contenga solo insiemi infiniti, ossia, insiemi con il complementare finito. Sappiamo dalla discussione dell'Esempio 1.1.15 che ogni intersezione finita di tali insiemi non è vuota. Quindi, ad M possiamo aggiungere un qualsiasi insieme A con il complementare finito senza violare la FI. Per massimalità di M deve, allora, valere $A \in M$ per ogni A con complementare finito. Quindi,

$$M \;=\; M_\infty \;:=\; \{A \subset \mathbb{N} \colon \#A^{\complement} < \infty\}.$$

Vediamo che, oltre all'insieme di partenza $\mathbb{N}$, l'insieme Ω degli elementi massimali di $FI(\mathcal{B})$ contiene esattamente un altro elemento: $\Omega = \{M_m \colon m \in \mathbb{N} \cup \{\infty\}\}$. Vogliamo vedere adesso quale sottoinsieme $\varphi(A)$ l'omomorfismo della dimostrazione assegna ad un elemento $A \in \mathcal{B}$. L'insieme $\varphi(A)$ contiene tutti gli $M \in \Omega$ tali che $A \in M$. Se $m \in A$, allora $A \in M_m$, e *vice versa*. Poi $A \in M_\infty$, se e solo se A è infinito. Allora

$$\varphi(A) \;=\; \begin{cases} \{M_m \colon m \in A\} & \#A < \infty, \\ \{M_m \colon m \in A\} \cup \{\infty\} & \#A^{\complement} < \infty. \end{cases}$$

Raccomandiamo come esercizio di verificare che questa funzione da $\mathcal{B}$ in $\mathcal{P}(\Omega)$ definisce veramente un omomorfismo.

Notiamo una differenza cruciale fra $\mathcal{B} \subset \mathcal{P}(\mathbb{N})$ e $\mathcal{A} = \varphi(\mathcal{B}) \subset \mathcal{P}(\mathbb{N} \cup \{\infty\})$. Mentre un sottoinsieme P di $\mathcal{B}$ che si gode della FI può avere l'intersezione $\bigcap_{A \in P} A$ vuota (per esempio $P = M_\infty$), un qualsiasi sottoinsieme Q di $\mathcal{A}$ che si gode della FI ha l'intersezione non vuota. Se Q è massimale, allora l'intersezione contiene esattamente un punto di Ω.

Appendice C

Probabilità booleane con la funzione di ripartizione

Spesso possiamo assegnare in modo naturale delle probabilità agli elementi $D \in \mathcal{D}$ di un sottoinsieme $\mathcal{D} \subset \mathcal{A}$ di un'algebra booleana $\mathcal{A} \subset \mathcal{P}(\Omega)$ di eventi di uno spazio campionario $\Omega \neq \emptyset$. Poi ci tocca ad estendere tale probabilità su tutto $\mathcal{A}$. Ci vediamo di fronte a due domande: Esiste un'estensione? Se esiste, è unica? Queste domande hanno risposte affermative, se $\mathcal{D}$ si gode delle proprietà di una *semialgebra booleana* che genera $\mathcal{A}$, e se la probabilità data su $\mathcal{D}$ soddisfa un minimo di condizioni di compatibilità.

Tratteremo la teoria generale nella Sezione C.1.

C.1 Probabilità su semialgebre booleane

Per prima cosa, se una funzione $P\colon \mathcal{D} \to \mathbb{R}$ possiede un'estensione come probabilità su tutto $\mathcal{A} \supset \mathcal{D}$, allora secondo la Proposizione 1.2.6 deve soddisfare

$$P(D_1 \cup \ldots \cup D_m) \;=\; P(D_1) + \ldots + P(D_m) \tag{C.1.1}$$

qualunque siano gli elementi $D_1, \ldots, D_m$ due a due disgiunti di $\mathcal{D}$. Se anche l'unione $D_1 \cup \ldots \cup D_m$ dovesse risultare un elemento di $\mathcal{D}$, allora questa è una condizione a la funzione P definita su $\mathcal{D}$. Se l'unione non è di $\mathcal{D}$, allora è comunque di $\mathcal{A}$, è non rimane altro che definire la probabilità di tale unione secondo la (C.1.1). Se ogni elemento A di $\mathcal{A}$ dovesse essere esprimibile come unione finita di elementi due a due disgiunti di $\mathcal{D}$, allora conoscere P su $\mathcal{D}$ determinerebbe univocamente la probabilità su $\mathcal{A}$. In prassi, dato un insieme $\mathcal{D}$ si sottoinsiemi di Ω si considera, infatti, l'insieme di tutti i sottoinsiemi di Ω che possono essere scritti come unione finita disgiunta di elementi di $\mathcal{D}$. Però, poiché questo sottoinsieme di $\mathcal{P}(\Omega)$ sia un'algebra

booleana su Ω, il sottoinsieme $\mathcal{D}$ di $\mathcal{P}(\Omega)$ deve soddisfare certe proprietà: $\mathcal{D}$ deve essere una *semialgebra booleana* su Ω. Ma un volta stabilito che $\mathcal{D}$ sia una semialgebra, basta poco più di (C.1.1) perché P possieda un'estensione.

C.1.1 Definizione. Sia $\mathcal{D} \subset \mathcal{P}(\Omega)$ una famiglia di sottoinsiemi di $\Omega \neq \emptyset$. Chiameremo $\mathcal{D}$ una ***semialgebra booleana*** su Ω, se soddisfa le seguenti proprietà:

1. $\Omega \in \mathcal{D}$.
2. Se $D, D' \in \mathcal{D}$, allora anche $D \cap D' \in \mathcal{D}$.
3. Se $D \in \mathcal{D}$ allora esistono $m \in \mathbb{N}$ e $D_1, \ldots, D_m \in \mathcal{D}$ due a due disgiunti tali che

$$D^{\complement} \;=\; D_1 \cup \ldots \cup D_m.$$

C.1.2 Proposizione. *Sia $\mathcal{D}$ una semialgebra booleana su $\Omega \neq \emptyset$. Allora*

$$\beta(\mathcal{D}) \;:=\; \{D_1 \cup \ldots \cup D_m \colon m \in \mathbb{N}; D_1, \ldots, D_m \in \mathcal{D}; i \neq j \Rightarrow D_i \cap D_j = \emptyset\}$$

*è un algebra booleana, l'algebra booleana **generata** da $\mathcal{D}$. (Evidentemente, non c'è un'algebra booleana più piccola su Ω che contenga $\mathcal{D}$.)*

Dimostrazione. Siano $A = D_1 \cup \ldots \cup D_m$ e $B = E_1 \cup \ldots \cup E_n$ ($D_i, E_j \in \mathcal{D}$) due elementi di $\beta(\mathcal{D})$. Allora, la loro intersezione

$$A \cap B \;=\; (D_1 \cup \ldots \cup D_m) \cap (E_1 \cup \ldots \cup E_n) \;=\; \bigcup_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} (D_i \cap E_j) \tag{C.1.2}$$

è l'unione finita degli elementi $D_i \cap E_j$ due a due disgiunti di $\mathcal{D}$, ossia, $A \cap B \in \beta(\mathcal{D})$.

Banalmente, il complementare $D^{\complement}$ di $D \in \mathcal{D}$ è di $\beta(\mathcal{D})$. Il complementare di un'unione $A = D_1 \cup \ldots \cup D_m$ di elementi di $\mathcal{D}$ è l'intersezione $D_1^{\complement} \cap \ldots \cap D_m^{\complement}$ dei complementari. I complementari sono di $\beta(\mathcal{D})$ e $\beta(\mathcal{D})$ è stabile sotto intersezioni, quindi, anche $A^{\complement}$ è di $\beta(\mathcal{D})$.

Questo dimostra (si veda l'Osservazione 1.1.13) che $\beta(\mathcal{D})$ è un'algebra booleana. ∎

C.1.3 Esempio. Gli intervalli semiaperti a sinistra

$$\mathcal{H} \;:=\; \Big\{(a, b], (a, \infty) \colon \; -\infty \leq a \leq b < \infty\Big\}$$

formano una semialgebra booleana su $\mathbb{R}$. (Esercizio!) Lo stesso vale per l'insieme degli intervalli semiaperti a destra e per l'insieme $\mathcal{I}$ di tutti gli intervalli di $\mathbb{R}$.

Gli insiemi

$$\mathcal{H}^n := \{H_1 \times \ldots \times H_n \colon H_1, \ldots, H_n \in \mathcal{H}\} \quad \text{e} \quad \mathcal{I}^n := \{I_1 \times \ldots \times I_n \colon I_1, \ldots, I_n \in \mathcal{I}\}$$

sono semialgebre booleane su $\mathbb{R}^n$. Quest'affermazione è un caso particolare della Proposizione C.3.1 che tratteremo alla fine di quest'appendice. Comunque, è consigliato di trattare il $\mathcal{H}^n$ (o, per lo meno, il caso $\mathcal{H}^2$) come esercizio.

Adesso ci occupiamo della domanda quando una funzione definita su $\mathcal{D}$ ammetta un'estensione su tutto $\beta(\mathcal{D})$.

C.1.4 Definizione. Una ***probabilità*** su una semialgebra booleana $\mathcal{D}$ su Ω è una funzione $P \colon \mathcal{D} \to \mathbb{R}$ che soddisfa le seguenti proprietà:

1. $P(D) \geq 0$ per ogni $D \in \mathcal{D}$.
2. $P(\Omega) = 1$.
3. Per qualunque scelta di elementi $D_1, \ldots, D_m \in \mathcal{D}$ due a due disgiunti, tali che anche l'unione $D_1 \cup \ldots \cup D_m$ è di $\mathcal{D}$, vale l'Equazione (C.1.1).

C.1.5 Proposizione. *Ogni probabilità su una semialgebra booleana $\mathcal{D}$ su $\Omega \neq \emptyset$ possiede un'unica estensione come probabilità booleana su $\beta(\mathcal{D})$.*

Dimostrazione. È chiaro che tale estensione è unica, perché ogni elemento $A \in \beta(\mathcal{D})$ può essere scritto come unione disgiunta $D_1 \cup \ldots \cup D_m$ di elementi di $\mathcal{D}$, e su un tale elemento non possiamo fare a meno di definire la probabilità $P(A)$ secondo la (C.1.1). La domanda è se tale definizione possa dipendere dalla decomposizione di A come unione disgiunta di elementi di $\mathcal{D}$. Sia, quindi, $A = E_1 \cup \ldots \cup E_n$ un'altra decomposizione di A in elementi $E_1, \ldots, E_n \in \mathcal{D}$ due a due disgiunti. Osserviamo, per ogni i gli insiemi $D_i \cap E_1, \ldots, D_i \cap E_n$ sono due a due disgiunti a la loro unione è $D_i \in \mathcal{D}$. Analogamente, per ogni j gli insiemi $D_1 \cap E_j, \ldots, D_m \cap E_j$ sono due a due disgiunti a la loro unione è $E_j \in \mathcal{D}$. Per la Proprietà (3) della Definizione C.1.4 otteniamo

$$\sum_{i=1}^{m} P(D_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} P(D_i \cap E_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} P(D_i \cap E_j) = \sum_{j=1}^{n} P(E_j).$$

Quindi, la nostra definizione del valore $P(A)$ non dipende dalla decomposizione di A in elementi due a due disgiunti di $\mathcal{D}$.

Rimane la verifica che $A \mapsto P(A)$ definisce una probabilità su $\beta(\mathcal{D})$. Chiaramente $P(A) \geq 0$ per ogni $A \in \beta(\mathcal{D})$ è $P(\Omega) = 1$. Siano allora $A = D_1 \cup \ldots \cup D_m$ e $B = E_1 \cup \ldots \cup E_n$ due eventi

disgiunti di $\beta(\mathcal{D})$ dati come unioni disgiunti di elementi $D_1, \ldots, D_m$, rispettivamente, $E_1, \ldots, E_n$ di $\mathcal{D}$. Dalla (C.1.2) concludiamo che $A \cap B$ può essere vuoto, se e solo se ogni $(D_i \cap E_j)$ è vuoto. Quindi, tutti gli $m + n$ eventi $D_1, \ldots, D_m, E_1, \ldots, E_n$ sono due a due disgiunti, di cui segue

$$P(A) + P(B) \;=\; P(D_1) + \ldots + P(D_m) + P(E_1) + \ldots + P(E_n) \;=\; P(A \cup B).$$

Allora $A \mapsto P(A)$ è una probabilità booleana su $\beta(\mathcal{D})$. ■

C.1.6 Osservazione. Notiamo che nella dimostrazione precedente la condizione $P(D) \geq 0$ è stata usata esclusivamente

C.2 Probabilità sugli intervalli

C.3 Prodotti di semialgebre booleane

EINGEKLEBT

Dimostrazione incompleta. Visto che (come nell'Osservazione 2.2.4)

$$(A_1 \times \ldots \times A_m) \cap (B_1 \times \ldots \times B_m) \;=\; (A_1 \cap B_1) \times \ldots \times (A_m \cap B_m),$$

con l'uso multiplo della legge distributiva vediamo che $\mathcal{A}$ è chiuso sotto intersezioni. Poiché

$$\left(\bigcup_{i=1}^{k} A_1^i \times \ldots \times A_m^i\right)^{\complement} \;=\; \bigcap_{i=1}^{k} (A_1^i \times \ldots \times A_m^i)^{\complement},$$

e $\mathcal{A}$ è chiuso sotto intersezioni, basta se dimostriamo che $(A_1 \times \ldots \times A_m)^{\complement} \in \mathcal{A}$. Osserviamo che

$$\begin{aligned} A_1 \times \ldots \times A_m \;=\; & (A_1 \times \Omega_2 \times \ldots \times \Omega_m) \\ & \cap \ldots \cap (\Omega_1 \times \ldots \times A_j \times \ldots \times \Omega_m) \\ & \cap \ldots \cap (\Omega_1 \times \ldots \times \Omega_{m-1} \times A_m). \end{aligned}$$

Come nella proposizione 2.2.5 abbiamo $(\Omega_1 \times \ldots \times A_j \times \ldots \times \Omega_m)^{\complement} = \Omega_1 \times \ldots \times A_j^{\complement} \times \ldots \times \Omega_m$. Allora, $(A_1 \times \ldots \times A_m)^{\complement}$ è l'unione degli insiemi $\Omega_1 \times \ldots \times A_j^{\complement} \times \ldots \times \Omega_m$ ($j = 1, \ldots, m$) e, quindi, un elemento di $\mathcal{A}$. Poiché $\mathcal{A}$ è chiuso sotto complementari ed intersezioni, secondo l'Osservazione 1.1.13 $\mathcal{A}$ è un'algebra booleana.

È sempre possibile scegliere gli $A_j^i \in \mathcal{A}_j$ della rappresentazione per un elemento di $\mathcal{A}$ tali che gli insiemi $(A_1^1 \times \ldots \times A_m^1)$, ... , $(A_1^k \times \ldots \times A_m^k)$ sono due a due disgiunti. (Per esempio,

$$\begin{aligned} & (A \times B) \cup (C \times D) \\ & \qquad = \; ((A \times B) \cap (C \times D)^{\complement}) \cup ((A \times B)^{\complement} \cap (C \times D)) \cup ((A \times B) \cap (C \times D)). \end{aligned}$$

Inserendo

$$(A \times B)^{\complement} \;=\; (A^{\complement} \times \Omega_2) \cup (\Omega_1 \times B^{\complement}) \;=\; (A^{\complement} \times B) \cup (A^{\complement} \times B^{\complement}) \cup (A \times B^{\complement})$$

e la formula analoga per $(C \times D)^{\complement}$, dopo qualche applicazione della legge distributiva otteniamo una decomposizione in insiemi prodotti disgiunti. Ci accontentiamo di questo ragionamento molto telegrafico.) Quindi, se esiste una probabilità, su un elemento di $\mathcal{A}$ dato in una decomposizione in prodotti disgiunti tale probabilità è unica e la dobbiamo definire

$$P\left(\bigcup_{i=1}^{k} A_1^i \times \ldots \times A_m^i\right) \;:=\; \sum_{i=1}^{k} P_1(A_1^i) \cdot \ldots \cdot P_m(A_m^i).$$

Rimane il problema di dimostrare che tale definizione non sia contraddittoria, cioè che il valore assegnato dalla formula non dipenda dalla decomposizione scelta. Una dimostrazione veloce ed elegante richiederebbe conoscenze del secondo semestre di un corso in algebra lineare. Una dimostrazione diretta è lunga e complicata. La omettiamo. ■

C.3.1 Proposizione. *Siano $\mathcal{D}_i$ semialegbre booleane su Ω_i ($i = 1, 2, \ldots$). Allora l'insieme*

$$\mathcal{D} := \{D_1 \times \ldots \times D_n \times \Omega_{n+1} \times \Omega_{n+1} \times \ldots \ : \ n \in \mathbb{N}, D_1 \in \Omega_1, \ldots, D_n \in \Omega_n\}$$

è una semialgebra booleana su $\Omega := \Omega_1 \times \Omega_2 \times \ldots$.

Dimostrazione. $\mathcal{D}$ contiene Ω ed è chiaramente stabile sotto intersezioni. Il complementare di un elemento $D_1 \times \ldots \times D_n \times \Omega_{n+1} \times \Omega_{n+1} \times \ldots$ di $\mathcal{D}$ è $(D_1 \times \ldots \times D_n)^{\complement} \times \Omega_{n+1} \times \Omega_{n+1} \times \ldots$ dove il complementare di $D_1 \times \ldots \times D_n$ è calcolato rispetto a $\Omega_1 \times \ldots \times \Omega_n$, come l'abbiamo già dimostrato nella Proposizione 2.2.5. Poi

$$\begin{aligned}(D_1 \times \ldots \times D_n)^{\complement} &= \Big[(D_1 \times \Omega_2 \times \ldots \times \Omega_n) \cap \ \ldots \ \cap (\Omega_1 \times \ldots \times \Omega_{n-1} \times D_n)\Big]^{\complement} \\ &= (D_1^{\complement} \times \Omega_2 \times \ldots \times \Omega_n) \cup \ \ldots \ \cup (\Omega_1 \times \ldots \times \Omega_{n-1} \times D_n^{\complement}) = E_1 \cup \ldots \cup E_n,\end{aligned}$$

dove poniamo $E_k := \Omega_1 \times \ldots \times \Omega_{k-1} \times D_k^{\complement} \times \Omega_{k+1} \times \ldots \times \Omega_n$. Notiamo la seguente procedura generale per trasformare un'unione in un'unione disgiunta:

$$\begin{aligned}E_1 \cup \ldots \cup E_n = \underbrace{\underbrace{E_1 \cup \Big[E_2 \cap E_1^{\complement}\Big]}_{=E_1 \cup E_2} \cup \Big[E_3 \cap (E_1 \cup E_2)^{\complement}\Big] \cup \ldots \cup \Big[E_n \cap (E_1 \cup \ldots \cup E_{n-1})^{\complement}\Big]}_{=E_1 \cup E_2 \cup E_3} \\ \underbrace{\qquad\qquad \ldots \ \ldots \ \ldots \ = \ \ldots \ \ldots \ \ldots \qquad\qquad}_{=E_1 \cup \ldots \cup E_n} \\ = E_1 \cup \Big[E_2 \cap E_1^{\complement}\Big] \cup \Big[E_3 \cap E_1^{\complement} \cap E_2^{\complement}\Big] \cup \ldots \cup \Big[E_n \cap E_1^{\complement} \cap \ldots \cap E_{n-1}^{\complement}\Big].\end{aligned}$$

(Esercizio: Verificare che gli $E_k \cap E_1^{\complement} \cap \ldots \cap E_{k-1}^{\complement}$ ($k = 1, \ldots, n$) siano veramente due a due disgiunti.) Per verificare che $E_1 \cup \ldots \cup E_n$ sia un'unione disgiunta di elementi di $\mathcal{D}$ basta di dimostrarlo per ogni $E_k \cap E_1^{\complement} \cap \ldots \cap E_{k-1}^{\complement}$. (Esercizio: Perché basta?) Per tali intersezioni troviamo

$$E_k \cap E_1^{\complement} \cap \ldots \cap E_{k-1}^{\complement} = D_1 \times \ldots \times D_{k-1} \times D_k^{\complement} \times \Omega_{k+1} \times \ldots \times \Omega_n.$$

$D_k^{\complement}$ è l'unione disgiunta di elementi di $\mathcal{D}_k$, quindi, $D_1 \times \ldots \times D_{k-1} \times D_k^{\complement} \times \Omega_{k+1} \times \ldots \times \Omega_n$ è l'unione disgiunta di elementi di $\mathcal{D}$. ∎

Bibliografia

[Bal98] P. Baldi, *Calcolo delle probabilità e statistica (2^a ed)*, McGraw Hill, 1998.

[Bal03] _____, *Introduzione alla probabilità — con elementi di statistica*, McGraw Hill, 2003.

[Gue03] Denis Guedj, *Il teorema del papagallo*, Tascabili degli Editori Associati S.p.A., 2003.

[Let93] G. Letta, *Probabilità elementare*, Zanichelli, 1993.

[Ske04] M. Skeide, *Algebra per l'informatica*,
available at: `http://serviziweb.unimol.it/pls/unimol/unimolise_docenti.h_preview?id_doc=2104&id _incarico=1360`, 2004,
Dispense della lezione del A.A. 2003/04.

Nessuno dei libri [Bal03, Let93, Bal98] può sostituire queste dispense. Però la loro lettura può aiutare alla comprensione degli argomenti, mettendoci luce da un punto di vista diverso e, forse, più standard. Il libro [Bal03] di Baldi è quello più elementare, mentre il libro [Bal98] dello stesso autore è indirizzato (anche) a studenti di matematica. (Contiene anche un capitolo esteso sull'inferenza statistica.) Il livello del libro [Let93] è più avanzato, in che l'autore introduce la teoria della misura e ne fa uso conseguente. Non vogliamo, comunque, nascondere il fatto che il libro di Letta, oltre a studenti di matematica, si riferisce anche esplicitamente a studenti d'informatica.

Il libro [Gue03] di Denis Guedj, un romanzo giallo di notevole livello letterario — e allo stesso tempo un compendio della storia matematica dagli inizi fino al diciannovesimo secolo, lo vorrei raccomandare a chiunque che abbia paura della matematica senza saperne il perché. (Chiaramente, le mie aspettative non sono che chi legge il libro dopo la lettura sappia perché ha paura della matematica, ma piuttosto che si renda conto che non c'è ragione di averne paura.)