

Metodi Probabilistici, Statistici e Processi Stocastici

F. Morandin

16 aprile 2001

1 Spazi di Probabilità

Uno spazio di probabilità è un *modello* matematico che permette di descrivere un esperimento casuale. Presume delle assunzioni a priori sull'esperimento stesso (spesso sono le probabilità elementari), ma non permette di fare verifiche sperimentali della correttezza delle assunzioni. È invece utile per dedurre probabilità complicate partendo da quelle elementari.

Osservazione. Ad esempio possiamo voler assumere che il lancio di una moneta abbia probabilità 0.5 di dare testa, e questo ci permette di calcolare le probabilità di molti eventi complessi. Tuttavia, anche dopo il verificarsi di eventi rari, come molte teste consecutive, non siamo autorizzati a dire *con assoluta certezza* che la moneta non è bilanciata e favorisce quella faccia; tutte le inferenze che vorremmo fare a questo proposito fanno parte in realtà degli studi *statistici*, e comunque hanno sempre un margine di incertezza.

1.1 Definizioni

Ogni spazio di probabilità è costituito da due oggetti.

1. L'insieme degli *esiti*, che di solito si indica con Ω .
2. La misura di probabilità P .

1.1.1 L'insieme degli esiti

Ω normalmente non ha una particolare struttura. Può essere un insieme numerico, come può non esserlo. La sua *cardinalità* può essere finita o infinita (e nel secondo caso può essere *numerabile* o *continua*); se essa è finita o numerabile si chiama anche *discreta*. Ci sono grandi differenze nella trattazione di questa materia a seconda se Ω sia discreto o continuo.

Un fatto di notevole importanza è che lo spazio degli esiti può essere scelto con una certa arbitrarietà. Tuttavia tipicamente esistono scelte più efficaci di altre. La scelta di Ω è quindi una parte fondamentale per la risoluzione del problema.

Esempio 1.1. Se l'esperimento casuale fosse pescare una carta a caso da un mazzo da poker, una possibile scelta per Ω sarebbe costituita dall'insieme delle 52 carte. Infatti l'esito dell'esperimento sarà una di queste.

1.1.2 La misura di probabilità

La misura di probabilità P , è in pratica una funzione che manda sottoinsiemi di Ω in numeri reali. I sottoinsiemi di Ω sono insiemi di esiti, e vengono denominati *eventi*¹. Ogni evento $A \subset \Omega$ ha una probabilità indicata da $P(A)$. Come ci si aspetterebbe e come vedremo, essa sarà un numero positivo minore o uguale a uno.

¹Solo nel caso in cui Ω sia continuo, i suoi sottoinsiemi possibili potrebbero essere troppi, e per vari tipi di problemi teorici bisogna ammettere che non tutti abbiano una probabilità. Verranno allora chiamati eventi solo quelli che ce l'hanno. Assumiamo da ora per semplicità che ogni volta che citeremo un insieme di esiti, questo sia un evento.

Esempio 1.2. *Nell'esempio del mazzo da poker, “Esce una carta di cuori”, “Esce un re”, “Esce una figura” o “Esce il 2 di picche” sono tutti eventi. Essi possono anche essere scritti come insiemi di esiti: rispettivamente abbiamo $\{A\heartsuit, 2\heartsuit, \dots, K\heartsuit\}$, $\{K\heartsuit, K\diamondsuit, K\clubsuit, K\spadesuit\}$, $\{A\heartsuit, K\heartsuit, Q\heartsuit, J\heartsuit, A\diamondsuit, \dots, J\spadesuit\}$ e $\{2\spadesuit\}$.*

1.1.3 Gli assiomi

Abbiamo già detto che è una richiesta sensata imporre che la probabilità di ogni evento sia un numero positivo non più grande di 1. Vi sono altre considerazioni immediate da fare affinché la teoria che costruiamo corrisponda alla nostra idea intuitiva di probabilità. Esse sono riassunte nei seguenti assiomi, che quindi vanno accettati come ipotesi di lavoro sensate e plausibili, e non necessitano di ulteriori giustificazioni.

1. Se A è un evento, $P(A) \geq 0$

2. $P(\Omega) = 1$

3. **Additività** Se A e B sono eventi disgiunti, $P(A \cup B) = P(A) + P(B)$

3b. **Additività numerabile** Se A_1, A_2, \dots è una successione di eventi a due a due disgiunti,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

L'assioma 2. garantisce che la probabilità degli eventi sia normalizzata a 1. In questo modo l'evento Ω (corrispondente alla frase “qualunque esito”) è certo, come è naturale aspettarsi.

L'assioma 3. a parole impone che se due eventi non sono compatibili (ad esempio, “Esce una figura” con “Esce il 2 di picche” dell'Esempio 1.2), la probabilità che avvenga uno qualunque dei due è la somma delle probabilità che avvengano l'uno e l'altro.

Osservazione. Come vedremo tra poco, l'assioma 3. è lo strumento che permette di ottenere molti dei risultati che presenteremo. Tra questi, vi è un'estensione (*additività finita*) a un numero qualsiasi di eventi (cfr. la Proposizione 1.6). Allora perché abbiamo bisogno di 3b.? Due sono le considerazioni da fare: a) i primi tre assiomi sono sufficienti a provare la Proposizione 1.6, ma non l'assioma 3b.; b) molto spesso invece, quando Ω è infinito, abbiamo proprio bisogno di unire infiniti eventi, quindi la Proposizione 1.6 non va bene.

Osservazione. Abbiamo detto che P deve associare una probabilità a ogni evento. Inoltre deve farlo compatibilmente con gli assiomi. Nell'affrontare qualunque problema, siamo noi che dobbiamo specificare questi numeri e questo può essere un compito molto arduo (si noti che se Ω ha n elementi, i suoi eventi, o sottoinsiemi possibili sono 2^n). Ci domandiamo allora: non potrebbe essere sufficiente assegnare la probabilità dei singoli esiti?

Seguendo infatti l'esempio del mazzo di carte, se abbiamo assegnato una probabilità di $\frac{1}{52}$ a ogni carta del mazzo, l'assioma 3. è sufficiente a imporre

che gli eventi dell'Esempio 1.2 abbiano rispettivamente probabilità $\frac{13}{52}$, $\frac{4}{52}$, $\frac{16}{52}$ e $\frac{1}{52}$.

Si può vedere in effetti che se Ω è finito (numerabile) l'assioma **3.** (**3b.**) è sufficiente a stabilire il valore di P su qualunque evento, una volta che esso sia stato deciso per i soli *eventi elementari*² in modo che la loro somma sia 1. Useremo molto questo fatto negli esempi a seguire.

Passando dal discreto al continuo, come ci si aspetta, le cose non vanno più così bene, e sarà necessario specificare P con mezzi diversi.

1.2 Proprietà elementari

Deduciamo dagli assiomi su P le seguenti proprietà elementari.

Proposizione 1.3. *Sia \emptyset l'insieme vuoto. Allora $P(\emptyset) = 0$.*

Dimostrazione. Siccome $\Omega \cap \emptyset = \emptyset$, Ω e \emptyset sono disgiunti, quindi si applica l'assioma **3.** e si trova

$$1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + P(\emptyset),$$

da cui la tesi, sottraendo 1. □

Proposizione 1.4. *Se $\bar{A} := \Omega \setminus A$ indica il complementare dell'evento A , allora $P(\bar{A}) = 1 - P(A)$.*

Dimostrazione. Poiché A e \bar{A} sono disgiunti e la loro unione è Ω ,

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

da cui la tesi. □

Proposizione 1.5. *Se A e B sono due eventi qualsiasi³, allora*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (1.1)$$

Dimostrazione. Sia $C := A \cap B$ e $A' := A \setminus B = A \setminus C$ (il disegno di un diagramma di Venn chiarisce molto la costruzione). Allora A' e C sono disgiunti e hanno unione A , inoltre A' e B sono disgiunti e hanno unione $A \cup B$, da cui applicando più volte gli assiomi,

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

da cui la tesi. □

Proposizione 1.6. Additività finita.⁴ *Siano dati n eventi A_1, A_2, \dots, A_n a due a due disgiunti, allora*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i). \quad (1.2)$$

²Si dice evento elementare un evento costituito da un solo esito. Si tratta quindi un singoletto (ad esempio $\{2\spadesuit\}$).

³In questo senso questa proposizione estende il risultato dell'assioma **3.**, correggendolo nel caso di eventi non disgiunti.

⁴Questa proprietà è ovviamente un banale corollario dell'assioma **3b.**, tuttavia essa può essere dimostrata anche a partire dal solo assioma **3.** ed è questo che proviamo.

Dimostrazione. La dimostrazione è fatta per induzione: per $n = 1$ essa si riduce a $P(A_1) = P(A_1)$ ed è quindi verificata banalmente.

Supponiamola ora vera per n e dimostriamola per $n + 1$. Sia $B_n := \bigcup_{i=1}^n A_i$. Per l'ipotesi induttiva, $P(B_n) = \sum_{i=1}^n P(A_i)$. È anche immediato che A_{n+1} e B_n sono disgiunti, quindi per il solito assioma,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P(B_n \cup A_{n+1}) = \\ &= P(B_n) + P(A_{n+1}) = \sum_{i=1}^n P(A_i) + P(A_{n+1}) = \sum_{i=1}^{n+1} P(A_i). \end{aligned}$$

Da questo segue immediatamente la tesi. \square

Proposizione 1.7. Formula di inclusione-esclusione. *Siano dati n eventi qualunque⁵ A_1, A_2, \dots, A_n ; allora*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{\substack{i,j=1 \\ i < j}}^n P(A_i \cap A_j) + \\ &+ \sum_{\substack{i,j,k=1 \\ i < j < k}}^n P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n). \quad (1.3) \end{aligned}$$

Anche questa proprietà si dimostra per induzione, tuttavia lo svolgimento è più tedioso che utile, e non viene incluso.

1.3 Esempi di spazi discreti

Viene di seguito esemplificato in cosa consista costruire uno spazio di probabilità partendo da un esercizio. Il caso di spazi continui verrà esemplificato solo in seguito.

Esempio 1.8. *Si lancia due volte un dado a quattro facce. Qual è la probabilità che la somma dei due tiri sia 5?*

Gli esiti di due lanci casuali sono coppie di numeri da 1 a 4. La scelta più semplice per Ω è quindi

$$\Omega = \{1, 2, 3, 4\} \times \{1, 2, 3, 4\} = \{(1, 1), (1, 2), \dots, (4, 4)\}.$$

È ovviamente un insieme finito, $\#\Omega = 16$, quindi per quanto detto in precedenza (cfr. seconda osservazione di Pag. 3), è sufficiente che assegnamo le probabilità agli eventi elementari (insomma, a ogni singola coppia), in modo che la loro somma sia 1. In mancanza per ora di strumenti ulteriori, accettiamo che è *sensato* che ogni coppia sia equiprobabile, per cui poniamo, con leggero abuso di notazione

$$P(i, j) = P(\{(i, j)\}) = \frac{1}{16}.$$

⁵In questo senso la formula di inclusione-esclusione estende la Proposizione 1.5 a un numero qualsiasi di eventi, oppure estende la Proposizione 1.6 a eventi non necessariamente disgiunti.

Questo completa la costruzione dello spazio di probabilità per questo esercizio.

Per procedere con lo svolgimento, analizziamo ora l'evento

$$A := \text{“somma dei due dadi pari a 5”},$$

che per essere più formali ed espliciti si scrive così:

$$A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}.$$

Siamo interessati a $P(A)$: mostriamo con dettaglio perfino eccessivo come si applica l'assioma **3.** per risolvere il problema.

$$\begin{aligned} P(A) &= P(\{(1, 4)\} \cup \{(2, 3)\} \cup \{(3, 2)\} \cup \{(4, 1)\}) \stackrel{\text{3.}}{=} \\ &\stackrel{\text{3.}}{=} P(1, 4) + P(2, 3) + P(3, 2) + P(4, 1) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4} \end{aligned}$$

Osservazione. Si noti per inciso come sia del tutto generale, quando Ω è finito e composto da eventi equiprobabili, che per un qualsiasi evento A , sia per forza $P(A) = \frac{\#A}{\#\Omega}$.

Esempio 1.9. Si estraggono due carte a caso da un mazzo da poker. Qual è la probabilità di ottenere una coppia (nel senso del punto del poker)?

Questo esercizio si presta bene a illustrare come lo spazio degli esiti Ω possa essere scelto in modi diversi. Vediamo tre approcci tra quelli possibili.

Sia innanzitutto \mathcal{M} l'insieme di tutte le 52 carte.

$$\mathcal{M} = \{A\heartsuit, 2\heartsuit, \dots, K\heartsuit, A\diamondsuit, \dots, K\diamondsuit, A\clubsuit, \dots, Q\clubsuit, K\clubsuit\} \quad (1.4)$$

a) Paia di carte ordinate. Ogni esito è costituito da un paio di carte, la prima e la seconda che vengono estratte. Poniamo quindi

$$\begin{aligned} \Omega = \{(c_1, c_2) \in \mathcal{M} \times \mathcal{M} \mid c_1 \neq c_2\} &= \{(A\heartsuit, 2\heartsuit), (A\heartsuit, 3\heartsuit), \dots, \\ &\quad (A\heartsuit, K\heartsuit), (2\heartsuit, A\heartsuit), (2\heartsuit, 3\heartsuit), \dots, (K\clubsuit, Q\clubsuit)\}. \end{aligned}$$

Questo è un insieme finito di $52 \cdot 51 = 2652$ elementi, siamo quindi autorizzati a dare la probabilità dei soli eventi elementari. Decidiamo in assenza di ulteriori informazioni che essi siano tutti equiprobabili (è sensato che non vi siano paia di carte più probabili di altre). Col solito abuso di notazione,

$$P(c_1, c_2) = 2652^{-1} \quad \forall (c_1, c_2) \in \Omega.$$

Finito lo studio di (Ω, P) passiamo all'evento richiesto dall'esercizio.

$$\begin{aligned} A := \{(A\heartsuit, A\diamondsuit), (A\heartsuit, A\clubsuit), (A\heartsuit, A\spadesuit), (2\heartsuit, 2\diamondsuit), \dots, \\ (A\diamondsuit, A\heartsuit), \dots, (K\spadesuit, K\clubsuit)\}. \end{aligned}$$

Questo evento è composto da tutte le paia di carte che costituiscono una coppia. Per contare tutti gli elementi senza dimenticarne e senza prenderne nessuno più volte, un metodo è quello illustrato dall'ordine con cui sono stati

scritti qui sopra. Si fissa una carta (inizialmente, l'asso di cuori), e si scorrono tutte le coppie che cominciano con essa (sono ovviamente tre, una per ogni altro seme). Poi si cambia la prima carta con la successiva (due di cuori), e si procede⁶.

Alla fine la cardinalità di A si deduce quindi essere $52 \cdot 3 = 156$, da cui si ricava immediatamente (sempre grazie all'assioma **3.**),

$$P(A) = \frac{156}{2652} = \frac{1}{17}.$$

b) Paia di carte senza ordine. Visto che nel primo caso abbiamo contato ogni coppia due volte, un'alternativa allettante consiste nel contare le paia e le coppie senza ordine.

Per fare questo, teniamo una sola versione di ogni paio, e decidiamo per convenzione di tenere quella che ha la prima carta che precede la seconda (rispetto all'ordine con cui le abbiamo scritte nella definizione di \mathcal{M} , equazione 1.4). Quindi ad esempio, tra $(A\heartsuit, A\diamondsuit)$ e $(A\diamondsuit, A\heartsuit)$ teniamo la prima; resta quindi inteso cosa intendiamo per $c_1 < c_2$ nella definizione seguente.

$$\Omega = \{(c_1, c_2) \in \mathcal{M} \times \mathcal{M} \mid c_1 < c_2\}.$$

In questo caso si vede bene che $\#\Omega = \binom{52}{2} = \frac{52 \cdot 51}{2} = 1326$. Di nuovo Ω è uno spazio finito. Di nuovo possiamo e vogliamo assegnare P sugli eventi elementari. Ogni paio è equiprobabile, e quindi,

$$P(c_1, c_2) = 1326^{-1} \quad \text{se } (c_1, c_2) \in \Omega.$$

Passando ad A , conviene contare le coppie tenendo prima fisso il valore della carta, facendo variare tutte le $\binom{4}{2} = 6$ combinazioni di due semi, e poi far variare il valore sui 13 possibili:

$$A = \{(A\heartsuit, A\diamondsuit), (A\heartsuit, A\clubsuit), (A\heartsuit, A\spadesuit), (A\diamondsuit, A\clubsuit), (A\diamondsuit, A\spadesuit), (A\clubsuit, A\spadesuit), (2\heartsuit, 2\diamondsuit), \dots, (K\clubsuit, K\spadesuit)\}, \quad (1.5)$$

e quindi $\#A = 13 \cdot 6 = 78$ e si ritrova $P(A) = \frac{78}{1326} = \frac{1}{17}$.

c) Ordinamenti per tutto il mazzo. Questo è un esempio da non seguire: possiamo definire Ω anche come l'insieme di tutti i possibili modi in cui ci si può presentare il mazzo dopo che abbiamo mescolato e prima di pescare le due carte.

Osservazione. Questo approccio è sicuramente eccessivo per il problema che stiamo risolvendo, ma ciò non inficia il procedimento, che porta comunque al risultato voluto, e in ogni caso una domanda ulteriore⁷ nel problema avrebbe potuto costringerci a procedere così fin dall'inizio.

⁶Si noti come a un certo punto si prende in considerazione anche la coppia ottenuta invertendo la prima. In realtà, questo succede per ogni coppia.

⁷Si pensi per esempio a questa richiesta. Supposto che il paio di carte pescate sia una coppia, qual è la probabilità che nelle tre carte successive ve ne sia almeno una che permetta di arrivare al tris?

Dunque un esito è una successione di 52 carte tutte diverse, in un ordine particolare, e Ω sarà l'insieme composto da tutte le *permutazioni* diverse di queste carte.

$$\Omega = \{(c_1, c_2, \dots, c_{52}) \in \mathcal{M}^{52} \mid c_i \neq c_j \quad \forall i \neq j\},$$

$$\#\Omega = 52 \cdot 51 \cdot \dots \cdot 2 \cdot 1 = 52!,$$

$$P(\omega) = \frac{1}{52!} \quad \forall \omega \in \Omega.$$

Gli elementi di A sono vettori di 52 carte che iniziano con una qualunque, a cui segue una delle tre che con essa fanno coppia, a cui seguono le altre 50 in un ordine qualsiasi.

$$\#A = 52 \cdot 3 \cdot 50!,$$

$$P(A) = \frac{\#A}{\#\Omega} = \frac{52 \cdot 51 \cdot \dots \cdot 2 \cdot 1 = 52!}{52 \cdot 3 \cdot 50!} = \frac{1}{17}.$$

Esempio 1.10. *Un dado a 12 facce viene lanciato ripetutamente fino a che non esce l'uno. Sia T il numero di lanci fatti. Qual è la probabilità che T sia pari?*

Dare subito una definizione per Ω non è facile. Come esiti si potrebbero scegliere le sequenze di numeri tirati, ma queste non hanno tutte la stessa lunghezza. In particolare se $n \geq 1$, l'insieme delle possibili sequenze di lunghezza n è

$$A_n = \{(d_1, d_2, \dots, d_n) \in \{1, 2, \dots, 12\}^n \mid d_i \neq 1, i < n, \quad d_n = 1\}.$$

Tuttavia per fortuna non occorre che Ω abbia una particolare struttura⁸, è lecito in particolare riunire tutti questi insiemi di esiti nello spazio finale.

$$\Omega = A_1 \cup A_2 \cup A_3 \cup \dots = \bigcup_{n=1}^{\infty} A_n.$$

Passando alla cardinalità, si vede facilmente che quella degli A_n è finita per ogni n :

$$\#A_1 = 1, \quad \#A_2 = 11, \quad \#A_3 = 121, \quad \dots, \quad \#A_n = 11^{n-1},$$

è però altrettanto ovvio che la cardinalità di Ω è infinita (numerabile), $\#\Omega = \sum_{n=1}^{\infty} \#A_n = \infty$.

Dobbiamo senz'altro aspettarci qualche fenomeno nuovo in questo caso, tuttavia per la stessa osservazione citata per i casi finiti, non cambia nulla per quanto riguarda l'assegnazione delle probabilità agli eventi di Ω . Di nuovo possiamo assegnarle ai soli eventi elementari (sempre in modo che la loro somma⁹

⁸Finora si era riusciti a definire Ω in una sola volta e in modo piuttosto elegante. Inoltre gli esempi riportati fino ad adesso erano sottoinsiemi di semplici prodotti cartesiani di insiemi. Questo non accade sempre, ma come prova questo esempio, la cosa non deve disturbare più di tanto.

⁹Anzi, serie.

sia pari a uno), e invocando l'assioma **3b.** averle definite anche su tutti gli altri eventi (finiti o infiniti).

Procediamo per piccoli passi. Se $\omega \in \Omega$, vuol dire che $\omega \in A_n$ per qualche numero positivo n . Fissato questo numero, è vero che $\omega = (d_1, d_2, \dots, d_n)$ è un vettore di lunghezza n le cui componenti sono numeri dal 2 al 12, tranne l'ultimo che è un 1.

La probabilità di avere lanciato in successione la particolare sequenza (d_1, d_2, \dots, d_n) è chiaramente $\left(\frac{1}{12}\right)^n$. Questo numero dipende da n , quindi in questo esempio gli eventi elementari *non* sono tutti equiprobabili¹⁰, essendolo solo a parità di lunghezza (mentre una successione qualsiasi di lunghezza 3 è sempre molto più probabile di una di lunghezza 5).

Raccogliendo queste considerazioni abbiamo

$$P(\{\omega\}) = \left(\frac{1}{12}\right)^n \quad \text{se } \omega \in A_n.$$

Verifichiamo per prudenza che la serie di questi numeri sia pari a uno.

$$P(A_n) = \#A_n \cdot \left(\frac{1}{12}\right)^n = \frac{1}{12} \left(\frac{11}{12}\right)^{n-1},$$

da cui

$$\sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{n=1}^{\infty} P(A_n) = \frac{1}{12} \sum_{n=1}^{\infty} \left(\frac{11}{12}\right)^{n-1} = \frac{1}{12} \sum_{n=0}^{\infty} \left(\frac{11}{12}\right)^n = 1,$$

e questo esaurisce lo studio dello spazio di probabilità.

Venendo all'evento $A := "T \text{ è pari}"$, esso si esplicita come segue:

$$A = \bigcup_{k=1}^{\infty} \{T = 2k\} = \bigcup_{k=1}^{\infty} A_{2k},$$

da cui

$$\begin{aligned} P(A) &= P\left(\bigcup_{k=1}^{\infty} A_{2k}\right) = \sum_{k=1}^{\infty} P(A_{2k}) = \frac{1}{12} \sum_{k=1}^{\infty} \left(\frac{11}{12}\right)^{2k-1} = \\ &= \frac{11}{144} \sum_{k=0}^{\infty} \left(\frac{121}{144}\right)^k = \frac{11}{23}. \end{aligned}$$

E questo conclude l'esempio.

1.4 Spazi di probabilità continui

Gli spazi più che numerabili sono di centrale importanza in teoria della probabilità. Infatti \mathbb{R} è un insieme di questo tipo, e quindi non appena il risultato

¹⁰È un fatto del tutto generale che negli spazi di probabilità numerabili gli eventi elementari non siano mai tutti equiprobabili. Perché? E in quelli continui può succedere?

di una prova è costituito da un numero reale casuale (pensiamo alle misurazioni soggette ad errore sperimentale, ad esempio), è assolutamente necessario ricorrervi.

In effetti \mathbb{R} è l'esempio di spazio continuo che studieremo più a lungo e con più cura, e quando lo lasceremo sarà solo per generalizzare allo studio di \mathbb{R}^d e dei suoi sottoinsiemi.

Proprio per questo, è opportuno ora dare un esempio diverso, e che siamo in grado di produrre costruttivamente con le informazioni a nostra disposizione. Esso sarà particolarmente utile per percepire le differenze e le similitudini con i casi discreti finora analizzati.

Esempio 1.11. *Una moneta con facce '0' e '1' viene lanciata infinite volte, prendendo nota della successione di cifre ottenute, che poi vengono usate come sviluppo binario – a destra della virgola – del numero reale casuale X ¹¹. Qual è la probabilità di ottenere un certo numero reale fissato? Qual è la probabilità che $X \in [a, b]$?*

Anche se X è un numero reale, lo spazio degli esiti più immediato non è \mathbb{R} . Prendiamo infatti per Ω l'insieme di tutte le successioni infinite di cifre '0' e '1'.

$$\Omega = \{(\omega_1, \omega_2, \dots, \omega_n, \dots) \in \{0, 1\}^{\mathbb{N}_+}\}.$$

Ricordiamo che nel caso di spazi continui¹² non è a priori sufficiente assegnare la probabilità degli eventi elementari, come lo era nel caso discreto.

Cominceremo ad assegnare le probabilità dagli insiemi per i quali essa è ovvia. Definiamo gli eventi $C_{\vec{a}}$ come segue:

$$C_0 = \{\omega \in \Omega \mid \omega = (0, \omega_2, \omega_3, \dots) \text{ per opportuni } \omega_2, \omega_3, \dots\},$$

$$C_1 = \{\omega \in \Omega \mid \omega = (1, \omega_2, \omega_3, \dots) \text{ per opportuni } \omega_2, \omega_3, \dots\},$$

$$C_{(0,0)} = \{\omega \in \Omega \mid \omega = (0, 0, \omega_3, \omega_4, \dots) \text{ per opportuni } \omega_3, \omega_4, \dots\},$$

e più in generale, se $\vec{a} = (a_1, a_2, \dots, a_n) \in \{0, 1\}^n$,

$$C_{\vec{a}} = \{\omega \in \Omega \mid \omega = (a_1, a_2, \dots, a_n, \omega_{n+1}, \dots) \text{ per opportuni } \omega_{n+1}, \dots\},$$

in modo tale che $C_{\vec{a}}$ sia l'insieme di tutte le successioni che hanno le prime n cifre fissate ai valori a_1, a_2, \dots, a_n .

È abbastanza semplice convincersi che se il numero di componenti di \vec{a} è n , allora dobbiamo porre $P(C_{\vec{a}}) = \frac{1}{2^n}$. Infatti ad esempio, C_0 è l'evento delle successioni che iniziano con 0; quindi cercare la sua probabilità equivale a domandarsi qual è la probabilità che il primo lancio di moneta abbia dato uno 0, senza porre condizioni sulle altre monete. Chiaramente questa quantità è $\frac{1}{2}$.

Più in generale, $C_{\vec{a}}$ è l'evento che fissa i primi n lanci di moneta ai valori assegnati a_1, a_2, \dots, a_n , quindi la sua probabilità sarà $\frac{1}{2^n}$.

¹¹Ad esempio, se la successione fosse $(1, 1, 1, 0, 1, 0, 0, 1, \dots)$, allora avremmo in rappresentazione binaria $X = 0.11101001\dots$, ovvero $X = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{32} + \frac{1}{256} + \dots$

¹²Si provi a dimostrare che Ω non è numerabile.

Osservazione. Fino a qui abbiamo assegnato le probabilità a tutti gli eventi $C_{\bar{a}}$. Questi sono un'infinità di sottoinsiemi che sminuzzano Ω in maniera via via più fine man mano che n cresce, e quindi forniscono già una analisi di Ω abbastanza dettagliata. Possiamo allora fermarci qui e cercare le probabilità che ci sono richieste applicando solo assiomi e proprietà, o dobbiamo sforzarci di assegnare probabilità ad altri eventi non raggiungibili altrimenti? Sicuramente tali eventi esistono (si veda la nota a piè di pagina 2), ma proprio per questo non si può sperare eliminarli tutti. La filosofia giusta è quella del risparmio: se si riesce a risolvere l'esercizio, non occorre fare altro.

Vediamo ora che l'esercizio è già risolvibile con i dati forniti fino a qui.

Probabilità di ottenere un reale x fissato. Notiamo intanto che $X \in [0, 1]$ per costruzione, quindi ha senso confrontarlo solo con gli $x \in [0, 1]$, essendo altrimenti $P(X = x) = 0$.

Sia allora ω una delle possibili¹³ successioni di cifre dello sviluppo binario di x , e sia $\omega = (\omega_1, \omega_2, \dots)$. Allora

$$\Omega \supset C_{\omega_1} \supset C_{(\omega_1, \omega_2)} \supset C_{(\omega_1, \omega_2, \omega_3)} \supset \dots \supset \{\omega\},$$

da cui usando la proprietà che $A \subset B \Rightarrow P(A) \leq P(B)$, si trova

$$P(\Omega) \geq P(C_{\omega_1}) \geq P(C_{(\omega_1, \omega_2)}) \geq \dots \geq P(\{\omega\}),$$

per cui, ricordando che $P(C_{(\omega_1, \dots, \omega_n)}) = \frac{1}{2^n}$, si ha, per ogni n , $P(\{\omega\}) \leq \frac{1}{2^n}$, e quindi

$$P(\{\omega\}) = 0.$$

Siccome x può avere una o due rappresentazioni binarie, l'evento $\{X = x\}$ o è un singoletto $\{\omega\}$, oppure l'unione di due singoletti $\{\omega', \omega''\}$. In entrambi i casi, $P(X = x) = 0$ perché ogni singoletto è trascurabile.

Probabilità che X cada tra due reali fissati. Ci chiediamo adesso quanto vale la quantità $P(X \in [s, t])$, per $0 \leq s \leq t \leq 1$.

È molto utile da qui in poi identificare Ω con $[0, 1]$. Sia allora $\varphi : \Omega \rightarrow [0, 1]$ l'applicazione che associa a ogni successione di cifre $\omega = (\omega_1, \omega_2, \dots)$ il numero reale con sviluppo binario $0.\omega_1\omega_2\dots$.

φ è surgettiva ma non iniettiva, perché ci sono punti x che hanno due sviluppi binari (uno che è sempre '0' da un certo punto in poi e uno che è sempre '1' da un certo punto in poi). Se però togliamo da Ω le successioni che sono definitivamente '1',

$$\tilde{\Omega} := \Omega \setminus \{\omega \in \Omega \mid \omega \text{ è definitivamente '1'}\},$$

si trova che la restrizione di φ a $\tilde{\Omega}$ ¹⁴ è *biunivoca* su $[0, 1]$ (che è il nuovo codominio).

¹³Vi sono numeri reali che hanno una sola rappresentazione binaria, e altri che ne hanno due. Infatti se x ammette uno sviluppo con un numero finito di cifre (ad esempio, $\frac{1}{4} = 0.01$), ne ammette anche un altro che è 1 da un certo punto in poi (nell'esempio, $\frac{1}{4} = 0.001111\dots$)

¹⁴Ci si potrebbe domandare se non abbiamo modificato gravemente lo spazio di probabilità, eliminando brutalmente una infinità di esiti. Si provi intanto per cominciare che $\Omega \setminus \tilde{\Omega}$ è numerabile: questo basterà a dedurre che $P(\Omega \setminus \tilde{\Omega}) = 0$, e quindi che "non abbiamo tolto poi molto" (si veda anche la sezione 1.4.1).

Vediamo cosa succede agli eventi $C_{\vec{a}}$. Se $\vec{a} = (a_1, a_2, \dots, a_n)$, allora $C_{\vec{a}}$ è costituito dalle successioni che cominciano con quelle n cifre, mentre $\varphi(C_{\vec{a}})$ è l'insieme dei punti reali i cui sviluppi binari cominciano con $0.a_1a_2 \cdots a_n$, e perciò sono tutti i punti da $0.a_1a_2 \cdots a_n 0000 \cdots$ fino a $0.a_1a_2 \cdots a_n 1111 \cdots$, ovvero, sono l'intervallo $[\alpha, \alpha + \frac{1}{2^n}]$, con $\alpha = 0.a_1a_2 \cdots a_n$.

Da quanto detto si deduce facilmente che $\varphi(\tilde{\Omega} \cap C_{\vec{a}}) = [\alpha, \alpha + \frac{1}{2^n})$ e ad esempio si ha

$$\varphi(\tilde{\Omega} \cap C_0) = \left[0, \frac{1}{2}\right),$$

$$\varphi(\tilde{\Omega} \cap C_1) = \left[\frac{1}{2}, 1\right),$$

$$\varphi(\tilde{\Omega} \cap C_{(0,0)}) = \left[0, \frac{1}{4}\right).$$

Ma allora

$$\begin{aligned} P\left(X \in \left[\alpha, \alpha + \frac{1}{2^n}\right)\right) &= P\left(\tilde{\Omega} \cap \left\{X \in \left[\alpha, \alpha + \frac{1}{2^n}\right)\right\}\right) = \\ &= P\left(\omega \in \tilde{\Omega} \mid \varphi(\omega) \in \left[\alpha, \alpha + \frac{1}{2^n}\right)\right) = P(\tilde{\Omega} \cap C_{\vec{a}}) = P(C_{\vec{a}}) = \frac{1}{2^n}, \end{aligned}$$

dove abbiamo sfruttato più volte il fatto che qualunque sia l'evento A , visto che $P(\tilde{\Omega}) = 1$, $P(A \cap \tilde{\Omega}) = P(A)$.

Siccome abbiamo provato che $\{X = x\}$ è trascurabile, si ha sempre

$$P(X \in [s, t]) = P(X \in [s, t)).$$

Quindi, se $[s, t)$ è del tipo $[\alpha, \alpha + \frac{1}{2^n})$, il problema è risolto, e vale

$$P(X \in [s, t]) = t - s.$$

Ma anche se s e t sono generici, l'intervallo $[s, t)$ si spezza facilmente¹⁵ in una unione (numerabile) disgiunta di intervalli del tipo richiesto:

$$[s, t) = \cdots \cup [\alpha_{-2}, \alpha_{-1}) \cup [\alpha_{-1}, \alpha_0) \cup [\alpha_0, \alpha_1) \cup [\alpha_1, \alpha_2) \cup \cdots,$$

con

$$\lim_{n \rightarrow -\infty} \alpha_n = s \quad \lim_{n \rightarrow \infty} \alpha_n = t,$$

¹⁵Facciamo un esempio per convincerci: sia $s = 1/3$ e $t = 2/3$; poniamo allora (in binario) $\alpha_0 = 0.1$, $\alpha_1 = 0.101$, $\alpha_2 = 0.10101$, eccetera, e poi poniamo $\alpha_{-1} = 0.011$, $\alpha_{-2} = 0.01011$, eccetera. In questo modo α_n tende a s e a t per n che tende a $-\infty$ e $+\infty$, inoltre gli α_n hanno sviluppo binario finito e la distanza tra due consecutivi è una potenza negativa di 2. Perciò ogni $[\alpha_n, \alpha_{n+1})$ è del tipo richiesto.

da cui usando l'assioma di additività numerabile,

$$\begin{aligned} P(X \in [s, t]) &= \\ &= \cdots + P(X \in [\alpha_{-1}, \alpha_0]) + P(X \in [\alpha_0, \alpha_1]) + P(X \in [\alpha_1, \alpha_2]) + \cdots = \\ &= \cdots + (\alpha_{-1} - \alpha_0) + (\alpha_0 - \alpha_1) + (\alpha_1 - \alpha_2) + \cdots = t - s. \end{aligned}$$

Per cui anche nel caso generale vale $P(X \in [s, t]) = t - s$, e questo risolve il problema.

Osservazione. Il lungo esempio che abbiamo appena concluso è molto istruttivo. Abbiamo visto che gli eventi elementari sono tutti trascurabili, e anche che assegnare la probabilità ad una famiglia opportuna di sottoinsiemi di Ω si è rivelato sufficiente alle esigenze dell'esercizio. Inoltre questi fatti sono stati *pro-vati* in questo caso particolare partendo dalle solite assunzioni sul lancio di una moneta. Di seguito essi saranno *assunti* per semplicità e per lavorare con buona generalità, tuttavia la loro correttezza è a questo punto evidente al lettore.

Osservazione. Se si legge l'esercizio dal punto di vista di X , esso ha come Ω alternativo l'intervallo reale $[0, 1]$. Siccome la probabilità che X cada in un qualsiasi sottointervallo è la lunghezza di quest'ultimo, e non dipende quindi dalla sua posizione, si parla in questo caso di *distribuzione uniforme su* $[0, 1]$.

1.4.1 Quando lo spazio degli esiti è \mathbb{R}

In tutta questa sezione, sarà $\Omega = \mathbb{R}$, e si analizzeranno i vari modi di assegnare i valori di P .

Gli eventi elementari sono trascurabili. Ovvero i punti hanno probabilità zero. Questa richiesta non è necessaria, ma semplifica molto le cose, anche perché altrimenti si scivola spesso nel caso discreto.

Si supponga ad esempio che $P(1) = P(2) = \frac{1}{2}$. Si deduce facilmente che in queste ipotesi tutti gli altri punti sono trascurabili, come pure tutti gli eventi a cui 1 e 2 non appartengono. In pratica, questo spazio è in realtà quello discreto costituito dal lancio di una moneta (con facce 1 e 2), dissimulato da spazio continuo.

Per evitare patologie di questo tipo, assumiamo quindi che $P(\{x\}) = 0$ per ogni $x \in \Omega$

Non è sufficiente assegnare la probabilità agli eventi elementari. Questa è una grossa differenza col caso discreto. Lì era sufficiente, perché *ogni* evento A , essendo sottoinsieme di Ω era al più numerabile, e quindi scrivendo

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right),$$

l'unione era fatta su un numero al più numerabile di singoletti, per cui applicando l'assioma di additività numerabile si ricavava

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

Qui invece un evento A può essere o non essere numerabile. Nel primo caso i ragionamenti appena fatti sono validi, solo che $P(\{\omega\}) = 0$ per ogni ω e quindi

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = 0,$$

e anche A risulta trascurabile. Nel secondo caso invece nulla si può dire.

Insomma l'assioma di additività numerabile – se si è costretti a partire dai singoletti – non permette che di raggiungere eventi di cardinalità numerabile, e in quanto tali trascurabili. Gli eventi interessanti possono essere quindi solo eventi di cardinalità continua, e in quanto tali la loro probabilità va assegnata per forza. O almeno, va assegnata ad una famiglia opportuna di eventi continui, usando poi gli assiomi per ricavare quella degli altri eventi di interesse. Nell'Esempio 1.11 questa famiglia era costituita dagli eventi $C_{\vec{a}}$. Quando si lavora su \mathbb{R} , varie scelte sono possibili. Quelle tipiche sono gli intervalli $[a, b]$ o le semirette $(-\infty, a]$.

Esercizio 1.12. Si dimostri che se si assegna la probabilità di una qualsiasi delle seguenti famiglie, la probabilità di tutte le altre si può calcolare usando solo gli assiomi assegnati e il fatto che i punti sono trascurabili:

- Gli intervalli chiusi $[a, b]$, con $a \leq b$.
- Gli intervalli aperti (a, b) , con $a < b$.
- Gli intervalli $[a, b)$, con $a < b$.
- Gli intervalli $(a, b]$, con $a < b$.
- Le semirette sinistre chiuse $(-\infty, a]$, con $a \in \mathbb{R}$.
- Le semirette sinistre aperte $(-\infty, a)$, con $a \in \mathbb{R}$.
- Le semirette destre chiuse $[a, \infty)$, con $a \in \mathbb{R}$.
- Le semirette destre aperte (a, ∞) , con $a \in \mathbb{R}$.

Come assegnare la probabilità agli intervalli. Una volta deciso che tutto quello che dobbiamo fare è specificare le quantità $P([a, b])$ al variare di a e b , bisogna però che lo facciamo rispettando gli assiomi. Ad esempio, se $a < b < c$, bisognerà che $P([a, c]) = P([a, b]) + P([b, c])$ ¹⁶.

Inoltre la quantità di probabilità che vanno assegnate è spaventosa, e verificare la compatibilità degli assiomi in tutte le combinazioni risulta proibitivo.

Fortunatamente esistono modi per assegnare in maniera compatta la probabilità a tutti gli insiemi di interesse, così che gli assiomi siano rispettati in modo automatico.

1. Supponiamo che sia data una funzione $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ che sia

- (a) positiva;
- (b) integrabile;
- (c) tale che $\int_{-\infty}^{+\infty} \alpha(t) dt = 1$.

¹⁶Perché?

Allora, se si pone $P([a, b]) = \int_a^b \alpha(t) dt$ per ogni a e b con $a \leq b$, le probabilità assegnate soddisfano tutti gli assiomi.

2. Supponiamo che sia data una funzione $\mu : \mathbb{R} \rightarrow [0, 1]$ che sia

- (a) non decrescente;
- (b) continua;
- (c) tale che $\lim_{x \rightarrow -\infty} \mu(x) = 0$ e $\lim_{x \rightarrow +\infty} \mu(x) = 1$.

Allora, se si pone $P((-\infty, a]) = \mu(a)$ per ogni $a \in \mathbb{R}$, oppure direttamente $P([a, b]) = P((a, b]) = P((-\infty, b]) - P((-\infty, a]) = \mu(b) - \mu(a)$ per ogni a e b con $a \leq b$, le probabilità assegnate soddisfano di nuovo tutti gli assiomi.

Osservazione. L'interpretazione che va data di queste funzioni è la seguente. Se si disegna il grafico di α , l'area sotto la curva è 1, e rappresenta la probabilità totale di Ω . Se si vuole la probabilità di un intervallo, bisogna calcolare l'area sotto la curva delimitata dagli estremi dell'intervallo. Se si cerca l'area di un qualunque altro evento, bisogna comunque guardare l'area corrispondente nello stesso modo.

In questo senso, i punti dove α è grande, sono zone ad alta probabilità, anche se i punti in sé sono tutti trascurabili. Le regioni di \mathbb{R} dove α è zero, avranno tutte probabilità nulla.

Si noti che è ben possibile che α sia maggiore di 1 in qualche punto, in quanto essa non rappresenta una probabilità, ma una quantità che integrata dà una probabilità.

Nell'Esempio 1.11, se avessimo voluto prendere come Ω tutto \mathbb{R} e guardato agli esiti come i valori di X , la scelta corrispondente sarebbe stata:

$$\alpha(x) := \begin{cases} 1 & \text{se } 0 \leq x \leq 1 \\ 0 & \text{altrimenti} \end{cases},$$

ovvero, usando la notazione standard delle funzioni indicatrici,

$$\alpha(x) := \mathbf{1}_{[0,1]}(x).$$

Nel caso in cui P sia data tramite α , essa può essere sempre data anche tramite μ , e vale

$$\mu(x) = \int_{-\infty}^x \alpha(t) dt,$$

quindi sempre nell'Esempio 1.11, si poteva assegnare P tramite una μ a rampa:

$$\mu(x) := \begin{cases} 0 & \text{se } x \leq 0 \\ x & \text{se } 0 \leq x \leq 1 \\ 1 & \text{se } 1 \leq x \end{cases}.$$

2 Variabili aleatorie

Ciò che accomuna molti degli esempi di spazi di probabilità dati fino a qui è il fatto che Ω sia un insieme numerico. Anche se questa non è una caratteristica del tutto generale (abbiamo dato esempi di insiemi di esiti non numerici) e molti argomenti di probabilità teorica sarebbero privi di senso in questo limitato contesto, è innegabile che per moltissimi fini pratici il caso numerico sia quello veramente interessante.

Precisamente questa osservazione porta al concetto di variabile aleatoria. Il modo esatto in cui ciò avviene non è però immediato. Procediamo per gradi.

Definizione. Si dice *variabile aleatoria* (spesso useremo l'abbreviazione v.a.) una qualsiasi¹⁷ applicazione $X : \Omega \rightarrow \mathbb{R}$.

Un nome alternativo e meno altisonante è *numero casuale*. Perché in realtà non si tratta di niente altro che di un valore numerico che è determinato dal caso, in maniera non prevedibile, ma con delle regole fisse.

Ad esempio per un dado a quattro facce, non si può forse dire che numero uscirà, ma si può certo sostenere che sarà un numero tra l'1 e il 4, e che tutti hanno la stessa probabilità.

Esempio 2.1. Si lancia due volte un dado da gioco a quattro facce. Si denota con X il punteggio ottenuto dal primo lancio, con Y il punteggio ottenuto col secondo lancio, e con Z la somma dei due precedenti. Si noti come X , Y e Z siano v.a. secondo la definizione data.

Si può scegliere come Ω il prodotto cartesiano $\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$, e assegnare $\frac{1}{16}$ di probabilità a ciascun evento elementare, proprio come nell'Esempio 1.8. In questo modo, se il caso ha scelto l'esito $\omega = (2, 4)$, significa senza alcuna possibilità di errore (è una deduzione *deterministica*) che avremo $X = 2$, $Y = 4$ e $Z = 6$. Analogamente per qualunque altro valore della coppia tra quelle contenute in Ω , vi sarà un solo valore possibile per le tre v.a. dell'esempio.

Più formalmente, se vogliamo rispettare le ipotesi, dovremo avere che se $\omega = (i, j)$, allora $X(\omega) = i$, $Y(\omega) = j$ e $Z(\omega) = i + j$.

Dovrebbe essere ora chiaro che con la definizione assiomatica che abbiamo dato di spazio di probabilità, ogni grandezza numerica casuale non è altro che una applicazione deterministica e ben precisa che ad ogni esito dell'esperimento associa un valore in \mathbb{R} . Così la definizione data è giustificata anche dal punto di vista intuitivo.

Osservazione. Nell'Esempio 2.1 abbiamo definito diverse v.a. su uno stesso spazio di probabilità. Inoltre il comportamento di ciascuna era legato alle altre.

Noi affronteremo questa situazione in dettaglio nella sezione 4, mentre per ora ci limitiamo allo studio di una sola v.a. o al limite allo studio di più v.a. prese però singolarmente ed estraniare dal loro contesto.

¹⁷Si dovrebbe anche fare la richiesta che i sottoinsiemi di Ω tipo $X^{-1}([a, b])$ siano *eventi* per ogni $a \leq b$, tuttavia questa proprietà va a ricadere su una più precisa definizione di evento che non vogliamo dare in questa sede. Come al solito assumeremo sempre, salvo specificazioni, di stare lavorando con v.a. che soddisfano questi requisiti.

2.1 Lo spazio immagine

Per spiegare a qualcuno come si comporta un dado a quattro facce, è sufficiente dire che può assumere i valori interi 1, 2, 3 e 4, ciascuno con $1/4$ di probabilità. Nello stesso modo, per descrivere formalmente come opera una singola variabile aleatoria, sarà opportuno specificare sia l'insieme dei suoi valori possibili, sia le probabilità con cui essi vengono assunti, ovvero la *legge* della variabile aleatoria.

2.1.1 Valori possibili

Se $X : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria, non dobbiamo aspettarci che per forza essa possa assumere tutti i valori reali. Ad esempio un dado a quattro facce è una v.a. che – vista come applicazione – ha un codominio di quattro elementi.

Di conseguenza chiamiamo insieme dei valori possibili di X il suo codominio, che a volte indicheremo con $X(\Omega)$, e che potrà anche contenere numeri con probabilità di uscita nulla. La cosa importante è che è quasi certo che esca un elemento di tale insieme.

Esercizio 2.2. Si dimostri che per qualunque v.a., i valori possibili che abbiano probabilità non nulla devono essere al più numerabili.

Definizione. Sia X una v.a.: a seconda dei suoi valori possibili, con le loro probabilità, distinguiamo vari casi; X si dirà:

discreta se i valori possibili che abbiano probabilità non nulla hanno tutti insieme probabilità 1;

continua se tutti i valori possibili di X hanno probabilità nulla. In questo caso necessariamente essi dovranno essere un'infinità più che numerabile¹⁸;

mista se i valori possibili che abbiano probabilità non nulla hanno tutti insieme probabilità diversa da 1 e da 0.

Non approfondiremo oltre il caso delle v.a. miste in quanto la loro modesta utilità non giustifica lo sforzo necessario ad includerle nella trattazione.

2.1.2 Legge

Dopo che abbiamo fissato l'esperimento – ovvero abbiamo fissato Ω , P , e una v.a. X – siamo già in grado di trovare delle probabilità di eventi dipendenti da X , come le seguenti:

$$P(X = 3), \quad P(X > 2), \quad P(X \leq 2 | 1 < X \leq 3),$$

ovvero, ancora più in generale, se S è un sottoinsieme di \mathbb{R} , possiamo calcolare $P(X \in S)$.

In pratica, ad ogni insieme di numeri reali S possiamo associare un numero, che è la probabilità che X cada in quell'insieme. Questo è possibile anche se S non è contenuto nei valori possibili di X ; semplicemente, l'insieme dei valori possibili avrà probabilità uno di contenere il valore scelto dal caso per X .

¹⁸Perché?

In qualche modo X permette di tramutare P (che era una probabilità su Ω) in una probabilità P' su \mathbb{R} , così definita,

$$P'(S) = P(X \in S).$$

Tale misura di probabilità è detta *legge* o *distribuzione* di X , e il fatto che soddisfi gli assiomi è semplicemente una verifica, che viene lasciata al lettore interessato.

Osservazione. Operativamente, la legge di una v.a. è spesso ciò che i problemi assegnano, mentre nessuna imposizione viene fatta su Ω e non viene quasi mai precisata X come funzione di Ω in \mathbb{R} . Ad esempio, quando ci viene detto che si lancia un dado bilanciato, quello che significa è che lo spazio dei valori possibili è $\{1, 2, 3, 4, 5, 6\}$, e che gli elementi di questo sottoinsieme finito di \mathbb{R} sono equiprobabili secondo P' . Non solo in qualunque modo noi volessimo definire Ω, P e X (compatibilmente con questa legge) otterremmo gli stessi risultati, ma addirittura non è affatto necessario definire questi enti, e possiamo lavorare direttamente usando la legge di X .

Osservazione. Non bisogna pensare che questo approccio renda meno utile il lavoro fatto sugli spazi di probabilità generali. Da adesso infatti sfrutteremo i risultati ottenuti in quel campo applicandoli alla legge di X , che è una probabilità su \mathbb{R} .

Una considerazione un po' avanzata, ma a volte illuminante: assegnare la legge di una v.a. X in realtà significa costruire un semplicissimo spazio degli esiti per ambientare l'esperimento, ovvero \mathbb{R} stesso, con la legge assegnata come misura di probabilità. Rispetto a questo spazio X sarà la funzione identità da \mathbb{R} in sé.

2.2 Funzione di ripartizione, densità di probabilità

Come si assegna la legge di una v.a.? Nello stesso modo in cui si assegna una misura di probabilità su \mathbb{R} , visto che di questo alla fine si tratta.

2.2.1 Variabili aleatorie continue

Cominciando dal caso continuo, che è un po' più sistematico, e vediamo gli oggetti analoghi delle funzioni α e μ definite nell'ultima parte della sezione 1.4.1.

Definizione. Se X è una v.a. reale, si chiama *funzione di ripartizione* (da ora in poi, FR) di X l'applicazione $F_X : \mathbb{R} \rightarrow [0, 1]$ così definita,

$$F_X(x) := P(X \leq x). \quad (2.1)$$

Proposizione 2.3. Ogni FR F_X è caratterizzata dalle seguenti proprietà che si deducono direttamente dalla definizione (2.1)

1. è non decrescente;
2. è in ogni caso continua a destra, mentre è addirittura una funzione continua se X è una v.a. continua;

3. valgono i seguenti limiti all'infinito,

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow +\infty} F_X(x) = 1. \quad (2.2)$$

Dimostrazione.

1. Se $a < b$, allora $(-\infty, a] \subset (-\infty, b]$, e anche $\{X \in (-\infty, a]\} \subset \{X \in (-\infty, b]\}$; quindi

$$F_X(a) = P(X \in (-\infty, a]) \leq P(X \in (-\infty, b]) = F_X(b),$$

che costituisce esattamente la non decrescenza di F_X .

2. Dire che F_X è continua a destra, significa dire che in ogni punto ha limite destro e sinistro, e il valore di F_X nel punto stesso è uguale al limite da destra. Denotiamo invece con $F_X(x_-)$ il limite sinistro:

$$F_X(x_-) := \lim_{y \uparrow x} F_X(y) \leq \lim_{y \downarrow x} F_X(y) =: F_X(x). \quad (2.3)$$

Non dimostreremo la proprietà in quanto essa richiede (semplici) elementi di teoria della misura che abbiamo tralasciato per non appesantire la trattazione.

Facciamo piuttosto notare che la differenza $F_X(x) - F_X(x_-)$ tra limite da destra e limite da sinistra è pari alla probabilità che $X = x$. Ecco quindi che F_X è continua se e solo se ogni punto ha probabilità zero, ovvero se e solo se X è a sua volta continua come variabile aleatoria.

3. Anche per questo risultato una dimostrazione rigorosa è al di sopra delle possibilità della teoria che abbiamo sviluppato, tuttavia non è difficile convincere il lettore della sua correttezza.

Infatti se x tende a $-\infty$ le probabilità $P(X \leq x)$ decrescono sempre più, perché corrispondono a eventi sempre più piccoli (uno contenuto nell'altro). Il fatto che tendano a zero non è ovvio, ma è vero.

Nello stesso modo si affronta l'altro limite, che corrisponde ad eventi sempre più grandi, tendenti a coprire tutto \mathbb{R} ; e \mathbb{R} ha probabilità uno.

□

Definizione. Se X è una v.a. continua, allora si chiama *funzione di densità di probabilità* o semplicemente densità (da ora in poi, FD) di X l'applicazione $f_X : \mathbb{R} \rightarrow \mathbb{R}_+$ così definita,

$$f_X(x) := \frac{d}{dx} F_X(x). \quad (2.4)$$

Proposizione 2.4. Ogni FD f_X è caratterizzata dalle seguenti proprietà che si deducono direttamente dalla definizione (2.4) e dalla Proposizione 2.3:

1. è non negativa;
2. è integrabile;

3. ha massa complessiva unitaria:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (2.5)$$

Gli insiemi S più frequenti di cui dobbiamo sapere calcolare $P(X \in S)$ sono gli intervalli. Il seguente enunciato evidenzia come sia possibile accedere a queste probabilità partendo dalla FR o dalla FD.

Proposizione 2.5. *Se X è una v.a. continua con FR F_X e densità f_X , e $a < b$ sono numeri reali qualsiasi, si ha*

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx. \quad (2.6)$$

Dove – proprio perché X è continua – nel primo termine gli estremi a e b possono essere inclusi o esclusi a piacimento senza cambiare il valore della probabilità.

Dimostrazione. Poiché l'evento $\{X \leq b\}$ è l'unione disgiunta degli eventi $\{X \leq a\}$ e $\{a < X \leq b\}$, per l'additività della misura di probabilità si ha

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b),$$

che riarrangiando i termini diventa la prima uguaglianza. La seconda segue dal fatto che F_X è una primitiva di f_X .

L'ultima considerazione infine, si deduce perché gli eventi $\{X = a\}$ e $\{X = b\}$ sono trascurabili, e quindi possono essere aggiunti o tolti senza modificare le probabilità. \square

Il seguente risultato viene dato senza dimostrazione.

Corollario. Se X è una v.a. continua con densità f_X , e S è un sottoinsieme di \mathbb{R} , si ha

$$P(X \in S) = \int_S f_X(x) dx. \quad (2.7)$$

Osservazione. Una conseguenza talvolta utile di questo risultato è il fatto che due FD f_1 e f_2 , che differiscano in un numero finito o anche numerabile di punti, rappresentano la stessa v.a., infatti si può dimostrare che per ogni sottoinsieme S di \mathbb{R} si avrà in quel caso

$$\int_S f_1(x) dx = \int_S f_2(x) dx.$$

2.2.2 Variabili aleatorie discrete

Quando X è una v.a. discreta, l'insieme dei valori possibili $X(\Omega)$ è finito o numerabile, inoltre i singoli punti hanno probabilità positiva (se ce ne sono che hanno probabilità nulla, si possono trascurare tutti senza modificare la legge di X).

La cosa più semplice allora è descrivere la legge in termini degli eventi elementari, ovvero assegnare il valore di $P(X = x_i)$ per ogni valore possibile x_i di X , con l'unica restrizione che tali probabilità devono avere somma unitaria.

Osservazione. Per risolvere un buon numero di problemi questo primo approccio è più che sufficiente, tuttavia avere un'impostazione così diversa per le v.a. discrete ci costringerebbe a sviluppare due teorie differenti da qui in poi.

Per evitare questo inconveniente ed anche per seguire un approccio che permetta di approssimare una v.a. con un'altra anche se magari appartengono alle due diverse categorie, ci costringiamo ad estendere il linguaggio delle FR e delle FD al caso discreto.

Per quanto riguarda la FR, la definizione data con la (2.1) è del tutto generale e compatibile anche con le v.a. discrete. Le differenze che si evidenziano sono riassunte nel seguente enunciato.

Proposizione 2.6. *Se X è una v.a. discreta con insieme dei valori possibili $S := X(\Omega)$, e se S non ha punti di accumulazione, allora $F_X(x)$ è una funzione “a scala” che è costante nei tratti in cui $x \notin S$, e ha delle discontinuità di tipo salto (crescenti) nei punti di S . Inoltre l'ampiezza dei salti è esattamente la probabilità del corrispondente valore di ascissa:*

$$\forall x \in S, P(X = x) = F_X(x) - F_X(x_-). \quad (2.8)$$

Più complicato passare alla densità, ovvero la derivata della funzione di ripartizione. Infatti, dove la FR di una v.a. continua era continua e crescente, cosa che la rendeva derivabile quasi ovunque (si può dare un significato preciso a questa frase), per la FR di una v.a. discreta la regola sono le discontinuità.

Per potere “derivare” tali discontinuità introduciamo un oggetto nuovo, che tratteremo grosso modo come se fosse una funzione, anche se le sue proprietà fanno capire che appartiene a una categoria diversa.

Definizione. Si chiama δ di Dirac, o *impulso*, quella particolare funzione generalizzata $\delta : \mathbb{R} \rightarrow [0, +\infty]$ tale che

$$\delta(x) = \begin{cases} 0 & x \neq 0 \\ +\infty & x = 0, \end{cases} \quad (2.9)$$

con l'intesa che $\int_{-\infty}^{+\infty} \delta(x) dx = 1$, e più in generale, se $A \subset \mathbb{R}$, allora $\int_A \delta(x) dx = 1$ se e solo se $0 \in A$.

Useremo diverse volte il seguente utile risultato di cui non diamo dimostrazione.

Proposizione 2.7. Se $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione continua in x_0 , allora

$$\int_{-\infty}^{+\infty} g(x) \delta(x - x_0) dx = g(x_0). \quad (2.10)$$

Notiamo inoltre che dalla definizione di impulso segue immediatamente che

$$\int_{-\infty}^x \delta(x' - x_0) dx' = \begin{cases} 0 & x < x_0 \\ 1 & x \geq x_0 \end{cases} = \mathbf{1}_{[x_0, +\infty)}(x), \quad (2.11)$$

per cui pare ragionevole sostenere che la derivata della funzione $\mathbf{1}_{[x_0, +\infty)}(x)$ (che è caratterizzata da una discontinuità di tipo salto), sia un impulso traslato in x_0 . Forti di questa intuizione, diamo la seguente definizione.

Definizione. Se X è una v.a. discreta, si chiama *funzione di densità di probabilità* di X la funzione generalizzata

$$f_X(x) := \sum_{x_i \in X(\Omega)} P(X = x_i) \delta(x - x_i). \quad (2.12)$$

Proposizione 2.8. Anche se X è una v.a. discreta, si ha che

$$P(X \in S) = \int_S f_X(x) dx, \quad \forall S \subset \mathbb{R}, \quad (2.13)$$

$$F_X(x) = \int_{-\infty}^x f_X(x') dx'. \quad (2.14)$$

Dimostrazione. Evidentemente la (2.13) implica la (2.14) con $S = (-\infty, x]$, quindi proviamo solo la prima.

$$\begin{aligned} \int_S f_X(x) dx &= \int_S \sum_{x_i \in X(\Omega)} P(X = x_i) \delta(x - x_i) dx = \\ &= \sum_{x_i \in X(\Omega)} P(X = x_i) \int_S \delta(x - x_i) dx = \sum_{\substack{x_i \in X(\Omega) \\ x_i \in S}} P(X = x_i) \mathbf{1}_S(x_i) = \\ &= \sum_{\substack{x_i \in X(\Omega) \\ x_i \in S}} P(X = x_i) = P(X \in S), \end{aligned}$$

dove si sono usate le proprietà della δ di Dirac. \square

2.3 Classi standard di variabili aleatorie reali

Elenchiamo di seguito le più comuni v.a. caratterizzandole con la loro legge, espressa tramite FR, FD, oppure tramite le probabilità degli eventi elementari nei casi discreti.

2.3.1 Classi di variabili aleatorie discrete

Bernoulliana. X è una v.a. bernoulliana se assume solo i valori 0 e 1. Di solito si denota con p la probabilità $P(X = 1)$, e tale parametro caratterizza completamente la legge di X .

$$\begin{aligned}P(X = 1) &= p \\P(X = 0) &= 1 - p \\P(X \neq 0, X \neq 1) &= 0.\end{aligned}\tag{2.15}$$

Binomiale. X è una v.a. binomiale di parametri $n \in \mathbb{N}_+$ e $p \in [0, 1]$ se può assumere solo i valori interi $0, 1, \dots, n$ con probabilità

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.\tag{2.16}$$

Come vedremo più avanti, se un esperimento prevede n v.a. bernoulliane indipendenti l'una dall'altra (in un senso che definiremo), X rappresenta il numero di bernoulliane che hanno dato esito 1.

Ad esempio, se tiriamo n volte un dado e denotiamo con X il numero di lanci che hanno visto uscire il numero 5, allora X avrà legge binomiale di parametri n e $\frac{1}{6}$.

Geometrica. X è una v.a. geometrica di parametro $p \in [0, 1]$ se può assumere solo i valori interi positivi (lo zero va escluso) con probabilità

$$P(X = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots.\tag{2.17}$$

Questo tipo di v.a. può ad esempio rappresentare il numero di esiti di bernoulliane indipendenti che dobbiamo attendere prima di vedere il primo 1.

Ad esempio se tiriamo un dado da gioco fino a che non esce il 5, il numero di lanci che dobbiamo fare ha legge geometrica di parametro $\frac{1}{6}$.

Poissoniana. X è una v.a. poissoniana di parametro $\mu > 0$ se può assumere solo i valori interi non negativi (lo zero incluso) con probabilità

$$P(X = k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots.\tag{2.18}$$

Questa v.a. si usa per approssimare alcuni casi di distribuzioni binomiali, e compare in problemi di questo tipo:

1. il numero di persone che entrano in un negozio in un intervallo di tempo;
 2. il numero di telefonate che passano per una centralina in un intervallo di tempo;
 3. il numero di gocce di pioggia che cadono in un secchio;
 4. il numero di fiori in un prato;
- ... e così via.

Esercizio 2.9. In ciascuno dei casi di v.a. discrete presentate si dimostri che la somma delle probabilità di tutti i valori possibili è 1.

2.3.2 Classi di variabili aleatorie continue

Gaussiana standard. X ha legge normale (o gaussiana) standard se può assumere tutti i valori di \mathbb{R} , e ha FD come segue,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad (2.19)$$

nel quale caso scriveremo $X \sim N(0, 1)$.

La costante moltiplicativa scelta è dovuta all'imposizione (2.5), e al fatto che $\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}$.

Per quanto riguarda la funzione di ripartizione F_X , essa non può essere ottenuta analiticamente. Per questo la si denota con un simbolo apposito, che di solito è Φ ; i suoi valori si trovano tabulati su molti testi di statistica, inoltre i più comuni programmi di calcolo matematico per computer sono in grado di dare un'approssimazione numerica, come anche molte calcolatrici scientifiche,

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx =: \Phi(x). \quad (2.20)$$

Il grafico di f_X è a campana, simmetrico attorno all'origine dove si concentra, con più del 99% della massa tra -3 e 3 .

Proprio la simmetria di f_X rispetto all'asse delle ordinate, fa sì che $\Phi(x)$ sia simmetrica rispetto al punto $(0, 1/2)$, in modo che vale la seguente utile relazione:

$$\Phi(-x) = 1 - \Phi(x). \quad (2.21)$$

Gaussiana. X ha legge normale (o gaussiana) di parametri $\mu \in \mathbb{R}$ e $\sigma > 0$ se può assumere tutti i valori di \mathbb{R} , e ha FD come segue,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (2.22)$$

nel quale caso scriveremo $X \sim N(\mu, \sigma^2)$.

Come proveremo nell'Esempio 2.13 e successiva osservazione, anche se la FR di una v.a. gaussiana non ha una espressione esplicita, può tuttavia essere ottenuta tramite quella di una gaussiana standard come segue,

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (2.23)$$

Dimostreremo inoltre che se Y è una gaussiana standard, allora ponendo $X := \sigma Y + \mu$, la v.a. X risultante è una gaussiana $N(\mu, \sigma^2)$. Questo fa intuire che X è una versione dilatata e traslata di Y , e in particolare che il grafico di f_X è a campana, simmetrico attorno all'asse di ascissa μ dove si concentra, con più del 99% della massa tra $\mu - 3\sigma$ e $\mu + 3\sigma$.

Le v.a. gaussiane e gaussiane standard, rappresentano tipicamente l'imprecisione casuale degli esperimenti di misura, o gli errori sperimentali, o più in generale delle quantità che sono somma di un gran numero di piccoli contributi casuali indipendenti tra loro.

Esponenziale. X è una v.a. esponenziale di *intensità* $\lambda > 0$, se può assumere solo valori reali positivi, sui quali ha densità

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad (2.24)$$

mentre $f_X(x) = 0$ se $x < 0$. La FR di X è allora,

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (2.25)$$

Una v.a. di questo tipo è tipicamente un tempo di attesa, e in particolare le eventualità che accadono dopo un tempo aleatorio esponenziale, sono avvenimenti ineluttabili che però non accelerano la loro venuta se anche sono attesi da molto tempo. In pratica, se valori tipici per X sono dell'ordine di t , e – nonostante questo – dopo un tempo t ciò che stiamo aspettando non è ancora accaduto, questo non avvicina statisticamente l'avvenimento, che potrà capitare in un ulteriore lasso di tempo dell'ordine di t a partire da ora.

Questa complicata spiegazione ha la sua esatta controparte matematica nella seguente proprietà¹⁹: se X è una v.a. esponenziale, allora

$$P(X \leq t + t_0 | X > t_0) = P(X \leq t). \quad (2.26)$$

Uno dei pochi esempi di v.a. che sono precisamente esponenziali è il tempo di decadimento di una qualunque particella instabile. Invece il tempo di rottura di una lampadina ad esempio, non va bene (è solo approssimativamente esponenziale), perché dopo molto tempo che essa è in funzione senza problemi, il filamento di tungsteno si sarà sicuramente deteriorato e assottigliato, e quindi dovremo aspettarci che il tempo che le resta da vivere sia più corto che se fosse nuova. Al contrario, le particelle radioattive non si deteriorano un po' alla volta, ma decadono senza preavviso.

Cauchy. X è una v.a. di Cauchy di parametro $c > 0$ se può assumere tutti i valori reali, e ha densità definita da

$$f_X(x) = \frac{1}{\pi} \frac{c}{c^2 + x^2}. \quad (2.27)$$

Mentre la sua FR si trova essere

$$F_X(x) = \int_{-\infty}^x \frac{1}{\pi} \frac{c dt}{c^2 + t^2} = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{c}. \quad (2.28)$$

¹⁹La si provi per esercizio. Si noti inoltre che essa vale anche nel discreto, per le v.a. geometriche. Si provi a dimostrarla anche in quel caso. Perché per le v.a. geometriche essa è intuitiva? Si spieghi infine perché questo argomento è una pratica confutazione delle cosiddette "teorie" ritardiste di alcuni giocatori del Lotto.

La densità di una v.a. di Cauchy è una curva a campana centrata sullo zero, similmente a una gaussiana standard. La grossa differenza è il decadimento di $f_X(x)$ all'infinito, che per la gaussiana è esponenziale (detto anche *decadimento veloce*) mentre per X è polinomiale (va come $\frac{1}{x^2}$). In caso di decadimento non esponenziale si parla anche di *densità a code pesanti*, nel senso che le code della distribuzione hanno elevata probabilità, e quindi non è tanto raro che X produca valori molto grandi o molto piccoli.

2.4 Le variabili aleatorie binomiali e le loro approssimazioni

Il testo tipico di un problema su v.a. binomiali è il seguente.

Esempio 2.10. *In una partita di $n = 1000$ moduli RAM, ognuno ha probabilità $p = 1/100$ di essere difettoso. Qual è la probabilità che i moduli difettosi siano esattamente 10? Qual è la probabilità che siano più di 15?*

Detto X il numero di moduli difettosi, con un po' di ragionamento ci si convince che X ha legge binomiale di parametri n e p , quindi se $k = 0, 1, 2, \dots, n$ si avrà

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (2.29)$$

Tuttavia questa non è una risposta conclusiva. Vediamo perché.

Approssimazione con una poissoniana. Si supponga infatti di volere rispondere alla prima domanda: si avrebbe

$$P(X = 10) = \binom{1000}{10} \frac{1}{100^{10}} \left(\frac{99}{100}\right)^{n-k},$$

un numero che è assai difficile da calcolare, anche con i moderni mezzi informatici.

Ci viene in aiuto il corollario alla seguente proposizione.

Proposizione 2.11. *Se p_n è una successione di numeri positivi che tende a zero in modo tale che $np_n \rightarrow \mu$, allora per ogni k ,*

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \rightarrow \frac{\mu^k}{k!} e^{-\mu}. \quad (2.30)$$

Corollario. Se $n \gg 1$ e np non è piccolo (dell'ordine di 1 o più grande va bene) allora per valori di k non tanto maggiori di np vale la seguente approssimazione,

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{(np)^k}{k!} e^{-np}. \quad (2.31)$$

Applicando il corollario all'esercizio si trova

$$P(X = 10) \simeq \frac{10^{10}}{10!} e^{-10} \simeq 12.5\%.$$

Usando l'approssimazione con una poissoniana quindi diventa possibile stimare $P(X = k)$ con normali metodi di calcolo. Come vedremo nel prossimo paragrafo però, ciò non è sempre sufficiente.

Approssimazione con una gaussiana. Proviamo ora a rispondere alla seconda domanda. La soluzione esatta – comunque la si guardi – si basa su una sommatoria dall'aspetto poco gestibile:

$$P(X > 15) = \sum_{k=16}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} = 1 - \sum_{k=0}^{15} \binom{n}{k} p^k (1-p)^{n-k}.$$

Nemmeno l'approssimazione poissoniana aiuta a sufficienza:

$$P(X > 15) \simeq e^{-np} \sum_{k=16}^{\infty} \frac{(np)^k}{k!} = 1 - e^{-np} \sum_{k=0}^{15} \frac{(np)^k}{k!},$$

infatti calcolare e sommare questi sedici addendi, anche se fattibile con un software di calcolo, è una procedura tediosa, e non è facile farsi rapidamente un'idea dell'ordine di grandezza del risultato.

L'approssimazione con una gaussiana però permette di trasformare la sommatoria in un integrale definito che sappiamo affrontare. Di nuovo introduciamo una proposizione il cui corollario darà una risposta al problema.

Proposizione 2.12. *Siano $0 < p < 1$, $q = 1 - p$ e $n \in \mathbb{N}$ tali che $npq \gg 1$, e $|k - np|$ sia dell'ordine di \sqrt{npq} o minore, allora vale la seguente approssimazione:*

$$\binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(k - np)^2}{2npq}\right) = f_Y(k), \quad (2.32)$$

con $Y \sim N(np, npq)$ v.a. gaussiana.

Corollario. Se X è una v.a. binomiale di parametri n e p ; se $q = 1 - p$, $npq \gg 1$ e $a < b$ sono interi tali che vale una qualsiasi delle seguenti condizioni,

1. a sta in un intorno di np di raggio dell'ordine di \sqrt{npq} ;
2. b sta in un intorno di np di raggio dell'ordine di \sqrt{npq} ;
3. $np \in (a, b)$,

allora vale la seguente approssimazione,

$$\begin{aligned} P(a < X \leq b) &\simeq \int_{a+\frac{1}{2}}^{b+\frac{1}{2}} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x - np)^2}{2npq}\right) dx = \\ &= \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a + \frac{1}{2} - np}{\sqrt{npq}}\right), \end{aligned} \quad (2.33)$$

ovvero, cosa a volte più utile,

$$\begin{aligned} P(a \leq X \leq b) &\simeq \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(x - np)^2}{2npq}\right) dx = \\ &= \Phi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\right). \end{aligned} \quad (2.34)$$

In entrambe le formule, i vari $\frac{1}{2}$ sommati o sottratti ad a e b possono essere ignorati competamente se rispettivamente, $|a - np| \gg 1$ e $|b - np| \gg 1$.

Vediamo in pratica come si applica questo risultato all'esempio che stavamo affrontando.

Nel nostro caso abbiamo $np = 10$, $npq = 9.9$, e poniamo $a = 0$, $b = 15$:

$$\begin{aligned} P(X > 15) &= 1 - P(0 \leq X \leq 15) \simeq 1 - \Phi\left(\frac{15.5 - 10}{\sqrt{9.9}}\right) + \Phi\left(\frac{-0.5 - 10}{\sqrt{9.9}}\right) \simeq \\ &\simeq 1 - \Phi(1.748) + \Phi(-3.337) \simeq 1 - \Phi(1.748) \simeq 4.0\%. \end{aligned}$$

2.5 Trasformazioni di una variabile aleatoria

Supponiamo di avere assegnata una v.a. X e una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$, in modo da definire una seconda v.a. come *trasformazione* di X tramite g :

$$Y := g(X).$$

Ci chiediamo allora in tutta generalità quale sia la legge di Y .

Esempio 2.13. Se $X \sim N(0, 1)$ è una v.a. normale standard e $g(x) := ax + b$, per due parametri reali a e b fissati, si trovino funzione di ripartizione e di densità di $Y := g(X)$.

Il modo più istruttivo per risolvere il problema, e anche quello più generale, è scrivere la FR di Y e svolgere tutti i passaggi algebrici necessari per metterla in relazione con quella di X , come segue,

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = \begin{cases} P\left(X \leq \frac{y-b}{a}\right) & \text{se } a > 0 \\ P\left(X \geq \frac{y-b}{a}\right) & \text{se } a < 0, \end{cases}$$

supponendo ad esempio che $a > 0$ (l'altro caso è abbastanza analogo), si ha

$$F_Y(y) = P\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right) = \Phi\left(\frac{y-b}{a}\right). \quad (2.35)$$

Infine, f_Y può essere trovata derivando F_X ,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{a} \Phi'\left(\frac{y-b}{a}\right) = \frac{1}{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-b)^2}{2a^2}\right). \quad (2.36)$$

Osservazione. Da qui, confrontando la (2.36) con la (2.22), si deduce che Y è gaussiana di parametri b e a , ovvero $Y \sim N(b, a^2)$, e si deduce anche che questo tipo di v.a. ha FR data dalla (2.35)

Se g non è una funzione monotona, questo approccio è un po' più difficile, in quanto ad un certo punto richiede di invertire g stessa. Il prossimo problema illustra la situazione.

Esempio 2.14. Sia ancora $X \sim N(0, 1)$ e $Y := g(X)$, ma questa volta con $g(x) := x^2$. Si trovi la FD di Y .

Conviene sempre passare per la FR. Se $y < 0$, ovviamente $F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = 0$. Se invece $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1, \end{aligned}$$

e di conseguenza, sugli y positivi,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{2}{2\sqrt{y}} \Phi'(\sqrt{y}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} \exp\left(-\frac{y}{2}\right). \quad (2.37)$$

Questo è un esempio particolare di v.a. di legge Gamma (si veda un testo di riferimento), e viene anche chiamata *chi-quadro* (χ^2) ad un grado di libertà²⁰.

I due esempi dati fino a qui servono a illustrare il procedimento, che non può essere formalizzato più di tanto in maniera generale. Quello che si può però dare è un enunciato (che non dimostriamo) che permette di ottenere molto rapidamente la densità $f_Y(y)$ in funzione di quella di X e della derivata di g .

Proposizione 2.15. *Sia X una v.a. reale di FD f_X , e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione derivabile. Si denoti con $g^{-1}(y)$ l'insieme delle controimmagini di y tramite g . Se questo è un insieme finito per ogni valore di y e si pone $Y := g(X)$, allora Y è una v.a. continua con densità di probabilità data dalla seguente formula:*

$$f_Y(y) = \sum_{x_i \in g^{-1}(y)} \frac{f_X(x_i)}{|g'(x_i)|}. \quad (2.38)$$

Esempio 2.16. *Si risolva nuovamente l'Esempio 2.14, sfruttando la Proposizione 2.15 anziché il metodo “con le mani” utilizzato in quella sede.*

La funzione $g(x) = x^2$ è sicuramente derivabile, inoltre per ogni $y > 0$ l'insieme $g^{-1}(y)$ è costituito esattamente da due elementi: $\pm\sqrt{y}$; mentre per $y < 0$ l'insieme delle controimmagini è vuoto. Perciò se $y < 0$, $f_Y(y) = 0$, mentre per $y > 0$

$$f_Y(y) = \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\sqrt{y}^2}{2}\right)}{2\sqrt{y}} + \frac{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(-\sqrt{y})^2}{2}\right)}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} \exp\left(-\frac{y}{2}\right). \quad (2.39)$$

Anche se $f_Y(0)$ non sarebbe definito possiamo porlo uguale a 0, tanto cambiare il valore di una FR in un insieme finito di punti non cambia la legge di Y .

²⁰Per il lettore curioso, una v.a. si chiama chi-quadro a n gradi di libertà se è la somma dei quadrati di n v.a. normali standard indipendenti. Per una definizione precisa dell'indipendenza tra variabili aleatorie si veda la sezione 4 di queste note.

3 Medie di variabili aleatorie

Quando non si può o non si vuole accedere alla legge di una variabile aleatoria in tutta la sua completezza, vi sono dei parametri alternativi e più compatti che descrivono solo in parte il suo comportamento ma che possono a volte essere più accessibili o più trattabili.

I più semplici tra questi sono la *speranza* e la *varianza*.

3.1 La speranza

Distinguiamo inizialmente il caso discreto da quello continuo.

Definizione. Se X è una v.a. discreta, tale che si abbia

$$\sum_{x_i \in X(\Omega)} |x_i| P(X = x_i) < \infty, \quad (3.1)$$

allora si dice che X è dotata di *speranza* (o anche *aspettazione*, *media*), che si indica con $\mathbf{E}[X]$ ed è pari alla seguente quantità²¹:

$$\mathbf{E}[X] := \sum_{x_i \in X(\Omega)} x_i P(X = x_i). \quad (3.2)$$

Osservazione. In pratica è molto comune chiamarla media, tuttavia il nome forse più chiarificatore è aspettazione: infatti essa è una media tra tutti i valori possibili della v.a. X , presi però con la frequenza con cui *sono attesi*.

Nel caso X sia il risultato ottenuto lanciando un dado a 3 facce equilibrate, ad esempio, ci aspettiamo $1/3$ delle volte 1, $1/3$ delle volte 2 e $1/3$ delle volte 3. La media è allora $\frac{1+2+3}{3}$ ovvero $\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3$.

Se il dado fosse truccato, e ci aspettassimo $1/6$ di 1, $1/3$ di 2 e $1/2$ di 3, la media calcolata sarebbe $\frac{1}{6} \cdot 1 + \frac{1}{3} \cdot 2 + \frac{1}{2} \cdot 3 = \frac{14}{6} = \frac{7}{3}$: giustamente un po' maggiore della precedente, ma ancora compresa tra il minore e il maggiore dei valori possibili, proprio come ci si aspetta da una qualunque *media pesata*²².

Definizione. Se X è una v.a. continua, con densità f_X e tale che si abbia

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < \infty, \quad (3.3)$$

allora si dice che X ha speranza; inoltre tale quantità è per definizione:

$$\mathbf{E}[X] := \int_{-\infty}^{+\infty} x f_X(x) dx. \quad (3.4)$$

²¹Si noti che si è perso il valore assoluto della condizione precedente. Con $X(\Omega)$ si intende l'immagine di Ω tramite X , ovvero l'insieme dei valori possibili di X .

²²Analogia fisica: se l'asse reale fosse un'asticella rigida di massa trascurabile, e noi attaccassimo nei punti di ascissa corrispondente a ciascun valore possibile x_i un pesetto di massa proporzionale a $P(X = x_i)$, allora $\mathbf{E}[X]$ cadrebbe sul baricentro esatto dell'asticella.

Osservazione. Questa seconda definizione è meno intuitiva della prima, tuttavia si può mostrare facilmente la loro reciproca compatibilità. Per essere più precisi, la seconda definizione è abbastanza generale da includere la prima, se si accetta di definire la “densità” come una somma di impulsi (operazione impropria, ma utile). Sia infatti X discreta, e si ponga

$$f_X(x) = \sum_{x_i \in X(\Omega)} P(X = x_i) \delta(x - x_i), \quad (3.5)$$

si può allora dedurre facilmente che $\int_{-\infty}^{+\infty} x f_X(x) dx = \mathbf{E}[X]$. Infatti

$$\begin{aligned} \int_{-\infty}^{+\infty} x f_X(x) dx &= \int_{-\infty}^{+\infty} x \sum_{x_i \in X(\Omega)} P(X = x_i) \delta(x - x_i) dx = \\ &= \sum_{x_i \in X(\Omega)} P(X = x_i) \int_{-\infty}^{+\infty} x \delta(x - x_i) dx = \sum_{x_i \in X(\Omega)} P(X = x_i) x_i = \mathbf{E}[X], \end{aligned}$$

dove si è sfruttata la fondamentale proprietà della δ di Dirac che se $x_0 \in A$ e se g è continua in x_0 , allora $\int_A g(x) \delta(x - x_0) dx = g(x_0)$.

3.1.1 Proprietà fondamentali

Elenchiamo di seguito alcune utili proprietà della speranza di una v.a., lasciandole per ora senza dimostrazione.

0. Non tutte le v.a. sono dotate di speranza²³.

Ad esempio è facile verificare che la condizione (3.3) non è vera nel caso di una X con distribuzione di Cauchy:

$$\begin{aligned} \int_{-\infty}^{+\infty} |x| f_X(x) dx &= \frac{1}{\pi} \int_{-\infty}^{+\infty} |x| \frac{c}{c^2 + x^2} dx = \frac{2c}{\pi} \int_0^{+\infty} \frac{x dx}{c^2 + x^2} = \\ &= \frac{c}{\pi} \int_0^{+\infty} \frac{d}{dx} \log(c^2 + x^2) dx = \frac{c}{\pi} [\log(c^2 + x^2)]_0^{+\infty} = +\infty. \end{aligned}$$

Questo accade perché le v.a. di Cauchy sono caratterizzate da quelle che si chiamano *code pesanti*, ovvero asintotiche della funzione di densità che pur tendendo a zero non lo fanno esponenzialmente, ma come polinomi (in questo caso, come $1/x^2$). Questo fa sì che sia frequente che si presentino esiti per X molto grandi in modulo ed è precisamente questo fatto che impedisce²⁴ di definire la speranza delle v.a. di Cauchy.

²³Questa come si capisce non è una proprietà di $\mathbf{E}[X]$ in senso stretto, inoltre quasi sempre le v.a. che studieremo saranno dotate di media; tuttavia è importante non dimenticarsi completamente che questa non è una cosa da dare per scontata. Da ora in poi tutte le volte che scriveremo $\mathbf{E}[X]$, supporremo che esso esista e faremo notare i rari casi in cui non sarà così.

²⁴Non tutte le v.a. con code pesanti presentano questo problema, comunque.

1. Linearità.

Se α e β sono numeri reali e X e Y sono v.a., si ha

$$\mathbf{E}[\alpha X + \beta Y] = \alpha \mathbf{E}[X] + \beta \mathbf{E}[Y], \quad (3.6)$$

che esprime più compattamente i due risultati:

$$(a) \quad \mathbf{E}[\alpha X] = \alpha \mathbf{E}[X].$$

$$(b) \quad \mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y].$$

2. Se X è costante ($X \equiv a$), oppure se $X = a$ quasi certamente²⁵, allora $\mathbf{E}[X] = a$.
3. La speranza $\mathbf{E}[X]$ è intermedia rispetto ai valori possibili di X ; infatti se a e b sono reali tali che $a \leq X \leq b$ sempre, oppure $a \leq X \leq b$ quasi certamente, allora $a \leq \mathbf{E}[X] \leq b$.

La proprietà di linearità è importantissima in molti problemi (di seguito sarà chiarito in qualche esempio), ma ne daremo dimostrazione solo in futuro, durante lo studio di come si comportano coppie o vettori di variabili aleatorie. Le proprietà 2. e 3. sono una conseguenza del fatto che

$$X \geq a, \text{ q.c.} \quad \Rightarrow \quad \mathbf{E}[X] \geq a. \quad (3.7)$$

Esercizio 3.1. Si provi a dimostrare la proprietà (3.7) nel caso di v.a. continue (varrà poi anche per quelle discrete usando il fatto che la definizione (3.4) si applica anche al loro caso).

3.1.2 Semplici esempi

Negli esempi che seguono mostreremo tra le altre cose quanto valgono le medie delle più comuni v.a. e come si usa tipicamente la *linearità* di $\mathbf{E}[\cdot]$.

Esempio 3.2. Si tira un dado a n facce, e X è il risultato che si ottiene. Si vuole calcolare $\mathbf{E}[X]$.

Si ha ovviamente che $X(\Omega) = \{1, 2, \dots, n\}$ e $P(X = x_i) = \frac{1}{n}$, $x_i = 1, 2, \dots, n$, quindi

$$\mathbf{E}[X] = \sum_{x_i \in X(\Omega)} x_i P(X = x_i) = \sum_{x_i \in \{1, 2, \dots, n\}} x_i \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Esempio 3.3. Si tirano questa volta due dadi a n facce, e X è il risultato che si ottiene sommando i due punteggi. Si vuole calcolare $\mathbf{E}[X]$.

L'espressione di $P(X = k)$ è piuttosto complicata, quindi applicare direttamente la definizione di speranza non è saggio. È molto meglio associare una v.a. a ciascuno dei due dadi tirati e usare poi la linearità.

Siano Y_1 e Y_2 i punteggi dei due dadi tirati, allora grazie all'Esempio 3.2, vale

$$\mathbf{E}[Y_1] = \mathbf{E}[Y_2] = \frac{n+1}{2};$$

²⁵Come si ricorderà ciò significa che $P(X = a) = 1$.

inoltre $X = Y_1 + Y_2$, da cui per la linearità,

$$\mathbf{E}[X] = \mathbf{E}[Y_1 + Y_2] = \mathbf{E}[Y_1] + \mathbf{E}[Y_2] = n + 1.$$

Esempio 3.4. Si tirano m dadi a n facce, e X è il risultato che si ottiene sommando tutti i punteggi. Si vuole calcolare $\mathbf{E}[X]$.

Questa volta scrivere e usare la legge di X direttamente sarebbe semplicemente proibitivo. Si può però generalizzare facilmente l'approccio dell'Esempio 3.3: siano Y_1, Y_2, \dots, Y_m i punteggi degli m dadi, allora

$$\mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^m Y_i\right] = \sum_{i=1}^m \mathbf{E}[Y_i] = m \frac{n+1}{2}.$$

Esempio 3.5. Sia X una v.a. bernoulliana di parametro p . Si provi che la media di X è p .

Direttamente dalla definizione,

$$\mathbf{E}[X] = \sum_{x_i \in X(\Omega)} x_i P(X = x_i) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0 + 1 \cdot p = p.$$

Esempio 3.6. Sia X una v.a. binomiale di parametri n e p . Si provi che la media di X è np .

Applicando la definizione si ha

$$\begin{aligned} \mathbf{E}[X] &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = p \sum_{k=1}^n k \frac{n!}{k! (n-k)!} p^{k-1} (1-p)^{n-k} = \\ &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)! (n-k)!} p^{k-1} (1-p)^{n-k} = np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}, \end{aligned}$$

da cui ponendo $j = k - 1$ e $m = n - 1$,

$$\mathbf{E}[X] = np \sum_{j=0}^m \binom{m}{j} p^j (1-p)^{m-j} = np.$$

Esempio 3.7. Sia X una v.a. geometrica di parametro p . Si provi che la media di X è $\frac{1}{p}$.

Ricordando che la legge di X è

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots,$$

si definisca, per $|z| < 1$,

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

in modo che

$$\varphi'(z) = \sum_{n=1}^{\infty} n z^{n-1} = \frac{d}{dz} \left(\frac{1}{1-z} \right) = \frac{1}{(1-z)^2};$$

allora

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1} = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = p \varphi'(1-p) = \frac{p}{p^2} = \frac{1}{p}.$$

Esempio 3.8. Sia X una v.a. poissoniana di parametro μ . Si provi che la media di X è μ .

Direttamente dalla definizione di aspettazione,

$$\mathbf{E}[X] = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = \mu \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} e^{-\mu} = \mu \sum_{j=0}^{\infty} \frac{\mu^j}{j!} e^{-\mu} = \mu,$$

dove si è posto $j = k - 1$ e si è sfruttato il fatto che $\sum_{j=0}^{\infty} \frac{\mu^j}{j!} e^{-\mu} = 1$.

Esempio 3.9. Sia X una v.a. normale standard. Si provi che la media di X è zero.

Usiamo la definizione per le v.a. continue:

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{d}{dx} \left(-e^{-\frac{x^2}{2}} \right) dx = \frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} = 0.$$

Esempio 3.10. Sia X una v.a. gaussiana di legge $N(\mu, \sigma^2)$. Si provi che la media di X è μ .

Come abbiamo già visto, uno dei modi per ottenere una v.a. $X \sim N(\mu, \sigma^2)$, è partire da una $Y \sim N(0, 1)$ e applicare una trasformazione lineare con i coefficienti fissati in questo modo:

$$X = \sigma Y + \mu. \quad (3.8)$$

Allora, visto che $\mathbf{E}[Y] = 0$ per l'Esempio 3.9, si avrà per linearità della media,

$$\mathbf{E}[X] = \mathbf{E}[\sigma Y + \mu] = \sigma \mathbf{E}[Y] + \mu = \mu.$$

Esempio 3.11. Sia X una v.a. esponenziale di intensità λ . Si provi che la media di X è $\frac{1}{\lambda}$.

Applicando la definizione si ha

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \left[-t e^{-\lambda t} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = 0 + \lambda^{-1} = \frac{1}{\lambda}.$$

3.1.3 Speranza di una funzione di variabile aleatoria.

La proprietà di linearità, come abbiamo visto, ci permette di gestire la situazione in cui $Y = g(X)$, nel caso in cui g sia una funzione lineare. Infatti, se ad esempio $g(x) = ax + b$, allora si ha $\mathbf{E}[Y] = \mathbf{E}[aX + b] = a\mathbf{E}[X] + b$.

Più in generale, se g è una funzione reale qualsiasi, ci viene in aiuto la seguente

Proposizione 3.12. Sia $g : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione²⁶ e X una v.a. reale con densità f_X ; posto $Y := g(X)$, se vale

$$\int_{-\infty}^{+\infty} |g(x)| f_X(x) dx < \infty, \quad (3.9)$$

allora Y ha aspettazione pari a

$$\mathbf{E}[Y] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx. \quad (3.10)$$

La dimostrazione di questo risultato esula dal livello di questo corso, tuttavia si noti come l'integrale che stiamo calcolando non sia altro che una media sui valori possibili di $g(X)$, pesati con la loro densità di probabilità f_X ; in linea quindi con la giustificazione intuitiva che abbiamo dato di (3.2).

Osservazione. Usando la Proposizione 3.12 possiamo riscrivere le condizioni (3.1) e (3.3) come

$$\mathbf{E}[|X|] < \infty, \quad (3.11)$$

e la condizione (3.9) come

$$\mathbf{E}[|g(X)|] < \infty, \quad (3.12)$$

Una delle funzioni g più interessanti di cui calcolare $\mathbf{E}[g(X)]$ è $g(x) = x^2$, ma anche x^n con n più generale è importante: diamo allora la seguente

Definizione. Se X è una v.a., si chiama *momento secondo* di X la quantità $\mathbf{E}[X^2]$. Più in generale, si chiama *momento n -esimo* di X il numero $\mathbf{E}[X^n]$.

Ovviamente il momento primo è la speranza, e questo fa capire che non è affatto detto che i vari momenti esistano sempre (dopotutto, abbiamo già visto che ci sono casi in cui la media non esiste). Quello che si può dire è che se esiste il momento n -esimo esistono anche i momenti con esponenti più piccoli²⁷, quindi ad esempio, se X ha momento secondo, avrà anche momento primo.

3.2 La varianza

La speranza di una v.a. è un valore reale che è intermedio rispetto ai valori assunti dalla v.a. stessa, spostato verso quelli più probabili. Indica sinteticamente qual è un valore tipico o l'ordine di grandezza della v.a. in questione.

La *varianza* di una v.a. invece, contiene informazione su quanto concentrata attorno alla media sia la distribuzione della v.a.; in pratica a una bassa varianza

²⁶Bisognerebbe chiedere che g sia *misurabile*, ma le funzioni che non lo sono sono talmente rare e complicate che probabilmente non verranno mai incontrare dallo studente, quindi evitiamo anche di definire cosa ciò significhi e supponiamo di essere in ipotesi di regolarità sufficiente.

²⁷Di questo risultato non diamo dimostrazione.

corrisponde una v.a. che con alta probabilità risulta vicina alla sua media, mentre ad una varianza alta corrisponde una v.a. che assume anche valori molto distanti dalla media, con elevata probabilità (si veda a questo proposito la Proposizione 3.19).

Definizione. Se X è una v.a. dotata di momento secondo, si chiama *varianza* di X la quantità

$$\text{Var}(X) := \mathbf{E} \left[(X - \mathbf{E}[X])^2 \right]. \quad (3.13)$$

Osservazione. Si noti come $\mathbf{E}[X]$ sia un semplice numero reale (diciamo μ), in modo tale che l'espressione (3.13) non è altro che la speranza di un polinomio di secondo grado in X :

$$\text{Var}(X) = \mathbf{E} \left[(X - \mu)^2 \right] = \mathbf{E} \left[X^2 - 2\mu X + \mu^2 \right].$$

Proviamo a giustificare questa scelta. È chiaro che $X - \mu$ è semplicemente una traslata di X , che è stata *centrata* sull'origine, in modo tale che la sua speranza sia zero:

$$\mathbf{E}[X - \mu] = \mathbf{E}[X] - \mu = \mu - \mu = 0.$$

La varianza di X è il momento secondo di questa traslata, per questo è anche detta *momento secondo centrato*. Questo fa anche capire bene perché $\text{Var}(X)$ dia una misura di quanto è concentrata X attorno alla sua media: quando X è molto diversa da μ , $(X - \mu)^2$ è grande, e viceversa quando X è vicina a μ , $(X - \mu)^2$ è piccolo. Se facciamo la media di questa quantità quindi troviamo il comportamento tipico e una stima in media di quanto X può allontanarsi da μ ²⁸.

3.2.1 Proprietà fondamentali

1. $\text{Var}(X)$ si può calcolare più semplicemente che con (3.13) tramite la seguente formula:

$$\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2. \quad (3.14)$$

Infatti posto $\mu := \mathbf{E}[X]$ e usando la linearità si ha:

$$\begin{aligned} \text{Var}(X) &= \mathbf{E} \left[(X - \mu)^2 \right] = \mathbf{E} \left[X^2 - 2\mu X + \mu^2 \right] = \\ &= \mathbf{E}[X^2] - 2\mu \mathbf{E}[X] + \mu^2 = \mathbf{E}[X^2] - \mu^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2. \end{aligned}$$

2. $\text{Var}(X) \geq 0$.

Infatti è la media di $(X - \mu)^2$, che è una v.a. positiva (si usi la (3.7) con $a = 0$).

²⁸Perché abbiamo usato il quadrato come potenza? Sostanzialmente per motivi di semplicità: si noti che il ragionamento appena fatto si adatterebbe altrettanto bene se non meglio alla quantità $\mathbf{E}[|X - \mathbf{E}[X]|]$, solo che se avessimo definito così la varianza ci saremmo trovati in difficoltà di ogni genere all'atto di fare i conti. Vi sono poi altri vantaggi più sottili che appariranno un po' alla volta.

3. La varianza non è un'operatore lineare (come era la media) ma *quadratico*, ovvero se a e b sono numeri reali qualsiasi,

$$\text{Var}(aX + b) = a^2 \text{Var}(X). \quad (3.15)$$

Infatti,

$$\begin{aligned} \text{Var}(aX + b) &= \mathbf{E} \left[(aX + b - \mathbf{E}[aX + b])^2 \right] = \\ &= \mathbf{E} \left[(aX + b - (a\mathbf{E}[X] + b))^2 \right] = \mathbf{E} \left[(aX - a\mathbf{E}[X])^2 \right] = \\ &= a^2 \mathbf{E} \left[(X - \mathbf{E}[X])^2 \right] = a^2 \text{Var}(X). \end{aligned}$$

In generale invece, nulla si può dire di $\text{Var}(aX + bY)$.

3.2.2 Alcuni semplici esempi

Calcoliamo di seguito la varianza delle più comuni variabili aleatorie.

Esempio 3.13. Sia X una v.a. normale standard. Si provi che la varianza di X è 1.

Usiamo la formula (3.14). Tenendo presente che $\mathbf{E}[X] = 0$ per l'Esempio 3.9, basta calcolare $\mathbf{E}[X^2]$.

$$\begin{aligned} \text{Var}(X) &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \mathbf{E}[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \\ &= \frac{1}{\sqrt{2\pi}} \left[-xe^{-\frac{x^2}{2}} \right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 0 + 1 = 1. \end{aligned}$$

Esempio 3.14. Sia X una v.a. gaussiana con legge $N(\mu, \sigma^2)$. Si provi che la varianza di X è σ^2 .

Abbiamo già fatto notare che X può essere vista come una trasformazione lineare di una gaussiana standard Y ,

$$X = \sigma Y + \mu,$$

allora per le proprietà della varianza,

$$\text{Var}(X) = \text{Var}(\sigma Y + \mu) = \sigma^2 \text{Var}(Y) = \sigma^2.$$

Osservazione. Abbiamo finalmente dimostrato che i parametri μ e σ^2 della gaussiana non sono altro che la sua media e la sua varianza. Da ora, quando diremo che X è una v.a. gaussiana con media μ e varianza σ^2 resterà inteso che $X \sim N(\mu, \sigma^2)$. Per inciso, anche il numero σ ha un nome: se X è una v.a. dotata di momento secondo, il numero $\sqrt{\text{Var}(X)}$ si chiama infatti *deviazione standard* di X .

Osservazione. La deviazione standard è la giusta grandezza con cui confrontare le deviazioni di X da $\mathbf{E}[X]$. L'applicazione più potente è senza dubbio la disuguaglianza di Chebyshev, che proveremo nella sezione 3.3; un altro risultato molto elegante comunque è il seguente. Se X è una v.a. gaussiana di media μ e varianza σ^2 , la probabilità che X cada in un intorno di μ di raggio σ o multipli è fissata indipendentemente dai parametri; ad esempio,

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &= 0.6827 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 0.9545 \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 0.9973. \end{aligned} \quad (3.16)$$

Queste probabilità discendono tutte dalla semplice formula

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right), \quad (3.17)$$

ovvero,

$$P(\mu - \eta\sigma \leq X \leq \mu + \eta\sigma) = \Phi(\eta) - \Phi(-\eta) = 2\Phi(\eta) - 1. \quad (3.18)$$

Esempio 3.15. Sia X una v.a. esponenziale di intensità λ . Si provi che la varianza di X è $\frac{1}{\lambda^2}$.

Calcoliamo innanzitutto il momento secondo. Dalla definizione, integrando due volte per parti, si trova

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

per cui, sfruttando anche il risultato dell'Esempio 3.11,

$$\text{Var}(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Esempio 3.16. Sia X una v.a. bernoulliana di media²⁹ p . Si provi che la varianza di X è $p(1-p)$.

Al solito passiamo dal momento secondo:

$$\mathbf{E}[X^2] = 0^2 \cdot (1-p) + 1^2 \cdot p = p,$$

per cui

$$\text{Var}(X) = p - p^2 = p(1-p).$$

Esempio 3.17. Sia X una v.a. geometrica di parametro p . Si provi che la varianza di X è $\frac{1-p}{p^2}$.

si definisca, per $|z| < 1$,

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

²⁹Da quanto abbiamo dimostrato nell'Esempio 3.5 sappiamo che il parametro libero della bernoulliana è uguale alla sua media.

in modo che

$$\varphi'(z) = \sum_{n=0}^{\infty} n z^{n-1} = \frac{1}{(1-z)^2}, \quad \varphi''(z) = \sum_{n=1}^{\infty} n(n-1) z^{n-2} = \frac{2}{(1-z)^3};$$

allora

$$\begin{aligned} \mathbf{E}[X^2 - X] &= \sum_{k=1}^{\infty} (k^2 - k) p(1-p)^{k-1} = \\ &= p(1-p) \sum_{k=1}^{\infty} k(k-1) (1-p)^{k-2} = p(1-p) \varphi''(1-p) = 2 \frac{1-p}{p^2}. \end{aligned}$$

Quindi

$$\begin{aligned} \text{Var}(X) &= \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \\ &= \mathbf{E}[X^2 - X] - \mathbf{E}[X]^2 + \mathbf{E}[X] = 2 \frac{1-p}{p^2} - \frac{1}{p^2} + \frac{1}{p} = \frac{1-p}{p^2}. \end{aligned}$$

3.3 Disuguaglianza di Markov, disuguaglianza di Chebyshev

Presentiamo di seguito due disuguaglianze fondamentali. Entrambe hanno una dimostrazione molto semplice, e soprattutto la seconda ha una grande utilità nella risoluzione dei problemi che coinvolgono variabili aleatorie di cui si sanno media e varianza, ma non si sa a che classe appartengono.

Proposizione 3.18 (Disuguaglianza di Markov). *Se X è una v.a. positiva dotata di media, e $\delta > 0$; allora*

$$P(X \geq \delta) \leq \frac{\mathbf{E}[X]}{\delta}. \quad (3.19)$$

Dimostrazione. Sia f_X la funzione densità di X . Allora

$$\mathbf{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x f_X(x) dx \geq \int_{\delta}^{+\infty} x f_X(x) dx,$$

e siccome $x \geq \delta$ in tutto il dominio di integrazione, si ha

$$\int_{\delta}^{+\infty} x f_X(x) dx \geq \delta \int_{\delta}^{+\infty} f_X(x) dx = \delta P(X \geq \delta),$$

da cui è immediato dedurre la (3.19). □

Osservazione. È importante notare che nelle ipotesi ci deve essere il fatto che X sia positiva.

Proposizione 3.19 (Disuguaglianza di Chebyshev). *Se X è una v.a. dotata di momento secondo, e $\delta > 0$; allora posto $\mu := \mathbf{E}[X]$ si ha*

$$P(|X - \mu| \geq \delta) \leq \frac{\text{Var}(X)}{\delta^2}. \quad (3.20)$$

Dimostrazione. Si noti intanto che

$$|X - \mu| \geq \delta \iff (X - \mu)^2 \geq \delta^2,$$

per cui

$$P(|X - \mu| \geq \delta) = P((X - \mu)^2 \geq \delta^2) \leq \frac{\mathbf{E}[(X - \mu)^2]}{\delta^2} = \frac{\text{Var}(X)}{\delta^2},$$

dove si è applicata la disuguaglianza di Markov alla v.a. $(X - \mu)^2$. \square

Osservazione. Si noti che questa disuguaglianza permette di stimare la probabilità che X sia vicino (o lontano) dalla sua media, sapendone solo la varianza. In particolare, non si suppone di sapere se X è gaussiana, o esponenziale o altro.

Ecco quindi che nei casi reali in cui poco o nulla si può sapere sulla distribuzione di una v.a. X , basta avere delle stime per $\mathbf{E}[X]$ e $\text{Var}(X)$ per poter inferire qualcosa di assolutamente rigoroso sulla probabilità di una deviazione dalla media.

Proprio per questa straordinaria generalità, non ci si può aspettare che le stime siano troppo buone. Si confrontino in particolare le seguenti probabilità con quelle date in (3.16) per una v.a. $X \sim N(\mu, \sigma^2)$. Si supponga ancora che σ sia la deviazione standard di X e μ la sua media, ma che null'altro si sappia di X . Allora, per la disuguaglianza di Chebyshev si ha

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\geq 0 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\geq 0.75 \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\geq 0.8889. \end{aligned} \tag{3.21}$$

Come si può notare, queste stime sono compatibili con i valori esatti calcolati per la gaussiana, ma essendo validi per ogni altra v.a. con le stesse media e varianza, sono abbastanza larghe, e in genere è preferibile quando si conosce la legge di X calcolare direttamente queste probabilità.

4 Vettori aleatori

Questa sezione è rivolta allo studio di un vettore di variabili aleatorie, ovvero all'analisi di due o più variabili aleatorie contemporaneamente.

Nei problemi pratici di probabilità capita sovente di dover analizzare il rapporto tra due v.a. distinte, ma magari legate fra loro. Come vedremo subito, analizzare separatamente ciascuna delle due con i mezzi ottenuti fino a qui non è quasi mai sufficiente: è necessario allora sviluppare degli strumenti nuovi, adatti a superare questa difficoltà.

Per semplicità verranno illustrati i risultati teorici solo nel caso di un vettore di due v.a., infatti la generalizzazione a tre o a n v.a. è del tutto immediata e verrà assunta implicitamente. Il caso di una successione di infinite v.a. (ovvero il caso dei *processi stocastici*) verrà invece accennato in questa sezione, e poi trattato approfonditamente più avanti.

4.1 Perché serve una nuova teoria?

Fino a qui abbiamo sempre sostenuto che sebbene una v.a. sia in realtà una applicazione $X : \Omega \rightarrow \mathbb{R}$, per dedurre risultati pratici sul suo comportamento non servisse altro che la sua legge.

In pratica, poiché eravamo interessati a quantità tipo

$$P(X < 5), \quad P(a < X \leq b), \quad P(\sin(X) < \epsilon | X^2 > \delta),$$

ci bastava conoscere per tutti i punti $x \in \mathbb{R}$ la funzione di ripartizione $F_X(x) := P(X \leq x)$, o la funzione densità di probabilità $f_X(x) := \frac{d}{dx}F_X(x)$ per potere ricavare – più o meno laboriosamente – tutte le probabilità di eventi che coinvolgessero solo X .

Quando invece il problema che affrontiamo richiede la definizione di due v.a. X e Y , e con esse ci chiede di dedurre probabilità di eventi che le coinvolgono entrambe, come $\{X < Y\}$ o $\{X > 2, Y^2 \leq 1\}$, anche se abbiamo le loro leggi non siamo in grado di dire nulla³⁰.

Gli esempi che seguono illustrano il fatto che sapere le leggi di X e di Y non permette di determinare univocamente le probabilità di eventi che le coinvolgono entrambe.

Costruiremo tre esperimenti casuali nei quali saranno definite due v.a. X e Y : ciascuna di esse si comporterà dal punto di vista statistico come un normale dado a sei facce bilanciato. Tuttavia, un evento come $\{X = 1, Y = 2\}$ avrà di volta in volta probabilità $\frac{1}{36}$, 0 e infine $\frac{1}{24}$.

Questo implicherà che le leggi di X e Y non contengono informazioni sufficienti a dedurre $P(X = 1, Y = 2)$.

Esempio 4.1. Si tira due volte un dado a sei facce; X è il primo risultato ottenuto e Y il secondo. Calcolare $P(X = i, Y = j)$, $i, j = 1, 2, \dots, 6$

È evidente che tutte le coppie di valori sono equiprobabili, di conseguenza si ha

$$P(X = i, Y = j) = \frac{1}{36}, \quad \forall i, j \tag{4.1}$$

³⁰Anche se siamo ancora in grado di studiare eventi che coinvolgano *separatamente* X o Y .

$$p_1 = \frac{1}{36},$$

P		X					
		1	2	3	4	5	6
Y	1	p_1	p_1	p_1	p_1	p_1	p_1
	2	p_1	p_1	p_1	p_1	p_1	p_1
	3	p_1	p_1	p_1	p_1	p_1	p_1
	4	p_1	p_1	p_1	p_1	p_1	p_1
	5	p_1	p_1	p_1	p_1	p_1	p_1
	6	p_1	p_1	p_1	p_1	p_1	p_1

(4.2)

Esempio 4.2. Ogni dado cubico da gioco, per motivi storici, ha sulle facce opposte numeri a somma costante (la faccia 1 è opposta alla faccia 6, la faccia 2 è opposta alla faccia 5 e la faccia 3 è opposta alla faccia 4); un dado di questo tipo viene lanciato, e si denotano con X e Y i punteggi della faccia superiore e inferiore dopo che si è fermato. Calcolare $P(X = i, Y = j)$, $i, j = 1, 2, \dots, 6$

Siccome si ha con probabilità 1 che $X + Y = 7$, la tabella delle probabilità è definita come segue:

$$P(X = i, Y = j) = \begin{cases} 1/6 & \text{se } X + Y = 7 \\ 0 & \text{altrimenti,} \end{cases} \quad (4.3)$$

$$p_2 = \frac{1}{6},$$

P		X					
		1	2	3	4	5	6
Y	1	0	0	0	0	0	p_2
	2	0	0	0	0	p_2	0
	3	0	0	0	p_2	0	0
	4	0	0	p_2	0	0	0
	5	0	p_2	0	0	0	0
	6	p_2	0	0	0	0	0

(4.4)

Esempio 4.3. Un dado da gioco come quello dell'Esempio 4.2 viene lanciato in una scatola rettangolare che è stata preventivamente inclinata di 45 gradi, in modo che il dado stesso si fermi lungo lo spigolo più in basso; due facce del dado sono a questo punto visibili dall'alto. Siano X e Y i punteggi di tali facce (diciamo, X quella di sinistra e Y quella di destra). Calcolare $P(X = i, Y = j)$, $i, j = 1, 2, \dots, 6$.

Ognuno dei 12 spigoli del cubo ha la stessa probabilità di finire in alto, e così è anche per la coppia di facce adiacenti che lo individua. Inoltre, le due orientazioni possibili che possono avere queste due facce (corrispondenti a scambiare X e Y) sono equiprobabili. Concludendo, due facce del dado adiacenti hanno una probabilità di $\frac{1}{24}$ di essere scelte come X e Y , mentre facce opposte e coppie di numeri uguali hanno probabilità 0:

$$P(X = i, Y = j) = \begin{cases} 1/24 & \text{se } X \neq Y \text{ e } X + Y \neq 7 \\ 0 & \text{altrimenti,} \end{cases} \quad (4.5)$$

$$p_3 = \frac{1}{24},$$

P		X					
		1	2	3	4	5	6
Y	1	0	p_3	p_3	p_3	p_3	0
	2	p_3	0	p_3	p_3	0	p_3
	3	p_3	p_3	0	0	p_3	p_3
	4	p_3	p_3	0	0	p_3	p_3
	5	p_3	0	p_3	p_3	0	p_3
	6	0	p_3	p_3	p_3	p_3	0

(4.6)

Osservazione. In tutti e tre gli esempi la legge di X e quella di Y sono uguali a quella di un normale dado bilanciato. Se questo non è intuitivo per il lettore, si noti che esse possono essere ricavate sommando per righe o per colonne i valori delle tabelle³¹,

$$P(X = i) = \sum_{j=1}^6 P(X = i, Y = j); \quad P(Y = j) = \sum_{i=1}^6 P(X = i, Y = j),$$
(4.7)

e nei nostri esempi la somma di tutte le colonne come quella di tutte le righe è sempre uguale a $1/6$, quindi per ogni i , $P(X = i) = P(Y = i) = 1/6$.

Osservazione. Anche se quelli che abbiamo mostrato sono solo esempi nel discreto, già contengono molti elementi della teoria generale. Riconosciamo nelle tabelle (4.2), (4.4) e (4.6) di cui sopra le *leggi congiunte* dei vettori (X, Y) , mentre le formule (4.7) anticipano il calcolo delle *leggi marginali* (ovvero le leggi di X e Y quando sono ricavate a partire dalla legge congiunta, anziché date come definizione di X e Y stesse).

4.2 Legge congiunta

Se X e Y sono due v.a. reali, la loro *legge congiunta* è la misura di probabilità che esse definiscono sul sottoinsieme di \mathbb{R}^2 costituito dai loro valori possibili.

In pratica, se è data una regione $D \subset \mathbb{R}^2$ la legge congiunta di X e Y è quella misura di probabilità P' su \mathbb{R}^2 per la quale vale

$$P'(D) = P((X, Y) \in D).$$

4.2.1 Definizioni

Gli strumenti attraverso cui noi studiamo la legge congiunta di X e Y sono le funzioni di ripartizione e di densità.

Definizione. Se X e Y sono due v.a. reali, la loro *funzione di ripartizione congiunta* (da ora in poi, FRC) è l'applicazione $F_{X,Y} : \mathbb{R}^2 \rightarrow [0, 1]$ definita da

$$F_{X,Y}(x, y) := P(X \leq x, Y \leq y) = P(\{X \leq x\} \cap \{Y \leq y\}). \quad (4.8)$$

Quello che suggerisce l'equazione (4.8) è che se $\{X \leq x\}$ e $\{Y \leq y\}$ sono eventi indipendenti per ogni scelta di x e y la FRC resta definita dando le sole leggi di X e Y . Diamo allora la seguente

³¹Si ricordi la dimostrazione della formula di probabilità totale.

Definizione. Due v.a. reali X e Y si dicono *indipendenti* se per ogni $x, y \in \mathbb{R}$ si ha che

$$F_{X,Y}(x, y) = F_X(x) F_Y(y), \quad (4.9)$$

ovvero che $\{X \leq x\}$ e $\{Y \leq y\}$ sono eventi indipendenti.

Definizione. Se X e Y sono due v.a. reali, la loro *funzione di densità di probabilità congiunta* (da ora in poi, FDC) è l'applicazione $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ definita da

$$f_{X,Y}(x, y) := \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} \quad (4.10)$$

4.2.2 Proprietà elementari

Deduciamo dalle definizioni della sezione precedente i seguenti immediati risultati.

Proposizione 4.4. *Ogni FRC $F_{X,Y}$ è caratterizzata dalle seguenti proprietà che si deducono immediatamente dalla definizione (4.8) con ragionamenti analoghi a quelli in dimensione uno.*

1. *è non decrescente in entrambe le variabili;*
2. *è continua a destra³² in entrambe le variabili;*
3. *valgono i seguenti limiti all'infinito, espressi con leggero abuso di notazione*

$$F_{X,Y}(t, -\infty) = F_{X,Y}(-\infty, t) = 0 \text{ per ogni } t \in \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} \quad (4.11)$$

$$F_{X,Y}(x, +\infty) = F_X(x), \quad F_{X,Y}(+\infty, y) = F_Y(y) \text{ per ogni } x, y \in \mathbb{R} \quad (4.12)$$

$$F_{X,Y}(+\infty, +\infty) = 1 \quad (4.13)$$

Proposizione 4.5. *Se $D \subset \mathbb{R}^2$ è un rettangolo definito da*

$$D := \{(x, y) \in \mathbb{R}^2 \mid a < x \leq b, c < y \leq d\},$$

allora

$$\begin{aligned} P((X, Y) \in D) &= P(a < X \leq b, c < Y \leq d) = \\ &= F_{X,Y}(a, c) + F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) = \int_a^b dx \int_c^d dy f_{X,Y}(x, y) \end{aligned} \quad (4.14)$$

Dimostrazione. La formula che usa la FRC si giustifica con un buon disegno di quali sono i settori interessati sul piano cartesiano, oppure algebricamente notando che

³²In pratica, molto più che nel caso monodimensionale, ci limiteremo al caso di coppie di v.a. continue per le quali $F_{X,Y}$ sarà quindi semplicemente *continua*.

- (a) $\{(X, Y) \in D\} \sqcup \{X \leq a \text{ oppure } Y \leq c\} = \{X \leq b, Y \leq d\}$,
dove con $A \sqcup B$ si denota l'unione disgiunta di A e B , ovvero la unione usuale, intendendo implicitamente però che $A \cap B = \emptyset$, da cui

$$\begin{aligned} P((X, Y) \in D) &= P(X \leq b, Y \leq d) - P(X \leq a \text{ oppure } Y \leq c) = \\ &= F_{X,Y}(b, d) - P(X \leq a \text{ oppure } Y \leq c); \end{aligned}$$

- (b) $\{X \leq a, Y \leq d\}$ e $\{X \leq b, Y \leq c\}$ hanno intersezione $\{X \leq a, Y \leq b\}$ e unione $\{X \leq a \text{ oppure } Y \leq c\}$, da cui

$$\begin{aligned} P(X \leq a \text{ oppure } Y \leq c) &= \\ &= P(X \leq a, Y \leq d) + P(X \leq b, Y \leq c) - P(X \leq a, Y \leq b) = \\ &= F_{X,Y}(a, d) + F_{X,Y}(b, c) - F_{X,Y}(a, c); \end{aligned}$$

mettendo insieme le due uguaglianze si prova la formula con la FRC.

Per quanto riguarda l'identità

$$F_{X,Y}(a, c) + F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) = \int_a^b dx \int_c^d dy f_{X,Y}(x, y),$$

essa è una diretta conseguenza del teorema fondamentale del calcolo e della definizione data in (4.10). \square

Corollario. Se D è un sottoinsieme qualsiasi di \mathbb{R}^2 allora

$$P((X, Y) \in D) = \iint_D f_{X,Y}(x, y) dx dy. \quad (4.15)$$

La dimostrazione di questo risultato va oltre le finalità di questo corso.

Proposizione 4.6. Se X e Y sono indipendenti,

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad (4.16)$$

inoltre ogni coppia A_X e A_Y di eventi definiti rispettivamente tramite³³ X e Y , sono tra loro indipendenti.

Dimostrazione. L'identità (4.16) si deduce immediatamente derivando l'identità (4.9) rispetto a x e y .

La seconda parte della tesi invece, discenderebbe dall'equazione (4.15) ma viene lasciata senza dimostrazione. \square

Proposizione 4.7. Se X e Y sono indipendenti e g, h sono due qualsiasi funzioni di \mathbb{R} in sé, allora anche le v.a. $g(X)$ e $h(Y)$ sono indipendenti.

Non si dà dimostrazione di questo risultato.

³³Si intendono eventi tipo $\{X \leq 5\}$, $\{X^2 > \log X\}$, $\{X < 2 < X^2\}$, o, più in generale $\{X \in S\}$ per un qualche S sottoinsieme di \mathbb{R} .

Proposizione 4.8 (Densità marginali). Se X e Y sono due v.a. reali con FDC $f_{X,Y}$, e con leggi marginali di densità f_X e f_Y , allora valgono le relazioni

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx \quad (4.17)$$

Dimostrazione. Proviamo ad esempio la prima identità:

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} F_{X,Y}(x, +\infty) = \\ &= \frac{d}{dx} \int_{-\infty}^x dx' \int_{-\infty}^{+\infty} dy f_{X,Y}(x', y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \end{aligned}$$

dove si è usata una delle (4.12)³⁴ □

Proposizione 4.9. Ogni FDC $f_{X,Y}$ è caratterizzata dalle seguenti proprietà che si deducono immediatamente dalla definizione e da quelle della FRC con ragionamenti analoghi a quelli in dimensione uno:

1. è non negativa;
2. è integrabile;
3. ha massa totale pari a uno: $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx dy = 1$.

4.2.3 Esempi

Esempio 4.10. Siano X e Y due v.a. di legge uniforme su $[0,1]$ e tra loro indipendenti. Si trovi la loro densità congiunta e le seguenti probabilità: (a) $P(X < 1/2)$, (b) $P(X \leq 1/3, Y \geq 1/2)$, (c) $P(X + Y < 1)$, (d) $P(X + Y < 1 | Y \geq 1/2)$, (e) $P(XY > 1/4)$.

La densità congiunta si trova immediatamente sfruttando che X e Y sono indipendenti e la Proposizione 4.6:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = \mathbf{1}_{[0,1] \times [0,1]}(x,y) = \begin{cases} 1 & \text{se } 0 \leq x \leq 1 \text{ e } 0 \leq y \leq 1 \\ 0 & \text{altrimenti.} \end{cases}$$

Le probabilità richieste si possono cercare usando l'equazione (4.15):

- (a) Si ha $P(X < 1/2) = P((X,Y) \in D_a)$, con D_a il rettangolo di vertici $(0,0)$ e $(1/2,1)$. Allora usando la (4.15),

$$P((X,Y) \in D_a) = \iint_{D_a} f_{X,Y}(x,y) dx dy = \iint_{D_a} dx dy = \text{Area}(D_a) = 1/2.$$

- (b) In questo caso $P(X \leq 1/3, Y \geq 1/2) = P((X,Y) \in D_b)$, con D_b il rettangolo di vertici $(0,1/2)$ e $(1/3,1)$. Quindi,

$$P((X,Y) \in D_b) = \iint_{D_b} f_{X,Y}(x,y) dx dy = \iint_{D_b} dx dy = \text{Area}(D_b) = 1/6.$$

³⁴Questa proposizione, con la successiva, traccia un'analogia tra le proprietà della FRC citate nella Proposizione 4.4 e quelle della FDC.

- (c) $P(X + Y < 1) = P((X, Y) \in D_c)$, con D_c il triangolo rettangolo di vertici $(0, 0)$, $(1, 0)$ e $(0, 1)$. Allora $P((X, Y) \in D_c) = \text{Area}(D_c) = 1/2$
- (d) $P(X + Y < 1 | Y \geq 1/2) = \frac{\text{Area}(D_d)}{\text{Area}(D'_d)}$, con D_d il triangolo rettangolo di vertici $(0, 1/2)$, $(1/2, 1/2)$ e $(0, 1)$ e con D'_d il rettangolo di vertici $(0, 1/2)$ e $(1, 1)$. Quindi $P(X + Y < 1 | Y \geq 1/2) = \frac{1/8}{1/2} = \frac{1}{4}$.
- (e) $P(XY > 1/4) = \text{Area}(D_e)$, dove D_e è la regione di piano cartesiano che sta al di sopra dell'iperbole di equazione $y = \frac{1}{4x}$, al di sotto della retta di equazione $y = 1$ e a sinistra della retta $x = 1$.

$$\text{Area}(D_e) = 1 - \int_0^1 \max\left(1, \frac{1}{4x}\right) dx = 1 - \int_0^{\frac{1}{4}} dx - \int_{\frac{1}{4}}^1 \frac{dx}{4x} = \frac{3}{4} - \frac{1}{4} \log 4.$$

Un buon disegno chiarisce molto il procedimento.

Esempio 4.11. Siano X e Y due v.a. esponenziali di intensità λ , tra loro indipendenti. Si trovi la loro densità congiunta e si trovi la funzione di ripartizione di $Z := \min(X, Y)$.

La densità congiunta si trova di nuovo grazie all'indipendenza e alla Proposizione 4.6:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \begin{cases} \lambda^2 e^{-\lambda(x+y)} & \text{se } x \geq 0 \text{ e } y \geq 0 \\ 0 & \text{altrimenti.} \end{cases}$$

Venendo a Z , possiamo ancora usare l'equazione (4.15):

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(\min(X, Y) \leq z) = \\ &= P((X, Y) \in D_z) = \iint_{D_z} f_{X,Y}(x, y) dx dy, \end{aligned}$$

dove abbiamo posto

$$D_z := \{(x, y) \in \mathbb{R}^2 \mid x \leq z \text{ oppure } y \leq z\}.$$

Sia ora $\overline{D_z}$ il complementare di D_z rispetto al primo quadrante del piano cartesiano: allora si ha $\overline{D_z} = \{(x, y) \in \mathbb{R}^2 \mid x > z, y > z\}$, e quindi

$$\begin{aligned} \iint_{D_z} f_{X,Y}(x, y) dx dy &= 1 - \iint_{\overline{D_z}} f_{X,Y}(x, y) dx dy = \\ &= 1 - \int_z^{+\infty} \int_z^{+\infty} f_{X,Y}(x, y) dx dy = 1 - \int_z^{+\infty} \lambda e^{-\lambda x} dx \int_z^{+\infty} \lambda e^{-\lambda y} dy = 1 - e^{-2\lambda z}. \end{aligned}$$

Quindi Z ha legge esponenziale di intensità 2λ .

Esempio 4.12. Sono date due v.a. X e Y con densità di probabilità congiunta

$$f_{X,Y}(x,y) = \begin{cases} e^{x-y} & \text{se } 0 \leq x \leq 1 \text{ e } y \geq x \\ 0 & \text{altrimenti;} \end{cases}$$

Si trovino le FD marginali di X e Y , si dica se X e Y sono indipendenti, si verifichi che abbiano integrale pari a uno e si calcolino $\mathbf{E}[X]$ e $\mathbf{E}[Y]$.

Le densità marginali si trovano applicando le formule (4.17).

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

questo integrale è nullo se $x \notin [0,1]$ perché in quel caso $f_{X,Y}$ è identicamente 0, mentre altrimenti si può – per lo stesso motivo – limitare l'integrale alla semiretta $[x, \infty)$, ed esso vale $e^x \int_x^\infty e^{-y} dy = e^x e^{-x} = 1$, per cui X ha legge $U[0,1]$:

$$f_X(x) = \mathbf{1}_{[0,1]}(x).$$

Venendo a Y , con ragionamenti analoghi si trova

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx = \begin{cases} e^{-y} \int_0^1 e^x dx & \text{se } 1 < y \\ e^{-y} \int_0^y e^x dx & \text{se } 0 < y \leq 1 \\ 0 & \text{se } y \leq 0, \end{cases}$$

da cui si deduce

$$f_Y(y) = \begin{cases} (e-1)e^{-y} & \text{se } 1 < y \\ e^{-y}(e^y-1) & \text{se } 0 < y \leq 1 \\ 0 & \text{se } y \leq 0 \end{cases} = \begin{cases} \min(1, e^{1-y}) - e^{-y} & \text{se } y > 0 \\ 0 & \text{se } y \leq 0. \end{cases}$$

Verifichiamo che la funzione trovata sia una FD valida,

$$\int_{-\infty}^{+\infty} f_Y(y) dy = \int_0^\infty \min(1, e^{1-y}) dy - \int_0^\infty e^{-y} dy = 2 - 1 = 1.$$

Per l'ultima parte dell'esempio, si vede subito che $\mathbf{E}[X] = 1/2$, mentre

$$\begin{aligned} \mathbf{E}[Y] &= \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_0^\infty y \min(1, e^{1-y}) dy - \int_0^\infty y e^{-y} dy = \\ &= \int_0^1 y dy + \int_1^\infty y e^{1-y} dy - \int_0^\infty y e^{-y} dy = \frac{1}{2} + \int_0^\infty (y+1) e^{-y} dy - 1 = \frac{3}{2}. \end{aligned}$$

Inseriamo qui un esempio un po' particolare (tanto per cambiare è nel discreto) il cui risultato tornerà utile in futuro.

Esempio 4.13. Fissati un numero naturale n e un reale $p \in [0,1]$, si considerino n v.a. Y_1, Y_2, \dots, Y_n , bernoulliane di media p e tutte indipendenti. Detta $X := \sum_{i=1}^n Y_i$ si dimostri che X ha legge binomiale di parametri n e p .

Detta $q = 1 - p$, bisogna provare per ogni n che

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad \forall k; \quad (*)$$

questo può essere fatto ad esempio per induzione.

Per $n = 1$, $X = Y_1$ è bernoulliana e si ha effettivamente

$$P(X = k) = \begin{cases} p & k = 1 \\ q & k = 0 \\ 0 & k \neq 0, k \neq 1 \end{cases} = p^k q^{1-k}, \quad k = 0, 1.$$

Supponiamo ora che la (*) sia vera per $n - 1$ e proviamola per n . Siano in particolare

$$X' := \sum_{i=1}^{n-1} Y_i, \quad X := \sum_{i=1}^n Y_i,$$

allora se $X = k$, a seconda se Y_n vale 0 o 1, X' valeva k o $k - 1$, e si applica la formula di probabilità totale:

$$\begin{aligned} P(X = k) &= \\ &= P(Y_n = 0 | X' = k) P(X' = k) + P(Y_n = 1 | X' = k - 1) P(X' = k - 1), \end{aligned}$$

da cui, ricordando che Y_n è indipendente dalle altre bernoulliane e quindi anche da X' , e usando l'ipotesi induttiva si trova

$$\begin{aligned} P(X = k) &= P(Y_n = 0) P(X' = k) + P(Y_n = 1) P(X' = k - 1) = \\ &= q \binom{n-1}{k} p^k q^{n-1-k} + p \binom{n-1}{k-1} p^{k-1} q^{n-k} = \\ &= \left[\binom{n-1}{k} + \binom{n-1}{k-1} \right] p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}. \end{aligned}$$

4.3 Momenti

Il calcolo dell'aspettazione di due v.a. non è in se stesso né difficile né interessante: semplicemente $\mathbf{E}[(X, Y)] = (\mathbf{E}[X], \mathbf{E}[Y])$.

Un oggetto molto più fruttuoso è invece la media di una v.a. ottenuta come funzione di X e Y . Vale il seguente risultato (analogo al caso monodimensionale) di cui non diamo dimostrazione.

Proposizione 4.14. *Se X e Y sono due v.a. con FDC $F_{X,Y}$, e se $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ è una funzione, allora posto $Z := g(X, Y)$ si ha*

$$\mathbf{E}[Z] = \mathbf{E}[g(X, Y)] := \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy, \quad (4.18)$$

dove questa uguaglianza ha senso solo se, con la stessa notazione, $\mathbf{E}[|Z|] < \infty$.

Siamo ormai in grado di provare la proprietà di linearità della aspettazione (equazione (3.6)).

Proposizione 4.15. *Se X e Y sono due v.a. con media finita, allora $\mathbf{E}[X + Y] = \mathbf{E}[X] + \mathbf{E}[Y]$.*

Dimostrazione. Poniamo $g(x, y) = x + y$ e applichiamo la proposizione 4.14.

$$\begin{aligned}\mathbf{E}[X + Y] &= \mathbf{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x + y) f_{X,Y}(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f_{X,Y}(x, y) dx dy + \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f_{X,Y}(x, y) dx dy.\end{aligned}$$

Ora, ricordando le formule (4.17) delle densità marginali,

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

e analogamente l'altro addendo; si ottiene in questo modo la tesi. \square

Di sapore simile il prossimo risultato, che analizza la speranza del prodotto di due variabili aleatorie.

Proposizione 4.16. *Se X e Y sono due v.a. indipendenti con media finita, allora $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.*

Dimostrazione. Poniamo questa volta $g(x, y) = x \cdot y$ e applichiamo nuovamente la proposizione 4.14, sfruttando che per l'indipendenza di X e Y , $f_{X,Y}(x, y) = f_X(x) f_Y(y)$,

$$\begin{aligned}\mathbf{E}[XY] &= \mathbf{E}[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{X,Y}(x, y) dx dy = \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_X(x) f_Y(y) dx dy = \int_{-\infty}^{+\infty} x f_X(x) dx \int_{-\infty}^{+\infty} y f_Y(y) dy = \mathbf{E}[X] \mathbf{E}[Y].\end{aligned}$$

\square

Si noti bene che mentre la proposizione 4.15 funziona per qualunque coppia di v.a. (purché le medie coinvolte esistano), la proposizione 4.16 è in generale falsa per una coppia arbitraria di v.a., essendo valida solo se X e Y sono indipendenti.

Ad esempio, se $Y = X$ è chiaro che le due v.a. non sono indipendenti, e infatti $\mathbf{E}[XY] = \mathbf{E}[X^2]$ mentre $\mathbf{E}[X]\mathbf{E}[Y] = \mathbf{E}[X]^2$; non solo queste due quantità non sono in generale uguali, ma abbiamo addirittura dato un nome alla loro differenza:

$$\mathbf{E}[X^2] - \mathbf{E}[X]^2 =: \text{Var}(X),$$

e abbiamo già avuto modo di fare notare come questa grandezza sia molto importante per lo studio di X .

4.3.1 Covarianza

Generalizzando il ragionamento appena fatto, potremmo dire che se X e Y sono indipendenti, allora $\mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y] = 0$, mentre in generale questa quantità potrebbe essere non nulla. È allora sensato dare dignità a questo oggetto e studiarne le proprietà.

Definizione. Se X e Y sono due variabili aleatorie dotate di momento secondo, si chiama *covarianza* di X e Y la quantità

$$\text{Cov}(X, Y) := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y], \quad (4.19)$$

dove la seconda forma è ottenuta dalla prima semplicemente svolgendo il prodotto all'interno di $\mathbf{E}[\cdot]$ e usando la linearità.

Proposizione 4.17. *L'operatore di covarianza gode delle seguenti proprietà:*

1. se X e Y sono indipendenti, allora $\text{Cov}(X, Y) = 0$;
2. è simmetrico, nel senso che $\text{Cov}(X, Y) = \text{Cov}(Y, X)$;
3. è lineare in entrambi gli argomenti,

$$\text{Cov}(\alpha_1 X_1 + \alpha_2 X_2, Y) = \alpha_1 \text{Cov}(X_1, Y) + \alpha_2 \text{Cov}(X_2, Y) \quad (4.20)$$

$$\text{Cov}(X, \beta_1 Y_1 + \beta_2 Y_2) = \beta_1 \text{Cov}(X, Y_1) + \beta_2 \text{Cov}(X, Y_2) \quad (4.21)$$

ovvero è bilineare:

$$\begin{aligned} \text{Cov}(\alpha_1 X_1 + \alpha_2 X_2, \beta_1 Y_1 + \beta_2 Y_2) &= \\ &= \alpha_1 \beta_1 \text{Cov}(X_1, Y_1) + \alpha_1 \beta_2 \text{Cov}(X_1, Y_2) + \\ &+ \alpha_2 \beta_1 \text{Cov}(X_2, Y_1) + \alpha_2 \beta_2 \text{Cov}(X_2, Y_2); \end{aligned} \quad (4.22)$$

4. la covarianza di una v.a. con se stessa è la sua varianza,

$$\text{Cov}(X, X) = \text{Var}(X); \quad (4.23)$$

5. la covarianza di due v.a. è limitata in modulo dal prodotto delle deviazioni standard,

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}. \quad (4.24)$$

Dimostrazione. Proviamo solo la 5. in quanto le altre proprietà si deducono immediatamente dalle definizioni.

Date due qualunque v.a. X_0 e Y_0 , si consideri il seguente polinomio di secondo grado in s ,

$$p(s) = s^2 \mathbf{E}[X_0^2] + 2s \mathbf{E}[X_0 Y_0] + \mathbf{E}[Y_0^2] = \mathbf{E}[(sX_0 + Y_0)^2] \geq 0;$$

siccome esso è sempre non negativo, non potrà avere due radici reali distinte, e in particolare il suo discriminante dovrà essere minore o uguale a zero,

$$0 \geq (2\mathbf{E}[X_0 Y_0])^2 - 4\mathbf{E}[X_0^2] \mathbf{E}[Y_0^2] = 4\mathbf{E}[X_0 Y_0]^2 - 4\mathbf{E}[X_0^2] \mathbf{E}[Y_0^2],$$

e quindi avremo $\mathbf{E}[X_0 Y_0]^2 \leq \mathbf{E}[X_0^2] \mathbf{E}[Y_0^2]$.

Sostituendo semplicemente $X_0 = X - \mathbf{E}[X]$ e $Y_0 = Y - \mathbf{E}[Y]$ si trova che $\text{Cov}(X, Y)^2 \leq \text{Var}(X) \text{Var}(Y)$, e questo risultato è equivalente alla tesi. \square

La covarianza di due v.a. è una grandezza molto importante, come vedremo in molte occasioni.

In sintesi, $\text{Cov}(X, Y)$ esprime quanto sono correlate le variazioni di X e Y dalla loro media; perciò se quando X è grande tende a esserlo anche Y , la covarianza sarà positiva, mentre se quando X è grande, Y è mediamente più piccolo di quanto lo sia quando X è piccolo, la covarianza sarà negativa.

Un altro interessante risultato è il seguente, che ci fornisce finalmente di una formula per la varianza della somma di due v.a.

Proposizione 4.18. *Se X e Y sono due v.a. dotate di momento secondo, si ha che*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \quad (4.25)$$

Ciò discende immediatamente dalla bilinearità della covarianza, scrivendo $\text{Var}(X + Y) = \text{Cov}(X + Y, X + Y)$ e svolgendoli conti.

Corollario. Se X e Y sono indipendenti, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Osservazione. Una utile applicazione di questo fatto è il calcolo della varianza di una v.a. binomiale.

Sia X una v.a. binomiale di parametri n e p . Siano Y_1, Y_2, \dots, Y_n v.a. bernoulliane di media p e tutte indipendenti. Sappiamo dall'Esempio 4.13 che X e $\sum_{k=1}^n Y_k$ hanno la stessa legge, quindi si ha

$$\text{Var}(X) = \text{Var}\left(\sum_{k=1}^n Y_k\right) = \sum_{k=1}^n \text{Var}(Y_k) = np(1-p)$$

4.3.2 Coefficiente di correlazione

Siccome il modulo di $\text{Cov}(X, Y)$ dipende fortemente dalla forza delle deviazioni di X e Y dalle loro medie, è molto utile riscalarne opportunamente la covarianza.

Definizione. Se X e Y sono due v.a. dotate di momento secondo, si chiama *coefficiente di correlazione lineare* di X e Y la grandezza

$$\varrho_{XY} := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}. \quad (4.26)$$

Le principali proprietà di ϱ_{XY} sono:

1. se X e Y sono indipendenti, $\varrho_{XY}=0$;
2. $-1 \leq \varrho_{XY} \leq 1$;
3. $\varrho_{XY} = \pm 1$ se e solo se X e Y sono in relazione lineare, ovvero se $X = aY + b$ per opportuni coefficienti a e b , inoltre il segno di a determina quello di ϱ_{XY} .

Le proprietà 1. e 2. sono ovvie conseguenze di quelle per la covarianza, mentre la 3. ha una dimostrazione non banale che tralasciamo.