

Lezioni di Calcolo delle probabilità

Antonelli Fabio

2021

Capitolo 1

INTRODUZIONE

I seguenti appunti sono stati elaborati nel corso delle Lezioni di Calcolo delle Probabilità per diversi corsi di laurea. Nello specifico sono pensati sia per un corso di laurea triennale in Ingegneria dell'Informazione sia per il corso di laurea triennale in Matematica. Per consentire l'uso degli appunti da parte di entrambi i corsi di laurea abbiamo evidenziato con la lettera M le parti più squisitamente dedicate agli studenti di Matematica, che invece gli studenti di Ingegneria possono tralasciare.

Alla fine di ogni paragrafo si trova una sezione di esercizi.

Capitolo 2

SPAZI DI PROBABILITÀ

2.1 Definizioni

La teoria della probabilità si è sviluppata per modellizzare fenomeni **aleatori** che presentano un esito incerto, ovvero diversi possibili risultati. Non essendoci una legge deterministica (una funzione) che governa l'output una volta stabilito l'input, occorre sviluppare una teoria che consenta di gestire questa incertezza.

Esempi 2.1.1. *Come esempi di fenomeni aleatori possiamo immediatamente pensare a*

1. *Il lancio di una moneta o di un dado;*
2. *l'estrazione di cinque carte da un mazzo;*
3. *il numero di telefonate che arrivano ad un centralino in un certo intervallo di tempo;*
4. *la lunghezza del lancio di un peso;*
5. *il tempo di attesa ad una fermata d'autobus.*

I primi tre fenomeni descritti sono cosiddetti **discreti** in quanto i possibili risultati vengono quantificati mediante un insieme finito o numerabile di valori. Gli ultimi due, invece, descrivono fenomeni **continui**, che possono assumere valori in un intervallo della retta reale. Per semplicità abbiamo citato fenomeni unidimensionali, ma si possono facilmente pensare fenomeni aleatori, sia discreti che continui, che vivono in spazi a più dimensioni, come ad esempio il numero di clienti in arrivo ad un servizio ed il numero di clienti serviti, oppure la posizione di un animale in una certa regione.

In un fenomeno aleatorio sappiamo che possono avvenire più risultati, che l'esito è incerto, ma come riuscire a quantificare questa incertezza? Ovvero come calcolare le possibilità a favore o contro un certo evento? In altre parole, come attribuire la probabilità che tale evento accada? Nel corso dei decenni vi sono stati diversi approcci alla teoria della probabilità: frequentista, soggettivista, assiomatico.

Noi presenteremo soltanto l'ultimo, che è quello più generalmente utilizzato.

Il primo passo è quello di definire un ambiente nel quale descrivere un qualsiasi fenomeno aleatorio e per farlo abbiamo bisogno di vari ingredienti.

1. Abbiamo bisogno di un insieme che definisca l'universo nel quale i risultati vengono determinati:

Ω detto **spazio degli eventi**.

2. Fissato Ω , abbiamo bisogno di dire cosa siamo in grado di riconoscere nel nostro universo. Per esempio, se si sta osservando come si addensano certe colonie di batteri, gli insiemi spaziali che potremo osservare dipenderanno dalla potenza del microscopio.

Abbiamo dunque bisogno di introdurre una famiglia di sottoinsiemi di Ω , che indichiamo con \mathcal{F} , che rappresenta l'assetto informativo riguardante un fenomeno aleatorio, ovvero che identifica quali insiemi di risultati possiamo osservare e riconoscere.

In questa famiglia ci aspettiamo di

- essere in grado di riconoscere il nulla e il tutto;
 - essere in grado, se possiamo osservare un evento, di dire anche cosa non abbiamo osservato;
 - poter riconoscere tanti eventi, contemporaneamente o alternativamente se sappiamo riconoscerli individualmente.
3. Infine, una volta elencati tutti i possibili eventi che siamo in grado di riconoscere, abbiamo bisogno di una regola, P , che quantifichi quanto è possibile l'avverarsi di ognuno di essi. Questa deve essere quindi una funzione che agisce sui sottoinsiemi in \mathcal{F} e deve dare delle regole per dedurre la possibilità di avverarsi di eventi alternativi o contemporanei.

Matematicamente quindi scriviamo la seguente definizione che formalizza i tre ingredienti fondamentali, appena descritti.

Definizione 2.1.1. *Uno **spazio di probabilità** è una terna (Ω, \mathcal{F}, P) , dove*

1. Ω è un insieme (spazio degli eventi).
2. \mathcal{F} è una σ -algebra su Ω , ovvero è una famiglia di sottoinsiemi di Ω tale che

$$(a) \emptyset, \Omega \in \mathcal{F};$$

$$(b) \text{ se } A \in \mathcal{F}, \text{ allora anche } A^c \in \mathcal{F};$$

$$(c) \text{ se } \{A_i\}_{i \in I} \subseteq \mathcal{F}, \text{ con } I \subseteq \mathbb{N} \text{ (una famiglia al più numerabile di insiemi di } \mathcal{F}), \text{ allora anche}$$

$$A_1 \cup A_2 \cup \dots \cup A_n \cup \dots = \bigcup_{i \in I} A_i \in \mathcal{F}.$$

3. $P : \mathcal{F} \rightarrow [0, 1]$ è una probabilità, ovvero una funzione di insieme tale che

$$(a) P(\Omega) = 1;$$

(b) (σ -additività) Per $I \subseteq \mathbb{N}$, se $\{A_i\}_{i \in I} \subseteq \mathcal{F}$, $A_i \cap A_j = \emptyset$ per ogni $i \neq j$, allora

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i) = P(A_1) + \dots + P(A_n) \dots$$

Osservazione 2.1.1. La σ -algebra \mathcal{F} rappresenta l'assetto informativo riguardante un fenomeno aleatorio, ovvero è formata da quegli insiemi di risultati che possiamo osservare e riconoscere.

Possiamo essere in grado di effettuare un'osservazione più o meno accurata. Pensiamo di avere due σ -algebre \mathcal{G} e \mathcal{F} che rappresentano due assetti informativi riguardo ad uno stesso fenomeno. Se l'osservazione realizzata da \mathcal{G} è più fine (ovvero contiene più informazioni) di quella determinata da \mathcal{F} , si scrive $\mathcal{F} \subseteq \mathcal{G}$ e ogni insieme di \mathcal{F} deve essere anche un insieme di \mathcal{G} , cioè gli insiemi in \mathcal{G} sono più piccoli degli insiemi in \mathcal{F} , perché si possono ricostruire gli insiemi (e quindi tutte le informazioni) di \mathcal{F} utilizzando gli insiemi di \mathcal{G} .

Le tre proprietà che definiscono una σ -algebra sono molto naturali e sono la traduzione matematica delle proprietà elencate prima per descrivere un assetto informativo.

Infine la probabilità rappresenta una misura dell'incertezza di ogni possibile risultato. Attribuiamo misura 1 alla certezza (100%, qualcosa deve pur verificarsi) e la seconda proprietà dice che le possibilità si sommano su eventi mutualmente esclusivi.

Diciamo cardinalità di un insieme A e la indichiamo con $|A|$ il numero di elementi, se finito, che esso contiene.

Possiamo estendere il concetto di cardinalità ad insiemi che sono in corrispondenza biunivoca con \mathbb{N} , in questo caso si dice che l'insieme ha la cardinalità del numerabile.

Se Ω è costituito da una quantità al più numerabile di elementi, si dice che è uno spazio discreto di probabilità. In particolare, se $|\Omega| = n < +\infty$, si dice che lo spazio è **finito** e di solito si assume l'insieme delle parti di Ω , $\mathcal{P}(\Omega)$, che è la famiglia di tutti i sottoinsiemi di Ω , come la σ -algebra di riferimento. In questi casi non si citerà esplicitamente la σ -algebra in quanto si sta assumendo di avere l'informazione più completa possibile.

Esempio 2.1.1. Supponiamo di stare studiando il fenomeno aleatorio rappresentato dal lancio di un dado.

In questo caso la naturale scelta per Ω è data dall'insieme $\{1, 2, 3, 4, 5, 6\}$, che rappresenta tutti i possibili risultati. Poiché siamo in un insieme di cardinalità pari a 6, come σ -algebra considereremo l'insieme delle parti di Ω , ovvero la famiglia di tutti i suoi sottoinsiemi

$$\mathcal{P}(\Omega) = \left\{ \emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{2, 3\}, \dots, \{1, 2, 3\}, \{1, 2, 4\}, \dots, \dots \right\},$$

che permette di rappresentare tutte le possibili combinazioni di risultati. Per esempio:

$$\{\text{esce un pari}\} = \{2, 4, 6\}, \text{ oppure } \{\text{esce un numero minore o uguale a 4}\} = \{1, 2, 3, 4\}.$$

Infine, sulla base della convinzione che il dado sia equo, attribuiamo stessa probabilità ad ogni singolo risultato, ovvero

$$P(\{i\}) = \frac{1}{6}, \quad i = 1, \dots, 6 \quad (\text{la probabilità che esce il numero } i).$$

Sfruttando quindi la proprietà (b) della probabilità possiamo attribuire la probabilità anche a tutti gli altri eventi. Per esempio

$$\begin{aligned} P(\{\text{esce un pari}\}) &= P(\{\text{esce 2 od esce 4 od esce 6}\}) = P(\{2, 4, 6\}) \\ &= P(\{2\} \cup \{4\} \cup \{6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

Facciamo notare quanto la proprietà (b) della probabilità sia importante che valga solo per insiemi disgiunti, infatti se si applicasse a qualsiasi famiglia di insiemi e non solo a quelle di insiemi disgiunti, arriveremmo ad una contraddizione. Consideriamo

$$A = \{1, 2, 3, 4\}, \quad B = \{3, 4, 5, 6\} \quad \Rightarrow \quad P(A) = \frac{2}{3} = P(B), \quad A \cap B \neq \emptyset.$$

Se dicessimo che anche in questo caso possiamo sommare le probabilità, arriveremmo ad un assurdo

$$1 = P(\{1, 2, 3, 4, 5, 6\}) = P(A \cup B) = P(A) + P(B) = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}.$$

Nel precedente esempio abbiamo visto che ci bastavano i 6 risultati per costruire qualsiasi altro evento. Nella realtà spesso le nostre osservazioni sono basate su una piccola famiglia di insiemi, sulla base dei quali costruiamo un'informazione più ampia. Matematicamente questo si traduce nella seguente

Definizione 2.1.2. Se \mathcal{A} è una famiglia di insiemi di Ω , viene detta la σ -algebra generata da \mathcal{A} , la più piccola σ -algebra contenente \mathcal{A} e si indica con $\sigma(\mathcal{A})$, ovvero

$$\sigma(\mathcal{A}) = \bigcap \{ \mathcal{F}, \sigma\text{-algebra}, \mathcal{A} \subseteq \mathcal{F} \}$$

La nozione di σ -algebra generata da una sottofamiglia di insiemi è compatibile con le proprietà di σ -algebra poiché l'intersezione di una qualsiasi famiglia di σ -algre (anche infinita) è ancora una σ -algebra. (**M:** dimostrarlo per esercizio)

Esempi 2.1.2. 1. Se $\mathcal{A} = \{A\}$ allora $\sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\}$.

2. Se $\mathcal{A} = \{A, B\}$ allora $\sigma(\mathcal{A}) = \{\emptyset, \Omega, A, A^c, B, B^c, A \cap B, A \cap B^c, A^c \cap B, A^c \cap B^c, A \cup B, A^c \cup B, A \cup B^c, A^c \cup B^c\}$.

3. $\mathcal{F} = \{[a, b), a, b \in \mathbb{R}\}$ non è una σ -algebra poiché il complementare di qualsiasi intervallo finito non è un intervallo finito e quindi non si trova in \mathcal{F} .

4. Dato $n \in \mathbb{N}$ e l'intervallo $[0, 1)$, la famiglia di sottoinsiemi

$$\mathcal{F} = \left\{ \emptyset, \left[0, \frac{1}{n}\right), \left[\frac{1}{n}, \frac{2}{n}\right), \dots, \left[\frac{n-1}{n}, 1\right) \right\} \text{ e tutte le loro unioni } \Bigg\}$$

è una σ -algebra.

5. La più piccola σ -algebra di \mathbb{R} che contiene gli intervalli aperti è denotata $\mathcal{B}(\mathbb{R})$ ed è chiamata la σ -algebra dei Boreliani di \mathbb{R} .

In buona sostanza per avere una σ -algebra è necessario che il complementare di un qualsiasi insieme della σ -algebra sia ancora nella σ -algebra, così come qualsiasi unione o intersezione al più numerabile di insiemi della σ -algebra. Si dice che le σ -algebre sono famiglie di insiemi chiuse rispetto alle operazioni insiemistiche di complemento, unione numerabile e intersezione numerabile.

Ricordiamo che

$$\begin{aligned}(A \cap B) \cup C &= (A \cup C) \cap (B \cup C), & (A \cup B) \cap C &= (A \cap C) \cup (B \cap C) \\ (A \cap B)^c &= A^c \cup B^c, & (A \cup B)^c &= A^c \cap B^c, & (A^c)^c &= A\end{aligned}$$

Osservazione 2.1.2. Usando la proprietà 3b, si ha per $A, B \in \mathcal{F}, \{A_n\}_n \subseteq \mathcal{F}$

1. $P(A^c) = 1 - P(A)$, poiché $\Omega = A \cup A^c, \emptyset = A \cap A^c$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, poiché possiamo scrivere in maniera disgiunta $A \cup B = (A \cap B^c) \cup (A \cap B) \cup (B \cap A^c)$, da cui

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(B \cap A^c) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B)$$

poiché $A = (A \cap B^c) \cup (A \cap B)$ e $B = (B \cap A^c) \cup (B \cap A)$ come unioni disgiunte.
3. $P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = 1 - P\left(\bigcup_{n \in \mathbb{N}} A_n^c\right)$
4. La 2. può essere generalizzata a più eventi (principio di esclusione - inclusione)

$$\begin{aligned}P(E_1 \cup \dots \cup E_n) &= \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \dots \\ &+ (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_r}) + \dots + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n)\end{aligned}$$

M: dimostrare per induzione, per esercizio.

La probabilità gode delle seguenti proprietà

Proposizione 2.1.1. Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Allora

1. per ogni $A, B \in \mathcal{F}$ tali che $A \subseteq B$ si ha $P(A) \leq P(B)$;
2. per ogni successione di insiemi in $\{A_n\} \subseteq \mathcal{F}$, tale che $A_1 \subseteq A_2 \subseteq \dots$ (o alternativamente $A_1 \supseteq A_2 \supseteq \dots$), si ha

$$\lim_{k \rightarrow +\infty} P(A_k) = P\left(\bigcup_{n=1}^{+\infty} A_n\right), \quad \left(\lim_{k \rightarrow +\infty} P(A_k) = P\left(\bigcap_{n=1}^{+\infty} A_n\right)\right)$$

3. (subadditività) per ogni $\{A_n\}_n \subseteq \mathcal{F}$

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) \leq \sum_{n=1}^{+\infty} P(A_n)$$

Dimostrazione: Poiché $A \subseteq B$ si ha $B = A \cup (B \cap A^c)$, che è un'unione disgiunta. Dalla seconda proprietà della probabilità si ha quindi

$$P(B) = P(A) + P(B \cap A^c) \Rightarrow P(A) \leq P(B),$$

essendo $P(B \cap A^c) \geq 0$ e la prima affermazione è dimostrata.

Per la seconda, costruiamo la seguente successione di insiemi in \mathcal{F}

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \dots, B_k = A_k \setminus A_{k-1}, \dots$$

Siccome gli A_i sono insiemi contenuti uno dentro l'altro, i B_i sono insiemi disgiunti e grazie a questa relazione di inclusione si ha anche

$$A_k = A_1 \cup \dots \cup A_k = B_1 \cup \dots \cup B_k, \quad \text{e} \quad \bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i.$$

Quindi

$$\begin{aligned} \lim_{k \rightarrow +\infty} P(A_k) &= \lim_{k \rightarrow +\infty} P(B_1 \cup \dots \cup B_k) = \lim_{k \rightarrow +\infty} \sum_{n=1}^k P(B_n) \\ &= \sum_{n=1}^{+\infty} P(B_n) = P\left(\bigcup_{n=1}^{+\infty} B_n\right) = P\left(\bigcup_{n=1}^{+\infty} A_n\right). \end{aligned}$$

Infine per dimostrare la terza affermazione, notiamo che per una generica famiglia di insiemi $\{A_n\}_n$, gli insiemi definiti come

$$B_1 = A_1, \dots, B_k = A_k \cap A_{k-1}^c \cap \dots \cap A_1^c$$

sono disgiunti e $A_1 \cup \dots \cup A_n = B_1 \cup \dots \cup B_n$, dunque per le due punti dimostrati precedentemente, si ha

$$\begin{aligned} P\left(\bigcup_{i=1}^{+\infty} A_i\right) &= \lim_{n \rightarrow +\infty} P\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{i=1}^n B_i\right) \\ &= \lim_{n \rightarrow +\infty} \sum_{i=1}^n P(B_i) \leq \lim_{n \rightarrow +\infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^{+\infty} P(A_i) \end{aligned}$$

completando la dimostrazione. \square

Le proprietà introdotte finora non dipendono dallo spazio di probabilità. Non abbiamo ancora affrontato il problema di come attribuire la probabilità ai vari eventi.

2.2 Spazi di probabilità equiprobabili

Se lanciamo un dado equilibrato, è naturale attribuire probabilità pari ad $\frac{1}{6}$ all'uscita del numero 6, pensando che si ha una possibilità su sei di ottenere il risultato desiderato. Come si arriva a questa conclusione?

Quel che stiamo pensando implicitamente è che il dado sia equo e che nessuna faccia abbia una probabilità maggiore di mostrarsi rispetto alle altre. Quindi l'unica possibilità è che tutte le facce siano equiprobabili, il che implica necessariamente che ognuna ha probabilità $\frac{1}{6}$ di mostrarsi.

In generale, supponiamo di avere un insieme Ω di cardinalità finita $|\Omega| = m \in \mathbb{N}$. Possiamo scrivere $\Omega = \{\omega_1, \dots, \omega_m\}$ e se attribuiamo uniformemente la probabilità, ovvero

$$P(\omega_i) = \frac{1}{m} = \frac{1}{|\Omega|}, \quad \forall \quad i = 1, \dots, m,$$

diciamo che $(\Omega, \mathcal{P}(\Omega), P)$ è uno spazio di probabilità **uniforme o equiprobabile**.

In questo caso, come detto prima, consideriamo $\mathcal{F} = \mathcal{P}(\Omega)$ e quindi anche i singoli elementi di ogni insieme $A \subseteq \Omega$ sono in \mathcal{F} . Questi sono tutti insiemi disgiunti, ognuno di probabilità $\frac{1}{|\Omega|}$, quindi sommando le probabilità (σ -additività di P) si ottiene che

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\# \text{ casi favorevoli}}{\# \text{ casi totali}}.$$

Quindi nel caso di fenomeni a probabilità uniforme, il passo fondamentale è individuare correttamente lo spazio Ω nel quale ci si muove poiché la sua cardinalità determinerà la probabilità di ogni sottoinsieme.

Osserviamo che lo spazio Ω può essere qualsiasi: numerico, alfabetico, qualitativo.

Esempi 2.2.1.

1. *Lancio di una moneta:* $\Omega = \{T, C\} \Rightarrow |\Omega| = 2 \Rightarrow P(T) = P(C) = \frac{1}{2}$.
2. *Lanci ripetuti di una moneta:* supponiamo di lanciare una moneta m volte, allora

$$\Omega = \{(\omega_1, \dots, \omega_m) : \omega_i = T, C, i = 1, \dots, m\},$$

ovvero tutte le m -ple le cui componenti possono prendere soltanto i valori T, C . Per ogni componente abbiamo due possibilità, quindi il numero di tutte le possibili m -ple così fatte è 2^m . Quindi la probabilità di un qualsiasi risultato di m lanci di una moneta è $\frac{1}{2^m}$

3. *Si lanciano due dadi e vogliamo capire con che probabilità si può ottenere la somma dei valori uguale a 8. Chiaramente non vi è motivo per cui un'uscita sia più probabile di un'altra e quindi siamo in uno spazio di probabilità uniforme, che dobbiamo individuare correttamente.*

Se facessimo il seguente ragionamento sarebbe sbagliato: le possibili somme che si possono ottenere dal lancio di due dadi sono $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, quindi

$$P(\text{somma} = 8) = \frac{1}{11}. \quad \textbf{ERRATO}$$

Così facendo staremmo perdendo di vista il contributo singolo dei due dadi, infatti il numero di combinazioni che formano una somma non è uguale per tutti i valori. Il modo corretto è

$$\Omega = \{(i, j), i, j = 1, \dots, 6\} \Rightarrow |\Omega| = 36, \Rightarrow P((i, j)) = \frac{1}{36}, \forall i, j = 1, \dots, 6.$$

Dunque le possibili combinazioni che danno somma 8 sono

$$A = \{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} \Rightarrow P(A) = \frac{5}{36}.$$

Diventa quindi importante saper contare correttamente la cardinalità degli insiemi che si devono considerare.

2.2.1 Pillole di combinatoria

In questa sezione richiamiamo alcuni cenni di calcolo combinatorio da utilizzare per stabilire la cardinalità di uno spazio degli eventi.

1. Se abbiamo due insiemi A e B allora definiamo **prodotto cartesiano** dei due insiemi, l'insieme delle coppie ordinate

$$A \times B = \{(i, j), : i \in A, j \in B\},$$

se i due insiemi sono finiti si ha

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

Il concetto di prodotto cartesiano può essere esteso a qualsiasi numero di insiemi, formando le triple, le quadruple, le n -ple ordinate e la cardinalità del prodotto cartesiano si ottiene moltiplicando le rispettive cardinalità.

2. Fattoriale.

Supponiamo di avere n scatole ed n palline numerate (ovvero distinguibili) da disporre nelle scatole, una per scatola. Quanti saranno i possibili modi per farlo?

Possiamo pensare di avere, man mano che si occupano le scatole:

n scelte per la prima pallina

$n - 1$ scelte per la seconda pallina

\vdots

2 scelte per la $n - 1$ -ma pallina

1 scelta per la n -ma pallina

per un totale di $n \cdot (n - 1) \cdots 2 \cdot 1$. Questo numero si indica con $n!$ (n fattoriale), è definito per $n \in \mathbb{N}$ ed indica anche il numero delle permutazioni della sequenza fondamentale $(1, 2, 3, \dots, n)$.

Altresì, il fattoriale può essere visto come il numero di funzioni iniettive da un insieme con n elementi in sè stesso.

Proprietà del fattoriale

- (a) $0! = 1$ (vi è un solo modo di non mettere palline nelle scatole);
- (b) $n! = n \cdot (n - 1)!$

3. Disposizioni.

Immaginiamo ora di avere m scatole ed n palline distinguibili con $n \leq m$ e di volerle disporre di nuovo una per scatola. Procedendo come prima si avrà:

m scelte per la prima pallina

$m - 1$ scelte per la seconda pallina

\vdots

$m - (n - 1)$ scelte per la n -ma pallina

ottenendo un totale di $m \cdot (m - 1) \cdots (m - n + 1) = \frac{m!}{(m - n)!}$ possibili casi, che rappresenta anche il numero di possibili estrazioni con ordinamento di n elementi da un insieme di m elementi.

4. Combinazioni.

Nel momento in cui ripensiamo le palline come indistinguibili, dobbiamo dividere il precedente numero per $n!$ in quanto vediamo soltanto le diverse occupazioni delle caselle ma non distinguiamo quali elementi le occupano, quindi una volta disposte in n caselle fissate qualsiasi n palline, poiché non si possono distinguere, le $n!$ permutazioni di questi elementi daranno luogo sempre alla stessa occupazione.

Abbiamo dunque che i possibili modi di disporre n palline indistinguibili in $m \geq n$ caselle è

$$\frac{m!}{n!(m - n)!}, \quad m, n \in \mathbb{N}$$

e rappresenta il numero delle possibili combinazioni di n elementi da un insieme di m o, alternativamente, il numero di possibili estrazioni senza ordinamento di n elementi da un insieme di m .

Questa quantità è detta **coefficiente binomiale** e si scrive

$$\binom{m}{n} = \frac{m!}{n!(m - n)!}$$

Proprietà del coefficiente binomiale

- (a) $\binom{m}{0} = 1, \binom{m}{m} = 1, \binom{m}{1} = m, \binom{m}{m-1} = m;$
- (b) $\binom{m}{n} = \binom{m}{m-n};$
- (c) $\binom{m}{n} = \binom{m-1}{n-1} + \binom{m-1}{n},$ poiché fissata una scatola come occupata da una pallina (primo caso) o vuota (secondo caso), si contano le possibili combinazioni sulle restanti scatole.
- (d) Binomio di Newton (**M**: dimostrare per induzione)

$$(a+b)^n = \sum_{n=0}^m \binom{m}{n} a^n b^{(m-n)}, \quad a, b \in \mathbb{R},$$

$$\text{da cui consegue anche } \sum_{n=0}^m \binom{m}{n} = 2^m, \quad \Rightarrow \quad \sum_{n=0}^m \binom{m}{n} \frac{1}{2^m} = 1.$$

5. Infine per calcolare il numero di possibili modi di mettere n palline in m caselle si ha:
- m scelte per la prima pallina
 - m scelte per la seconda pallina
 - \vdots
 - m scelte per la n -ma pallina
- e quindi un totale di m^n possibili scelte, che coincide anche con il numero di tutte le possibili funzioni da un insieme con n oggetti ad un insieme con m oggetti.

Esempi 2.2.2.

1. *Vengono formate targhe con 7 simboli: due lettere da un insieme alfabetico di 26 e 5 numeri da un insieme di 10 cifre. Quante sono le possibili targhe che si possono formare se*
 - (a) *i simboli si possono ripetere;*
 - (b) *i simboli non si possono ripetere e le combinazioni sono considerate con ordinamento;*
 - (c) *i simboli non si possono ripetere e le combinazioni sono considerate senza ordinamento.*
2. *Si lanciano due dadi. Calcolare la probabilità dei seguenti eventi*
 - (a) $E = \{ \text{somma dei risultati} = \text{dispari} \};$
 - (b) $F = \{ \text{almeno uno dei due risultati} = 1 \};$

- (c) $G = \{ \text{somma dei risultati} = 5 \};$
 (d) $E \cap F, E \cup F, F \cap G, E \cap F^c, E \cap F \cap G$
3. Nel gioco del poker con un mazzo di carte francesi (52 carte), calcolare la probabilità di avere un poker servito in prima mano.

Soluzioni:

1. Nel primo caso abbiamo 26 simboli da cui scegliere per la prima lettera ed altrettanti per la seconda, analogamente per i 5 numeri abbiamo 10 possibilità per ognuno, quindi il numero totale di possibili targhe che si possono formare è

$$26^2 10^5.$$

Nel secondo caso abbiamo 26 scelte per la prima lettera e 25 per la seconda, mentre abbiamo 10 scelte per il primo numero quindi 9 per il secondo, 8 per il terzo, 7 per il quarto ed infine 6 per il quinto, dunque il numero totale di targhe è dato da $26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6$.

Nel terzo caso non possiamo più distinguere la posizione degli elementi perciò e il precedente numero va diviso per $2!$ (lettere) e per $5!$ (numeri), ottenendo

$$\frac{26 \cdot 25}{2!} \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5!} = \binom{26}{2} \binom{10}{5}$$

2. Nel secondo esempio abbiamo che $\Omega = \{(i, j) : i, j = 1, \dots, 6\}$ con cardinalità 36.

Le coppie in E sono quelle costituite da un dispari e da un pari, che devono essere la metà dei casi totali ovvero 18, da cui $P(E) = \frac{1}{2}$. Per quel che riguarda F , se 1 è uscito sul primo dado allora abbiamo 6 possibilità per il secondo dado, viceversa se 1 è uscito sul secondo dado abbiamo 5 possibilità distinte per il secondo, poiché la coppia (1,1) è stata già contata, abbiamo quindi un totale di 11 casi favorevoli, da cui $P(F) = \frac{11}{36}$.

$$G = \{(1, 4), (2, 3), (3, 2), (4, 1)\} \Rightarrow P(G) = \frac{4}{36}.$$

$$E \cap F = \{(1, 2), (2, 1), (1, 4), (4, 1), (1, 6), (6, 1)\} \Rightarrow P(E \cap F) = \frac{6}{36}.$$

$$E \cup F = E \cup \{(1, 1), (3, 1), (1, 3), (1, 5), (5, 1)\} \Rightarrow P(E \cup F) = \frac{23}{36}.$$

$$E \cap F^c = E \setminus (E \cap F) \Rightarrow P(E \cap F^c) = \frac{12}{36}.$$

$$E \cap F \cap G = F \cap G = \{(1, 4), (4, 1)\} \Rightarrow P(E \cap F \cap G) = P(F \cap G) = \frac{2}{36}.$$

3. Nel gioco del poker, pensando di essere serviti per primi, si estraggono senza reinserimento 5 carte da un mazzo di 52 carte, quindi il numero di casi possibili è $\binom{52}{5}$. Per contare il numero di casi favorevoli, notiamo innanzitutto che la probabilità di ottenere un poker di assi deve essere la stessa di quella di ottenere un poker di K e via dicendo quindi vuole dire che abbiamo 13 differenti poker, ognuno con eguale probabilità, quindi $P(\text{poker}) = 13 P(\text{poker assi})$. D'altra parte il poker di assi si ottiene estraendo tutti gli assi (1 sola possibilità) ed una qualsiasi delle altre 48 carte, quindi vi sono $\binom{48}{1}$ casi favorevoli. Concludendo

$$P(\text{poker}) = 13 \frac{\binom{48}{1}}{\binom{52}{5}}.$$

2.3 Revisione di teoria

1. Dare la definizione di σ -algebra.
2. Se A, B sono due sottoinsiemi di Ω , qual è la σ -algebra generata da essi?
3. Dimostrare che

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

4. Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Dimostrare che se $\{A_n\}_{n \in \mathbb{N}}$ in \mathcal{F} è una successione crescente di eventi, allora

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n)$$

5. Dire quali delle seguenti è vera
 - (a) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
 - (b) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
 - (c) $(A \cup B) \cup C = A \cup (B \cup C)$
 - (d) $(A \cup B) \cap C = A \cup (B \cap C)$
 - (e) $(A \cup B) \cap C = (A \cup C) \cap (B \cup C)$
 - (f) Se A e B sono due insiemi, qual è la σ -algebra generata da $A \cup B$?
6. **M:** Dimostrare per induzione la formula di inclusione-esclusione.
7. **M:** Dimostrare per induzione la formula del binomio di Newton.

2.4 Esercizi

1. Quante combinazioni di lettere di lunghezza 5 si possono formare da un alfabeto di 21 lettere se si ammettono ripetizioni? E se vi possono essere al massimo due lettere uguali? Considerare il caso con e senza ordinamento.
2. Ad un tavolo tondo vi sono 10 posti. Gli ospiti si siedono a caso. Qual è la probabilità che A e B siedano vicini?
3. In una classe con banchi da due vi sono $2n$ alunni ed esattamente n banchi. In quanti modi possono essere disposti, facendoli sedere casualmente?
4. Vi sono sei tazze con piattini coordinati, due rosse, due bianche e due blu. Le tazze vengono messe a caso sui piattini. Calcolare la probabilità che nessuna tazza sia accordata con il piattino.
5. Si lanciano due dadi equi. Calcolare
 - (a) la probabilità $d(A) = \{ \text{la differenza tra i due risultati è in modulo uguale a 2} \}$.
 - (b) Siano $B = \{ \text{esce un dispari su almeno uno dei due dadi} \}$, $C = \{ \text{esce un dispari sul dado 1} \}$. Calcolare $P(A \cap B)$ e $P(A \cap C)$.
6. Una password è costituita da 8 caratteri e viene generata a caso servendosi di un alfabeto di 26 lettere. Si possono usare sia lettere maiuscole che minuscole. Calcolare la probabilità di generare una password
 - (a) con sole maiuscole se sono ammesse ripetizioni;
 - (b) con sole maiuscole se non sono ammesse ripetizioni. Considerare il caso con e senza ordinamento.
7. **M:** Una password, per essere generata deve contenere un minimo di 8 caratteri fino ad un massimo di 16 e deve contenere almeno un numero, una lettera minuscola ed una lettera maiuscola. Quante sono le combinazioni possibili, considerando che l'ordine conta?
8. Pescando 5 lettere senza reinserimento da un sacchetto che contiene le 21 lettere dell'alfabeto italiano, qual è la probabilità di pescare solo vocali?
9. Si gioca a sette e mezzo e si è il/la primo/a a giocare. Dal mazzo di 40 carte (4 semi, tre figure e i numeri da 1 a 7 per ciascun seme) si estraggono due carte. Considerando il re di denari come le altre figure, calcolare la probabilità di avere un sette e mezzo servito (ovvero che esca un sette ed una figura). Le figure valgono mezzo punto e i numeri il loro valore.

Capitolo 3

PROBABILITÀ CONDIZIONATA ED INDIPENDENZA

3.1 Probabilità condizionata

Introduciamo ora uno dei concetti più importanti della teoria della probabilità. Iniziamo con un esempio.

Esempio 3.1.1. *Abbiamo un'urna con 6 palline nere e 4 bianche e si effettuano 2 estrazioni senza reinserimento. Vogliamo valutare la probabilità che alla seconda estrazione esca una pallina bianca.*

Non possiamo rispondere in maniera immediata a questa domanda, ma se ci viene detto cosa è successo nella prima estrazione, invece sì. Infatti se alla prima estrazione è uscita una pallina bianca, allora la probabilità di estrarne un'altra pallina bianca è $\frac{3}{9}$, altrimenti è di $\frac{4}{9}$.

Come sfruttare queste informazioni parziali per rispondere definitivamente al precedente quesito?

Definizione 3.1.1. *siano $A, B \in \mathcal{F}$ con $P(B) > 0$, si dice **probabilità di A condizionata a B***

$$(3.1) \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

La precedente definizione va interpretata come la formula per riassegnare la probabilità ad ogni singolo evento, una volta che si sa che l'evento B è avvenuto. Infatti la formula ci dice che la probabilità di A dato B è la probabilità che entrambi gli eventi siano accaduti, rinormalizzata (ovvero riscalata) usando la probabilità di B . Notiamo che $P(B|B) = 1$ ed è quindi come se B fosse diventato l'evento certo, l'universo nel quale adesso ci si muove.

Osservazione 3.1.1.

1. La funzione di insieme $P(\cdot|B)$ è effettivamente una probabilità, infatti

$$(a) \quad 0 \leq P(A|B) \leq 1 \text{ per ogni } A \in \mathcal{F} \text{ e } P(\Omega|B) = 1.$$

(b) Se $\{A_n\}_n \subseteq \mathcal{F}$ è una successione di eventi disgiunti e si ha

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n | B\right) &= \frac{P\left(\bigcup_{n=1}^{\infty} A_n \cap B\right)}{P(B)} = \frac{\sum_{n=1}^{\infty} P(A_n \cap B)}{P(B)} \\ &= \sum_{n=1}^{\infty} \frac{P(A_n \cap B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n | B), \end{aligned}$$

poiché anche $\{A_n \cap B\}_n \subseteq \mathcal{F}$ è una successione di eventi disgiunti.

2. La formula (3.1) può essere riletta come

$$(3.2) \quad P(A \cap B) = P(A|B)P(B),$$

facendo capire che la probabilità condizionata diventa uno strumento per assegnare la probabilità all'evento $A \cap B$. Spesso le probabilità condizionate sono molto più facili da valutare rispetto alle probabilità dirette.

Alla luce di quanto detto, riesaminiamo il nostro esempio. Indichiamo con

$$\begin{aligned} B_i &= \{\text{esce una pallina bianca alla } i\text{-esima estrazione}\} \\ N_i &= \{\text{esce una pallina nera alla } i\text{-esima estrazione}\}, \end{aligned} \quad i = 1, 2.$$

Chiaramente $B_1 \cup N_1 = \Omega$ e quindi possiamo decomporre l'evento B_2 nei due eventi disgiunti

$$B_2 = (B_2 \cap B_1) \cup (B_2 \cap N_1),$$

dunque si ha

$$P(B_2) = P(B_2 \cap B_1) + P(B_2 \cap N_1) = P(B_2|B_1)P(B_1) + P(B_2|N_1)P(N_1) = \frac{3}{9} \frac{4}{10} + \frac{4}{9} \frac{6}{10} = \frac{2}{5}$$

Vediamo ora un altro esempio che fa capire l'importanza e l'impiego della probabilità condizionata.

Esempio 3.1.2. Una certa malattia ha incidenza del 3% in una popolazione. Indichiamo con M l'evento $\{\text{si è malati}\}$ e quindi $M^c = \{\text{non si è malati}\}$. Dunque $P(M) = 3\%$ e $P(M^c) = 97\%$.

Si è sviluppato un test per diagnosticare questa malattia. Nella fase sperimentale, si è amministrato il test ad una popolazione nota e si sono registrati risultati ottenuti. Il test ha dato risultato positivo al 98% su individui malati e negativo al 96% su individui sani, ovvero si sono rilevate le seguenti probabilità condizionate

$$P(+|M) = 98\%, \quad P(-|M) = 2\%, \quad P(-|M^c) = 96\%, \quad P(+|M^c) = 4\%.$$

Nella fase di utilizzo del test si vuole calcolare la probabilità che, preso un individuo a caso, il test risulti positivo e la probabilità di aver effettivamente contratto la malattia sapendo che il test è risultato positivo. Si devono cioè calcolare $P(+)$ e $P(M|+)$.

Per la prima si nota che l'evento $+$ può essere decomposto nei due eventi disgiunti $+\cap M$ e $+\cap M^c$, quindi

$$P(+) = P(+\cap M) + P(+\cap M^c) = P(+|M)P(M) + P(+|M^c)P(M^c) = \frac{98 \cdot 3 + 4 \cdot 97}{10^4}$$

per la seconda notiamo

$$P(M|+) = \frac{P(+\cap M)}{P(+)} = \frac{P(+|M)P(M)}{P(+)} = \frac{98 \cdot 3}{98 \cdot 3 + 4 \cdot 97} = \frac{294}{682},$$

ovvero una probabilità inferiore al 50%, quindi il test non è per nulla affidabile se vogliamo stabilire se si ha la malattia sulla base della positività del test.

Quanto abbiamo utilizzato prima per calcolare le probabilità desiderate può essere formalizzato in maniera generale.

Proposizione 3.1.1. Formula di Bayes Siano $A, B \in \mathcal{F}$ con $P(A) > 0, P(B) > 0$, allora

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

Dimostrazione: Basta notare che

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \quad \square$$

Data la seguente

Definizione 3.1.2. Una famiglia di insiemi $\{E_i\}_{i=1}^n$ si dice **partizione finita** di Ω se

1. $E_i \cap E_j = \emptyset$ per ogni $i \neq j$, $i, j = 1, \dots, n$;

2. $\Omega = \bigcup_{i=1}^n E_i$

vale la seguente

Proposizione 3.1.2. Formula delle probabilità totali Sia $\{E_i\}_{i=1}^n \subseteq \mathcal{F}$ una partizione finita di Ω , allora per ogni $A \in \mathcal{F}$

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

Dimostrazione: Basta notare che quando si intersecano insiemi disgiunti con un ulteriore insieme, gli insiemi risultanti sono ancora disgiunti, quindi usando la σ -additività della probabilità si ottiene

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(A \cap \bigcup_{i=1}^n E_i\right) = P\left(\bigcup_{i=1}^n (A \cap E_i)\right) \\ &= \sum_{i=1}^n P(A \cap E_i) = \sum_{i=1}^n P(A|E_i)P(E_i) \end{aligned} \quad \square$$

Osserviamo che la formula vale anche se $P(E_i) = 0$, qualora si definisce un valore arbitrario per $P(A|E_i)$.

3.2 Indipendenza di eventi

Introduciamo un'ulteriore nozione fondamentale della teoria della probabilità, collegata in maniera diretta alla probabilità condizionata.

Definizione 3.2.1. Due eventi A e B si dicono **indipendenti** se

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

Dato $I \subseteq \mathbb{N}$, una famiglia di eventi $\{A_i\}_{i \in I}$ si dice **indipendente** se comunque scelti un intero $k \leq \text{card}(I)$, $i_1, \dots, i_k \in I$ e A_{i_1}, \dots, A_{i_k} insiemi distinti della famiglia, risulta

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

$\{A_i\}_{i \in I}$ è una famiglia di eventi **indipendenti a due a due** se per ogni $i \neq j$ risulta

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

Intuitivamente l'indipendenza tra eventi significa che l'avverarsi di uno non influenza le possibilità di accadere degli altri. Per esempio non vi è ragione per cui il lancio di una moneta e quelli di un dado debbano avere una rilevanza gli uni per gli altri e quindi si può supporre indipendenza di questi eventi. Analogamente avviene per i lanci di due o più dadi o monete o, equivalentemente, per i lanci ripetuti di una moneta o di un dado.

Se, per esempio, vogliamo calcolare la probabilità che esca lo stesso numero sui due dadi, si può procedere nel seguente modo

$$\begin{aligned} P(\text{stesso numero}) &= \sum_{i=1}^6 P(\text{sul primo dado esce } i \text{ e sul secondo dado esce } i) \\ &= \sum_{i=1}^6 P(\text{sul primo dado esce } i)P(\text{sul secondo dado esce } i) = \frac{6}{36} = \frac{1}{6}. \end{aligned}$$

Osserviamo che l'indipendenza di una famiglia di eventi implica direttamente l'indipendenza a due a due della stessa, ma il contrario non è vero come mostra il seguente esempio.

Esempio 3.2.1. Consideriamo uno spazio di probabilità equiprobabile costituito da 4 elementi $\Omega = \{1, 2, 3, 4\}$ e gli eventi $A_1 = \{1, 2\}$, $A_2 = \{1, 3\}$, $A_3 = \{1, 4\}$. Allora si ha

$$P(A_i) = \frac{1}{2} \quad \forall i = 1, 2, 3; \quad P(A_i \cap A_j) = \frac{1}{4} = P(A_i)P(A_j) \quad \forall i \neq j = 1, 2, 3$$

$$\text{ma } P(A_1 \cap A_2 \cap A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \frac{1}{8}.$$

Se A e B sono eventi indipendenti allora $P(A|B) = P(A)$. Infine evidenziamo che **eventi disgiunti non sono indipendenti**. Infatti

$$0 = P(\emptyset) = P(A \cap A^c) \neq P(A)P(A^c) = P(A)[1 - P(A)]$$

e l'uguaglianza è verificata solo se A o B è l'insieme vuoto, che è indipendente da qualsiasi altro insieme.

Osservazione 3.2.1. *In realtà l'indipendenza di eventi è una nozione che riguarda le σ -algebre e si può definire in maniera generale.*

Notiamo che se A, B sono due eventi indipendenti tra loro, allora lo stesso è vero per A^c e B o A e B^c o A^c e B^c , poiché

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c),$$

da cui abbiamo stabilito che se prendiamo

$$\sigma(A) = \{\emptyset, A, A^c, \Omega\} \quad \sigma(B) = \{\emptyset, B, B^c, \Omega\}$$

allora per ogni $C \in \sigma(A)$ e $D \in \sigma(B)$ vale

$$P(C \cap D) = P(C)P(D).$$

In particolare vale la seguente

Definizione 3.2.2. *se \mathcal{G} e \mathcal{F} sono due σ -algebre su Ω diciamo che sono indipendenti se*

$$P(A \cap B) = P(A)P(B), \quad \forall A \in \mathcal{G}, B \in \mathcal{F},$$

nozione che si può generalizzare ad un qualsiasi numero di σ -algebre.

Quando diciamo che due eventi sono indipendenti, stiamo dicendo che le σ -algebre da essi generate sono tra loro indipendenti. In altre parole, indipendenza vuol dire che i due set informativi sono indipendenti tra loro.

Osservazione 3.2.2. ATTENZIONE: *l'osservazione precedente è tesa ad evitare un comune errore che gli studenti spesso commettono definendo la nozione di indipendenza di per $n \geq 3$ eventi nella seguente maniera ERRATA:*

$$\boxed{A_1, \dots, A_n \text{ sono indipendenti se } P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n)}.$$

Infatti se così fosse, ad una qualsiasi famiglia di insiemi A_1, \dots, A_n potremo sempre aggiungere un ulteriore insieme $A_{n+1} = \emptyset$ e la precedente uguaglianza sarebbe verificata

$$\begin{aligned} P(A_1 \cap \dots \cap A_n \cap A_{n+1}) &= P(A_1 \cap \dots \cap A_n \cap \emptyset) = P(\emptyset) = 0 \\ &= P(A_1) \dots P(A_n)P(\emptyset) = P(A_1) \dots P(A_n)P(A_{n+1}) \end{aligned}$$

concludendo che A_1, \dots, A_{n+1} e conseguentemente A_1, \dots, A_n sono indipendenti, cioè che qualsiasi famiglia di eventi è indipendente: chiaramente un assurdo!

Viceversa la seguente definizione è **CORRETTA**

$$(3.3) \quad \boxed{A_1, \dots, A_n \text{ sono indipendenti se } P(B_1 \cap \dots \cap B_n) = P(B_1) \cdots P(B_n) \text{ per ogni scelta di } B_1 \in \sigma(A_1), \dots, B_n \in \sigma(A_n)}$$

poiché possiamo ricostruire le intersezioni di qualsiasi sottogruppo di eventi della famiglia come nella definizione 3.2.1 scegliendo A_{i_1}, \dots, A_{i_k} e ponendo in (7.1) tutti gli altri insiemi uguali ad Ω .

La nozione di indipendenza viene espressa rispetto ad una probabilità. Possiamo quindi pensare di definire anche una proprietà di indipendenza rispetto alla probabilità condizionata.

Definizione 3.2.3. Gli eventi $A_1, \dots, A_n \in \mathcal{F}$ si dicono **condizionatamente indipendenti** dato l'evento B con $P(B) > 0$, se per ogni $k \leq n$ e $i_1 \neq \dots \neq i_k \in \{1, \dots, n\}$ si ha

$$P(A_{i_1} \cap \dots \cap A_{i_k} | B) = P(A_{i_1} | B) \cdots P(A_{i_k} | B).$$

Le due definizioni di indipendenza e di indipendenza condizionale non hanno nessuna implicazione l'una con l'altra.

Esempio 3.2.2. Abbiamo una moneta equa ed un'urna con 5 palline bianche e 6 nere.

Si lancia la moneta e se esce Testa si mette una pallina bianca nell'urna, se esce Croce se ne mettono due.

Una volta completata l'urna si fanno due estrazioni con reinserimento e denotiamo

$A_1 =$ esce bianco in prima estrazione, $A_2 =$ esce bianco in seconda estrazione.

Se abbiamo messo esattamente una pallina bianca nell'urna l'evento $A_1 \cap A_2$ avrà probabilità $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, se invece sono state messe due palline bianche nell'urna lo stesso evento avrà probabilità $\frac{7}{13} \cdot \frac{7}{13} = \frac{49}{169}$.

In altre parole, matematicamente stiamo dicendo che gli eventi A_1 ed A_2 sono condizionatamente indipendenti

$$P(A_1 \cap A_2 | T) = P(A_1 | T) \cdot P(A_2 | T), \quad P(A_1 \cap A_2 | C) = P(A_1 | C) \cdot P(A_2 | C).$$

Non sono invece indipendenti, infatti si ha

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1 \cap A_2 | T)P(T) + P(A_1 \cap A_2 | C)P(C) \\ &= P(A_1 | T) \cdot P(A_2 | T)P(T) + P(A_1 | C) \cdot P(A_2 | C) = \frac{1}{2} \left(\frac{1}{4} + \frac{49}{169} \right) = \frac{365}{1352}, \end{aligned}$$

mentre

$$P(A_1) = P(A_2) = P(A_1 | T)P(T) + P(A_1 | C)P(C) = \frac{1}{2} \left(\frac{1}{2} + \frac{7}{13} \right) = \frac{27}{52}$$

da cui

$$P(A_1) \cdot P(A_2) = \frac{729}{2704} \neq \frac{365}{1352} = P(A_1 \cap A_2).$$

Neanche l'implicazione contraria è vera

Esempio 3.2.3. Si effettuano due lanci di una moneta equa. Chiamiamo

$$A = \{TT\} \cup \{TC\}, \quad B = \{TT\} \cup \{CT\}, \quad D = \{TC\} \cup \{CT\},$$

cioè l'evento A vuol dire che esce testa al primo lancio, il B che esce testa al secondo lancio e D che esce una sola testa su due lanci.

Chiaramente A e B sono indipendenti poiché

$$P(A) = P(B) = \frac{1}{2} \Rightarrow P(A \cap B) = P(\{TT\}) = \frac{1}{4} = P(A) \cdot P(B).$$

I due eventi non sono, invece, indipendenti condizionatamente a D . Infatti

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{1}{2} = P(B|D),$$

ma

$$P(A \cap B|D) = \frac{P(A \cap B \cap D)}{P(D)} = \frac{1}{P(D)} P(\{TT\} \cap \{ \text{esce una sola testa} \}) = 0.$$

3.3 Revisione di Teoria

1. Dare la definizione di indipendenza di n eventi A_1, \dots, A_n .
2. Dare un esempio di eventi indipendenti a due a due, ma non indipendenti.
3. Dimostrare la formula delle probabilità totali.
4. Dimostrare la formula di Bayes.
5. Se A e B sono due eventi indipendenti con $P(A) = \frac{1}{4}$ e $P(B) = \frac{1}{5}$, quanto vale $P(A^c|B^c)$?
6. Siano A, B, C tre eventi tali che $P(A|B \cap C) = \frac{2}{5}$ e $P(B|C) = \frac{1}{3}$. Calcolare $P(A \cap B|C)$.
7. Se $P(B) = \frac{1}{3}$, $P(A|B) = \frac{1}{4}$, $P(C|B \cap A) = \frac{1}{5}$, quant'è $P(A \cap B \cap C)$?

3.4 Esercizi

1. A, B estraggono ognuno due carte insieme, una volta sola, da un mazzo formato da due J, due K e due Q. Inizia A e vince chi estrae almeno un J, se l'avversario non ne estrae nessuno, altrimenti la partita è patta.
Qual è la probabilità di vincita di A , di B e di un pareggio?

2. Si devono indovinare gli ultimi due caratteri di una password e si sa che il primo è una lettera (da un alfabeto italiano di 21 lettere) e il secondo un numero. Si hanno tre tentativi, ovviamente ad ogni tentativo si escludono i caratteri che non hanno funzionato. La correttezza della lettera e quella del numero si possono osservare separatamente.

Calcolare le probabilità di indovinare al primo, al secondo ed al terzo tentativo.

3. Da un mazzo di 36 carte da poker (6,7,8,9,10,J,Q,K,A) vengono tolte quattro carte scelte a caso

Assumendo di essere i primi a giocare, che probabilità vi è di avere un poker di assi servito nell'estrazione di cinque carte dal mazzo così modificato?

4. In un torneo di tennis si gioca su 3 campi contemporaneamente in maniera indipendente. Ogni campo vede due partecipanti e ci si ferma al primo giocatore che ottiene tre set su cinque. Chiamiamo A_i, B_i $i = 1, 2, 3$ i giocatori sui quattro campi e ad ogni set, indipendentemente dagli altri, si ha

$$P(A_1 \text{ vince}) = \frac{1}{2}, \quad P(A_2 \text{ vince}) = \frac{2}{3}, \quad P(A_3 \text{ vince}) = \frac{1}{4}$$

- (a) Calcolare la probabilità che il gioco si fermi in al massimo 4 set sul campo 1, sul campo 2 e sul campo 3 rispettivamente.
- (b) Calcolare la probabilità che il gioco si fermi in al massimo 4 set su tutti e tre i campi.
5. Nello stesso contesto dell'esempio 3.1.2, calcolare anche la probabilità di essere malati sulla base di un test negativo e di essere malati sulla base di un test negativo.
6. Un televisore viene prodotto in due stabilimenti A e B , che coprono rispettivamente il 40% ed il 60% della produzione. La probabilità di un certo difetto di fabbricazione nello stabilimento A è di 0.05, mentre nello stabilimento B è di 0.15. I rivenditori vengono riforniti casualmente
- (a) Comprando un tv di questa marca, qual è la probabilità del difetto?
- (b) Sapendo che il difetto si è verificato, qual è la probabilità che il tv provenga dallo stabilimento B ?
7. In una popolazione di 100 individui, 45 hanno gruppo sanguigno A , 55 gruppo B . L'incidenza di un certo gene nel gruppo sanguigno A è dell'12%, mentre nel gruppo B è del 10%.

- (a) Prendendo a caso un individuo dalla popolazione, qual è la probabilità che sia portatore del gene?
- (b) Sapendo di avere estratto un portatore del gene, qual è la probabilità che abbia gruppo sanguigno A ?

8. Vi sono due monete, A equa e B che dà testa con una probabilità pari a $\frac{1}{4}$. Non si sa quale moneta si stia lanciando.
 - (a) Calcolare la probabilità che esca testa al primo lancio.
 - (b) Se nei primi due lanci è uscita testa, qual è la probabilità che si stia lanciando la moneta A ?
9. I componenti prodotti da una ditta possono presentare due tipi di difetti, con percentuali del 3% e del 7% rispettivamente. I due difetti si possono produrre indipendentemente.
 - (a) qual è la probabilità che un componente abbia entrambi i difetti?
 - (b) Qual è la probabilità che si presenti almeno uno dei due difetti?
 - (c) Qual è la probabilità che si presenti il difetto 1 sapendo che il componente è difettoso?
10. Un computer viene assemblato in due stabilimenti A e B , che coprono rispettivamente il 45% ed il 55% della produzione. La probabilità di rottura della ventola inserita dallo stabilimento A è di 0.05, mentre La probabilità di rottura della ventola inserita dallo stabilimento B è di 0.15. I rivenditori vengono riforniti casualmente
 - (a) Comprando un computer di questa marca, qual è la probabilità di rottura della ventola?
 - (b) Sapendo che la ventola si è rotta, qual è la probabilità che il computer provenga dallo stabilimento B ?
11. Il dado A ha 4 facce rosse e 2 blu, il dado B 2 rosse e 4 blu. Si fa il seguente gioco. Prima si lancia una moneta equa, se viene testa si lancia ripetutamente il dado A , altrimenti il dado B . Voi non sapete quale dado è stato estratto, ma vi vengono comunicati gli esiti di ogni lancio del dado.
 - (a) Mostrare che la probabilità del rosso ad ogni lancio è $\frac{1}{2}$.
 - (b) Se i primi due lanci hanno dato rosso, qual è la probabilità che venga rosso anche al terzo lancio.
12. Si sceglie a caso di lanciare un dado a 6 facce, numerate da 1 a 6, oppure un tetraedro a 4 facce, numerate da 1 a 4. Viene comunicato il risultato senza mostrare cosa si è usato.
 - (a) Qual è la probabilità che esca 4 in un lancio?
 - (b) Qual è la probabilità che si sia usato il tetraedro sapendo che su due lanci è uscito due volte 4?

Capitolo 4

VARIABILI ALEATORIE DISCRETE

4.1 Definizione e densità di probabilità

Descrivere i fenomeni aleatori a parole è spesso lungo, faticoso e soggetto ad ambiguità linguistiche che sottovalutiamo poiché spesso il significato delle frasi dipende dal contesto in cui si usano le parole, dando quindi per scontato molti dettagli che vengono sottintesi.

È quindi utile riuscire a descrivere qualsiasi fenomeno aleatorio in maniera quantitativa, al fine di eliminare le ambiguità linguistiche e la lunghezza delle descrizioni, cioè vogliamo costruire delle funzioni che associno gli eventi a dei valori numerici.

Questo è il concetto di **variabile aleatoria**, di cui diamo la definizione generale ed in questo capitolo tratteremo solo quelle discrete.

Definizione 4.1.1. *Sia (Ω, \mathcal{F}, P) uno spazio di probabilità. Una funzione*

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

*è detta **variabile aleatoria (v.a.)** se per ogni $t \in \mathbb{R}$ si ha che $\{\omega \in \Omega : X(\omega) \leq t\} = X^{-1}((-\infty, t]) \in \mathcal{F}$*

Questa definizione dice che la controimmagine di qualsiasi intervallo $(-\infty, t]$ attraverso la funzione X è un insieme della nostra σ -algebra, in altre parole possiamo sempre stabilire, sulla base del nostro set informativo, se la v.a. supera o no un qualsiasi livello t . Per convenzione, in probabilità si omette l'argomento della funzione a meno che non sia strettamente necessario, per cui l'insieme $\{\omega \in \Omega : X(\omega) \leq t\}$ viene scritto $\{X \leq t\}$.

In altre parole, una funzione X è una v.a. se siamo in grado di confrontarla con qualsiasi valore assegnato, ovvero siamo sempre in grado di osservare se si trova al di sopra o al di sotto del valore.

Osservazione 4.1.1. (M) *In Analisi Matematica. la stessa nozione definisce le funzioni misurabili. Ovvero una funzione $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ si dice misurabile se per ogni $t \in \mathbb{R}$, l'insieme $\phi^{-1}((-\infty, t])$ si trova nella σ -algebra dei Boreliani $\mathcal{B}(\mathbb{R})$, che è la σ -algebra su \mathbb{R} generata da tutti i sottoinsiemi aperti.*

Poiché ogni insieme aperto può essere scritto come unione numerabile di intervalli aperti e siccome una σ -algebra è chiusa rispetto alle unioni numerabili, $\mathcal{B}(\mathbb{R})$ è equivalentemente generata da tutti gli intervalli aperti.

D'altra parte, preso un qualsiasi intervallo aperto (a, b) , si ha

$$\begin{aligned}(a, b) &= (-\infty, b) \cap (a, +\infty) \\ (a, +\infty) &= (-\infty, a]^c \\ (-\infty, b) &= \bigcup_{n=1}^{+\infty} \left(-\infty, b - \frac{1}{n} \right]\end{aligned}$$

e una σ -algebra è chiusa rispetto ad ognuna delle precedenti operazioni di insieme, quindi possiamo concludere che $\mathcal{B}(\mathbb{R})$ è equivalentemente generata da tutti gli intervalli del tipo $(-\infty, t]$.

Osserviamo ancora che anche gli insiemi del tipo $\{t_0\}$, costituiti da un singolo punto, appartengono a $\mathcal{B}(\mathbb{R})$, infatti

$$\{t_0\} = \bigcap_{n=1}^{+\infty} \left(t_0 - \frac{1}{n}, t_0 + \frac{1}{n} \right).$$

Diciamo v.a. **discreta**, una qualsiasi v.a. che assume con probabilità positiva al più una quantità numerabile di valori. In questo caso, se $\{x_1, x_2, \dots\}$ è l'insieme dei possibili valori che può assumere X , avremo necessariamente

$$\{X \leq t\} = \bigcup_{x_i \leq t} \{X = x_i\}, \text{ quindi } \Omega = \{X \in \mathbb{R}\} = \bigcup_{i=1}^{\infty} \{X = x_i\}$$

Definizione 4.1.2. Per una v.a. discreta X definiamo densità (discreta) di probabilità la funzione

$$\begin{aligned}p: \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longrightarrow P(X = x)\end{aligned}$$

tale che

a. $p(x) = 0$ tranne al più per una quantità numerabile di valori di $\{x_i\}_{i \in I} \subseteq \mathbb{R}$, dove $I \subseteq \mathbb{N}$.

b. $\sum_{x \in \mathbb{R}} p(x) = \sum_{i \in I} p(x_i) = 1$.

Quindi una densità di una v.a. discreta è una successione od un vettore di valori che verifica le due proprietà elencate nella definizione.

Esempio 4.1.1. Dire, giustificando la risposta se la seguente è una densità di probabilità di una v.a. che assume i valori 1, 2, 3, 4 oppure no.

$$P(X = 1) = \frac{1}{4}, P(X = 2) = \frac{2}{5}, P(X = 3) = \frac{1}{10}, P(X = 4) = \frac{3}{10}.$$

Soluzione: questa non è una densità di probabilità poiché

$$\frac{1}{4} + \frac{2}{5} + \frac{1}{10} + \frac{3}{10} = \frac{21}{20} > 1.$$

Analogamente se avessimo avuto

$$P(X=1) = \frac{1}{4}, P(X=2) = \frac{3}{4}, P(X=3) = \frac{1}{10}, P(X=4) = -\frac{1}{10}$$

anche questa funzione non sarebbe stata una densità, in quanto nessun valore può essere negativo.

Come conseguenza abbiamo che qualsiasi probabilità relativa ad X può essere calcolata mediante la densità, infatti per ogni sottoinsieme A di \mathbb{R} abbiamo

$$P(X \in A) = \sum_{x_i \in A} p(x_i).$$

Un primo esempio di v.a. discreta è dato dalla **funzione indicatrice di un insieme** $A \in \mathcal{F}$, definita come

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c, \end{cases}$$

la sua densità è quindi data da $P(\mathbf{1}_A = 1) = P(A)$ e $P(\mathbf{1}_A = 0) = P(A^c) = 1 - P(A)$.

Questo tipo di v.a. è detto di Bernoulli, cioè una v.a. X è detta di Bernoulli se

$$X = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } 1 - p \end{cases}$$

dove $0 < p < 1$. D'ora in poi scriveremo $X \sim \text{Bin}(1, p)$ dicendo che X “segue una distribuzione di Bernoulli”.

Quindi, quando si studia un fenomeno aleatorio che può essere rappresentato mediante una v.a., la prima cosa da capire è qual è l'insieme di valori che la v.a. può assumere.

Esempio 4.1.2. Se abbiamo 10 palline in un'urna di cui 4 bianche ed effettuiamo 5 estrazioni senza reinserimento e siamo interessati a capire le probabilità di estrarre le palline bianche, possiamo definire la v.a.

$$X = \text{numero di palline bianche estratte nelle 5 estrazioni}$$

ed è chiaro che i possibili valori che X assume con probabilità positiva sono 0,1,2,3,4. Se le palline bianche nell'urna fossero state 6 allora i possibili valori della v.a. sarebbero stati 0,1,2,3,4,5.

Ancora, se dalla stessa urna ripetiamo estrazioni con reinserimento e ci fermiamo la prima volta che esce una pallina bianca ed indichiamo con

$$T = \text{il numero di estrazioni necessarie per ottenere la prima pallina bianca,}$$

allora T è una v.a. che può prendere come valori tutti i numeri naturali.

Possiamo pensare di avere dei fenomeni, la cui rappresentazione richiede che siano usate più v.a., per esempio potremmo lanciare più monete o più dadi contemporaneamente e voler studiare la probabilità di ottenere i risultati congiunti. Per farlo abbiamo bisogno di estendere la nozione di densità. D'ora in poi le intersezioni di insiemi descritti attraverso v.a. verranno indicate mediante la virgola, ovvero, se X ed Y sono due v.a., scriviamo

$$P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A, Y \in B).$$

Definizione 4.1.3. Sia (X_1, \dots, X_n) un vettore di v.a. discrete, diciamo densità congiunta delle v.a. X_1, \dots, X_n , la famiglia di probabilità data da

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) := P(X_1 = x_1, \dots, X_n = x_n),$$

al variare di $x_1, \dots, x_n \in \mathbb{R}$.

Esempio 4.1.3. Si ha un'urna con 6 palline nere e 4 bianche, si effettua un'estrazione e si rimette la pallina estratta nell'urna insieme ad un'altra pallina dello stesso colore, quindi si effettua una seconda estrazione. Se

$$X_i = \begin{cases} 1 & \text{se esce bianco all'estrazione } i \\ 0 & \text{se esce nero all'estrazione } i, \end{cases} \quad i = 1, 2$$

calcolare la probabilità che sia uscita una pallina bianca in entrambe le estrazioni, $P(X_1 = 1, X_2 = 1)$.

Soluzione: la precedente domanda si deve risolvere per condizionamento.

$$P(X_1 = 1, X_2 = 1) = P(X_2 = 1 | X_1 = 1)P(X_1 = 1) = P(X_2 = 1 | X_1 = 1) \frac{4}{10} = \frac{5}{11} \frac{4}{10} = \frac{2}{11}$$

e analogamente potremmo calcolare

$$P(X_1 = 0, X_2 = 0), P(X_1 = 1, X_2 = 0), P(X_1 = 0, X_2 = 1)$$

ottenendo così la densità congiunta di X_1 ed X_2 .

Quando abbiamo un vettore aleatorio (X_1, \dots, X_n) , chiamiamo densità marginali le densità delle singole componenti

$$p_{X_1}(\cdot), \dots, p_{X_n}(\cdot)$$

Se abbiamo la densità congiunta delle n v.a. è sempre possibile calcolare tutte le densità marginali, semplicemente saturando tutte le altre componenti.

Prendendo $n = 2$, supponiamo che le v.a. X, Y assumano valori rispettivamente in $\{x_i\}_{i \in I}$ e $\{y_j\}_{j \in J}$, per qualche $I, J \subseteq \mathbb{N}$, allora per ogni $i \in I$ si ha

$$\begin{aligned} p_X(x_i) &= P(X = x_i) = P(X = x_i, \Omega) = P(X = x_i, Y \in \mathbb{R}) = P\left(X = x_i, \bigcup_{j \in J} \{Y = y_j\}\right) \\ &= \sum_{j \in J} P(X = x_i, Y = y_j) = \sum_{j \in J} p_{X,Y}(x_i, y_j) \\ p_Y(y_j) &= \sum_{i \in I} p_{X,Y}(x_i, y_j) \end{aligned}$$

e più in generale per qualsiasi n v.a. con valori in rispettivi $\{x_{i_1}\}_{\{i_1 \in I_1\}}, \dots, \{x_{i_n}\}_{\{i_n \in I_n\}}$, con $I_1, \dots, I_n \subseteq \mathbb{N}$, si ha

$$p_{X_k}(x_k) = \sum_{i_1 \in I_1} \cdots \sum_{k-1 \in I_{k-1}} \sum_{i+1 \in I_{i+1}} \cdots \sum_{i_n} p_{X_1, \dots, X_n}(x_1, \dots, x_n).$$

La precedente tecnica può essere generalizzata per ottenere la densità congiunta di una qualsiasi sottofamiglia delle v.a. X_1, \dots, X_n , saturando tutte le componenti della densità congiunta che non interessano.

Esempio 4.1.4. Siano X ed Y due v.a. discrete a valori rispettivamente in $\{-1, 0, 1, 2\}$ ed in $\{-2, -1, 1\}$. La loro densità congiunta $p_{X,Y}(x, y)$ è descritta dalla tabella

Y/X	$X = -1$	$X = 0$	$X = 1$	$X = 2$
$Y = -2$	$\frac{1}{8}$	$\frac{1}{12}$	α	0
$Y = -1$	0	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{8}$
$Y = 1$	$\frac{1}{6}$	0	0	$\frac{1}{12}$

con $\alpha \in \mathbb{R}$. Calcolare α e le densità marginali p_X e p_Y .

Soluzione: Per trovare α basta ricordare che una densità deve sommarsi ad 1 quando si considerano tutti i possibili valori. Quindi sommando i valori in tabella ed eguagliandoli ad 1 si deduce che $\alpha = \frac{1}{6}$.

Quando si rappresenta una congiunta per mezzo di una tabella (nel caso che vi siano soltanto due v.a.) basta sommare per colonna e per riga per ottenere le marginali. Si ha dunque

$$\begin{aligned} P(X = -1) &= \frac{1}{8} + \frac{1}{6} = \frac{7}{24}, \quad P(X = 0) = \frac{1}{12} + \frac{1}{6} = \frac{1}{4}, \quad P(X = 1) = \frac{1}{6} + \frac{1}{12} = \frac{1}{4}, \\ P(X = 2) &= \frac{1}{8} + \frac{1}{12} = \frac{5}{24}, \\ P(Y = -2) &= \frac{1}{8} + \frac{1}{12} + \frac{1}{6} = \frac{9}{24}, \quad P(Y = -1) = \frac{1}{6} + \frac{1}{12} + \frac{1}{8} = \frac{9}{24}, \quad P(Y = 1) = \frac{1}{6} + \frac{1}{12} = \frac{1}{4}. \end{aligned}$$

Quindi, se si ha la densità congiunta, si hanno automaticamente tutte le densità marginali. Il contrario non è altrettanto vero, nel senso che le stesse densità marginali possono provenire da differenti densità congiunte.

Esempio 4.1.5. Consideriamo le seguenti densità congiunte riassunte nelle seguenti due tabelle.

Y/X	$X = -1$	$X = 0$	$X = 2$
$Y = -2$	$\frac{1}{4}$	$\frac{1}{8}$	0
$Y = 1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$Y = 0$	$\frac{1}{8}$	0	$\frac{1}{8}$

Y/X	$X = -1$	$X = 0$	$X = 2$
$Y = -2$	$\frac{1}{4}$	$\frac{1}{8}$	0
$Y = 1$	$\frac{1}{4}$	0	$\frac{1}{8}$
$Y = 0$	0	$\frac{1}{8}$	$\frac{1}{8}$

In entrambi i casi otteniamo (sommando prima sulle colonne e quindi sulle righe)

$$\begin{aligned} P(X = -1) &= \frac{1}{2}, \quad P(X = 0) = \frac{1}{4}, \quad P(X = 2) = \frac{1}{4} \\ P(Y = -2) &= \frac{3}{8}, \quad P(Y = 1) = \frac{3}{8}, \quad P(Y = 0) = \frac{1}{4} \end{aligned}$$

e quindi, conoscendo soltanto le marginali, non potremmo decidere a quale densità congiunta corrispondono.

L'unico caso in cui le densità marginali identificano univocamente la densità congiunta è quando ci troviamo in condizioni di indipendenza, come nel caso dell'intersezione di eventi.

Definizione 4.1.4. Siano X_1, \dots, X_n v.a. discrete, diciamo che sono **indipendenti** se

$$(4.1) \quad P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n)$$

per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$.

Una famiglia numerabile di v.a. $\{X_k\}_{k \in \mathbb{N}}$ si dice indipendente se (4.1) è verificata per ogni $n \in \mathbb{N}$.

Esempi di v.a. indipendenti sono: i risultati dei lanci ripetuti di un dado, i risultati dei lanci di più dadi contemporaneamente, le v.a. di Bernoulli associate al lancio ripetuto di una moneta.

Siamo ora in condizioni di introdurre dei modelli generali che si potranno adattare a molte situazioni.

Osservazione 4.1.2. (M) Come abbiamo visto precedentemente, la nozione di indipendenza è in realtà una nozione che riguarda le σ -algebre. Questo è vero anche nel caso delle v.a. Definiamo σ -algebra generata da una v.a. X :

$$\sigma(X) = \sigma(\{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\})$$

ovvero la σ -algebra generata da tutte le contro immagini dei Boreliani attraverso la v.a. X . Se la v.a. X è discreta allora basteranno le contro immagini dei valori assunti da X con probabilità positiva per generare questa σ -algebra. Allora l'indipendenza di due o più v.a. X ed Y si può riscrivere dicendo che X ed Y sono indipendenti se le σ -algebre da esse generate sono indipendenti. In effetti, la (4.1) verifica esattamente questo per le controimmagini dei valori assunti dalle v.a.

4.2 Modelli estrattivi

Supponiamo di avere un'urna con m palline totali di cui r bianche e $m-r$ nere ed effettuiamo n estrazioni da quest'urna, con $r \leq m, n$ numeri interi.

Indichiamo con

$$X_i = \begin{cases} 1 & \text{se esce bianco all'estrazione } i \\ 0 & \text{se esce nero all'estrazione } i, \end{cases} \quad i = 1, \dots, n$$

le v.a. che rappresentano gli esiti ad ogni singola estrazione. Queste v.a. si dicono anche **schema di successo-insuccesso**.

Per valutare la probabilità di qualsiasi evento che riguarda una serie di estrazioni abbiamo bisogno di capire la densità delle X_i che può essere determinata soltanto se si specifica se le estrazioni sono:

- a) con reinserimento
- b) senza reinserimento.

4.2.1 Estrazioni con reinserimento

Si effettuano le estrazioni e ogni volta che si estrae una pallina, ne si osserva il colore e quindi la si reinserisce nell'urna, che viene opportunamente mescolata.

Questo vuol dire che tutte le estrazioni si svolgono sempre nelle medesime condizioni. Inoltre non vi è ragione per cui ciò che succede in un'estrazione debba influenzare quel che succede in un'altra.

Quel che stiamo dicendo è che tutte le v.a. hanno la stessa densità ovvero sono **identicamente distribuite** e inoltre sono indipendenti. Questa situazione si abbrevia scrivendo che le v.a. sono **i.i.d.**

In questo caso se indichiamo con $p = \frac{r}{m}$ e quindi $1 - p = \frac{m - r}{m}$, avremo

$$X_i = \begin{cases} 1 & \text{con probabilità } p \\ 0 & \text{con probabilità } 1 - p \end{cases} \quad i = 1, \dots, n.$$

Se vogliamo calcolare delle probabilità di insiemi del tipo “escono 4 palline bianche in n estrazioni”, possiamo definire l'ulteriore v.a.

$$(4.2) \quad S_n = \sum_{i=1}^n X_i = \# \text{ di palline bianche uscite nelle } n \text{ estrazioni,}$$

allora questa v.a. può prendere tutti i valori $k = 0, 1, \dots, n$ e la domanda precedente può essere riassunta come

$$P(S_n = 4),$$

quindi è chiaro che il nostro fine è di calcolare la densità della v.a. S_n , cioè

$$P(S_n = k), \quad k = 0, 1, \dots, n.$$

Procediamo nella seguente maniera. L'evento $\{S_n = k\}$ vuol dire che sono uscite **esattamente** k palline bianche nelle n estrazioni, ma non stiamo considerando le posizioni in cui sono uscite. D'altra parte non vi è ragione per cui la successione in cui tutte le k palline bianche siano uscite nelle prime k estrazioni sia più o meno probabile della successione in cui le k palline bianche sono uscite consecutivamente nelle ultime k estrazioni o di qualsiasi altra successione di estrazioni che vede uscire esattamente k palline bianche, cioè stiamo

dicendo che questi eventi hanno tutti la stessa probabilità. Quindi

$$\begin{aligned}
 & P(S_n = k) \\
 &= (\# \text{ modi di estrarre } k \text{ bianche su } n) P(k \text{ palline bianche nelle prime } k \text{ estrazioni}) \\
 &= \binom{n}{k} P(X_1 = 1, X_2 = 1, \dots, X_k = 1, X_{k+1} = 0, \dots, X_{n-1} = 0, X_n = 0) \\
 &= \binom{n}{k} P(X_1 = 1) P(X_2 = 1) \cdots P(X_k = 1) P(X_{k+1} = 0) \cdots P(X_{n-1} = 0) P(X_n = 0) \\
 &= \binom{n}{k} p^k (1-p)^{n-k}
 \end{aligned}$$

dove abbiamo utilizzato l'indipendenza e l'identica distribuzione delle v.a. X_i .

Definizione 4.2.1. Una v.a. X che prende valori $k = 0, 1, \dots, n$ tale che

$$(4.3) \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

è detta v.a. **binomiale di parametri n e p** e si scrive $X \sim \text{Bin}(n, p)$.

La costruzione della v.a. S_n , data da (4.2), inoltre dimostra che una v.a. $\text{Bin}(n, p)$ può essere vista come somma di n v.a. di Bernoulli di parametro p indipendenti. Una Bernoulliana di parametro p è indicata con $\text{Bin}(1, p)$.

Esempio 4.2.1. Un libro è costituito da 400 pagine ed in ogni pagina, indipendentemente da tutte le altre, la probabilità di riscontrarvi almeno un errore è del 2%. Indichiamo con S_{400} il numero di pagine errate del libro (cioè che contengono almeno un errore). Qual è la probabilità che vi siano almeno 4 pagine errate nel libro?

Soluzione: questa situazione si può ricondurre ad un modello estrattivo con reinserimento. Le 400 pagine sono le estrazioni che si ripetono in maniera identica ed indipendente, mentre il successo è rappresentato dal fatto che ci sia almeno un errore nella pagina (pallina bianca), mentre l'insuccesso è quando non ve ne è nessuno (pallina nera). Stiamo quindi chiedendo la probabilità che il numero dei successi sia almeno 4, ovvero $P(S_{400} \geq 4)$. Poiché stiamo supponendo indipendenza ed identica distribuzione, sappiamo che $S_{400} \sim \text{Bin}(400, 0.02)$ poiché viene contato 1 ogni volta che si incontra una pagina che contiene almeno un errore. Allora si ha

$$P(S_{400} \geq 4) = \sum_{k=4}^{400} P(S_{400} = k) = \sum_{k=4}^{400} \binom{400}{k} 0.02^k 0.98^{400-k}$$

che è molto lunga da calcolare. In questo caso conviene calcolare la probabilità del complementare

$$\begin{aligned}
 P(S_{400} \geq 4) &= 1 - P(S_{400} \leq 3) = 1 - \sum_{k=0}^3 \binom{400}{k} 0.02^k 0.98^{400-k} \\
 &= 1 - (0.98^{400} + 400 \cdot 0.02 \cdot 0.98^{399} + 200 \cdot 399 \cdot 0.02^2 \cdot 0.98^{398} + 200 \cdot 133 \cdot 398 \cdot 0.02^3 \cdot 0.98^{397})
 \end{aligned}$$

Generalizzazione. (M)

Nel modello estrattivo sopraesposto abbiamo considerato un'urna composta da soltanto due tipi di palline. Si può generalizzare il modello a più tipi.

Se abbiamo una popolazione di m individui ripartita in l sottopopolazioni di

m_1 individui di tipo 1,

m_2 individui di tipo 2,

\vdots

m_l individui di tipo l

con $m_1 + \dots + m_l = m$ e si effettuano n estrazioni con reinserimento. Indicando i risultati dell' i -esima estrazione con

$$X_i^j = \begin{cases} 1 & \text{se viene estratto un individuo di tipo } j \text{ con probabilità } p_j = \frac{m_j}{m} \\ 0 & \text{altrimenti con probabilità } 1 - p_j \end{cases}$$

per $j = 1, \dots, l$ e con

$$S_n^j = \sum_{i=1}^n X_i^j$$

il numero di individui di tipo j estratti nelle n estrazioni, adesso le v.a. X_i^j sono tutte indipendenti. Ogni successione di estrazioni che vede esattamente n_1 elementi estratti di tipo 1, n_2 elementi estratti di tipo 2, ..., n_l elementi estratti di tipo l , con $n_1 + \dots + n_l = n$, ha uguale probabilità di realizzarsi poiché l'ordine con cui sono stati estratti non conta. Quindi possiamo considerare la successione che vede le prime n_1 estrazioni di tipo 1, le seconde n_2 estrazioni di tipo 2 fino alle ultime l di tipo l , valutare la sua probabilità e quindi moltiplicare per il numero di possibili estrazioni che vedono n_1 posti occupati dal tipo 1, n_2 posti occupati dal tipo 2, ..., n_l posti occupati dal tipo l , cioè

$$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-n_2-\dots-n_{l-1}}{n_l} = \frac{n!}{n_1!n_2!\dots n_l!},$$

mentre per indipendenza abbiamo

$$\begin{aligned} & P(X_1^1 = 1, \dots, X_{n_1}^1 = 1, X_{n_1+1}^2 = 1, \dots, X_{n_1+\dots+n_{l-1}+1}^l = 1, \dots, X_n^l = 1) \\ &= P(X_1^1 = 1) \dots P(X_{n_1}^1 = 1) P(X_{n_1+1}^2 = 1) \dots P(X_{n_1+\dots+n_{l-1}+1}^l = 1) \dots P(X_n^l = 1) \\ &= p_1^{n_1} \dots p_l^{n_l}. \end{aligned}$$

In conclusione possiamo scrivere

$$P(S_n^1 = n_1, \dots, S_n^l = n_l) = \frac{n!}{n_1!n_2!\dots n_l!} p_1^{n_1} \dots p_l^{n_l}$$

Esempio 4.2.2. Si ha un'urna con 5 palline bianche, 3 verdi e 6 rosse e si effettuano 5 estrazioni con reinserimento. Vogliamo sapere la probabilità che nelle 5 estrazioni compaiano esattamente due palline rosse e due bianche.

Ad ogni estrazione si estrae una pallina bianca con probabilità $\frac{5}{14}$, una verde con probabilità $\frac{3}{14}$ ed una rossa probabilità $\frac{6}{14}$. Applicando la formula prima scritta si ha che

$$P(S_5^1 = 2, S_5^2 = 1, S_5^3 = 2) = \frac{5!}{2!1!2!} \left(\frac{5}{14}\right)^2 \left(\frac{3}{14}\right) \left(\frac{6}{14}\right)^2 = 30 \frac{2700}{14^5}.$$

4.2.2 Estrazioni senza reinserimento

Siamo esattamente nelle stesse condizioni di prima (stessa composizione iniziale dell'urna), ma le palline non vengono reinserite e vogliamo calcolare la densità della v.a.

$$S_n = \sum_{i=1}^n X_i = \# \text{ di palline bianche uscite nelle } n \text{ estrazioni}.$$

Prima di tutto occorre capire quali valori può assumere questa v.a. Chiaramente se vi sono più palline bianche che estrazioni, noi avremo la possibilità di estrarre anche tutte palline bianche, ma se è vero il contrario su n estrazioni vi sarà sempre almeno una pallina nera, viceversa se $m - r \geq n$ vi sarà la possibilità di estrarre tutte palline nere e nessuna bianca, in caso contrario nelle estrazioni si presenterà sempre almeno una pallina bianca, quindi S_n prende i valori $k = \max(0, n - (m - r)), \dots, \min(n, r)$. Analizziamo il caso in cui $r, (m - r) \geq n$.

Poichè le estrazioni sono senza reinserimento, ciò che accade in una estrazione influenza l'esito delle altre, questo vuol dire che le v.a. X_i non sono indipendenti e quindi l'unica maniera in cui possiamo attribuire la probabilità è per condizionamento rispetto a ciò che è successo in precedenza.

Di nuovo non vi è motivo per cui una successione che vede k palline bianche estratte sia più o meno probabile di un'altra con lo stesso numero di palline bianche estratte. Da cui $P(S_n = k) = (\# \text{ modi di estrarre } k \text{ bianche su } n) \cdot P(k \text{ bianche nelle prime } k)$

estrazioni) che è uguale a

$$\begin{aligned}
& \binom{n}{k} P(X_1 = 1, X_2 = 1, \dots, X_k = 1, X_{k+1} = 0, \dots, X_{n-1} = 0, X_n = 0) \\
&= \binom{n}{k} P(X_n = 0, X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_2 = 1, X_1 = 1) \\
&= \binom{n}{k} P(X_n = 0 | X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&= \binom{n}{k} P(X_n = 0 | X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_{n-1} = 0 | X_{n-2} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_{n-2} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&= \dots \\
&\quad \vdots \\
&= \binom{n}{k} P(X_n = 0 | X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_{n-1} = 0 | X_{n-2} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdots P(X_{k+1} = 0 | X_k = 1, \dots, X_2 = 1, X_1 = 1) P(X_k = 1 | X_{k-1} = 1, \dots, X_1 = 1) \\
&\quad \cdots P(X_2 = 1 | X_1 = 1) P(X_1 = 1).
\end{aligned}$$

Dunque, riscrivendo la formula precedente partendo dalla prima estrazione, si ottiene

$$\begin{aligned}
P(S_n = k) &= \binom{n}{k} P(X_1 = 1) P(X_2 = 1 | X_1 = 1) \cdots P(X_k = 1 | X_{k-1} = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_{k+1} = 0 | X_k = 1, \dots, X_2 = 1, X_1 = 1) \\
&\quad \cdots P(X_{n-1} = 0 | X_{n-2} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&\quad \cdot P(X_n = 0 | X_{n-1} = 0, \dots, X_{k+1} = 0, X_k = 1, \dots, X_1 = 1) \\
&= \binom{n}{k} \frac{r}{m} \frac{r-1}{m-1} \cdots \frac{r-(k-1)}{m-(k-1)} \frac{m-r}{m-k} \frac{m-r-1}{m-(k+1)} \cdots \frac{m-r-(n-k-1)}{m-(n-1)}
\end{aligned}$$

che, sfruttando la definizione dei coefficienti binomiali, diventa

$$\begin{aligned}
P(S_n = k) &= \binom{n}{k} \frac{\frac{r!}{(r-k)!} \frac{(m-r)!}{[(m-r)-(n-k)]!}}{\frac{m!}{(m-n)!}} = \frac{n!}{k!(n-k)!} \frac{\frac{r!}{(r-k)!} \frac{(m-r)!}{[(m-r)-(n-k)]!}}{\frac{m!}{(m-n)!}} \\
&= \frac{\frac{r!}{(r-k)!k!} \frac{(m-r)!}{[(m-r)-(n-k)]!(n-k)!}}{\frac{m!}{(m-n)!n!}} = \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}}.
\end{aligned}$$

La scrittura finale è suggestiva in quanto ci dice che la probabilità di estrarre k palline bianche da un campione di n in una popolazione ripartita in due tipi (bianco e nero) rispettivamente di dimensione r e $m-r$ è data dal numero di modi di estrarre (senza ordinamento)

esattamente k individui di tipo 1 dalla popolazione di dimensione r , moltiplicato per il numero di modi di estrarre (senza ordinamento) un campione di $n - k$ individui di tipo 2 da una popolazione di dimensione $m - r$, diviso per il numero di modi di estrarre un campione di dimensione n da una popolazione totale di m individui.

Definizione 4.2.2. diciamo che una v.a. X segue una distribuzione *ipergeometrica di parametri* (r, m, n) , se essa prende valori $k = 0, \dots, \min(r, n)$ e

$$(4.4) \quad P(X = k) = \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}}.$$

La precedente formula (4.4) si può generalizzare a più tipi. Se abbiamo una popolazione di m individui ripartita in l sottopopolazioni di

m_1 individui di tipo 1,

m_2 individui di tipo 2,

\vdots

m_l individui di tipo l

con $m_1 + \dots + m_l = m$ e si effettuano n estrazioni, indicando i risultati dell' i -esima estrazione con

$$X_i^j = \begin{cases} 1 & \text{se viene estratto un individuo di tipo } j \\ 0 & \text{altrimenti} \end{cases}$$

per $j = 1, \dots, l$ e con

$$S_n^j = \sum_{i=1}^n X_i^j,$$

il numero di individui di tipo j estratti nelle n estrazioni, si ha per indici $n_1 + \dots + n_l = n$

$$P(S_n^1 = n_1, \dots, S_n^l = n_l) = \frac{\binom{m_1}{n_1} \dots \binom{m_l}{n_l}}{\binom{m}{n}}$$

ovvero abbiamo espresso la densità congiunta delle v.a. S_n^1, \dots, S_n^l .

Esempio 4.2.3. Nel gioco del poker con 52 carte francesi, 4 semi, 13 valori, calcolare la probabilità di avere in prima mano servita:

1. esattamente una coppia;
2. esattamente una doppia coppia;
3. esattamente un tris;
4. un full;
5. colore;
6. una scala;

7. una scala colore.

Soluzione: Per tutte le combinazioni considerate, il numero di casi totali è sempre $\binom{52}{5}$.

1. Poiché vogliamo ottenere esattamente una coppia, una volta scelto un valore per la coppia le altre tre carte devono risultare di valori tutti differenti e differenti dal valore della coppia. Quindi, per contare i casi favorevoli teniamo conto che la nostra popolazione è suddivisa in 13 sottopopolazioni di quattro carte e per una singola specifica coppia abbiamo $\binom{4}{2}\binom{4}{1}\binom{4}{1}\binom{4}{1} = 6 \cdot 4^3$ possibilità e quindi per una coppia qualsiasi abbiamo

$$P(\text{esattamente una coppia}) = \binom{13}{4} \frac{\binom{4}{2}\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}},$$

dove il coefficiente $\binom{13}{4}$ viene dalla scelta dei 4 valori distinti qualsiasi.

2. In questo caso dobbiamo scegliere un valore per la prima coppia, un secondo e diverso valore per la seconda coppia ed infine un qualsiasi altro valore per la carta singola. Per il primo valore dobbiamo estrarre due carte, lo stesso per il secondo valore, per il terzo una sola. Tutte le doppie coppie sono equiprobabili. Abbiamo quindi

$$P(\text{esattamente una doppia coppia}) = \binom{13}{3} \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{\binom{52}{5}}.$$

3. Per un tris abbiamo bisogno di tre carte con lo stesso valore e quindi altre due con valori diversi dal tris e tra loro, dunque

$$P(\text{esattamente un tris}) = \binom{13}{3} \cdot \frac{\binom{4}{3}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}.$$

4. Un full presenta soltanto due valori, uno per il tris ed uno per la coppia, quindi

$$P(\text{esattamente un full}) = \binom{13}{2} \cdot \frac{\binom{4}{3}\binom{4}{2}}{\binom{52}{5}}.$$

5. Il colore presenta 5 carte dello stesso seme, quindi abbiamo 4 semi, cioè 4 sottopopolazioni di 13 carte ciascuna

$$P(\text{esattamente colore}) = \binom{4}{1} \cdot \frac{\binom{13}{5}}{\binom{52}{5}}.$$

6. Per la scala dobbiamo avere 5 valori consecutivi senza riguardo del seme, calcolando che una scala può anche finire con l'asso, abbiamo 10 punti di possibile inizio della scala e quindi 4 possibilità per ogni valore, da cui

$$P(\text{scala}) = 10 \cdot \frac{\binom{4}{1}\binom{4}{1}\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}}$$

questa probabilità comprende anche la scala colore (ovvero una scala con carte dello stesso seme).

7. La probabilità della scala colore si calcola come prima con l'unica differenza che, una volta stabilito da dove parte, la scala deve avere tutte carte dello stesso seme. Quindi abbiamo 4 possibilità di scelta solo per la prima carta, poi il seme delle rimanenti deve essere lo stesso, perciò

$$P(\text{scala reale}) = 10 \cdot \frac{\binom{4}{1} \binom{1}{1} \binom{1}{1} \binom{1}{1} \binom{1}{1}}{\binom{52}{5}}.$$

4.3 Altre distribuzioni

1. Una v.a. X che prende valori $\{x_i\}_{i=1}^n$ in un insieme con $n \in \mathbb{N}$ elementi si dice **uniforme** se

$$P(X = x_i) = \frac{1}{n}, \quad i = 1, \dots, n$$

2. Un'altra distribuzione discreta si ottiene dal seguente fenomeno estrattivo. Supponiamo come prima di avere un'urna con m palline di cui r bianche e le altre nere e che si effettuino estrazioni con reinserimento finché non si estrae la prima pallina bianca. Come abbiamo visto prima, ad ogni estrazione, la probabilità di avere una pallina bianca è $p := \frac{r}{m}$. Se indichiamo con $\{X_i\}$ la successione delle v.a. indipendenti ed identicamente distribuite (**i.i.d.**) che esprimono lo schema di successo-insuccesso per l'estrazione del bianco, la v.a.

$$\tau = \inf\{n : X_n = 1\}$$

rappresenterà il tempo di primo successo. Chiaramente questa v.a. può assumere tutti i valori naturali a partire da 1 e vogliamo capire la sua densità. Si ha per $k = 1, 2, \dots$

$$\begin{aligned} (4.5) \quad P(\tau = k) &= P(X_1 = 0, X_2 = 0, \dots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0)P(X_2 = 0) \cdots P(X_{n-1} = 0)P(X_n = 1) = (1-p)^{k-1}p. \end{aligned}$$

Una v.a. che ha questo tipo di densità si dice che segue una distribuzione **geometrica** (modificata) di parametro $0 < p < 1$.

Notiamo che il fatto di sapere che la (4.5) è in effetti una densità di probabilità ha come conseguenza che

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

risultato confermato dalla convergenza della serie geometrica di ragione minore di 1.

Esempio 4.3.1. In un'urna con 6 palline bianche si introducono N palline nere, dove N è una v.a. uniforme su $\{4, 5, 6\}$. Una volta inserite le palline, si effettuano estrazioni con reinserimento fermandosi la prima volta, T , che si estrae una pallina bianca. Calcolare la distribuzione di T

Anche in questo caso i possibili valori di T sono tutti i naturali \mathbb{N} , ma non possiamo calcolare direttamente la densità di T in quanto non sappiamo quante palline nere siano state introdotte, dobbiamo dunque condizionare rispetto agli esiti precedenti. Se sono state introdotte 4 palline nere, la probabilità di estrarre una bianca ad ogni estrazione è $\frac{3}{5}$, se 5 è $\frac{6}{11}$, se 6 è $\frac{1}{2}$. Quindi a seconda della composizione dell'urna si ottiene una v.a. geometrica con parametri differenti. Applicando la formula delle probabilità totali, abbiamo dunque

$$\begin{aligned} P(T = k) &= \\ &= P(T = k|N = 4)P(N = 4) + P(T = k|N = 5)P(N = 5) + P(T = k|N = 6)P(N = 6) \\ &= \frac{1}{3}[P(T = k|N = 4) + P(T = k|N = 5) + P(T = k|N = 6)] \\ &= \frac{1}{3}\left[\frac{3}{5}\left(\frac{2}{5}\right)^{k-1} + \frac{6}{11}\left(\frac{5}{11}\right)^{k-1} + \frac{1}{2}\left(\frac{1}{2}\right)^{k-1}\right] \end{aligned}$$

È possibile calcolare abbastanza facilmente la coda della densità geometrica

$$\begin{aligned} P(T \geq k) &= \sum_{h=k}^{\infty} p(1-p)^{h-1} = \sum_{h=k}^{\infty} p(1-p)^{h-k+k-1} = (1-p)^{k-1} \sum_{h=k}^{\infty} p(1-p)^{h-k} \\ &= (1-p)^{k-1} \sum_{j=0}^{\infty} p(1-p)^j = (1-p)^{k-1} \sum_{i=1}^{\infty} p(1-p)^{i-1} = (1-p)^{k-1}, \end{aligned}$$

dove nell'ultimo passaggio, abbiamo sfruttato il fatto che la densità si somma ad 1.

La densità geometrica ha anche l'importante proprietà di **perdita della memoria**: sapendo che non vi è stato ancora nessun successo ad un tempo prefissato n , calcolare la probabilità che avvenga il successo in un tempo successivo $n + k$ è equivalente alla probabilità di aver un successo al tempo k , come se si fosse fatto ripartire da zero l'orologio, infatti per $h \geq 1$ si ha

$$\begin{aligned} P(T = n + k|T > n) &= \frac{P(T = n + k, T > n)}{P(T > n)} = \frac{P(T = n + k)}{P(T \geq n + 1)} = \frac{p(1-p)^{n+k-1}}{(1-p)^n} \\ &= p(1-p)^{k-1} = P(T = k). \end{aligned}$$

Esempio 4.3.2. Abbiamo due sacchetti della tombola, uno completo ed uno a cui mancano i numeri dall'82 all'86 compresi. Vogliamo effettuare estrazioni con reinserimento finché non esce un numero maggiore od uguale a 85. I due sacchetti sono indistinguibili e per minimizzare l'errore decidiamo di seguire la seguente strategia. Ad ogni estrazione si sceglie un sacchetto a caso, si effettua l'estrazione ed eventualmente ci si ferma se si è avuto successo. Se, invece, si deve continuare si reinserisce il

numero nel sacchetto usato e si sceglie nuovamente un sacchetto a caso per effettuare la successiva estrazione.

Se indichiamo con τ il tempo di primo successo, che distribuzione segue?

Soluzione:

Notiamo che ogni volta l'estrazione si ripete esattamente nelle stesse condizioni ed indipendentemente da quanto avvenuto in precedenza, questo vuol dire che si ha una successione di v.a. di Bernoulli i.i.d.

$$X_i = \begin{cases} 1 & p \\ 0 & (1-p) \end{cases}$$

e non ci resta che determinare p , ovvero la probabilità di estrarre un numero maggiore od uguale di 85 in una singola estrazione. Per calcolare questa probabilità dobbiamo usare il condizionamento. Se indichiamo con A il sacchetto completo e con B quello incompleto, avremo

$$p = P(X_i = 1) = P(X_i = 1|A)P(A) + P(X_i = 1|B)P(B) = \frac{1}{2} \left[\frac{6}{90} + \frac{4}{85} \right]$$

e quindi $\tau \sim \text{geom}(p)$.

3. Ultima distribuzione discreta che andiamo a considerare è la v.a. di Poisson. Diciamo che X segue una distribuzione di **Poisson di parametro** $\lambda > 0$ e scriviamo $X \sim \text{Poisson}(\lambda)$, se X prende valori $k = 0, 1, 2, \dots$ e

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Come prima, il fatto che questa è una densità di probabilità implica che

$$1 = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \Rightarrow e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!},$$

confermata dalla convergenza della serie esponenziale.

La distribuzione di Poisson è indicata per modellizzare il numero di arrivi in un determinato periodo. Infatti questa distribuzione attribuisce un'alta probabilità all'evento che vi sia un numero piccolo di arrivi, mentre diventa esponenzialmente piccola per numeri grandi.

La distribuzione di Poisson può essere preziosa anche per approssimare la distribuzione binomiale che può coinvolgere dei calcoli piuttosto lunghi e complessi.

Proposizione 4.3.1. Se abbiamo una v.a. $X \sim \text{Bin}(n, p_n)$ tale che $np_n \rightarrow \lambda > 0$ per $n \rightarrow +\infty$, allora per $k = 0, \dots, n$ si ha

$$P(X = k) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

Dimostrazione: (M). Dimostriamo il risultato nel caso più semplice quando $p_n = \frac{\lambda}{n}$.

$$\begin{aligned} P(X = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

poiché valgono i due limiti notevoli

$$\begin{aligned} \left(1 - \frac{\lambda}{n}\right)^n &\longrightarrow e^{-\lambda} \\ \frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} &\longrightarrow 1 \end{aligned}$$

□

Esempio 4.3.3. Dieci minuti prima dell'orario di apertura di una banca si forma una fila all'entrata seguendo una distribuzione di Poisson(6). Per decidere quanti sportelli aprire inizialmente il direttore osserva la fila che si \tilde{A} formata. Se ci sono almeno 12 persone, i soliti due cassieri non basteranno.

Calcolare la probabilità di dover aggiungere un cassiere.

Soluzione:

Se N indica la v.a. che esprime la formazione della fila, il problema chiede semplicemente di valutare $P(N \geq 12)$ e dalla definizione di densità di Poisson si ha

$$P(N \geq 12) = 1 - P(\leq 11) = 1 - \sum_{k=0}^{11} e^{-6} \frac{6^k}{k!}.$$

4.4 Densità condizionate e indipendenza

Come abbiamo visto, nel caso delle estrazioni senza reinserimento si generano delle v.a. tra loro dipendenti, le cui densità congiunte possono essere dedotte soltanto mediante condizionamento. In realtà questo è un metodo generale che coinvolge la nozione di densità condizionale.

In questa sezione ci riduciamo al caso di due v.a. per semplicità, ma quanto presentato si estende direttamente a più v.a.

Come abbiamo visto nelle sezioni precedenti, se X ed Y hanno densità congiunta

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

si possono sempre ricavare le densità marginali mediante saturazione, ovvero sommando su tutti i possibili valori della v.a. che non interessa

Abbiamo anche visto che dalle marginali non è sempre possibile risalire alla congiunta. L'unico caso in cui siamo in grado di ricostruire univocamente la densità congiunta a partire dalle marginali è quando le v.a. sono indipendenti, poiché in questo caso vale l'identità

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \quad \forall x, y \in \mathbb{R}.$$

Osserviamo infine che l'indipendenza è una proprietà che si trasferisce alle funzioni di v.a.

Se abbiamo due v.a. X ed Y indipendenti lo è anche qualsiasi coppia di funzioni di queste v.a. cioè se $\phi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ sono due funzioni allora anche $\phi(X)$ e $\psi(Y)$ sono v.a. indipendenti, infatti (M)

$$\begin{aligned} P(\phi(X) = z, \psi(Y) = v) &= P(X \in \phi^{-1}(z), Y \in \psi^{-1}(v)) \\ &= P\left(\bigcup_{x \in \phi^{-1}(z)} \{X = x\}, \bigcup_{y \in \psi^{-1}(v)} \{Y = y\}\right) \\ &= \sum_{x \in \phi^{-1}(z)} \sum_{y \in \psi^{-1}(v)} P(X = x, Y = y) = \sum_{x \in \phi^{-1}(z)} \sum_{y \in \psi^{-1}(v)} P(X = x)P(Y = y) \\ &= \sum_{x \in \phi^{-1}(z)} P(X = x) \sum_{y \in \psi^{-1}(v)} P(Y = y) = P\left(\bigcup_{x \in \phi^{-1}(z)} \{X = x\}\right)P\left(\bigcup_{y \in \psi^{-1}(v)} \{Y = y\}\right) \\ &= P(\phi(X) = z)P(\psi(Y) = v) \end{aligned}$$

Esempio 4.4.1. Consideriamo le v.a. indipendenti X ed Y con rispettive densità marginali

$$\begin{aligned} P(X = -2) &= \frac{1}{4}, & P(X = -1) &= \frac{1}{4}, & P(X = 1) &= \frac{1}{4}, & P(X = 2) &= \frac{1}{4} \\ P(Y = -4) &= \frac{1}{2}, & P(Y = -1) &= \frac{1}{8}, & P(Y = 1) &= \frac{1}{8}, & P(Y = 3) &= \frac{1}{4}, \end{aligned}$$

allora anche X^2 ed Y^2 sono indipendenti, infatti

$$P(X^2 = 1) = P(X^2 = 4) = \frac{1}{2}, \quad P(Y^2 = 1) = \frac{1}{4}, \quad P(Y^2 = 9) = \frac{1}{4}, \quad P(Y^2 = 16) = \frac{1}{2}$$

e abbiamo

$$\begin{aligned} &P(X^2 = 1, Y^2 = 1) \\ &= P(X = 1, Y = 1) + P(X = -1, Y = 1) + P(X = 1, Y = -1) + P(X = -1, Y = -1) \\ &= P(X = 1)P(Y = 1) + P(X = -1)P(Y = 1) + P(X = 1)P(Y = -1) + P(X = -1)P(Y = -1) \\ &= \frac{1}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{1}{8} = \frac{1}{8} = \frac{1}{4} \cdot \frac{1}{2} = P(X^2 = 1)P(Y^2 = 1). \end{aligned}$$

Verificando tutti gli altri possibili valori si giunge alla conclusione.

Lo stesso risultato può essere esteso a più dimensioni

Proposizione 4.4.1. Siano $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_m)$ due vettori aleatori indipendenti e $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ e $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ due funzioni deterministiche, allora anche le v.a. definite da $Z = \phi(\mathbf{X})$ e $W = \psi(\mathbf{Y})$ sono v.a. indipendenti.

Nel caso in cui non siamo in condizioni di indipendenza, l'unica strada che ci rimane disponibile è quella del condizionamento

$$p_{X,Y}(x, y) = P(X = x, Y = y) = P(Y = y|X = x)P(X = x),$$

se è possibile conoscere $P(Y = y|X = x)$. Abbiamo quindi la seguente definizione

Definizione 4.4.1. *Siano X ed Y due v.a. con densità congiunta $p_{X,Y}(x, y)$ e rispettive densità marginali $p_X(x)$ e $p_Y(y)$, definiamo densità di Y condizionata ad X*

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x, y)}{p_X(x)}, \quad \text{dove } p_X(x) \neq 0$$

La densità condizionale si comporta esattamente come una densità di probabilità per ogni valore dato della v.a. rispetto alla quale si sta condizionando. Per ogni x fissato tale che $p_X(x) \neq 0$

a) $p_{Y|X}(\cdot|x) \geq 0$;

b) $\sum_y p_{Y|X}(y|x) = 1$.

Un esempio di tale situazione è dato dalle estrazioni senza reinserimento quando vogliamo guardare la distribuzione congiunta degli esiti delle estrazioni. Situazioni di dipendenza possono accadere in numerosi contesti.

Esempio 4.4.2. *Supponiamo che il numero di jobs in arrivo ad un server dipenda dal numero di terminali che si collegano. Indichiamo il numero di terminali che si collegano con la v.a. X $\sim \text{Bin}(6, \frac{1}{2})$. Se si collegano k terminali allora il numero dei jobs in arrivo sarà una v.a. che segue una distribuzione di Poisson di parametro $100k$. Indichiamo con N la v.a. che esprime il numero di jobs in fila.*

Allora la densità congiunta, per $k = 0, \dots, 6$, $h \in \mathbb{N}$ sarà data da

$$P(X = k, N = h) = P(N = h|X = k)P(X = k) = \binom{6}{k} \left(\frac{1}{2}\right)^6 \frac{(100k)^h}{h!} e^{-100k}$$

4.5 Calcoli con densità congiunte

In molte situazioni dobbiamo descrivere fenomeni aleatori mediante funzioni di più v.a.

Quando le dimensioni dei problemi aumentano, i calcoli, che coinvolgono le densità congiunte, diventano molto più complessi ed a volte ingestibili. Se si è in condizioni di indipendenza, almeno per alcuni tipi di trasformazioni, si possono fare delle considerazioni generali e riuscire a dare delle risposte più esplicite.

In quel che segue, trattiamo specificamente i fenomeni che riguardano il massimo, il minimo e la somma di v.a., che non sono in generale facili da caratterizzare, ma che si semplificano in alcuni casi in condizioni di indipendenza.

Date X ed Y v.a. possiamo esplicitare le densità per le v.a.

$$Z = \min(X, Y), \quad \max(X, Y), \quad X + Y$$

1. Se $Z = \min(X, Y)$, Se X ed Y sono v.a. discrete anche Z lo è e si ha

$$P(Z = z) = P(\min(X, Y) = z) = P(X = z, Y \geq z) + P(X > z, Y = z),$$

che nel caso dell'indipendenza diventa

$$P(Z = z) = P(X = z)P(Y \geq z) + P(X > z)P(Y = z).$$

Esempio 4.5.1. Come nell'esempio precedente supponiamo di voler effettuare estrazioni finchè non esce il primo 4 su uno qualsiasi tra il dado ed il tetraedro. Quindi se T_1 indica il tempo di primo successo del dado e T_2 indica il tempo di primo successo del tetraedro, il numero di lanci necessari per ottenere il primo 4 su uno dei due è rappresentato da $T = \min(T_1, T_2)$ e le v.a. sono indipendenti.

In generale siano $T_1 \sim \text{geom}(p)$ e $T_2 \sim \text{geom}(q)$ due v.a. indipendenti e vogliamo calcolare la densità di $T = \min(T_1, T_2)$, applicando le formule della densità geometrica, si ha

$$\begin{aligned} P(T = k) &= P(T_1 = k)P(T_2 \geq k) + P(T_1 > k)P(T_2 = k) \\ &= p(1-p)^{k-1}(1-q)^{k-1} + (1-p)^k q(1-q)^{k-1} \\ &= (1-p)^{k-1}(1-q)^{k-1}[p + q(1-p)] \\ &= [p + q - pq][(1-p)(1-q)]^{k-1} = [p + q - pq][1 - (p + q - pq)]^{k-1} \end{aligned}$$

ovvero $T \sim \text{geom}(p + q - pq)$

(M) La precedente formula può essere estesa a qualsiasi numero di v.a. geometriche indipendenti, generando una formula simile a quella di inclusione-esclusione per eventi. Trovarla per esercizio.

2. Sia $Z = \max(X, Y)$. Se X ed Y sono v.a. discrete anche Z lo è ed abbiamo

$$P(Z = z) = P(\max(X, Y) = z) = P(X = z, Y \leq z) + P(X < z, Y = z)$$

e se abbiamo anche indipendenza, la precedente diventa

$$P(Z = z) = P(X = z)P(Y \leq z) + P(X < z)P(Y = z)$$

Esempio 4.5.2. Si lanciano un dado ed un tetraedro equi e si continuano i lanci finchè su entrambi non è apparso un numero maggiore od uguale a 4. Cioè, non appena appare un numero maggiore od uguale a 4 su uno dei due ci si ferma e si continua con l'altro. Calcolare la densità della v.a. che esprime il numero di lanci necessari per aver ottenuto un numero maggiore od uguale a 4 su entrambi.

Se indichiamo con T_1 = "tempo di primo successo per ottenere un numero maggiore od uguale a 4 sul dado", allora sappiamo che $T_1 \sim \text{geom}(\frac{1}{2})$, analogamente se T_2 = "tempo di primo successo per ottenere maggiore od uguale a 4 sul tetraedro", allora $T_2 \sim$

$\text{geom}(\frac{1}{4})$ e le due v.a. sono indipendenti. Indichiamo con $p = \frac{1}{2}$ e $q = \frac{1}{4}$. Dunque il tempo T necessario per ottenere un numero maggiore od uguale a 4 su entrambi è $T = \max(T_1, T_2)$ ed è una v.a. discreta che prende valori su $k = 1, 2, \dots$ con densità data da

$$\begin{aligned}
 P(T = k) &= P(\max(T_1, T_2) = k) = P(T_1 = k, T_2 \leq k) + P(T_1 < k, T_2 = k) \\
 &= P(T_1 = k)P(T_2 \leq k) + P(T_1 < k)P(T_2 = k) \\
 &= P(T_1 = k)[1 - P(T_2 \geq k + 1)] + [1 - P(T_1 \geq k)]P(T_2 = k) \\
 &= p(1 - p)^{k-1}(1 - (1 - q)^k) + (1 - (1 - p)^{k-1})q(1 - q)^{k-1} \\
 &= p(1 - p)^{k-1} + q(1 - q)^{k-1} - (p + q - pq)(1 - [p + q - pq])^{k-1} \\
 &= \left(\frac{1}{2}\right)^k + \frac{1}{4}\left(\frac{3}{4}\right)^{k-1} - \frac{5}{8}\left(\frac{3}{8}\right)^{k-1}
 \end{aligned}$$

La formula generale, che possiamo leggere nella penultima riga per due v.a. indipendenti $X \sim \text{geom}(p)$ e $Y \sim \text{geom}(q)$ può essere scritta in maniera più suggestiva come

$$p_{\max(X,Y)}(k) = p_X(k) + p_Y(k) - p_{\min(X,Y)}(k).$$

3. Analizziamo ora il caso $Z = X + Y$. Chiaramente, quanto viene detto per due v.a. può essere esteso alla somma di un qualsiasi numero di v.a., iterando la somma.

Se vogliamo calcolare la densità di Z , abbiamo

$$P(Z = z) = P(X + Y = z) = \sum_x P(X + Y = z, X = x) = \sum_x P(Y = z - x, X = x)$$

e se siamo in condizioni di indipendenza

$$P(Z = z) = \sum_x P(Y = z - x)P(X = x).$$

Vediamo due casi particolari

Siano $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ tra loro indipendenti. Per esempio potrebbero rappresentare il numero di telefonate che arrivano ad un centralino in un'ora ed il numero di telefonate che arrivano allo stesso centralino all'ora successiva che supponiamo indipendenti poiché gli utenti non sono collegati tra loro. Allora $X + Y$ rappresenterà il numero totale di chiamate arrivate al centralino nelle due ore. Vogliamo calcolare la densità di $X + Y$.

Questa v.a. prende ancora valori $k = 0, 1, 2, \dots$ e si ha

$$P(X + Y = k) = \sum_{h=0}^{\infty} P(Y = k - h)P(X = h) = \sum_{h=0}^k P(Y = k - h)P(X = h),$$

poiché la probabilità di una Poisson è diversa da zero soltanto per valori non negativi. Dunque

$$\begin{aligned} P(X + Y = k) &= \sum_{h=0}^k P(Y = k - h)P(X = h) = \sum_{h=0}^k e^{-\lambda} \frac{\lambda^h}{h!} e^{-\mu} \frac{\mu^{k-h}}{(k-h)!} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{h=0}^k \frac{k!}{h!(k-h)!} \mu^{k-h} \lambda^h \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{h=0}^k \binom{k}{h} \mu^{k-h} \lambda^h = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!}, \end{aligned}$$

ovvero $X + Y$ è una v.a. di Poisson($\lambda + \mu$).

La stessa proprietà si riscontra anche per le distribuzioni binomiali.

Siano $X \sim \text{Bin}(n, p)$ ed $Y \sim \text{Bin}(m, p)$, $m, n \in \mathbb{N}$ due v.a. binomiali indipendenti allora $Z = X + Y \sim \text{Bin}(m + n, p)$, infatti X può essere vista come somma di n v.a. di Bernoulli indipendenti di parametro p , mentre Y può essere vista come somma di m v.a. di Bernoulli indipendenti di parametro p , allora

$$Z = X_1 + \dots + X_n + Y_1 + \dots + Y_m = Z_1 + \dots + Z_{m+n}$$

4.6 Revisione di Teoria

1. Dire, giustificando la risposta se la seguente è una densità di probabilità oppure no.

$$P(X = 1) = \frac{1}{4}, P(X = 2) = \frac{2}{5}, P(X = 3) = \frac{1}{10}, P(X = 4) = \frac{3}{10}.$$

2. Dare la definizione di funzione di densità di probabilità per v.a. discrete.
3. Dare la definizione di funzione di densità congiunta.
4. Dare la definizione di una famiglia finita di v.a. indipendenti.
5. Dare la definizione di una famiglia numerabile di v.a. indipendenti.
6. Dare la definizione di densità geometrica per una v.a. T .
7. Dare la definizione di densità di Poisson.
8. Sia data una v.a. X che segue una $\text{Bin}(400, \frac{1}{700})$. Come si può approssimare $P(X \geq 2)$?
9. Sia T una v.a. che segue una legge geometrica di parametro p . Valutare

$$P(T = k + n | T \geq n)$$

4.7 Esercizi

1. Un treno ha 8 carrozze. Ogni carrozza può, indipendentemente dalle altre, avere un difetto con probabilità 0.1. Le carrozze vengono controllate indipendentemente le une dalle altre da 2 controllori, ciascuno dei quali, se esiste un difetto lo scopre in maniera indipendente con probabilità 0.7. Il treno parte in ritardo se anche un solo difetto è scoperto. Calcolare la probabilità dei seguenti eventi:
 - (a) $A_1 = \{\text{un difetto c'è e viene scoperto sulla carrozza 1}\}$,
 - (b) $B = \{\text{il treno parte in ritardo}\}$.
2. Vi sono 3 carte numerate da 5 a 7 ed un'urna con 6 palline bianche. Si estrae a caso una carta e si mettono nell'urna tante palline nere quanto è il valore della carta. Si incominciano a effettuare estrazioni con reinserimento fermandosi la prima volta T che si estrae una pallina nera. Calcolare la densità di T .
3. Da un mazzo con 4 re, 4 regine, 4 fanti e 4 assi, A e B estraggono a turno una carta. A vince se estrae per primo un re, mentre B vince se estrae per primo un asso. Ogni volta la carta viene reimmessa nel mazzo ed il mazzo mescolato prima che il turno passi all'avversario. A gioca per primo.
 - (a) Calcolare la probabilità che vinca A e la probabilità che vinca B.
 - (b) Sia D la v.a. che indica la durata del gioco. Che densità ha D ?
4. Viene lanciato un dado equo e ad ogni lancio si introduce in un'urna un numero di palline pari al risultato del dado se questo è pari, zero se dispari. Indichiamo con N_2 il numero di palline presenti nell'urna dopo due lanci. Calcolare la densità di N_2 .

Capitolo 5

MEDIA E MOMENTI

La conoscenza di una v.a. che descrive un fenomeno si riassume nella densità di probabilità, ma questa potrebbe essere piuttosto complessa. È quindi utile avere delle quantità riassuntive di una v.a. che consentano di avere un'idea del suo comportamento.

5.1 La Media Matematica

Poiché la funzione di densità di probabilità (per una v.a. discreta) corrisponde in fisica al concetto di densità di massa, è ragionevole aspettarsi che vi sia un concetto corrispondente al concetto di baricentro. Questo è dato dalla

Definizione 5.1.1. *Sia X una v.a. discreta a valori in $\{x_i\}_{i \in I}$, dove $I \subseteq \mathbb{N}$, allora definiamo **media** di X la quantità*

$$(5.1) \quad \mathbb{E}(X) = \sum_{i \in I} x_i P(X = x_i) = \sum_{i \in I} x_i p_X(x_i),$$

sempre che la quantità $\sum_{i \in I} |x_i| P(X = x_i)$ risulti finita.

Altri possibili nomi sono: valore atteso, attesa, valore medio, aspettazione. Se risulta $\mathbb{E}(X) = 0$ allora la v.a. si dice **centrata**

Se $\phi : \mathbb{R} \rightarrow \mathbb{R}$ è una funzione reale e $Y = \phi(X)$ è una v.a. formata da X tramite ϕ , non è necessario conoscere la densità p_Y per poter calcolare la media di Y , infatti vale la

Proposizione 5.1.1. *Siano X una v.a. discreta e $\phi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione reale tali che*

$$\sum_{i \in I} |x_i| P(X = x_i) < +\infty, \quad \sum_{i \in I} |\phi(x_i)| P(X = x_i) < +\infty$$

allora

$$\mathbb{E}(\phi(X)) = \sum_{i \in I} \phi(x_i) P(X = x_i)$$

Dimostrazione: **(M).** Indichiamo con $\{y_j\}_{j \in J}$ i valori assunti da Y e poniamo $A_j = \phi^{-1}(y_j)$, da cui

$$P(Y = y_j) = P(X \in A_j) = \sum_{x_i \in A_j} P(X = x_i),$$

che implica

$$\begin{aligned} \mathbb{E}(\phi(X)) &= \sum_{j \in J} y_j \sum_{x_i \in A_j} P(X = x_i) = \sum_{j \in J} \sum_{x_i \in A_j} y_j P(X = x_i) \\ &= \sum_{j \in J} \sum_{x_i \in A_j} \phi(x_i) P(X = x_i) = \sum_{i \in I} \phi(x_i) P(X = x_i) \end{aligned}$$

poiché gli $\{A_j\}_{j \in J}$ costituiscono una partizione di Ω . □

La precedente proposizione vale anche se $\mathbf{X} = (X_1, \dots, X_n)$ è un vettore aleatorio, usando la densità congiunta delle v.a. Se $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$, si ha

$$\mathbb{E}(\phi(\mathbf{X})) = \sum_{x_1} \cdots \sum_{x_n} \phi(x_1, \dots, x_n) p_{\mathbf{X}}(x_1, \dots, x_n).$$

Osserviamo che dalla proposizione precedente si può riscrivere la definizione di media dicendo che $\mathbb{E}(X)$ esiste finita per X se $\mathbb{E}(|X|) < +\infty$.

5.1.1 Proprietà della media

1. Disuguaglianza triangolare: $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Dimostrazione: Dalla definizione di media, se $\{x_i\}_{i \in I}$ sono i valori assunti da X , si ha

$$|\mathbb{E}(X)| = \left| \sum_{i \in I} x_i P(X = x_i) \right| \leq \sum_{i \in I} |x_i| P(X = x_i) = \mathbb{E}(|X|),$$

dove abbiamo usato la disuguaglianza triangolare e la positività delle probabilità.

2. Linearità: se X ed Y sono due v.a. dotate di media e $\alpha, \beta \in \mathbb{R}$ allora

$$\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y).$$

Dimostrazione: Supponiamo che $|\mathbb{E}(\alpha X + \beta Y)| < +\infty$, allora se $\{x_i\}_{i \in I}$ e $\{y_j\}_{j \in J}$ sono i rispettivi valori di X ed Y , dalla definizione di media si ha

$$\begin{aligned} \mathbb{E}(\alpha X + \beta Y) &= \sum_{i \in I} \sum_{j \in J} (\alpha x_i + \beta y_j) P(X = x_i, Y = y_j) \\ &= \sum_{i \in I} \sum_{j \in J} (\alpha x_i + \beta y_j) P(X = x_i, Y = y_j) \\ &= \alpha \sum_{i \in I} \sum_{j \in J} x_i P(X = x_i, Y = y_j) + \beta \sum_{i \in I} \sum_{j \in J} y_j P(X = x_i, Y = y_j) \\ &= \alpha \sum_{i \in I} x_i P(X = x_i) + \beta \sum_{j \in J} y_j P(Y = y_j) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y), \end{aligned}$$

questa proprietà insieme con la precedente dà anche la buona definizione della media della somma poiché

$$|\mathbb{E}(\alpha X + \beta Y)| \leq |\alpha| \mathbb{E}(|X|) + |\beta| \mathbb{E}(|Y|) < +\infty.$$

Notiamo anche che la definizione di media implica che $\mathbb{E}(\alpha) = \alpha$ per ogni costante α .

3. Monotonia: se $X \geq 0$ allora $\mathbb{E}(X) \geq 0$.

Viene direttamente dalla definizione di media.

Prese v.a. X ed Y tali che $P(X \leq Y) = 1$, allora si ha che $\mathbb{E}(X) \leq \mathbb{E}(Y)$, poiché basta applicare la monotonia e la linearità della media alla v.a. $Y - X \geq 0$.

Infine osserviamo che se X è una v.a. limitata $|X| \leq M$, allora per la proprietà di monotonia si ha anche $|\mathbb{E}(X)| \leq \mathbb{E}(|X|) \leq M$.

5.1.2 Calcolo delle medie delle principali distribuzioni

1. $X \sim \text{Bernoulli}(p)$, allora

$$\mathbb{E}(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p = p$$

2. X uniforme su $\{x_1, \dots, x_m\}$, allora

$$\mathbb{E}(X) = \sum_{i=1}^m x_i P(X = x_i) = \frac{1}{m} \sum_{i=1}^m x_i$$

ovvero coincide con la media aritmetica dei valori.

3. $X \sim \text{Bin}(n, p)$. Possiamo vedere la v.a. come $X = X_1 + \dots + X_n$, dove X_i sono Bernoulli(p) i.i.d. Allora per linearità

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n) = np$$

4. $X \sim \text{ipergeom}(r, m, n)$, allora

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} = \sum_{k=1}^n k \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} = \frac{r}{m} n \sum_{k=1}^n \frac{\binom{r-1}{k-1} \binom{m-1-(r-1)}{n-1-(k-1)}}{\binom{m-1}{n-1}} \\ &= \frac{r}{m} n \sum_{h=0}^{n-1} \frac{\binom{r-1}{h} \binom{m-1-(r-1)}{n-1-h}}{\binom{m-1}{n-1}} = \frac{r}{m} n \end{aligned}$$

5. $X \sim \text{geom}(p)$, allora

$$\begin{aligned}
 \mathbb{E}(X) &= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = \sum_{k=1}^{+\infty} (k-1+1)p(1-p)^{k-1} \\
 &= \sum_{k=1}^{+\infty} (k-1)p(1-p)^{k-1} + \sum_{k=1}^{+\infty} p(1-p)^{k-1} \\
 &= \sum_{k=2}^{+\infty} (k-1)p(1-p)^{k-1-1}(1-p) + 1 = (1-p) \sum_{h=1}^{+\infty} hp(1-p)^{h-1} + 1 \\
 &= (1-p)\mathbb{E}(X) + 1
 \end{aligned}$$

e risolvendo l'equazione ottenuta, si ha $\mathbb{E}(X) = \frac{1}{p}$

6. $X \sim \text{Poisson}(\lambda)$, allora

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{h=0}^{+\infty} e^{-\lambda} \frac{\lambda^h}{h!} = \lambda.$$

Evidenziamo che, per tutte le distribuzioni considerate, la media è sempre collegata ai parametri che le caratterizzano.

Ricordiamo che la media è un numero ed in un certo senso esprime una prima approssimazione della v.a.

Osserviamo, infine, che se una v.a. Z è espressa come funzione di un vettore di v.a. (X_1, \dots, X_n) allora la media utilizza la densità congiunta del vettore e quindi può dover sfruttare anche le densità condizionali.

Esempio 5.1.1. Supponiamo di avere una successione di v.a. $\{X_i\}_{i \geq 0}$ di v.a. i.d. dotate di media e sia N una v.a. di Poisson di parametro λ da esse indipendente.

Consideriamo la somma aleatoria

$$S_N = \sum_{i=1}^N X_i.$$

Vogliamo calcolare la densità e la media di S_N .

Soluzione: Per ogni x si ha

$$\begin{aligned}
 P(S_N = x) &= \sum_{i=0}^{\infty} P(S_N = x | N = i) P(N = i) = \sum_{i=0}^{\infty} P(S_i = x | N = i) P(N = i) \\
 &= \sum_{i=0}^{\infty} P(S_i = x) P(N = i).
 \end{aligned}$$

Fortunatamente non abbiamo bisogno di calcolare la densità di S_N per calcolare la sua media

$$\begin{aligned}\mathbb{E}(S_N) &= \sum_x xP(S_N = x) = \sum_x x \sum_{i=0}^{\infty} P(S_i = x)P(N = i) \\ &= \sum_{i=0}^{\infty} P(N = i) \sum_x xP(S_i = x) = \sum_{i=0}^{\infty} P(N = i)E(S_i) \\ &= \sum_{i=0}^{\infty} P(N = i)iE(X_0) = E(X_0) \sum_{i=0}^{\infty} iP(N = i) = E(X_0)E(N).\end{aligned}$$

5.2 Momenti, Varianza e Covarianza

La proposizione 5.1.1 ci viene in aiuto per definire

Definizione 5.2.1. Siano $k \in \mathbb{N}$ e X una v.a. tale che $\mathbb{E}(|X|^k) < \infty$, allora diciamo **momento di ordine k** di X , $\mathbb{E}(X^k)$ e **momento centrato di ordine k** , la quantità $\mathbb{E}[(X - \mathbb{E}(X))^k]$.

I secondi sono detti momenti centrati poiché, data X , la v.a. $Y = X - \mathbb{E}(X)$ è tale che $\mathbb{E}(Y) = 0$.

La precedente definizione è generale, per una v.a. discreta con valori $\{x_i\}_{i \in I}$, esplicitamente diventa

$$\mathbb{E}(X^k) = \sum_{i \in I} x_i^k P(X = x_i), \quad \mathbb{E}([X - \mathbb{E}(X)]^k) = \sum_{i \in I} (x_i - \mathbb{E}(X))^k P(X = x_i),$$

prendendo $\phi(x) = x^k$ oppure $\phi(x) = (x - \mathbb{E}(X))^k$ nella proposizione 5.1.1.

È importante notare che i momenti esprimono l'ordine di integrabilità della v.a.

Proposizione 5.2.1. Se X è una v.a. con momento di ordine k , allora possiede tutti i momenti di ordine $0 \leq r \leq k$.

Se X ed Y possiedono il momento di ordine k anche $X + Y$ possiede il momento di ordine k .

Dimostrazione: poiché vale la seguente disuguaglianza per $0 \leq r \leq k$

$$|x|^r \leq 1 + |x|^k, \quad \forall x \in \mathbb{R},$$

applicandola alla media, si ha

$$\mathbb{E}(|X|^r) = \sum_{i \in I} |x_i|^r P(X = x_i) \leq \sum_{i \in I} (1 + |x_i|^k) P(X = x_i) = 1 + \mathbb{E}(|X|^k) < +\infty.$$

Per la seconda affermazione notiamo che vale la disuguaglianza ((**M**): dimostrarla per induzione per esercizio)

$$|x + y|^k \leq 2^{k-1}(|x|^k + |y|^k), \quad \forall x, y \in \mathbb{R}$$

da cui

$$\begin{aligned}
\mathbb{E}(|X + Y|^k) &= \sum_{i \in I} \sum_{j \in I} |x_i + y_j|^k P(X = x_i, Y = y_j) \\
&\leq \sum_{i \in I} \sum_{j \in I} 2^{k-1} (|x_i|^k + |y_j|^k) P(X = x_i, Y = y_j) \\
&= \sum_{i \in I} \sum_{j \in I} 2^{k-1} |x_i|^k P(X = x_i, Y = y_j) + \sum_{i \in I} \sum_{j \in I} 2^{k-1} |y_j|^k P(X = x_i, Y = y_j) \\
&= 2^{k-1} \left[\sum_{i \in I} |x_i|^k P(X = x_i) + \sum_{j \in I} |y_j|^k P(Y = y_j) \right] \\
&= 2^{k-1} [\mathbb{E}(|X|^k) + \mathbb{E}(|Y|^k)] < +\infty
\end{aligned}$$

dove nel penultimo passaggio abbiamo scambiato le sommatorie nella seconda serie. \square

Nel caso in cui $\mathbb{E}(X^2) < +\infty$, ricopre particolare interesse il momento centrato di ordine 2, che viene chiamato **Varianza** e si scrive $\text{Var}(X)$.

Osservando la precedente formula, la varianza è l'errore quadratico medio (pesato secondo la densità di X), che si commette sostituendo la v.a. X con la sua media. Può quindi essere considerato un indice della dispersione dei valori della X intorno alla sua media. A parità di pesi della densità, una v.a. con valori maggiormente dispersi lungo l'asse reale avrà una varianza maggiore rispetto ad una v.a. con valori più concentrati. È una distanza quadratica media tra X e $\mathbb{E}(X)$.

Esempio 5.2.1. Consideriamo le v.a.

$$X = \begin{cases} -1 & \frac{1}{3} \\ 0 & \frac{1}{3} \\ 1 & \frac{1}{3} \end{cases}, \quad Y = \begin{cases} -5 & \frac{1}{3} \\ 0 & \frac{1}{3} \\ 5 & \frac{1}{3} \end{cases}, \quad Z = \begin{cases} -1 & \frac{4}{9} \\ 0 & \frac{1}{9} \\ 1 & \frac{4}{9} \end{cases}$$

in tutti e tre i casi si ha $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$, ma quando andiamo a calcolare le varianze, che in tutti e tre i casi coincidono con il momento secondo, si nota che

$$\mathbb{E}(X^2) = 1 \frac{1}{3} + (-1)^2 \frac{1}{3} = \frac{2}{3}, \quad \mathbb{E}(Y^2) = 5^2 \frac{1}{3} + (-5)^2 \frac{1}{3} = \frac{25}{3}, \quad \mathbb{E}(Z^2) = 1 \frac{4}{9} + (-1)^2 \frac{4}{9} = \frac{8}{9}$$

ed è quindi chiaro che la varianza aumenta ogni volta che, a parità di pesi, la dispersione aumenta e, a parità di dispersione, se aumentano i pesi.

Proprietà della varianza

1. $\text{Var}(X) \geq 0$ per ogni X .
2. $\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$, infatti

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}(X))^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}(X) + \mathbb{E}[(\mathbb{E}(X))^2]] \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + [\mathbb{E}(X)]^2 = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2
\end{aligned}$$

3. $\text{Var}(aX) = a^2 \text{Var}(X)$, per qualsiasi $a \in \mathbb{R}$ infatti

$$\begin{aligned} E[(aX - \mathbb{E}(aX))^2] &= E[(aX - a\mathbb{E}(X))^2] \\ &= E[a^2(X - \mathbb{E}(X))^2] = a^2 E[(X - \mathbb{E}(X))^2] = a^2 \text{Var}(X) \end{aligned}$$

4. $\text{Var}(X + c) = \text{Var}(X)$, $c \in \mathbb{R}$, infatti dalla definizione si ha immediatamente

$$\text{Var}(X + c) = E[(X + c) - \mathbb{E}(X + c)]^2 = E[(X + c - \mathbb{E}(X) - c)^2] = \text{Var}(X).$$

Le due proprietà precedenti ci dicono che se trasliamo una v.a., la dispersione dei suoi valori (ovvero quanto sono distanti l'uno dall'altro) non cambia, mentre questa non è invariante per cambiamenti di scala, ovvero se cambiamo unità di misura nella misurazione di un certo fenomeno aleatorio, questa influenza quadraticamente la dispersione della v.a. che lo descrive.

Esempio 5.2.2. *L è una v.a. che esprime il numero di centimetri di stoffa usati per confezionare un abito e si è rilevato che la sua varianza è pari a $\sigma_{cm}^2 = 49 \text{cm}^2$. Allora, passando dal sistema metrico decimale al sistema inglese, poiché $1 \text{ cm} = 0.39 \text{ inches}$ si ha che $\sigma_{in}^2 = 49 \cdot 0.39^2 \text{in}^2 = 7.477 \text{in}^2$. Ovviamente non vi è stata una riduzione effettiva dell'errore, come potrebbe sembrare, ma semplicemente l'uso di un'unità di misura più grande per misurarlo.*

Poiché abbiamo visto che alcune v.a. vengono definite mediante la somma di altre v.a. e la varianza non si comporta linearmente diventa importante capire come calcolare $\text{Var}(X + Y)$ per due v.a. X ed Y dotate di momento secondo.

Applicando la definizione abbiamo

$$\begin{aligned} \text{Var}(X + Y) &= E[(X + Y)^2] - (E(X + Y))^2 = E(X^2 + 2XY + Y^2) - [\mathbb{E}(X) + \mathbb{E}(Y)]^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - [\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)^2] \\ &= E(X^2) - \mathbb{E}(X)^2 + E(Y^2) - \mathbb{E}(Y)^2 + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)], \end{aligned}$$

dove, grazie alla proposizione 5.1.1 si ha, utilizzando la densità congiunta di X ed Y ,

$$\mathbb{E}(XY) = \sum_x \sum_y xy P(X = x, Y = y).$$

La quantità $[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)]$ ha un significato ed un nome preciso

Definizione 5.2.2. *Date due v.a. X ed Y tali che $\text{Var}(X), \text{Var}(Y) < +\infty$ diciamo **covarianza** di X ed Y*

$$(5.2) \quad \text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Esplicitando la precedente espressione (5.2) nel caso delle v.a. discrete otteniamo

$$\text{cov}(X, Y) = \sum_{i \in I} \sum_{j \in J} (x_i - \mathbb{E}(X))(y_j - \mathbb{E}(Y))P(X = x_i, Y = y_j),$$

da cui si capisce che la covarianza esprime una misura di quanto in media le due v.a. varino insieme. Svolgendo i calcoli si nota immediatamente che un'espressione equivalente è data da

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

da cui possiamo riscrivere

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y).$$

Questa formula può essere iterata a qualsiasi numero di v.a. X_1, \dots, X_n dotate di momento secondo

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i \in I} \sum_{j > i} \text{cov}(X_i, X_j).$$

Se le v.a. sono indipendenti abbiamo la seguente

Proposizione 5.2.2. *Siano X ed Y due v.a. indipendenti dotate di media, allora anche XY è dotata di media e si ha*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Dimostrazione: Siano $\{x_i\}_{i \in I}$ e $\{y_j\}_{j \in J}$ i valori assunti rispettivamente dalle due v.a., grazie alla condizione di indipendenza sappiamo che

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall i \in I, j \in J.$$

Dunque si ha

$$\begin{aligned} \mathbb{E}(|XY|) &= \sum_{i \in I} \sum_{j \in J} |x_i y_j| P(X = x_i, Y = y_j) = \sum_{i \in I} \sum_{j \in J} |x_i y_j| P(X = x_i) P(Y = y_j) \\ &= \sum_{i \in I} |x_i| P(X = x_i) \sum_{j \in J} |y_j| P(Y = y_j) = \mathbb{E}(|X|) \mathbb{E}(|Y|) < +\infty \end{aligned}$$

Analogamente si dimostra la tesi. □

La precedente proposizione implica che per due v.a. indipendenti vale

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Questa formula può essere iterata a qualsiasi numero di v.a. indipendenti X_1, \dots, X_n dotate di momento secondo, ottenendo

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i).$$

L'indipendenza è una condizione molto forte ed infatti abbiamo la seguente

Definizione 5.2.3. Due v.a. X ed Y , dotate di momento secondo si dicono **scorrelate** se $cov(X, Y) = 0$.

Chiaramente due v.a. indipendenti sono anche scorrelate, ma il contrario non è vero

Esempio 5.2.3. Sia

$$X = \begin{cases} 1 & p \\ -1 & 1-p \end{cases}$$

e consideriamo $Y = X^2 \equiv 1$. Allora X ed Y sono evidentemente dipendenti, ma sono scorrelate poiché $0 = \mathbb{E}(X) = \mathbb{E}(X^3) = \mathbb{E}(X \cdot X^2)$ e quindi $\mathbb{E}(XY) = 0 = \mathbb{E}(X)\mathbb{E}(Y)$.

Perché la covarianza di due v.a. sia definita, è necessario che le due v.a. siano dotate di momento secondo come si vede dalla seguente

Proposizione 5.2.3. Siano X ed Y due v.a. tali che $Var(X), Var(Y) < +\infty$, allora si ha

$$|\mathbb{E}(XY)|^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

Dimostrazione: Si consideri la v.a. $X + \theta Y$ per $\theta \in \mathbb{R}$, per monotonia della media abbiamo

$$0 \leq \mathbb{E}((X + \theta Y)^2) = \mathbb{E}(X^2 + 2\theta XY + \theta^2 Y^2) = \mathbb{E}(X^2) + 2\theta \mathbb{E}(XY) + \theta^2 \mathbb{E}(Y^2),$$

che è un polinomio di secondo grado con coefficiente quadratico positivo ovvero rappresenta una parabola rivolta verso l'alto che rimane sempre non negativa, quindi il polinomio o non ha radici reali o ne ha due coincidenti e il suo discriminante ridotto deve essere

$$\Delta = [\mathbb{E}(XY)]^2 - \mathbb{E}(X^2)\mathbb{E}(Y^2) \leq 0$$

da cui la tesi. □

Osservazione 5.2.1. La covarianza tra due v.a. è una forma bilineare simmetrica, ovvero $\forall \alpha, \beta \in \mathbb{R}$ si ha

$$cov(X, Y) = cov(Y, X), \quad cov(\alpha X + \beta Y, Z) = \alpha cov(X, Z) + \beta cov(Y, Z),$$

come si dimostra direttamente dalla definizione. Quindi anche la covarianza non è invariante per cambi di unità di misura, dipende dai fattori di scala.

Per questo motivo può essere utile introdurre il **coefficiente di correlazione**

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}},$$

il quale verifica

1. $\rho_{X,Y} = 0$ se e soltanto se $cov(X, Y) = 0$, quindi continua ad essere un indice della scorrelazione;
2. $\rho_{aX,bY} = \rho_{X,Y}$, quindi è invariante per cambiamenti di scala;
3. dalla proposizione 5.2.3 abbiamo che $-1 \leq \rho_{X,Y} \leq 1$, cioè è sempre limitato.

Notiamo che le proprietà della varianza sono state dimostrate indipendentemente dall'uso della densità delle v.a. e quindi rimarranno valide anche quando si introdurranno le v.a. continue.

5.2.1 Calcolo delle varianze delle principali distribuzioni

Come per le medie, è utile saper riconoscere le varianze delle principali distribuzioni discrete. Useremo la formula $\mathbb{E}(X^2) - [\mathbb{E}(X)]^2$.

1. Sia $X \sim \text{Bin}(1, p)$, allora $X^2 \equiv X$ e $\mathbb{E}(X^2) = p$, dunque

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

2. $X \sim \text{Bin}(n, p)$, allora come al solito sappiamo che la possiamo rappresentare come $X = X_1 + \dots + X_n$ con $X_i \sim \text{Bin}(1, p)$ indipendenti, dunque grazie alla formula per la somma di v.a. indipendenti

$$\text{Var}(X) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = np(1 - p)$$

3. Per X uniforme su $\{x_i\}_{i \in I}$ si applica soltanto la definizione

$$\text{Var}(X) = \frac{1}{m} \sum_{i \in I} \left(x_i - \left(\frac{1}{m} \sum_{i \in I} x_i \right) \right)^2 = \frac{1}{m} \sum_{i \in I} x_i^2 - \left(\frac{1}{m} \sum_{i \in I} x_i \right)^2$$

4. $X \sim \text{Poisson}(\lambda)$ si ha

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} k(k-1+1) e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} + \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} \\ &= \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{h=0}^{\infty} e^{-\lambda} \frac{\lambda^h}{h!} + \lambda \end{aligned}$$

$$\text{quindi } \text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

5. $X \sim \text{geom}(p)$, allora

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} = \sum_{k=1}^{\infty} (k-1+1)^2 p(1-p)^{k-1} \\ &= \sum_{k=1}^{\infty} [(k-1)^2 + 2(k-1) + 1] p(1-p)^{k-1} \\ &= (1-p) \sum_{k=2}^{\infty} (k-1)^2 p(1-p)^{k-2} + 2(1-p) \sum_{k=2}^{\infty} (k-1) p(1-p)^{k-2} \\ &\quad + \sum_{k=1}^{\infty} p(1-p)^{k-1}, \end{aligned}$$

da cui ponendo $h = k - 1$, si ottiene

$$\begin{aligned} \mathbb{E}(X^2) &= (1-p) \sum_{h=1}^{\infty} h^2 p(1-p)^{h-1} + 2(1-p) \sum_{h=1}^{\infty} h p(1-p)^{h-1} + 1 \\ &= (1-p) \mathbb{E}(X^2) + 2(1-p) \mathbb{E}(X) + 1 = (1-p) \mathbb{E}(X^2) + 2(1-p) \frac{1}{p} + 1 \end{aligned}$$

e risolvendo l'equazione concludiamo

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

6. $X \sim \text{ipergeom}(m, r, n)$. Procedendo come nel caso della media, si ha

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^n k^2 \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} = \sum_{k=0}^n k(k-1+1) \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} \\ &= \sum_{k=2}^n k(k-1) \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} + \sum_{k=1}^n k \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}} \\ &= \frac{r(r-1)}{m(m-1)} n(n-1) \sum_{k=2}^n \frac{\binom{r-2}{k-2} \binom{m-2-(r-2)}{n-2-(k-2)}}{\binom{m-2}{n-2}} + \frac{r}{m} n \\ &= \frac{r(r-1)}{m(m-1)} n(n-1) \sum_{h=0}^{n-2} \frac{\binom{r-2}{h} \binom{m-2-(r-2)}{n-2-h}}{\binom{m-2}{n-2}} + \frac{r}{m} n \\ &= \frac{r(r-1)}{m(m-1)} n(n-1) + \frac{r}{m} n, \end{aligned}$$

da cui

$$\text{Var}(X) = \frac{r}{m} n \left[\frac{(r-1)}{(m-1)} (n-1) - \frac{r}{m} n + 1 \right].$$

Abbiamo detto che la varianza misura la dispersione della v.a. rispetto alla sua media, la seguente disuguaglianza da una stima di questa distanza attraverso la varianza.

Proposizione 5.2.4. (*disuguaglianza di Chebyshev*) Sia X una v.a. dotata di momento del secondo ordine (o equivalentemente $\text{Var}(X) < +\infty$). Allora, denotando con $\mu := \mathbb{E}(X)$ e $\sigma^2 := \text{Var}(X)$, per ogni $\epsilon > 0$ vale

$$(5.3) \quad P(|X - \mathbb{E}(X)| > \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$

Dimostrazione: Dato $\epsilon > 0$, consideriamo l'evento $A = \{|X - \mathbb{E}(X)| > \epsilon\}$, da cui $A^c = \{|X - \mathbb{E}(X)| \leq \epsilon\}$. Si ha

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[(X - \mathbb{E}(X))^2(\mathbf{1}_A + \mathbf{1}_{A^c})] > \epsilon^2 P(A) + 0$$

da cui la tesi. □

Esempio 5.2.4. Si lancia ripetutamente una moneta che non si sa se equa. Se rappresentiamo con uno schema di successo insuccesso i lanci della moneta

$$X_i = \begin{cases} 1 & \text{se } T \text{ con prob. } p \\ 0 & \text{se } C \text{ con prob. } 1-p, \end{cases}$$

allora la moneta è equa se $p = \frac{1}{2}$. Supponiamo di effettuare 100 lanci, allora $S_{100} = X_1 + \dots + X_{100}$ è il numero di teste uscite sui 100 lanci e sappiamo che $S_{100} \sim \text{Bin}(100, \frac{1}{2})$, dunque $\mathbb{E}(S_{100}) = 100 \cdot \frac{1}{2} = 50$ e $\text{Var}(S_{100}) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25$, applicando la disuguaglianza di Chebyshev quindi si ottiene

$$P(|S_{100} - 50| > 10) \leq \frac{25}{100} = 0.25,$$

e dividendo per 100

$$P(|\frac{S_{100}}{100} - \frac{1}{2}| > \frac{1}{10}) \leq \frac{25}{100} = 0.25,$$

cioè, ripetendo una successione di 100 lanci più volte, ci aspettiamo che la media aritmetica dei valori di ciascuna sequenza da 100 si discosti da $\frac{1}{2}$ per più di un decimo, in meno di un quarto dei casi.

5.3 Retta di regressione lineare

Diamo ora un' applicazione di quanto introdotto finora, che in realtà prescinde dalla natura discreta delle v.a. e quindi rimarrà valida, nel suo impianto teorico, anche nel caso continuo.

Nel caso in cui due v.a. siano tra loro dipendenti sarebbe importante capire che tipo di dipendenza le lega, per esempio se una è una funzione dell'altra ed in caso quale funzione. In assenza di informazioni in tal senso, si può cercare di trovare la migliore dipendenza di tipo lineare che approssima l'una sulla base dell'altra. In altre parole, date X ed Y v.a. si vogliono determinare i parametri a e $b \in \mathbb{R}$ che minimizzano la distanza

$$\mathbb{E}[(Y - aX - b)^2].$$

Riscriviamo la precedente quantità come

$$\mathbb{E}[(Y - aX - b)^2] = \mathbb{E}[(Y - \mathbb{E}(Y)) - a(X - \mathbb{E}(X)) - (b - \mathbb{E}(Y) + a\mathbb{E}(X))]^2$$

e chiamiamo $c = (b - \mathbb{E}(Y) + a\mathbb{E}(X))$, quindi abbiamo

$$\begin{aligned} \mathbb{E}[(Y - aX - b)^2] &= \mathbb{E}[(Y - \mathbb{E}(Y))^2] + a^2\mathbb{E}[(X - \mathbb{E}(X))^2] - 2a\mathbb{E}[(Y - \mathbb{E}(Y))(X - \mathbb{E}(X))] + c^2 \\ &\quad - 2c\mathbb{E}(Y - \mathbb{E}(Y)) - 2ac\mathbb{E}(X - \mathbb{E}(X)) \\ &= \text{Var}(Y) + a^2\text{Var}(X) - 2a\text{cov}(X, Y) + c^2 + 0 + 0 =: f(a, c) \end{aligned}$$

la precedente uguaglianza definisce un paraboloide non negativo rivolto verso l'alto nelle variabili a, c , quindi per trovare il minimo di questa funzione è sufficiente individuare il punto critico, perciò derivando f in a e c e ponendo le uguaglianze a 0 si ottiene

$$(5.4) \quad a = \frac{\text{cov}(X, Y)}{\text{Var}(X)}, \quad c = 0 \Rightarrow b = \mathbb{E}(Y) - a\mathbb{E}(X) = \mathbb{E}(Y) - \frac{\text{cov}(X, Y)}{\text{Var}(X)}\mathbb{E}(X),$$

quindi concludiamo dicendo che la **retta di regressione lineare** di Y su X è data da $aX + b$ con a e b scelti come in (5.4).

La retta di regressione lineare è molto utile per cercare di stabilire la dipendenza tra due set di dati statistici, ma è anche usata per modellizzare fenomeni in senso teorico.

Esempio 5.3.1. (esempio 2.55 di Baldi - Calcolo delle Probabilità 2^a ed.)

Supponiamo che un segnale Y segua una distribuzione con media μ e varianza σ^2 (che si conoscono in quanto si conosce la fonte della trasmissione).

Poiché la ricevente è difettosa, in ricezione si osserva un segnale leggermente diverso X , che si suppone essere dato da $X = Y + W$, dove W rappresenta l'errore di misurazione causato dall'apparecchio ricevente. Si suppone che W abbia media 0, varianza ρ^2 e che sia indipendente da Y .

Come stimare Y sulla base dell'osservazione X ? Notiamo che noi non possiamo incorporare l'errore dal messaggio ricevuto.

Poiché si è ipotizzata una dipendenza di tipo lineare di X da Y , è ragionevole pensare che la retta di regressione lineare di Y su X possa dare una buona stima del segnale inviato.

Calcoliamo dunque

$$\begin{aligned}\text{Var}(X) &= \text{Var}(Y) + \text{Var}(W) = \sigma^2 + \rho^2 \\ \text{cov}(X, Y) &= \text{cov}(Y + W, Y) = \text{cov}(Y, Y) + \text{cov}(W, Y) = \text{Var}(Y) + 0 = \sigma^2\end{aligned}$$

l'ultimo passaggio è giustificato poiché W ed Y sono indipendenti.

Da (5.4) otteniamo i coefficienti

$$a = \frac{\text{cov}(X, Y)}{\text{Var}(X)} = \frac{\sigma^2}{\sigma^2 + \rho^2}, \quad b = \mathbb{E}(Y) - a\mathbb{E}(X) = \mu - \frac{\sigma^2}{\sigma^2 + \rho^2}\mu = \frac{\rho^2}{\sigma^2 + \rho^2}\mu,$$

dunque la stima cercata è

$$aX + b = \frac{\sigma^2}{\sigma^2 + \rho^2}X + \frac{\rho^2}{\sigma^2 + \rho^2}\mu = X - \frac{\rho^2}{\sigma^2 + \rho^2}(X - \mu),$$

che significa che bisogna correggere l'osservazione X spostandola nella direzione della media di Y .

5.4 Revisione di Teoria

1. Sia X una v.a., dare la definizione di varianza di X e dire di quali proprietà gode.
2. Calcolare la media e la varianza delle principali densità discrete.
3. Dimostrare la disuguaglianza di Chebyshev.
4. Se $X \sim \text{geom}(\frac{1}{3})$, qual è la varianza di $2X - 4$?
5. Se $X \sim \text{Poisson}(3)$, qual è la varianza di $\frac{1}{\sqrt{3}}X$?
6. Siano X ed Y due v.a. con medie $\mathbb{E}(X) = \mu_X$, $\mathbb{E}(Y) = \mu_Y$, coefficiente di correlazione ρ e varianze $\text{Var}(X) = \sigma_X^2$, $\text{Var}(Y) = \sigma_Y^2$. Quanto vale $\mathbb{E}((X + Y)^2)$?
7. Siano X_1, \dots, X_n v.a. indipendenti con distribuzione di Poisson (λ). Sia $Y = X_1 + \dots + X_n$, quanto vale $\mathbf{Var}(Y)$?

8. Dare la definizione di covarianza di due v.a. X ed Y e dire di quali proprietà gode.
9. Dare la definizione di indice di correlazione di due v.a. X ed Y e dire di quali proprietà gode.
10. Dare la definizione di v.a. scorrelate. Dimostrare che v.a. indipendenti sono scorrelate, ma che non è vero il viceversa
11. Dare la definizione di retta di regressione lineare di una v.a. Y su X , entrambe con media e varianza finite.
12. Siano X_1, \dots, X_n v.a. $\text{Bin}(1, p)$ indipendenti. Sia $Y = X_1 + \dots + X_n$, qual è la densità di Y ? $\mathbb{E}(Y)$, $\text{Var}(Y)$?
13. Se $\mathbb{E}(X) = \mu$, $\text{Var}(X) = \sigma$, $\mathbb{E}(Y) = \lambda$ e X ed Y sono indipendenti, quanto vale $\mathbb{E}(X^2 Y)$?

5.5 Esercizi

Riprendendo alcuni esercizi della Sezione precedente.

1. Da un mazzo con 4 re, 4 regine, 4 fanti e 4 assi, A e B estraggono a turno una carta. A vince se estrae per primo un re, mentre B vince se estrae per primo un asso. Ogni volta la carta viene reimmessa nel mazzo ed il mazzo viene mescolato prima che il turno passi all'avversario. Sia D la v.a. che indica la durata del gioco. Calcolarne la densità e la media.
2. A, B estraggono, senza reinserimento, due carte ciascuno da un mazzo formato da due J, due K e due Q. Inizia A e vince chi estrae almeno un J, se l'avversario non ne estrae nessuno, altrimenti la partita è patta.
Si ripete la partita finché uno dei giocatori non vince ed sia D la durata del gioco. Calcolare $\mathbb{E}(D)$?
3. Viene lanciato un dado equo e ad ogni lancio si introduce in un'urna un numero di palline pari al risultato del dado se questo è pari, zero se dispari. Indichiamo con N_2 il numero di palline presenti nell'urna dopo due lanci.
Calcolare $E(N_2)$ e $\text{Var}(N_2)$.
4. In una classe i $\frac{2}{5}$ della componente femminile ha avuto un risultato alto in una prova, altri $\frac{2}{5}$ un risultato medio e le restanti un risultato basso. La componente maschile della classe ha invece ottenuto: $\frac{1}{4}$ alto, $\frac{1}{2}$ medio, i restanti basso. I $\frac{3}{5}$ della classe sono femmine, i $\frac{2}{5}$ maschi. Sia X la v.a. che assume 1 se si ha una studentessa e 0 se si ha uno studente e sia Y la v.a. che assume il valore 1 se la prova dello/a studente/ssa ha ottenuto un risultato basso, 2 se medio e 3 se alto.
(a) Scrivere la densità discreta congiunta delle variabili X, Y

- (b) Calcolare le densità marginali, $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\text{Var}(X)$, e $\text{Var}(Y)$.
- (c) Calcolare $\text{cov}(X, Y)$.
5. Si lancia una moneta equa 3 volte e in un'urna si introducono 3 palline bianche ogni volta che esce testa o 4 rosse, ogni volta che esce croce. Siano N_B ed N_R rispettivamente il numero di palline bianche e di palline rosse introdotte nell'urna
- (a) Calcolare la densità di N_B e di N_R .
- (b) Calcolare $\mathbb{E}(N_B)$ ed $\mathbb{E}(N_R)$.
- (c) Calcolare la densità di $N_B + N_R$.
6. Siano X ed Y due v.a. discrete a valori rispettivamente in $\{-2, 0, 1\}$ ed in $\{0, 2\}$. La loro densità congiunta $p_{X,Y}(x, y)$ è descritta dalla tabella

Y/X	$X = -2$	$X = 0$	$X = 1$
$Y = 0$	$\frac{1}{4}$	$\frac{1}{8}$	α
$Y = 2$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

con $\alpha \in \mathbb{R}$.

- (a) Calcolare α e le densità marginali p_X e p_Y .
- (b) Calcolare $\mathbb{E}(X)$, $\mathbb{E}(Y)$ e $\mathbb{E}(XY)$.
7. Due dadi equi vengono lanciati separatamente più volte. Sia X il numero di lanci del primo dado per ottenere il primo 2 e sia Y il numero di lanci necessari del secondo dado per ottenere 5 o 6 per la prima volta.
- (a) Qual è la legge di X ? Di Y ?
- (b) Sia $Z = \max(X, Y)$. Scrivere una formula per la densità di Z .
8. Un gruppo WWF è composto da 9 ragazzi spagnoli di cui 3 uomini e 6 donne, 3 ragazzi francesi, di cui 2 uomini e una donna, e 6 ragazzi greci di cui 4 uomini e 2 donne. Viene selezionato un rappresentante a caso. Definiamo le due variabili aleatorie X , Y , dove $X \in \{1, 2, 3\}$ indica se viene selezionato rispettivamente uno spagnolo, un francese o un greco e $Y \in \{0, 1\}$ se viene selezionato un uomo o una donna.
- (a) Dire qual è la densità discreta congiunta delle variabili X, Y
- (b) Dire se le variabili X, Y sono indipendenti
- (c) Calcolare la densità di probabilità discreta della variabile $Z = X + Y$, il suo valore atteso e la sua varianza.
- (d) Dire quanto vale $P(X = 1|Y = 1)$

- (e) Supponiamo che ci siano N gruppi WWF, ognuno con la stessa percentuale di spagnoli, francesi e greci, e ognuno con la stessa percentuale di donne e uomini per ogni nazionalità, e supponiamo che ogni gruppo selezioni a caso un rappresentante. Indichiamo con $Y_i \in \{0, 1\}$ la variabile aleatoria che indica se il rappresentante selezionato dal gruppo i -esimo è rispettivamente un uomo o una donna. Dire qual'è la densità discreta della variabile $S = \sum_{i=1}^N Y_i$ e dare un'espressione (senza calcolarla) per la probabilità che siano stati selezionati più di $\frac{3}{4}N$ donne. Usando la disuguaglianza di Chebyshev dare una stima di tale probabilità.
9. Siano X ed Y due v.a. discrete a valori rispettivamente in $\{-1, 0, 1\}$ ed in $\{-1, 1\}$. La loro densità congiunta $p_{X,Y}(x, y)$ è descritta dalla tabella

Y/X	$X = -1$	$X = 0$	$X = 1$
$Y = -1$	$\frac{1}{4}$	$\frac{1}{8}$	α
$Y = 1$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

con $\alpha \in \mathbb{R}$.

- (a) Calcolare α e le densità marginali p_X e p_Y .
- (b) Dire se X ed Y sono indipendenti, scorrelate o correlate.
10. Vi sono due monete, A equa e B che dà testa con una probabilità pari a $\frac{1}{4}$. Sia N_1 il numero di lanci necessari per ottenere la prima testa sulla moneta A e sia N_2 il numero di lanci necessari per ottenere la prima testa sulla moneta B .
- (a) Calcolare la probabilità che esca testa sulla moneta A prima che sulla moneta B .
- (b) Calcolare il numero medio di lanci per ottenere la prima testa (su A o su B)
11. Siano X, Y, Z tre v.a. indipendenti tutte distribuite secondo una Poisson di parametro $\lambda = 3$. Calcolare la retta di regressione lineare di $Y + Z$ su $X + Y$.

Capitolo 6

ESERCIZI DI RIEPILOGO

1. Siano A, B due eventi indipendenti con $P(A) = \frac{1}{3}$ e $P(B) = \frac{1}{5}$, allora è vero che $P(A|B) = \frac{8}{15}$? Giustificare la risposta.
2. Se $P(A|B) = \frac{1}{3}$ e $P(B) = \frac{1}{4}$, quanto vale $P(B \cap A^c)$?
3. Dire se la seguente è una densità di probabilità, giustificare la risposta.

$$P(X = 0) = \frac{1}{4}, \quad P(X = -1) = \frac{1}{3}, \quad P(X = 1) = \frac{1}{6}, \quad P(X = 2) = \frac{1}{6}$$

4. Siano X, Y, Z tre v.a. geometriche rispettivamente di parametri p_1, p_2, p_3 , Calcolare

$$\mathbb{E}(\min(X, Y)Z^2)$$

5. Siano X_1, X_2 , due v.a. indipendenti con medie μ_1 e μ_2 , e varianze σ_1^2 e σ_2^2 . Calcolare $\mathbb{E}(3X_1X_2^2)$.
6. Sia X una v.a. di Poisson di parametro $\lambda > 0$. Calcolare $\mathbb{E}(e^X)$.
7. Si hanno due mazzi uguali di 40 carte con 4 semi numerati da 1 a 4. Si pescano a caso 4 carte da ciascun mazzo e si scambiano. Sia X la v.a. che esprime il numero di carte uguali che troviamo nel primo mazzo dopo lo scambio.
 - (a) Calcolare la legge di X .
 - (b) Calcolare $E(X)$ e $\text{Var}(X)$.

8. Una popolazione presenta un genotipo di tipo 1, 2 e 3 e fenotipo di tipo 1, 2, 3. Gli individui con fenotipo 1 rappresentano $\frac{1}{5}$ della popolazione, quelli con fenotipo 2 i $\frac{2}{5}$. Se si ha il fenotipo 1, allora con probabilità $\frac{3}{5}$ si presenta il genotipo 1, con probabilità $\frac{2}{5}$ il genotipo 2. Se invece si ha il fenotipo 2, con probabilità $\frac{1}{10}$ si ha il genotipo 1, con probabilità $\frac{3}{10}$ il genotipo 2, mentre il resto ha il genotipo 3. Infine non vi sono individui che con fenotipo 3 hanno genotipo 1, ma $\frac{2}{5}$ di questi hanno il genotipo 2. Se denotiamo con X la v.a. che indica il genotipo e con Y quella che indica il fenotipo, calcolare

- (a) la distribuzione congiunta di X ed Y ;
- (b) la covarianza di X ed Y .

9. Siano X ed Y due v.a. con densità congiunta $p_{X,Y}(x,y)$ descritta da

Y/X	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	15%	5%	0
$Y = 2$	5%	35%	10%
$Y = 3$	0	20%	10%

- (a) Calcolare le marginali p_X e p_Y , $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$.
 - (b) Dire se X ed Y sono scorrelate.
10. Vi sono due laghetti A e B . In A la probabilità di pescare un pesce rosso è di $\frac{1}{4}$ mentre nel laghetto B è di $\frac{1}{8}$. Si sceglie un laghetto a caso e quindi si continua pescare finchè non si trova un pesce rosso, ributtando nel laghetto i pesci se non sono rossi.
- (a) Calcolare la densità di probabilità del numero di tentativi necessario per pescare un pesce rosso.
 - (b) In media quanti tentativi saranno necessari per ottenere un pesce rosso?
11. Si devono indovinare tre cifre di una combinazione. I numeri si possono ripetere. Iniziando dalla prima, si procede in successione e si hanno tre tentativi per ogni cifra. Ad ogni tentativo viene comunicato se si è trovata la cifra esatta oppure no.
- (a) Calcolare la probabilità di indovinare la prima cifra.
 - (b) Se si indovinano tutte e tre le cifre si vincono 400 euro, 200 se si riesce ad indovinarne 2, 100 per 1 e 0 se non se ne indovina nessuna. Sia V la variabile aleatoria che indica la vincita. Calcolare la densità di V .
 - (c) Calcolare la media di V .
12. Un ragazzo disordinato non tiene separati nei suoi tre cassetti le sue 12 paia di calzini, 12 paia di mutande e 12 canottiere. In ogni cassetto vi sono esattamente 12 pezzi di biancheria (consideriamo i calzini piegati insieme a formare un unico pezzo). Nel primo cassetto vi sono 6 paia di mutande, nel secondo 2 paia e nel terzo le rimanenti. La mattina il ragazzo è particolarmente assonnato, apre a caso un cassetto e pesca a caso tre pezzi di biancheria.
- (a) Qual è la probabilità che non peschi alcuna mutanda?
 - (b) Sapendo di non aver pescato alcun paio di mutande, qual è la probabilità che abbia aperto il secondo cassetto?
13. In una popolazione $\frac{1}{10}$ degli individui ha i capelli rossi e gli occhi verdi, $\frac{1}{15}$ capelli rossi e occhi azzurri, $\frac{1}{5}$ capelli biondi e occhi verdi, $\frac{1}{10}$ capelli biondi e occhi azzurri,

- $\frac{1}{15}$ capelli scuri e occhi verdi, $\frac{2}{15}$ capelli scuri e occhi azzurri, $\frac{4}{15}$ occhi e capelli scuri, non vi sono individui con i capelli rossi e gli occhi scuri, non si sa quanti siano con i capelli biondi e gli occhi scuri. Indichiamo con $X = 1; 2; 3$ rispettivamente capelli rossi, biondi e scuri e con $Y = 1; 2; 3$ rispettivamente occhi verdi, azzurri e scuri
14. Si è stimato che $\frac{1}{500}$ della popolazione bovina è affetto da BSE. Un test per la BSE dà un positivo su un animale malato con probabilità 0.99 e negativo su un animale sano con probabilità 0.98
 - (a) Effettuando un test su un animale scelto a caso, qual è la probabilità che risulti positivo?
 - (b) Sapendo che il test è risultato positivo, qual è la probabilità che l'animale sia effettivamente malato?
 15. Vi è un'urna con 6 palline bianche. Si lancia un dado e si mettono nell'urna tante palline nere quanto è il pari più vicino, maggiore o uguale al risultato. Stabilita la composizione dell'urna, si incominciano a effettuare estrazioni con reinserimento fermandosi la prima volta T che si estrae una pallina nera.
 - (a) Calcolare la densità di T .
 - (b) Calcolare $E(T)$.
 16. Una banca è aperta dalle 9 alle 11. Ogni ora il numero di clienti che arrivano è una Poisson (10) indipendentemente da quanto succede nell'altra ora.
 - (a) Qual è la probabilità che il numero totale di clienti arrivati alla banca sia maggiore o uguale a 15?
 - (b) In media quante persone arriveranno?
 17. Due tetraedri equi con le facce numerate da 1 a 4 vengono lanciati. Sia X la v.a. che indica il massimo risultato uscito,
 - (a) qual è la densità di X ?
 - (b) Quanto valgono $E(X)$ e $\text{Var}(X)$?
 18. Un giocatore partecipa al seguente gioco. Si lanciano tre dadi equi ed il giocatore vince $1 \leq i \in \mathbb{N}$ euro se il 6 appare i volte, perde 1 euro se il 6 non appare su alcun dado. Indicando con V la v.a. vincita/perdita
 - (a) Calcolare la densità di V .
 - (b) Calcolare la vincita/perdita media.
 19. In un torneo di tennis si gioca su 3 campi contemporaneamente in maniera indipendente. Ogni campo vede due partecipanti e ci si ferma al primo giocatore che ottiene

tre set su cinque. Chiamiamo A_i, B_i $i = 1, 2, 3$ i giocatori sui quattro campi e ad ogni set, indipendentemente dagli altri, si ha

$$P(A_1 \text{ vince}) = \frac{1}{2}, \quad P(A_2 \text{ vince}) = \frac{2}{3}, \quad P(A_3 \text{ vince}) = \frac{1}{4}$$

- (a) Calcolare la probabilità che il gioco si fermi in al massimo 4 set sul campo 1, sul campo 2 e sul campo 3 rispettivamente.
 - (b) Calcolare la probabilità che il gioco si fermi in al massimo 4 set su tutti e tre i campi.
20. Vi sono due mazzi di carte francesi indistinguibili (4 semi, 13 valori per seme), il mazzo A contiene anche due jolly, il mazzo B contiene anche 4 jolly. Si sceglie un mazzo a caso e si estraggono 5 carte e si è interessati al poker d'assi, considerando che i jolly possono assumere qualsiasi valore si desidera.
- (a) Qual è la probabilità che si possa formare un poker d'assi utilizzando esattamente due assi e due jolly?
 - (b) Sapendo di aver formato un poker d'assi utilizzando esattamente due assi e due jolly, qual è la probabilità che si stia usando il mazzo B?

Capitolo 7

VARIABILI ALEATORIE CONTINUE

7.1 Funzioni di ripartizione e di densità

Molti fenomeni aleatori, come il tempo di arrivo di un certo mezzo, il tempo di sopravvivenza di un macchinario, il taglio di una stoffa non possono essere descritti mediante v.a. discrete, poiché prendono naturalmente valori in un intervallo della retta reale.

In questi casi, l'attribuzione della probabilità ad un singolo valore diventa più problematica, poiché, a causa della natura continua del fenomeno, non possiamo dire che la lampadina si fulminerà esattamente dopo 215,56 ore di funzionamento. In altre parole è naturale aspettarsi che la probabilità che la nostra v.a. assuma esattamente questo valore sia nulla, diversamente da quanto capitava nel caso discreto. La probabilità è attribuita in maniera diffusa ed abbiamo bisogno di uno strumento più adatto a descriverla.

Sicuramente potremo attribuire probabilità all'evento che la lampadina funzioni ancora dopo 100 ore di funzionamento, oppure all'evento che un treno sia arrivato prima delle cinque. Introduciamo quindi la

Definizione 7.1.1. Sia X una v.a. su (Ω, \mathcal{F}, P) , definiamo **funzione di ripartizione** di X , la funzione

$$F_X(t) = P(X \leq t) = P(X^{-1}((-\infty, t])) \quad t \in \mathbb{R}.$$

Proprietà della funzione di ripartizione F_X :

1. $\lim_{t \rightarrow +\infty} F_X(t) = 1$; $\lim_{t \rightarrow -\infty} F_X(t) = 0$;
2. è non decrescente;
3. è continua a destra con limiti a sinistra.

La prima e la terza proprietà si dimostrano mediante alle proprietà di continuità della probabilità per famiglie di insiemi crescenti o decrescenti, che abbiamo visto come conseguenza della definizione di probabilità, mentre la seconda è dovuta alla proprietà di monotonia della probabilità.

Dimostrazione: (M)

1. Sia $\{t_n\}_n$ una qualsiasi successione di numeri reali che tende ad $+\infty$ per $n \rightarrow \infty$. Ponendo $A_n = \{X \leq t_n\}$, si ha che $A_n \subseteq A_{n+1}$, quindi applicando la continuità della probabilità si ha

$$\begin{aligned}\lim_{n \rightarrow \infty} P(A_n) &= P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} X^{-1}((-\infty, t_n])\right) = P\left(X^{-1}\left(\bigcup_{n=1}^{\infty} (-\infty, t_n]\right)\right) \\ &= P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.\end{aligned}$$

Valendo per ogni successione tendente ad infinito, la proprietà è dimostrata. Analogamente si dimostra il secondo limite scegliendo $\lim_{n \rightarrow +\infty} t_n = -\infty$.

2. Se $s \leq t$, $\{X \leq s\} \subseteq \{X \leq t\}$, da cui per la monotonia della probabilità

$$F_X(s) = P(X \leq s) \leq P(X \leq t) = F_X(t).$$

3. Sia $\{t_n\}_n$ una qualsiasi successione decrescente che tende a $t \in \mathbb{R}$ per $n \rightarrow \infty$. Ponendo $A_n = \{X \leq t_n\}$, si ha che $A_{n+1} \subseteq A_n$, quindi applicando la continuità della probabilità si ha

$$\begin{aligned}\lim_{n \rightarrow \infty} P(A_n) &= P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} X^{-1}((-\infty, t_n])\right) = P(X^{-1}\left(\bigcap_{n=1}^{\infty} (-\infty, t_n]\right)) \\ &= P(X^{-1}((-\infty, t])) = F_X(t).\end{aligned}$$

Valendo per ogni successione decrescente tendente a t , la continuità da destra è dimostrata. L'esistenza del limite sinistro è similmente dimostrata scegliendo $t_n \uparrow t$.

Osservazione 7.1.1. *Le tre proprietà della funzione di ripartizione implicano che F_X può avere al più una quantità numerabile di discontinuità di salto.*

Ogni volta che la funzione di ripartizione presenta un salto in un punto t_0 , la probabilità che $X = t_0$ è non nulla e pari a $F_X(t_0) - F_X(t_0-)$, in questo caso si dice che t_0 è un **atomo** per la legge di X ; nel caso in cui invece la funzione di ripartizione risulti continua su tutta la retta reale, allora si dice che essa è **non atomica**.

Nel caso delle v.a. discrete, la funzione di ripartizione risulterà una funzione costante a tratti ed in ogni salto si leggerà la probabilità che la v.a. assuma il corrispondente valore sull'asse delle x .

Definizione 7.1.2. *Se una funzione di ripartizione F_X è derivabile in \mathbb{R} , allora la sua derivata, f_X , è detta **funzione di densità** di X e gode delle seguenti due proprietà:*

1. $f_X(t) \geq 0$, per ogni $t \in \mathbb{R}$;
2. $\int_{\mathbb{R}} f_X(t) dt = 1$.

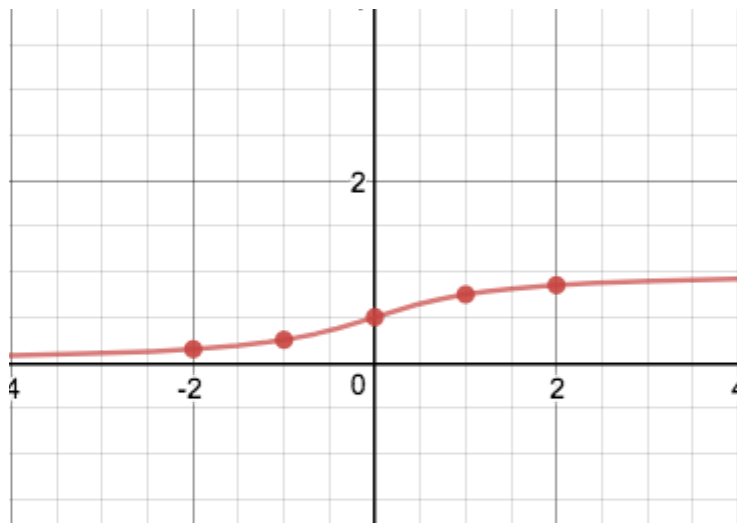


Figura 7.1: una funzione di ripartizione derivabile

Dal teorema fondamentale del calcolo integrale abbiamo dunque che

$$F_X(x) = \int_{-\infty}^x f_X(t)dt, \quad P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t)dt, \quad \forall x, a \leq b \in \mathbb{R}.$$

Il concetto di densità di probabilità è analogo al concetto di densità di massa per i corpi. Questa non è più concentrata in punti isolati, ma diffusa in tutto un sottoinsieme contenente degli intervalli. Come detto prima, nel caso di v.a. discrete la funzione di ripartizione diventa una funzione costante a tratti (la massa è concentrata in punti).

Esempio 7.1.1. Analogamente al caso discreto, possiamo definire una v.a. **uniforme** su un intervallo della retta reale $[a, b]$. In questo caso vogliamo che la probabilità sia concentrata su questo intervallo e che sia uguale su tutti gli elementi che compongono l'intervallo. Queste considerazioni portano a definire la densità di probabilità

$$f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

La funzione verifica le due proprietà che definiscono una funzione di densità e la sua antiderivata, ovvero la funzione di distribuzione, è (nel caso in cui $a = 0$ e $b = 1$)

$$F_X(x) = x \mathbf{1}_{[0,1)}(x) + \mathbf{1}_{[1,+\infty)}(x)$$

È sufficiente che una funzione sia non negativa ed integrabile per definire una densità in quanto la si può sempre rinormalizzare per rendere l'integrale finito pari ad 1, come mostriamo nel seguente esempio.

Esempio 7.1.2. Sia

$$f(x) = C(4x - 2x^2) \mathbf{1}_{(0,2)}(x), \quad C \in \mathbb{R}_+,$$

Determinare la costante C che rende f una densità.

Prima di tutto osserviamo che la funzione $f(x)$ vale $C(4x - 2x^2)$ nell'intervallo $(0, 2)$ e 0 altrimenti, quindi è non negativa, verificando la prima proprietà della funzione di densità.

Non rimane che imporre

$$\int_{\mathbb{R}} f(x) dx = 1,$$

che vuol dire che

$$C \int_0^2 (4x - 2x^2) dx = 1 \Rightarrow C \left(2x^2 - \frac{2}{3}x^3 \right) \Big|_0^2 \Rightarrow C \frac{8}{3} = 1 \Rightarrow C = \frac{3}{8}.$$

Abbiamo introdotto la distribuzione uniforme su un intervallo, vi sono molti altri esempi di distribuzioni continue, che presenteremo a breve.

Prima di farlo vogliamo estendere i concetti di indipendenza di v.a., di densità congiunta e condizionale introdotti precedentemente per il caso discreto.

Definizione 7.1.3. : Sia $\mathbf{X} = (X_1, \dots, X_k)$ un vettore aleatorio. Allora chiamiamo funzione di ripartizione congiunta di X_1, \dots, X_k

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= F_{X_1, \dots, X_k}(x_1, \dots, x_k) \\ &= P((X_1, \dots, X_k) \in (-\infty, x_1] \times \dots \times (-\infty, x_k]) = P(X_1 \leq x_1, \dots, X_k \leq x_k) \end{aligned}$$

per $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$.

Osserviamo che anche in questo caso

1. $F_{\mathbf{X}}$ è non decrescente in ogni variabile;
2. $F_{\mathbf{X}}$ è continua da destra nel senso che

$$\lim_{h \downarrow 0} F_{\mathbf{X}}(x_1 + h, \dots, x_k + h) = F_{\mathbf{X}}(x_1, \dots, x_k)$$

3. $F_{\mathbf{X}}(x_1, \dots, x_k) \rightarrow 1$ se ogni $x_i \rightarrow +\infty$ e $F_{\mathbf{X}}(x_1, \dots, x_k) \rightarrow 0$ se $x_i \rightarrow -\infty$ per qualche $i = 1, \dots, k$

Se la funzione di ripartizione congiunta risulta derivabile n volte allora possiamo definire la funzione di densità congiunta data da

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \frac{\partial^k}{\partial x_1 \dots \partial x_k} F_{X_1, \dots, X_k}(x_1, \dots, x_k)$$

Valgono proprietà analoghe al caso unidimensionale

1. $f_{X_1, \dots, X_k}(x_1, \dots, x_k) \geq 0$, per ogni $(x_1, \dots, x_k) \in \mathbb{R}^k$;
2. $\underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{k \text{ volte}} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_1 \dots dx_k = 1$.

Saturando un qualsiasi numero di variabili si ottengono le distribuzioni marginali, in particolare saturando tutte le variabili tranne una si ottengono le distribuzioni marginali unidimensionali.

$$f_{X_i}(x_i) = \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}_{k-1 \text{ volte}} f_{X_1, \dots, X_k}(x_1, \dots, x_k) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_k.$$

Esempio 7.1.3. Siano X, Y due v.a. con densità congiunta data da

$$f_{X,Y}(x, y) = \frac{3}{31}(6 - x^2 - xy)\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,2]}(y)$$

Calcolare le densità marginali f_X e f_Y .

Procediamo per saturazione. Per ogni $x \in [0, 1]$ (nel primo caso) e per $y \in [0, 2]$ (nel secondo) fissato si ha

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dy = \int_0^2 \frac{3}{31}(6 - x^2 - xy) dy \\ &= \frac{3}{31} \left[(6 - x^2)y - x \frac{y^2}{2} \right]_0^2 = \frac{6}{31}(6 - x^2 - x) \\ f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \int_0^1 \frac{3}{31}(6 - x^2 - xy) dx \\ &= \frac{3}{31} \left[6x - \frac{x^3}{3} - \frac{x^2}{2} y \right]_0^1 = \frac{17}{31} - \frac{3}{62}y. \end{aligned}$$

Trovare le marginali non è sempre così facile. Per calcolarle occorre eseguire integrali multipli, il cui calcolo dipende dal dominio di definizione della densità congiunta.

In quel che segue non supereremo mai la dimensione 2 nel calcolo esplicito degli integrali, ma quanto viene mostrato rimane valido anche per dimensioni maggiori.

Nel seguente esempio, mostriamo che anche nel caso di una funzione integranda relativamente semplice il calcolo delle marginali non risulta immediato a causa della forma del dominio di definizione.

Esempio 7.1.4. Supponiamo di avere una densità uniforme sulla regione del piano rappresentata dal semicerchio superiore con centro l'origine e raggio 1. In questo caso la funzione di densità è data da

$$f(x, y) = \frac{\mathbf{1}_A}{\text{area}(A)}$$

e vogliamo trovare le densità marginali.

Chiaramente $\text{area}(A) = \frac{\pi}{2}$, ma la regione individuata non è un rettangolo quindi dobbiamo decidere in che forma rappresentarla, ovvero quale delle due variabili vogliamo usare come variabile indipendente. In altri termini, ricordiamo che questo vuol dire passare ad una rappresentazione normale della regione rappresentata dal semicerchio

$$A = \{(x, y) \in \mathbb{R}^2 : y \geq 0, x^2 + y^2 \leq 1\}$$

Se scegliamo una rappresentazione normale rispetto alla variabile x , possiamo scrivere:

$$A = \{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, 0 \leq y \leq \sqrt{1-x^2}\},$$

oppure in rappresentazione normale rispetto alla variabile y abbiamo:

$$A = \{(x, y) \in \mathbb{R}^2 : 0 \leq y \leq 1, -\sqrt{1-y^2} \leq x \leq \sqrt{1-y^2}\}.$$

La prima rappresentazione viene usata quando vogliamo trovare la densità marginale della X , la seconda per la densità marginale della Y . Si ha allora

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x, y) dy = \int_0^{\sqrt{1-x^2}} \frac{\mathbf{1}_A}{\text{area}(A)} dy = \frac{2}{\pi} \sqrt{1-x^2}, & -1 \leq x \leq 1 \\ f_Y(y) &= \int_{\mathbb{R}} f(x, y) dx = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{\mathbf{1}_A}{\text{area}(A)} dx = \frac{4}{\pi} \sqrt{1-y^2}, & 0 \leq y \leq 1 \end{aligned}$$

Quando guardiamo delle v.a. congiuntamente, è senz'altro un'informazione preziosa sapere se rappresentano fenomeni indipendenti, abbiamo dunque bisogno della definizione

Definizione 7.1.4. Una famiglia finita di variabili aleatorie $\{X_i\}_{i=1}^n$ si dice **indipendente** se

$$(7.1) \quad P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n),$$

per ogni scelta di $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, dove con $\mathcal{B}(\mathbb{R})$ abbiamo indicato la σ -algebra dei Boreliani, ovvero la più piccola σ -algebra di \mathbb{R} che contiene gli intervalli aperti.

Qualora la famiglia fosse illimitata, per avere indipendenza è necessario che la condizione (7.1) sia verificata per ogni $n \in \mathbb{N}$.

Se prendiamo $B_1 = (-\infty, x_1], \dots, B_n = (-\infty, x_n]$, la precedente condizione diventa

$$(7.2) \quad P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n),$$

per ogni scelta di $x_1, \dots, x_n \in \mathbb{R}$, ovvero

$$(7.3) \quad F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n)$$

e nel caso esista la funzione di densità congiunta, per differenziazione lo stesso accade per le funzioni di densità

$$(7.4) \quad f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n).$$

Qualora la famiglia fosse illimitata, per avere indipendenza è necessario che la condizione (7.2) sia verificata per ogni $n \in \mathbb{N}$.

Osservazione 7.1.2. Le due condizioni di indipendenza (7.1) e (7.2) (e quindi (7.3)) sono totalmente equivalenti, poiché $\mathcal{B}(\mathbb{R})$ è generata dagli intervalli del tipo $(-\infty, x]$. Perciò, d'ora in poi useremo direttamente la condizione (7.3) o (7.4) se in presenza di densità.

Come nel caso discreto, se vale l'indipendenza, siamo in grado di ricostruire la funzione di ripartizione (o quella di densità) congiunta a partire dalle marginali.

Le due condizioni (7.3) e (7.4) diventano anche un criterio per stabilire l'indipendenza di due o più v.a. Se accade che la funzione di densità congiunta può essere scritta come

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n),$$

ovvero si può fattorizzare in funzioni separate ognuna (non necessariamente di densità) di una distinta variabile, allora le v.a. sono indipendenti, in quanto l'unica cosa che rimane da fare per individuare le marginali è calcolare le giuste costanti di normalizzazione per ogni marginale.

Esempio 7.1.5. *Le v.a. con funzioni di densità date nell'esempio 7.1.4 non sono indipendenti, infatti per $x, y \in A$ abbiamo*

$$f_{X,Y}(x, y) = \frac{2}{\pi} \neq \frac{2}{\pi} \sqrt{1-x^2} \frac{4}{\pi} \sqrt{1-y^2} = f_X(x) f_Y(y).$$

Viceversa consideriamo la funzione di densità congiunta data da

$$f(x, y) = \frac{32e^{\frac{9}{2}}}{75} xy^2 e^{-\frac{1}{2}x-4y} \mathbf{1}_{\{x \geq 1, y \geq 1\}}$$

possiamo immediatamente dire che è la congiunta di due v.a. indipendenti in quanto scrivibile come $\frac{32e^{\frac{9}{2}}}{75} f_1(x) f_2(y)$, dove abbiamo scelto

$$f_1(x) = x e^{-\frac{1}{2}x} \mathbf{1}_{\{x \geq 1\}}, \quad f_2(y) = y^2 e^{-4y} \mathbf{1}_{\{y \geq 1\}},$$

quindi integrando

$$\int_1^{+\infty} x e^{-\frac{1}{2}x} dx = \frac{6}{\sqrt{e}},$$

identifichiamo più precisamente le marginali

$$f_X(x) = \frac{\sqrt{e}}{6} x e^{-\frac{1}{2}x} \mathbf{1}_{\{x \geq 1\}}, \quad f_Y(y) = \frac{64e^4}{25} y^2 e^{-4y} \mathbf{1}_{\{y \geq 1\}}.$$

Osservazione 7.1.3. *Dalla definizione di indipendenza discende immediatamente che se X_1, \dots, X_n sono delle v.a. indipendenti e $\phi_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, n$ delle funzioni (misurabili), allora anche le v.a. definite da $Y_i = \phi_i(X_i)$ sono indipendenti.*

Lo dimostriamo per $n = 2$: presi $y_1, y_2 \in \mathbb{R}$ arbitrari fissati, si ha

$$\begin{aligned} P(Y_1 \leq y_1, Y_2 \leq y_2) &= P(\phi_1(X_1) \leq y_1, \phi_2(X_2) \leq y_2) \\ &= P(X_1 \in \phi_1^{-1}((-\infty, y_1]), X_2 \in \phi_2^{-1}((-\infty, y_2])) \\ &= P(X_1 \in \phi_1^{-1}((-\infty, y_1])) P(X_2 \in \phi_2^{-1}((-\infty, y_2])) \\ &= P(\phi_1(X_1) \leq y_1) P(\phi_2(X_2) \leq y_2) = P(Y_1 \leq y_1) P(Y_2 \leq y_2). \end{aligned}$$

Quando non si è in presenza di v.a. indipendenti, si può ricorrere al condizionamento. Se due v.a. X ed Y sono dotate di densità congiunta $f_{X,Y}$ e hanno rispettive densità marginali f_X e f_Y , analogamente a quanto fatto nel caso delle v.a. discrete, possiamo definire la densità di Y condizionata ad X (o viceversa) come

$$(7.5) \quad f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \text{dove } f_X(x) \neq 0$$

da cui si ottiene

$$(7.6) \quad f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x).$$

Esempi 7.1.1.

1. Prendiamo un v.a. bidimensionale (X,Y) distribuita uniformemente nel triangolo Δ del piano cartesiano di vertici $(0,0)$, $(1,0)$ e $(1,2)$, ovvero

$$f_{X,Y}(x,y) = \begin{cases} 1 & (x,y) \in \Delta \\ 0 & (x,y) \in \Delta^c. \end{cases}$$

Qual è la distribuzione di Y dato $X = \frac{1}{3}$?

Sappiamo che se abbiamo le funzioni di densità, allora

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \mathbf{1}_{\{f_X(x) > 0\}}$$

e quindi nel nostro caso si ottiene per $x \in [0,1]$

$$\begin{aligned} f_X(x) &= \int_0^{2x} dy = 2x \\ f_{Y|X}(y|x) &= \frac{1}{2x} \mathbf{1}_{\{0 \leq y \leq 2x\}} \end{aligned}$$

distribuzione uniforme su $[0,2x]$, dove $x \in (0,1]$ è fissato. Quindi per rispondere alla domanda basta sostituire $x = \frac{1}{3}$ nella formula.

2. Sia R una v.a. distribuita uniformemente nell'intervallo $[0,4]$. Dato $R = r$, la presenza di un animale in un'area è rappresentata da una v.a. $\mathbf{X} = (X_1, X_2)$ distribuita uniformemente nel cerchio di centro 0 e raggio r . Calcolare la densità di \mathbf{X} .

Soluzione: Abbiamo

$$f_R(r) = \frac{1}{4} \mathbf{1}_{[0,4]}(r), \quad f_{\mathbf{X}|R}(\mathbf{x}|r) = \frac{1}{\pi r^2} \mathbf{1}_{\{x_1^2 + x_2^2 \leq r^2\}}$$

da cui

$$f_{\mathbf{X},R}(\mathbf{x},r) = f_{\mathbf{X}|R}(\mathbf{x}|r)f_R(r) = \frac{1}{\pi r^2} \mathbf{1}_{\{x_1^2 + x_2^2 \leq r^2\}} \frac{1}{4} \mathbf{1}_{[0,4]}(r)$$

e, saturando rispetto ad r , si ottiene per $x_1^2 + x_2^2 \leq 4$

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_{\mathbb{R}} \frac{1}{\pi r^2} \mathbf{1}_{\{x_1^2 + x_2^2 \leq r^2\}} \frac{1}{4} \mathbf{1}_{[0,4]}(r) dr = \frac{1}{4\pi} \int_{x_1^2 + x_2^2}^4 \frac{1}{r^2} dr \\ &= -\frac{1}{4\pi} \frac{1}{r} \Big|_{x_1^2 + x_2^2}^4 = -\frac{1}{4\pi(x_1^2 + x_2^2)} - \frac{1}{16\pi} \end{aligned}$$

7.2 Trasformazioni di variabili aleatorie

Come nel caso delle v.a. discrete, possiamo generare nuove v.a. applicando delle funzioni alle v.a. Queste nuove v.a. danno luogo a diverse funzioni di ripartizione che possono essere indentificate. Infatti se X è una v.a. e $\phi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione, allora la v.a. $Y = \phi(X)$ ha funzione di distribuzione data

$$F_Y(y) = P(Y \leq y) = P(\phi(X) \leq y) = P(X \in \phi^{-1}((-\infty, y])).$$

Se ϕ è una funzione invertibile, per esempio consideriamo una funzione strettamente crescente, possiamo esplicitare la precedente formula ulteriormente ed ottenere

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(\phi(X) \leq y) = P(X \in \phi^{-1}((-\infty, y])) \\ &= P(X \in (\phi^{-1}(-\infty), \phi^{-1}(y))) = F_X(\phi^{-1}(y)). \end{aligned}$$

Se inoltre sia F_X che ϕ sono derivabili, dal teorema della derivata della funzione composta si ottiene che la precedente relazione si traduce direttamente per le funzioni di densità nella seguente

$$f_Y(y) = \frac{dF_Y}{dy}(\phi^{-1}(y)) = f_X(\phi^{-1}(y)) \frac{d}{dy}(\phi^{-1}(y)) = f_X(\phi^{-1}(y)) \frac{1}{\phi'(\phi^{-1}(y))}.$$

Esempio 7.2.1. Supponiamo di avere una v.a. X distribuita in maniera uniforme sull'intervallo $(0, 1]$ e definiamo $Y = -\ln X$. Allora la v.a. Y ha valori nell'intervallo $[0, +\infty)$. Inoltre la sua funzione di ripartizione è data da, per $y \geq 0$

$$F_Y(y) = P(Y \leq y) = P(-\ln X \leq y) = P(\ln X \geq -y) = P(X \geq e^{-y}) = 1 - e^{-y}$$

da cui segue, visto che la funzione risulta derivabile, che la sua funzione di densità è

$$f_Y(y) = e^{-y} \mathbf{1}_{[0, +\infty)}(y).$$

Una qualsiasi v.a. Y che abbia funzione di densità di probabilità del tipo

$$f_Y(y) = \lambda e^{-\lambda y} \mathbf{1}_{[0, +\infty)}(y), \quad \lambda > 0$$

è detta v.a. **esponenziale di parametro** λ (e si scrive $X \sim \exp(\lambda)$) ed è spesso usata per descrivere il tempo di vita di un macchinario o un tempo di attesa. Quindi la v.a. precedentemente determinata è un'esponenziale di parametro 1.

Un secondo esempio importante viene dalle trasformazioni lineari

Esempio 7.2.2. Sia X una v.a. con densità f_X e poniamo $Y = aX + b$, $a \neq 0, b \in \mathbb{R}$, abbiamo allora

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = \begin{cases} P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a}), & a > 0 \\ P(X \geq \frac{y-b}{a}) = 1 - F_X(\frac{y-b}{a}), & a < 0 \end{cases}$$

da cui derivando

$$f_Y(y) = \begin{cases} \frac{1}{a} f_X\left(\frac{y-b}{a}\right), & a > 0 \\ -\frac{1}{a} f_X\left(\frac{y-b}{a}\right), & a < 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Non abbiamo necessariamente bisogno di avere funzioni ϕ totalmente invertibili per poter applicare la trasformazione di v.a., basta che siano invertibili a tratti.

Esempio 7.2.3. Supponiamo che X sia una v.a. dotata di densità f_X e consideriamo $Y = X^2$ e ne vogliamo determinare la densità. Sempre partendo dalla funzione di ripartizione, si ha per $y \geq 0$ (un quadrato prende solo valori non negativi)

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

da cui derivando si ottiene

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})).$$

La precedente formula deve però tenere anche conto di dove la v.a. assume valori con probabilità positiva. Per esempio, se abbiamo una v.a. $X \sim \exp(\lambda)$ e consideriamo $Y = X^4$, ci rendiamo conto che, poiché X è concentrata su valori reali non negativi, si ha

$$F_Y(y) = P(Y \leq y) = P(0 \leq X^4 \leq y) = P(0 \leq X \leq \sqrt[4]{y}) = F_X(\sqrt[4]{y}),$$

da cui

$$f_Y(y) = \frac{1}{4} y^{\frac{1}{4}-1} \lambda e^{-\lambda \sqrt[4]{y}}.$$

Analogamente possiamo dedurre la formula per la densità di una v.a. trasformazione di un vettore di v.a. Lo facciamo vedere nel caso di due v.a.

Siano X ed Y due v.a. dotate di densità congiunta $f_{X,Y}(x,y)$ e sia $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ una funzione vettoriale invertibile e differenziabile con continuità (diffeomorfismo). Sia $(Z, W) = \phi(X, Y) = (\phi_1(X, Y), \phi_2(X, Y))$, allora

$$\begin{aligned} P(Z \leq z, W \leq w) &= P\left((X, Y) \in \phi^{-1}((-\infty, z] \times (-\infty, w])\right) \\ &= \int \int_{\phi^{-1}((-\infty, z] \times (-\infty, w])} f_{X,Y}(x,y) dx dy \\ &= \int_{-\infty}^z \int_{-\infty}^w f_{X,Y}(\phi^{-1}(u,v)) |\det J_{\phi^{-1}}(u,v)| du dv, \end{aligned}$$

dove abbiamo indicato con J_{ϕ} la matrice Jacobiana

$$J_{\phi} = \begin{pmatrix} \frac{\partial \phi_1}{\partial x} & \frac{\partial \phi_1}{\partial y} \\ \frac{\partial \phi_2}{\partial x} & \frac{\partial \phi_2}{\partial y} \end{pmatrix}.$$

Allora si ha che il vettore $(Z, W) = \phi(X, Y)$ avrà densità

$$(7.7) \quad f_{Z,W}(z, w) = \frac{1}{|\det J_\phi(\phi^{-1}(z, w))|} f_{X,Y}(\phi^{-1}(z, w))$$

Esempio 7.2.4. *Trasformazione di Box Müller (rivisitata)*

Sia (R, Θ) un vettore aleatorio bidimensionale distribuito uniformemente sul rettangolo $(0, 1] \times [0, 2\pi)$ secondo la densità data da

$$f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} \mathbf{1}_{[0,1]}(r) \mathbf{1}_{(0,2\pi)}(\theta).$$

Consideriamo ora le v.a. definite dal seguente diffeomorfismo $(r, \theta) \rightarrow (z, w)$ sui due sottoinsiemi $(0, 1] \times [0, \pi]$ e $(0, 1] \times [\pi, 2\pi)$

$$(z, w) = \phi(r, \theta) = (\cos \theta \sqrt{-2 \ln r}, \sin \theta \sqrt{-2 \ln r}),$$

da cui si ottiene la funzione inversa

$$\phi^{-1}(z, w) = (e^{-\frac{z^2+w^2}{2}}, \operatorname{arccotan}(\frac{z}{w})).$$

Lo Jacobiano è dato da

$$J_\phi = \begin{pmatrix} \cos \theta \frac{-2}{2r\sqrt{-2 \ln r}} & -\sin \theta \sqrt{-2 \ln r} \\ \sin \theta \frac{-2}{2r\sqrt{-2 \ln r}} & \cos \theta \sqrt{-2 \ln r} \end{pmatrix},$$

che permette di calcolare

$$|\det J_\phi(\phi^{-1}(z, w))| = |-\frac{1}{r} \cos^2 \theta - \frac{1}{r} \sin^2 \theta| = \frac{1}{r} = \frac{1}{e^{-\frac{z^2+w^2}{2}}}.$$

Dalla formula (7.7) otteniamo la densità di Z, W

$$\begin{aligned} f_{Z,W}(z, w) &= f_{R,\Theta}(\phi^{-1}(z, w)) = \frac{1}{2\pi} \mathbf{1}_{[0,1]}(r) \mathbf{1}_{(0,2\pi)}(\theta) r \\ &= \frac{e^{-\frac{z^2+w^2}{2}}}{2\pi} \mathbf{1}_{[0,1]}(e^{-\frac{z^2+w^2}{2}}) \{ \mathbf{1}_{(0,\pi]}(\operatorname{arccotan}(\frac{z}{w})) + \mathbf{1}_{[\pi,2\pi)}(\operatorname{arccotan}(\frac{z}{w})) \} \\ &= \frac{1}{2\pi} e^{-\frac{z^2+w^2}{2}} \mathbf{1}_{\mathbb{R}}(z) \{ \mathbf{1}_{(0,+\infty)}(w) + \mathbf{1}_{(-\infty,0)}(w) \} \\ &= \frac{1}{2\pi} e^{-\frac{z^2+w^2}{2}}, \quad z, w \in \mathbb{R} \end{aligned}$$

e poiché questa è una densità congiunta abbiamo

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{Z,W}(z, w) dz dw = 1 \Rightarrow \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{2\pi} e^{-\frac{z^2+w^2}{2}} dz dw = 1,$$

da cui discende anche

$$1 = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{2\pi} e^{-\frac{z^2+w^2}{2}} dz dw = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{z^2}{2}} dz \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{w^2}{2}} dw = \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{z^2}{2}} dz \right)^2,$$

ovvero

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{z^2}{2}} dz = 1.$$

Diciamo che Z ha densità Normale o Gaussiana standard e si scrive che $Z \sim \mathcal{N}(0, 1)$ se

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}.$$

Riprendendo l'esempio 7.2.2, se abbiamo $X \sim \mathcal{N}(0; 1)$ e $Y = \sigma X + \mu$ con $\sigma > 0$ e $\mu \in \mathbb{R}$, allora

$$f_Y(y) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

e diciamo che $Y \sim \mathcal{N}(\mu, \sigma^2)$ segue una Gaussiana (o Normale) con parametri μ e σ^2 .

7.3 Media e momenti

In questa sezione estendiamo i concetti di media e momenti già esposti per le v.a. discrete.

Definizione 7.3.1. Sia X una v.a. con densità f_X , definiamo **media o aspettazione matematica** di X la quantità

$$(7.8) \quad \mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

sempre che la quantità

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx$$

risulti finita, ovvero che la funzione $x f_X(x)$ risulti integrabile.

Essendo un integrale, le stesse proprietà di monotonia e linearità e disuguaglianza triangolare della media continuano a valere e analogamente a prima vale

Proposizione 7.3.1. Siano X una v.a. con densità f_X e $\phi : \mathbb{R} \rightarrow \mathbb{R}$ una funzione reale tali che

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

allora

$$\mathbb{E}(\phi(X)) = \int_{-\infty}^{+\infty} \phi(x) f_X(x) dx$$

Questa proposizione vale anche se $\mathbf{X} = (X_1, \dots, X_n)$ è una v.a. vettoriale, usando la densità congiunta del vettore. Se \mathbf{X} ha densità congiunta $f_{\mathbf{X}}(x_1, \dots, x_n)$ e $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ sono tali che

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} |\phi(x_1, \dots, x_n)| f_{\mathbf{X}}(x_1, \dots, x_n) dx_1, \dots, dx_n < +\infty$$

allora

$$\mathbb{E}(\phi(X_1, \dots, X_n)) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \phi(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1, \dots, dx_n.$$

Applicando la proposizione 7.3.1 a $\phi(x) = |x|$ e $\phi(x) = x^k$, $\phi(x) = (x - \mathbb{E}(X))^k$ possiamo dire che la media esiste qualora $\mathbb{E}(|X|) < +\infty$ e possiamo definire i momenti di ordine k centrati e non. Analogamente, nel caso multidimensionale la proposizione può essere usata per definire la media della somma o del prodotto, del massimo o del minimo di più variabili aleatorie.

Definizione 7.3.2. Siano $k \in \mathbb{N}$ e X una v.a. tale che $\mathbb{E}(|X|^k) < \infty$, allora diciamo **momento di ordine k di X** la quantità $\mathbb{E}(X^k)$ e **momento centrato di ordine k** la quantità $\mathbb{E}[(X - \mathbb{E}(X))^k]$.

Ovviamente il momento centrato di ordine 2 è la varianza che verifica le proprietà elencate prima. Notiamo infatti che erano state dimostrate a prescindere dalla forma della densità.

Anche in questo contesto, se X ed Y sono v.a. dotate di densità congiunta $f_{X,Y}$, usando la versione multidimensionale della proposizione 7.3.1 si possono definire la covarianza e il coefficiente di correlazione

$$\text{cov}(X, Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) = \int_{\mathbb{R}} \int_{\mathbb{R}} [x - \mathbb{E}(X)][y - \mathbb{E}(Y)] f_{X,Y}(x, y) dx dy.$$

Di nuovo tutte le proprietà illustrate precedentemente rimangono valide, compresa la formulazione alternativa

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \int_{\mathbb{R}} \int_{\mathbb{R}} xy f_{X,Y}(x, y) dx dy - \mathbb{E}(X)\mathbb{E}(Y)$$

ed il fatto che due v.a. indipendenti sono anche scorrelate come si vede dalla seguente

Proposizione 7.3.2. Siano X ed Y due v.a. indipendenti dotate di media, allora anche XY è dotata di media e si ha

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

Dimostrazione: Infatti si ha

$$\begin{aligned} \mathbb{E}(|XY|) &= \int_{\mathbb{R}} \int_{\mathbb{R}} |xy| f_{X,Y}(x, y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} |x| |y| f_X(x) f_Y(y) dx dy = \int_{\mathbb{R}} |y| f_Y(y) \int_{\mathbb{R}} |x| f_X(x) dx dy \\ &= \int_{\mathbb{R}} |x| f_X(x) dx \int_{\mathbb{R}} |y| f_Y(y) dy = \mathbb{E}(|X|)\mathbb{E}(|Y|). \end{aligned}$$

Abbiamo così verificato l'integrabilità di XY e si procede analogamente per il prodotto senza valore assoluto. \square

7.4 Le principali densità assolutamente continue e i loro momenti

7.4.1 La densità esponenziale

Ricordiamo che diciamo che una v.a. $T \sim \exp(\lambda)$, con $\lambda > 0$ se

$$f_T(t) = \lambda e^{-\lambda t} \mathbf{1}_{(0,+\infty)}(t).$$

Integrando si ottiene immediatamente che la sua funzione di ripartizione è data da

$$(7.9) \quad P(T \leq t) = F_T(t) = \int_{-\infty}^t \lambda e^{-\lambda s} \mathbf{1}_{(0,+\infty)}(s) ds = (1 - e^{-\lambda t}) \mathbf{1}_{(0,+\infty)}(t).$$

La densità esponenziale viene spesso usata per descrivere un tempo di sopravvivenza poiché è concentrata sui valori reali positivi e la probabilità che sia molto grande è esponenzialmente piccola, infatti la probabilità complementare di (7.9) è

$$P(T > t) = e^{-\lambda t}.$$

Inoltre, come la geometrica, esibisce la proprietà di perdita della memoria

$$\begin{aligned} P(T > s+t | T > t) &= \frac{P(T > s+t, T > t)}{P(T > t)} = \frac{P(T > s+t)}{P(T > t)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s), \end{aligned}$$

ovvero se sappiamo che al tempo t il nostro macchinario è ancora in funzione, la probabilità che continui a funzionare per almeno un ulteriore tempo s è equivalente alla probabilità che duri almeno un tempo s a partire dall'istante iniziale. In un certo senso, condizionatamente al fatto che si è ancora in funzione, è come se l'orologio che misura il tempo di sopravvivenza si azzerasse. È chiaro che usare una v.a. esponenziale per modellizzare il tempo di vita di un macchinario che subisce usura non è completamente realistico, vedremo in seguito che esistono altre densità che sono concentrate sulla semiretta reale positiva e che quindi possono essere utilizzate per modellizzare dei tempi di vita.

La media e la varianza di una distribuzione esponenziale sono facilmente calcolabili.

$$\begin{aligned} \mathbb{E}(X) &= \int_0^\infty x \lambda e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = 0 + \frac{1}{\lambda} \int_0^\infty \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \\ \text{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} \\ &= -x^2 e^{-\lambda x} \Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x} dx = 0 + \frac{1}{\lambda} \int_0^\infty 2x \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \end{aligned}$$

Osserviamo che anche questa volta il parametro che caratterizza la distribuzione è collegato alla media della v.a. (ed anche alla varianza).

7.4.2 La densità Gaussiana

Diciamo che una v.a. X segue una distribuzione Gaussiana o Normale Standard e scriviamo $X \sim \mathcal{N}(0, 1)$, se la sua densità è data da

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Essendo una densità, come abbiamo visto nell'esempio 7.2.4, sappiamo che

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 \quad \text{e} \quad \forall a < b \in \mathbb{R}, \quad P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

L'ultimo integrale è però calcolabile esplicitamente solo su alcuni intervalli specifici, che implica non abbiamo una forma esplicita della funzione di ripartizione della Gaussiana standard, usualmente indicata con $\phi(x)$

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \quad x \in \mathbb{R}.$$

Questa funzione è stata tabulata al variare dei valori di x tra -3 e 3 ad intervalli di 0,01, calcolando gli integrali attraverso un'approssimazione numerica. Nella tabella della funzione di ripartizione della Normale standard, riportata in qualsiasi libro di Probabilità o di Statistica, in colonna troviamo il numero x fino alla prima cifra decimale ed in riga la seconda cifra decimale, all'incrocio si trova il valore $\phi(x)$.

Di solito non vengono tabulati valori oltre il 3 ed il -3 in quanto l'area sottostante la curva della densità in questo intervallo rappresenta il 99% della probabilità.

Comunque alcune considerazioni per valori sulla funzione di distribuzione si possono fare. La funzione di densità è simmetrica rispetto all'asse delle y e quindi così è l'area sottesa alla curva. Di conseguenza abbiamo

$$\phi(0) = \frac{1}{2}, \quad \phi(-x) = 1 - \phi(x)$$

e questo permette di calcolare i valori di ϕ soltanto per valori positivi dell'argomento e dedurre gli altri dalle identità precedenti.

Possiamo calcolare la media e la varianza di una Normale standard abbastanza facilmente, infatti abbiamo

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

essendo la funzione integranda una funzione dispari. Di conseguenza la varianza può essere calcolata per parità e per parti

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= -2x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_0^{+\infty} + 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \frac{1}{2} = 1. \end{aligned}$$

Abbiamo quindi visto che lo 0 e l'1 che caratterizzano la distribuzione Gaussiana standard corrispondono alla media ed alla varianza della stessa.

Una distribuzione Gaussiana standard può essere adatta a rappresentare un errore che si commette in una misurazione, ma che in media è zero, ovvero che si commette per eccesso o difetto con eguale probabilità. Inoltre si richiede che questo errore possa essere di ampiezza unitaria con alta probabilità, mentre può essere grande con probabilità esponenzialmente piccola.

Si potrebbero però avere fenomeni dove è necessario modellizzare un errore che in media non è nullo e che in valore assoluto può avvenire con probabilità alta in intervalli diversi da quello unitario. In realtà le distribuzioni Gaussiane rappresentano tutta una famiglia di distribuzioni e offrono tutta la flessibilità desiderabile per modellizzare qualsiasi tipo di errore che abbia un decadimento esponenziale rispetto al suo valore assoluto.

Consideriamo infatti la seguente trasformazione lineare $Y = \sigma X + \mu$ dove $X \sim \mathcal{N}(0, 1)$ e $\sigma \neq 0$, allora la densità di Y è data da

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

che si riduce ad una Gaussiana standard nel caso $\mu = 0, \sigma^2 = 1$. Chiamiamo questa densità Gaussiana di media μ e varianza σ^2 e si scrive $Y \sim \mathcal{N}(\mu, \sigma^2)$ ed in effetti abbiamo

$$\mathbb{E}(Y) = \mathbb{E}(\sigma X + \mu) = \sigma \cdot 0 + \mu = \mu, \quad \text{Var}(Y) = \text{Var}(\sigma X) = \sigma^2 \text{Var}(X) = \sigma^2.$$

Viceversa possiamo concludere che se $Y \sim \mathcal{N}(\mu, \sigma^2)$ allora $X = \frac{Y - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Esempio 7.4.1. *La temperatura in gradi di un certo ambiente è una v.a. $T \sim \mathcal{N}(30, 4)$. È importante che questo ambiente rimanga ad una temperatura stabile per consentire la riproduzione di una certa colonia di batteri che muore a temperature maggiori di 34 gradi. Vogliamo quindi calcolare $P(T > 34)$.*

Soluzione. *Le uniche tavole disponibili sono quelle della Gaussiana standard e quindi dobbiamo standardizzare la nostra v.a. T per poterle usare. D'altra parte sappiamo che $\mathbb{E}(T) = 30$ e $\text{Var}(T) = 4$, quindi*

$$P(T > 34) = P\left(\frac{T - 30}{2} > \frac{34 - 30}{2}\right) = P(\mathcal{N}(0, 1) > 2) = 1 - \phi(2) = 1 - 0.9772 = 0.0228.$$

In generale diciamo che **standardizziamo** una v.a. X dotata di momento secondo quando consideriamo la sua trasformata lineare

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}},$$

poiché si nota subito che $\mathbb{E}(Y) = 0$ e $\text{Var}(Y) = 1$ e diciamo che Y è la **standardizzata** di X .

7.4.3 La densità Gamma

Introduciamo un'altra famiglia di densità di probabilità che presenta particolare flessibilità per modellizzare moltissimi fenomeni che prendono valori positivi.

Diciamo che una v.a. X segue una distribuzione $\Gamma(\alpha, \lambda)$ per α e $\lambda > 0$, se la sua funzione di densità è data da

$$f_X(x) = \frac{1}{C} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{\{x>0\}},$$

dove C è un'opportuna costante, che stabiliremo in seguito, affinché $\int_{\mathbb{R}} f_X(x) dx = 1$.

Anche questa densità è concentrata sulla semiretta reale positiva e quindi è adatta a modellizzare, per esempio, fenomeni temporali. A seconda della scelta dei parametri α e λ , la forma della funzione di densità può cambiare notevolmente permettendo di adattarsi a varie situazioni.

Per calcolare la costante C basta risolvere l'uguaglianza

$$\frac{1}{C} \int_0^{+\infty} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = 1,$$

ovvero dobbiamo calcolare l'integrale. Applicando la sostituzione $y = \lambda x$ ci rendiamo conto che questo integrale e conseguentemente la costante C dipendono solo da α , infatti tenendo conto che $dy = \lambda dx$ abbiamo

$$\int_0^{+\infty} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = \int_0^{+\infty} y^{\alpha-1} e^{-y} dy.$$

Questo integrale, che denotiamo con $\Gamma(\alpha)$, non è calcolabile con metodi elementari per qualsiasi valore di α , ma solo numericamente. Scriviamo quindi la densità come

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{\{x>0\}}.$$

L'integrale definisce la cosiddetta **funzione gamma**, che gode di alcune proprietà e può essere calcolata esplicitamente per alcuni valori particolari di α .

1. La funzione $\Gamma(\alpha)$ è ben definita per qualsiasi valore $\alpha > 0$, infatti

$$\int_0^{+\infty} y^{\alpha-1} e^{-y} dy \leq \int_0^1 y^{\alpha-1} e^{-y} dy + \int_1^{+\infty} y^{\alpha-1} e^{-y} dy \leq \int_0^1 y^{\alpha-1} dy + \int_1^{+\infty} y^\alpha e^{-y} dy < +\infty.$$

2. Notiamo che

$$\Gamma(1) = \int_0^{+\infty} e^{-y} dy = -e^{-y} \Big|_0^{+\infty} = 1.$$

3. Per $\alpha > 0$, vale la seguente proprietà iterativa, grazie all'integrazione per parti

$$\begin{aligned} \Gamma(\alpha + 1) &= \int_0^{+\infty} y^\alpha e^{-y} dy = -e^{-y} y^\alpha \Big|_0^{+\infty} + (\alpha - 1) \int_0^{+\infty} y^{\alpha-1} e^{-y} dy \\ &= 0 + \alpha \Gamma(\alpha) \end{aligned}$$

4. Conseguentemente, dai due punti precedenti abbiamo per $n \in \mathbb{N}$,

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \dots = n(n-1)(n-2) \dots 1\Gamma(1) = n!$$

Le densità gamma sono una famiglia di funzioni di densità piuttosto ampia e flessibile e comprendono alcune densità chiamate diversamente in letteratura.

1. Un'esponenziale di parametro λ può essere vista come una $\Gamma(1, \lambda)$.
2. Se abbiamo una v.a. $X \sim \mathcal{N}(0; \sigma^2)$ e consideriamo la densità di $Y = X^2$, applicando la formula per la densità del quadrato si ottiene per $y \geq 0$

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})] = 2\frac{1}{2}\frac{1}{\sqrt{2\pi\sigma^2y}}e^{-\frac{y}{2\sigma^2}} = \frac{1}{\sqrt{\pi}}\left(\frac{1}{2\sigma^2}\right)^{\frac{1}{2}}y^{\frac{1}{2}-1}e^{-\frac{1}{2\sigma^2}y},$$

ovvero una $\Gamma(\frac{1}{2}, \frac{1}{2\sigma^2})$. Essendo questa una densità, abbiamo necessariamente che $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ e per iterazione possiamo concludere per ogni multiplo dispari di $\frac{1}{2}$

$$\Gamma\left(\frac{2n+1}{2}\right) = \frac{2n-1}{2}\Gamma\left(\frac{2n-1}{2}\right) = \dots = \frac{2n-1}{2}\frac{2n-3}{2} \dots \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{(2n-1)!!}{2^n}\sqrt{\pi},$$

dove abbiamo definito

$$(2n-1)!! = (2n-1)(2n-3) \dots 3 \cdot 1, \quad (2n)!! = (2n)(2n-2) \dots 4 \cdot 2.$$

Osservazione 7.4.1. La precedente equivalenza è utile anche per calcolare i momenti di qualsiasi ordine di una Gaussiana standard.

Infatti se $X \sim \mathcal{N}(0; 1)$, per i momenti dispari sappiamo immediatamente che

$$\mathbb{E}(X^{2n-1}) = 0, \quad n \in \mathbb{N},$$

poiché la funzione di densità è pari e la potenza dispari, mentre per i momenti pari si ha, ponendo $Y = X^2$,

$$\begin{aligned} \mathbb{E}(X^{2n}) &= E(Y^n) = \int_0^{+\infty} \frac{1}{\sqrt{\pi}} y^n \left(\frac{1}{2}\right)^{\frac{1}{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y} dy = \frac{1}{\sqrt{\pi}} 2^n \int_0^{+\infty} \left(\frac{1}{2}\right)^{n+\frac{1}{2}} y^{n+\frac{1}{2}-1} e^{-\frac{1}{2}y} dy \\ &= \frac{1}{\sqrt{\pi}} 2^n \Gamma\left(\frac{2n+1}{2}\right) \int_0^{+\infty} \frac{1}{\Gamma\left(\frac{2n+1}{2}\right)} \left(\frac{1}{2}\right)^{n+\frac{1}{2}} y^{\frac{2n+1}{2}-1} e^{-\frac{1}{2}y} dy \\ &= \frac{1}{\sqrt{\pi}} 2^n \frac{(2n-1)!!}{2^n} \sqrt{\pi} = (2n-1)!! \end{aligned}$$

Riassumendo, per $n \in \mathbb{N}$ abbiamo una forma completamente esplicita per $X \sim \Gamma(n, \lambda)$

$$f_X(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} \mathbf{1}_{(0, +\infty)}(x)$$

e per $X \sim \Gamma(n + \frac{1}{2}, \lambda)$

$$f_X(x) = \frac{2^n \lambda^{n+\frac{1}{2}}}{(2n-1)!! \sqrt{\pi}} x^{n-\frac{1}{2}} e^{-\lambda x} \mathbf{1}_{(0,+\infty)}(x).$$

La densità $\Gamma(\frac{n}{2}, \frac{1}{2})$ ha una particolare importanza in statistica e prende il nome di **chi quadro con n gradi di libertà** e si scrive $X \sim \chi^2(n)$. In teoria dell'informazione, la distribuzione $\Gamma(n, \lambda)$ è invece nota come **distribuzione di Erlang a n stadi**.

Infine il calcolo generale dei momenti di una densità gamma non è difficile, grazie alla proprietà iterativa della funzione Γ .

Per $n = 1, 2, \dots$ si ha

$$\begin{aligned} \mathbb{E}(X^n) &= \int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{n+\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\lambda^n} \int_0^{+\infty} \frac{\lambda^{n+\alpha}}{\Gamma(n+\alpha)} x^{n+\alpha-1} e^{-\lambda x} dx \\ &= \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\lambda^n} \cdot 1 = \frac{(n-1+\alpha)(n-2+\alpha)\dots\alpha\Gamma(\alpha)}{\Gamma(\alpha)\lambda^n} = \frac{\overbrace{\alpha(\alpha+1)\dots(\alpha+n-1)}^{n \text{ volte}}}{\lambda^n}, \end{aligned}$$

da cui si deduce

$$\mathbb{E}(X) = \frac{\alpha}{\lambda}, \quad \mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2} \quad \Rightarrow \quad \text{Var}(X) = \frac{\alpha}{\lambda^2}$$

7.5 Applicazioni

1. Processo di Poisson.

In questa sezione vediamo un paio di applicazioni delle v.a. assolutamente continue. In particolare dalla prima si comprende come nasce la distribuzione di Poisson. Qui si chiama processo perché è una famiglia di v.a. che dipende anche dal tempo.

Supponiamo di avere un intervallo di tempo $[0, t]$ e di voler contare il numero di arrivi ad un servizio in questo intervallo di tempo. Indicheremo con N_t la v.a. che conta tale numero di arrivi. Chiaramente è a valori interi e quindi vogliamo determinare $P(N_t = n)$, per $n \in \mathbb{N}$. Per farlo abbiamo bisogno di modellizzare il fenomeno più dettagliatamente. Indichiamo con T_i i cosiddetti tempi di interarrivo, ovvero i tempi che intercorrono tra un arrivo ed un altro, da cui il tempo di n -mo arrivo è dato da $S_n = T_1 + \dots + T_n$. Quindi nell'intervallo $[0, t]$ il numero di arrivi è maggiore od uguale a n se e soltanto se vi sono stati almeno n arrivi in questo intervallo di tempo, cioè

$$\{N_t \geq n\} \iff \{T_1 + \dots + T_n \leq t\} = \{S_n \leq t\},$$

da cui calcolare $P(N_t \geq n)$ è equivalente a determinare la funzione di ripartizione di S_n .

In molti casi è ragionevole pensare che questi tempi di interarrivo siano indipendenti ed identicamente distribuiti; se inoltre supponiamo che siano esponenziali di parametro

λ , allora dalle proprietà della densità gamma, (che vedremo nel prossimo capitolo) sappiamo che $S_n \sim \Gamma(n, \lambda)$, da cui

$$\begin{aligned} P(N_t \geq n) &= P(S_n \leq t) = \int_0^t \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} dx \\ &= -\frac{\lambda^{n-1}}{(n-1)!} x^{n-1} e^{-\lambda x} \Big|_0^t + \int_0^t \frac{\lambda^{n-1}}{(n-2)!} x^{n-2} e^{-\lambda x} dx \\ &= -\frac{\lambda^{n-1}}{(n-1)!} t^{n-1} e^{-\lambda t} + \int_0^t \frac{\lambda^{n-1}}{(n-2)!} x^{n-2} e^{-\lambda x} dx \\ &= \dots = -\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} + 1 \end{aligned}$$

da cui

$$P(N_t = n) = P(N_t \geq n) - P(N_t \geq n+1) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

ovvero $N_t \sim \text{Poisson}(\lambda t)$. La famiglia di v.a. indicizzata su \mathbb{R}_+ , $\{N_t\}_{t \geq 0}$ è detta **processo di Poisson**.

Questo è un primo esempio di modello di formazione delle code, tali modelli sono necessari quando si intende pianificare un servizio (un server, uno sportello, un centralino). Notiamo che tale costruzione si è basata su tre ipotesi: indipendenza, identica distribuzione e densità esponenziale. Si può pervenire a modelli di code più generali di quello considerato rilassando una qualsiasi delle precedenti ipotesi.

2. Una seconda applicazione è data dai tempi di attesa. Si intende valutare la probabilità

$$P(T > t + h | T > t),$$

che un certo evento non accada per un dato lasso di tempo sapendo che non si è verificato fino ad oggi.

Nel caso della distribuzione esponenziale ciò si traduceva nella perdita di memoria ovvero il fenomeno agiva facendo ripartire l'orologio da zero. Chiaramente questo non è sempre il caso, infatti se T rappresenta il tempo di vita di un macchinario soggetto ad usura, il modello esponenziale non può più essere valido.

In generale supponiamo che T sia una v.a. positiva con funzione di densità f_T e funzione di ripartizione F_T , allora possiamo scrivere

$$P(T \leq t + h | T > t) = \frac{P(T \leq t + h, T > t)}{P(T > t)} = \frac{P(t < T \leq t + h)}{P(T > t)} = \frac{\int_t^{t+h} f_T(x) dx}{1 - F_T(t)}.$$

Derivando la precedente espressione rispetto ad h e chiamando tasso istantaneo di guasto la funzione

$$r(t) = \frac{f_T(t)}{1 - F_T(t)} \quad \Rightarrow \quad \frac{d}{dt} F_T(t) = r(t)(1 - F_T(t))$$

da cui si ottiene la seguente equazione differenziale per la funzione di sopravvivenza $G(t) := 1 - F_T(t)$

$$G'(t) = -r(t)G(t), \quad G(0) = 1 \Rightarrow G(t) = e^{-\int_0^t r(s)ds} \Rightarrow F_T(t) = 1 - e^{-\int_0^t r(s)ds}$$

e

$$(7.10) \quad f_T(t) = r(t)e^{-\int_0^t r(s)ds}.$$

La densità esponenziale ha un tasso istantaneo di guasto costante.

Quindi l'idea è quella di modellizzare il tasso istantaneo di guasto e di ricavare la densità compatibile con esso.

Per esempio supponiamo che il tasso istantaneo di guasto abbia un comportamento polinomiale

$$r(t) = \frac{\lambda}{\beta} t^{\beta-1}, \quad \lambda, \beta > 0$$

allora da (7.10) otteniamo che

$$f_T(t) = \lambda \beta t^{\beta-1} e^{-\lambda t^\beta}, \quad t > 0$$

detta **distribuzione di Weibull di parametri λ e β** e che si può ottenere attraverso la trasformazione $T = S^{\frac{1}{\beta}}$, dove $S \sim \exp(\lambda)$.

Per esercizio calcolare la media e la varianza della distribuzione di Weibull.

7.6 Revisione di Teoria

1. Dare la definizione di funzione di ripartizione.
2. Dare la definizione di v.a. scorrelate. Dimostrare che v.a. indipendenti sono scorrelate, ma che non è vero il viceversa.
3. Dimostrare la perdita di memoria della distribuzione esponenziale.
4. Sia X una v.a. con funzione di densità f_X , calcolare la funzione di densità di X^2 .
5. Dare la definizione di densità congiunta di due v.a. continue X ed Y .
6. Dare la definizione di densità condizionale di Y data X .
7. Calcolare le medie e le varianze delle principali distribuzioni.

7.7 Esercizi

1. Sia X una v.a. con funzione di distribuzione

$$F(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{50}t^2 & 0 \leq t \leq 5 \\ -\frac{1}{50}t^2 + \frac{2}{5}t - 1 & 5 \leq t \leq 10 \\ 1 & t \geq 10 \end{cases}$$

- (a) In quali intervalli è concentrata la X ?
(b) Dire se X possiede una densità ed in caso calcolarla.
2. Per $\lambda > 1$, consideriamo la funzione

$$f(x) = C \frac{1}{x^\lambda + 1} \mathbf{1}_{[1, +\infty)}(x)$$

- (a) Determinare C in modo tale che f sia una densità di probabilità.
(b) Calcolare la legge di $Y = \ln(X)$.
3. Se $X \sim \mathcal{N}(3; 4)$ che distribuzione segue $Y = 3X - 2$?
4. Se $X \sim \mathcal{N}(3; 3)$ calcolare $P(X \leq -1)$
5. Sia X una v.a. $\mathcal{N}(2, 1)$, calcolare la legge di $Y = e^X$.
6. Sia X una v.a. $\mathcal{N}(\mu, \sigma^2)$, calcolare la legge di $Y = X^2$.
7. Sia X una v.a. uniformemente distribuita su $[-1, 1]$. Determinare $P(|X| > \frac{1}{2})$ e la densità di probabilità di $|X|$.
8. Il tempo di funzionamento in anni di un computer è distribuito secondo un'esponenziale di parametro $\lambda = \frac{1}{4}$. Comprando un computer di questo tipo, usato ma funzionante, qual è la probabilità che continui a funzionare per i seguenti 4 anni?
9. Calcolare il coefficiente di correlazione di $X + Y$ e $X - Y$ dove $X \sim \Gamma(3, \frac{1}{4})$ e $Y \sim \mathcal{N}(-1; \frac{1}{9})$ sono indipendenti tra loro.
10. Calcolare la media di e^X , dove $X \sim \mathcal{N}(0, 1)$.
11. Sia $Y \sim \Gamma(\alpha, \lambda)$. Calcolare $\mathbb{E}(Y e^{-2Y})$.
12. Si consideri una v.a. X con densità

$$f(x) = 4xe^{-2x^2} \mathbf{1}_{\{x>0\}}.$$

Calcolare la media e la varianza di X .

13. Sia X una v.a. con densità di probabilità

$$f(x) = \frac{1}{4}xe^{-x/2}\mathbf{1}_{[0,+\infty)}(x).$$

Calcolare $\mathbb{E}(X)$ e $\text{Var}(X)$.

14. Il risultato di un test di ammissione nazionale è distribuito secondo una Gaussiana $\mathcal{N}(100, 100)$. I risultati degli studenti partecipanti sono indipendenti. Si scelgono 5 studenti a caso che hanno affrontato il test. Calcolare la probabilità che

- (a) i loro punteggi siano tutti inferiori a 80;
- (b) esattamente tre punteggi siano superiori a 84.

15. Siano $X \sim \mathcal{N}(1, 1)$ e $Y \sim \mathcal{N}(2, 2)$ tra loro indipendenti. Calcolare la retta di regressione lineare di $W = XY$ su $Z = X - Y$.

16. Il peso in grammi di un certo prodotto alimentare in uscita da una catena di confezionamento è distribuito secondo una Gaussiana $\mathcal{N}(500, 30)$.

Calcolare la probabilità che il peso di una confezione sia minore o uguale a 485 grammi.

Si prendono 10 pacchetti del prodotto, confezionati dalla stessa catena tutti indipendentemente. Calcolare la probabilità

- (a) che almeno uno dei pacchetti pesi meno di 485 grammi;
- (b) che al massimo due pacchetti pesino meno di 485 grammi.

17. Sia X una v.v. con densità di probabilità

$$f(x) = \frac{3}{4}(1 - x^2)\mathbf{1}_{[-1,1]}(x).$$

Calcolare $\mathbb{E}(X)$ e $\text{Var}(X)$.

Capitolo 8

VARIABILI ALEATORIE A PIÙ DIMENSIONI

In generale trattare v.a. a più dimensioni può diventare molto difficile. Sono però uno strumento importante per generare e simulare nuove distribuzioni che trovano poi riscontro, per esempio in Statistica, per rappresentare le distribuzioni di probabilità di dati raccolti da un qualche fenomeno fisico.

8.1 Alcuni esempi

Esempio 8.1.1. Sia (X, Y) una v.a. bidimensionale distribuita uniformemente sul cerchio unitario, C .

Questo vuol dire che la funzione di densità congiunta è data da

$$f_{X,Y}(x, y) = \frac{1}{\text{area } C} \mathbf{1}_C(x, y) = \frac{1}{\pi} \mathbf{1}_{\{x^2+y^2 \leq 1\}}$$

e ci interessa sapere quando la v.a. aleatoria Y supera di almeno 0.5 la v.a. X , cioè siamo interessati a

$$P(Y \geq X + \frac{1}{2}).$$

Dobbiamo quindi calcolare il seguente integrale

$$P(Y \geq X + \frac{1}{2}) = \int \int_{\{y \geq x + \frac{1}{2}\}} f_{X,Y}(x, y) dx dy = \frac{1}{\pi} \int \int_{\{y \geq x + \frac{1}{2}\} \cap \{x^2+y^2 \leq 1\}} dx dy$$

e dobbiamo riscrivere il dominio di integrazione in forma normale rispetto alla x o alla y .

RisolviAMO quindi il sistema per trovare gli estremi di integrazione per i due integrali

$$\begin{cases} x^2 + y^2 = 1 \\ y = x + 0.5 \end{cases} \Rightarrow \begin{cases} 2x^2 + x - \frac{3}{4} = 0 \\ y = x + 0.5 \end{cases} \Rightarrow \begin{cases} x = 0.41 \\ y = 0.91 \end{cases}, \quad \begin{cases} x = -0.91 \\ y = -0.41 \end{cases}.$$

Da cui

$$\begin{aligned} & \{y \geq x + \frac{1}{2}\} \cap \{x^2 + y^2 \leq 1\} \\ = & \{-1 \leq x \leq -0.91, -\sqrt{1-x^2} \leq y \leq \sqrt{1-x^2}\} \cup \{-0.91 \leq x \leq 0.41, x + 0.5 \leq y \leq \sqrt{1-x^2}\} \end{aligned}$$

e l'integrale diventa

$$\begin{aligned}
& \int \int_{\{y \geq x + \frac{1}{2}\} \cap \{x^2 + y^2 \leq 1\}} dx dy = \int_{-1}^{-0.91} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dx dy + \int_{-0.91}^{0.41} \int_{x+0.5}^{\sqrt{1-x^2}} dx dy \\
&= \int_{-1}^{-0.91} 2\sqrt{1-x^2} dx + \int_{-0.91}^{0.41} \left[\sqrt{1-x^2} - x - \frac{1}{2} \right] dx \\
&= \left[x\sqrt{1-x^2} + \arcsin x \right]_{-1}^{-0.91} + \frac{1}{2} \left[(x\sqrt{1-x^2} + \arcsin x) - x^2 - x \right]_{-0.91}^{0.41}.
\end{aligned}$$

In altre situazioni, pur conoscendo le densità congiunta, diventa molto difficile calcolare la densità di una qualche trasformazione delle v.a.

Esempio 8.1.2. Siano $X \sim \mathcal{N}(0, 1)$ e $Y \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ indipendenti e consideriamo la v.a. $Z = X - Y$, qual è la densità di Z ?

Per $z \in \mathbb{R}$ abbiamo che la funzione di ripartizione è

$$F_Z(z) = P(X - Y \leq z) = \int_0^{+\infty} \int_{-\infty}^{y+z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} dx dy,$$

che non sappiamo integrare esplicitamente.

L'ultimo esempio dimostra che determinare esplicitamente la funzione di distribuzione può essere molto laborioso. Viceversa, nella precedente potremmo derivare rispetto ad x ed y per avere un'espressione esplicita della densità. Ricordando che un integrale può essere calcolato numericamente, poter quindi dare un'espressione della densità congiunta diventa molto importante per le applicazioni.

Nei seguenti esempi, usando le trasformazioni bidimensionali introduciamo alcune densità importanti per la statistica

Esempi 8.1.1. (M)

1. Sia $X \sim \Gamma(\alpha, \lambda)$ e consideriamo la v.a. $Y = aX$ per qualche $a > 0$. Allora, come trasformazione lineare di X , sappiamo subito che per $y > 0$, si ha

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y}{a}\right) = \frac{1}{a} \frac{\lambda^\alpha}{\Gamma(\alpha)} \left(\frac{y}{a}\right)^{\alpha-1} e^{-\lambda \frac{y}{a}} = \left(\frac{\lambda}{a}\right)^\alpha \frac{1}{\Gamma(\alpha)} y^{\alpha-1} e^{-\frac{\lambda}{a}y} = \Gamma\left(\alpha, \frac{\lambda}{a}\right)$$

2. (Distribuzione di Fisher)

Utilizziamo il risultato al punto precedente per definire la Distribuzione di Fisher.

Osservazione 8.1.1. Siano $X \sim \Gamma(\alpha, \lambda)$ e $Y \sim \Gamma(\beta, \lambda)$ due v.a. indipendenti e applichiamo la trasformazione $\phi(x, y) = (x, \frac{x}{y})$, che ha inversa data da $\phi^{-1}(u, v) = (u, \frac{u}{v})$. Sfruttando la formula per il cambio di variabili (7.7), abbiamo

$$J_\phi = \begin{pmatrix} 1 & 0 \\ \frac{1}{y} & -\frac{x}{y^2} \end{pmatrix}.$$

e quindi $|\det J_\phi(\phi^{-1}(u, v))| = \frac{v^2}{u}$, da cui per $(U, V) = \phi(X, Y)$ si ha

$$\begin{aligned} f_{U,V} &= \frac{f_{X,Y}(\phi^{-1}(u, v))}{|\det J_\phi(\phi^{-1}(u, v))|} = \frac{\lambda^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\lambda u} \frac{\lambda^\beta}{\Gamma(\beta)} \left(\frac{u}{v}\right)^{\beta-1} e^{-\frac{\lambda}{v} u} \frac{u}{v^2} \mathbf{1}_{\{u>0\}} \mathbf{1}_{\{v>0\}} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{v^{\beta+1}} u^{\alpha+\beta-1} e^{-\lambda(1+\frac{1}{v})u} \mathbf{1}_{\{u>0\}} \mathbf{1}_{\{v>0\}} \end{aligned}$$

e possiamo calcolare la densità marginale della V per $v > 0$, data da

$$\begin{aligned} f_V(v) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{v^{\beta+1}} \left(1 + \frac{1}{v}\right)^{-(\alpha+\beta)} \int_0^{+\infty} \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} \left(1 + \frac{1}{v}\right)^{\alpha+\beta} u^{\alpha+\beta-1} e^{-\lambda(1+\frac{1}{v})u} du \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{v^{\alpha-1}}{(v+1)^{\alpha+\beta}}. \end{aligned}$$

Quando $X \sim \chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$ e $Y \sim \chi^2(m) = \Gamma(\frac{m}{2}, \frac{1}{2})$ then $\frac{X}{n} \sim \Gamma(\frac{n}{2}, \frac{n}{2})$ and $\frac{Y}{m} \sim \Gamma(\frac{m}{2}, \frac{m}{2})$ e la densità di $V = \frac{X/n}{Y/m}$ è chiamata una **densità di Fisher con n, m gradi di libertà** ed è esplicitamente ottenuta come

$$f_V(v) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \frac{v^{\frac{n}{2}-1}}{(v+1)^{\frac{n+m}{2}}}$$

e si scrive $\frac{X}{Y} \frac{m}{n} \sim F(n, m)$.

3. (*t di Student*) Siano $X \sim \mathcal{N}(0, 1)$ e $Y \sim \chi^2(n)$ due v.a. indipendenti e definiamo la trasformazione bidimensionale $\phi(x, y) = (x, \sqrt{n} \frac{x}{\sqrt{y}})$, che ha inversa data da $\phi^{-1}(u, v) = \left(u, n \left(\frac{u}{v}\right)^2\right)$, perciò lo Jacobiano è

$$J_\phi = \begin{pmatrix} 1 & 0 \\ \frac{1}{y} & -\frac{\sqrt{n}x}{2y^{\frac{3}{2}}} \end{pmatrix}.$$

e quindi $|\det J_\phi(\phi^{-1}(u, v))| = \frac{v^3}{2nu^2}$, da cui per $(U, V) = \phi(X, Y)$ si ha

$$\begin{aligned} f_{U,V} &= \frac{f_{X,Y}(u, n(\frac{u}{v})^2)}{\frac{v^3}{2nu^2}} = \frac{2nu^2}{v^3} f_X(u) f_Y\left(n \frac{u^2}{v^2}\right) \\ &= \frac{2nu^2}{v^3} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \left(n \frac{u^2}{v^2}\right)^{\frac{n}{2}-1} e^{-\frac{1}{2} n (\frac{u}{v})^2} \\ &= \frac{1}{2^{\frac{n}{2}-1} \Gamma(\frac{n}{2})} \frac{1}{v} \left(\frac{nu^2}{v^2}\right)^{\frac{n}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2} (1+\frac{n}{v^2})} \end{aligned}$$

e, come prima, possiamo calcolare la densità marginale della V per saturazione della u (per esercizio), ottenendo

$$\begin{aligned} f_V(v) &= \int_{-\infty}^{+\infty} \frac{1}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} \frac{1}{v} \left(\frac{nu^2}{v^2}\right)^{\frac{n}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}(1+\frac{n}{v^2})} du \\ &= \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{v^2}{n}\right)^{-\frac{n+1}{2}}, \quad v \in \mathbb{R} \end{aligned}$$

chiamata **distribuzione t di Student con n gradi di libertà** e si scrive $U \sim t(n)$.

Notiamo che anche la t di Student è una distribuzione simmetrica, come la Gaussiana standard, ma con un decadimento polinomiale anziché esponenziale per $|v| \rightarrow +\infty$.

Quindi i calcoli che coinvolgono la funzione di densità congiunta possono diventare abbastanza complicati. L'indipendenza aiuta ed a volte si riesce ad arrivare a formule esplicite. Nella prossima sezione analizziamo alcune specifiche trasformazioni bidimensionali che coinvolgono due v.a. continue ed indipendenti e vediamo di analizzare il caso particolare delle distribuzioni esponenziali, Gamma e Gaussiane. Le trattiamo soltanto nel caso bidimensionale per questioni di semplicità espositiva, ma si possono estendere facilmente a più dimensioni.

8.2 Trasformazioni bidimensionali: max, min e somma

Siano X ed Y due v.a. continue indipendenti con rispettive funzioni di densità f_X ed f_Y e funzioni di distribuzioni F_X e F_Y . consideriamo tre casi $\min(X, Y)$, $\max(X, Y)$, $X + Y$.

1. $Z = \min(X, Y)$. Invece di cercare subito la funzione di ripartizione di Z , in questo caso è più conveniente guardare la funzione di sopravvivenza $G(z) = P(Z > z)$, infatti abbiamo

$$\begin{aligned} P(Z > z) &= P(\min(X, Y) > z) = P(X > z, Y > z) = P(X > z)P(Y > z) \\ &= (1 - F_X(z))(1 - F_Y(z)), \end{aligned}$$

dove abbiamo applicato l'indipendenza delle v.a., da cui possiamo concludere

$$(8.1) \quad F_Z(z) = F_X(z) + F_Y(z) - F_X(z)F_Y(z)$$

e derivando si ottiene

$$(8.2) \quad f_Z(z) = f_X(z)(1 - F_Y(z)) + f_Y(z)(1 - F_X(z)).$$

Nel caso di v.a. esponenziali $X \sim \exp(\lambda)$ e $Y \sim \exp(\mu)$ si ha

$$f_Z(z) = \lambda e^{-\lambda z} e^{-\mu z} + \mu e^{-\mu z} e^{-\lambda z} = (\lambda + \mu) e^{-(\lambda + \mu)z}$$

ovvero un'esponenziale di parametro $\lambda + \mu$.

Esempio 8.2.1. Siano T_1 e T_2 i tempi di vita di due componenti posti in serie. Assumiamo che ognuno dei tempi di vita segua una $\exp(0.5)$ e che siano indipendenti. Qual è il tempo medio di vita dell'intero sistema?

In questo caso, poiché i componenti sono posti in serie la v.a. che rappresenta il tempo di vita del sistema è data da $T = \min(T_1, T_2)$, visto che il primo componente che si rompe blocca tutto il sistema. Perciò, per quanto detto prima sappiamo che $T \sim \exp(0.5 + 0.5) = \exp(1)$ e quindi la sua media sarà pari ad 1. Avendo posto i componenti in serie, il tempo di vita medio del sistema è più piccolo rispetto al tempo di vita del singolo componente.

Nel caso di X ed Y Gaussiane o Gamma la formula (8.2) non si può rendere maggiormente esplicita in quanto non possiamo calcolare in forma chiusa le funzioni di ripartizione.

2. Consideriamo ora $Z = \max(X, Y)$. In questo caso, partendo dalla funzione di ripartizione si ha, grazie all'indipendenza,

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(\max(X, Y) \leq z) = P(X \leq z, Y \leq z) \\ &= P(X \leq z)P(Y \leq z) = F_X(z)F_Y(z), \end{aligned}$$

da cui derivando si ottiene

$$(8.3) \quad f_Z(z) = f_X(z)F_Y(z) + f_Y(z)F_X(z).$$

Anche questa volta possiamo esplicitare la formula nel caso dell'esponenziale non pervenendo però ad una distribuzione nota; se $X \sim \exp(\lambda)$ e $Y \sim \exp(\mu)$ si ha

$$\begin{aligned} f_Z(z) &= \lambda e^{-\lambda z}(1 - e^{-\mu z}) + \mu e^{-\mu z}(1 - e^{-\lambda z}) = \lambda e^{-\lambda z} + \mu e^{-\mu z} - (\lambda + \mu)e^{-(\lambda + \mu)z} \\ &=^d f_X(z) + f_Y(z) - f_{\min(X, Y)}(z), \end{aligned}$$

dove l'ultima uguaglianza è da intendersi "in distribuzione", ovvero tra funzioni di densità e non tra v.a.

Esempio 8.2.2. Siano T_1 e T_2 i tempi di vita di due componenti posti in parallelo ognuno con distribuzione $\exp(0.5)$. Possiamo supporre che le due v.a. siano indipendenti. Qual è il tempo medio di vita dell'intero sistema?

In questo caso, poiché i componenti sono posti in parallelo, la v.a. che rappresenta il tempo di vita del sistema è data da $T = \max(T_1, T_2)$, visto che il sistema si blocca solo se si rompono entrambi. Per quanto detto prima, abbiamo

$$\begin{aligned} \mathbb{E}(T) &= \int_0^{+\infty} t \left[\frac{1}{2} e^{-\frac{1}{2}t} (1 - e^{-\frac{1}{2}t}) + \frac{1}{2} e^{-\frac{1}{2}t} (1 - e^{-\frac{1}{2}t}) \right] dt \\ &= 2 \int_0^{+\infty} t \frac{1}{2} e^{-\frac{1}{2}t} dt - \int_0^{+\infty} t e^{-t} dt = 2 \cdot 2 - 1 = 3 \end{aligned}$$

In questo caso, chiaramente, avendo posto i componenti in parallelo, il tempo di vita medio del sistema si allunga rispetto a quello del singolo componente.

Analogamente a quanto succede con il minimo, se X ed Y fossero delle v.a. Gaussiane o Gamma la formula (8.3) non potrebbe essere esplicitata oltre, sempre per la presenza di funzioni di ripartizione che non si conoscono in forma algebrica.

3. Sia $Z = X + Y$. In questo caso dobbiamo esprimere la funzione di ripartizione della Z tramite la funzione di densità congiunta che, in caso di indipendenza è

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

Abbiamo dunque

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X + Y \leq z) = \int \int_{x+y \leq z} f_X(x)f_Y(y)dydx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_X(x)f_Y(y)dydx = \int_{-\infty}^{+\infty} f_X(x)F_Y(z-x)dx \end{aligned}$$

e derivando (passando la derivata sotto il segno di integrale grazie all'assoluta integrabilità delle densità) si ottiene

$$(8.4) \quad f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx.$$

Anche stavolta possiamo specializzare la precedente formula al caso di esponenziali indipendenti $X \sim \exp(\lambda)$ e $Y \sim \exp(\mu)$, ma dobbiamo distinguere il caso $\lambda \neq \mu$ e $\lambda = \mu$. Nel primo caso, supponendo $\lambda > \mu$ e applicando la formula (8.4) arriviamo a

$$\begin{aligned} f_Z(z) &= \int_0^z \lambda e^{-\lambda x} \mu e^{-\mu(z-x)} dx = \lambda \mu e^{-\mu z} \int_0^z e^{-(\lambda-\mu)x} dx \\ &= \frac{\lambda \mu}{\lambda - \mu} e^{-\mu z} (1 - e^{-(\lambda-\mu)z}) = \frac{\lambda \mu}{\lambda - \mu} (e^{-\mu z} - e^{-\lambda z}), \end{aligned}$$

qualora si fosse nel caso $\mu > \lambda$ la stessa formula si potrebbe applicare con i ruoli di X ed Y scambiati.

È invece diverso il caso $\lambda = \mu$, infatti si ha

$$f_Z(z) = \int_0^z \mu e^{-\mu x} \mu e^{-\mu(z-x)} dx = \mu^2 z e^{-\mu z},$$

ovvero una $\Gamma(2, \mu)$.

Altrettanto interessante è il caso di v.a. Gaussiane. Siano $X \sim \mathcal{N}(0, \sigma^2)$ ed $Y \sim \mathcal{N}(0, \eta^2)$ indipendenti, applichiamo la (8.4) e scriviamo

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\eta^2}} e^{-\frac{(z-x)^2}{2\eta^2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\eta^2}} e^{-\frac{z^2-2zx+x^2}{2\eta^2}} dx \\ &= \frac{1}{2\pi\sqrt{\sigma^2\eta^2}} e^{-\frac{z^2}{2\eta^2}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma^2} - \frac{x^2-2zx}{2\eta^2}} dx. \end{aligned}$$

Concentriamoci per un attimo sull'esponente $-\frac{x^2}{2\sigma^2} - \frac{x^2-2zx}{2\eta^2}$. Seguendo una procedura di completamento dei quadrati otteniamo

$$\begin{aligned} -\frac{x^2}{2\sigma^2} - \frac{x^2-2zx}{2\eta^2} &= -\frac{(\eta^2 + \sigma^2)x^2 - 2\sigma^2zx}{2\sigma^2\eta^2} = -\frac{(\eta^2 + \sigma^2)[x^2 - 2\frac{\sigma^2}{\eta^2+\sigma^2}zx]}{2\sigma^2\eta^2} \\ &= -\frac{x^2 - 2\frac{\sigma^2}{\eta^2+\sigma^2}zx + \frac{\sigma^4}{(\eta^2+\sigma^2)^2}z^2 - \frac{\sigma^4}{(\eta^2+\sigma^2)^2}z^2}{2\frac{\sigma^2\eta^2}{\eta^2+\sigma^2}} \\ &= -\frac{(x - \frac{\sigma^2}{\eta^2+\sigma^2}z)^2}{2\frac{\sigma^2\eta^2}{\eta^2+\sigma^2}} + \frac{\sigma^2}{\eta^2(\eta^2 + \sigma^2)}\frac{z^2}{2} \end{aligned}$$

e sostituendo nella formula precedente si ha

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi\sqrt{\sigma^2\eta^2}} e^{-\frac{z^2}{2\eta^2} + \frac{\sigma^2}{\eta^2(\eta^2+\sigma^2)}\frac{z^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(x - \frac{\sigma^2}{\eta^2+\sigma^2}z)^2}{2\frac{\sigma^2\eta^2}{\eta^2+\sigma^2}}} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2(\eta^2+\sigma^2)}} \frac{1}{\sqrt{\eta^2 + \sigma^2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\frac{\sigma^2\eta^2}{\eta^2+\sigma^2}}} e^{-\frac{(x - \frac{\sigma^2}{\eta^2+\sigma^2}z)^2}{2\frac{\sigma^2\eta^2}{\eta^2+\sigma^2}}} dx \\ &= \frac{1}{\sqrt{2\pi(\eta^2 + \sigma^2)}} e^{-\frac{z^2}{2(\eta^2+\sigma^2)}}, \end{aligned}$$

e l'integrale risulta uguale ad 1 essendo l'integrale esteso a \mathbb{R} di una densità Gaussiana $\mathcal{N}(\frac{\sigma^2}{\eta^2+\sigma^2}z, \frac{\sigma^2\eta^2}{\eta^2+\sigma^2})$, quindi possiamo concludere che $Z = X + Y \sim \mathcal{N}(0, \eta^2 + \sigma^2)$.

Se la media non è nulla, $X \sim \mathcal{N}(\mu, \sigma^2)$ ed $Y \sim \mathcal{N}(\nu, \eta^2)$, allora possiamo scrivere $X = W + \mu$ e $Y = U + \nu$, dove $W \sim \mathcal{N}(0, \sigma^2)$ e $U \sim \mathcal{N}(0, \eta^2)$. Dunque $Z = X + Y = W + U + \mu + \nu$ e quindi deve essere una $\mathcal{N}(\mu + \nu, \eta^2 + \sigma^2)$. Possiamo riassumere dicendo che la somma di due v.a. Gaussiane indipendenti è ancora una Gaussiana con media la somma delle medie e varianza la somma delle varianze.

Iterando il precedente procedimento possiamo scrivere la proposizione generale

Proposizione 8.2.1. *Siano X_i v.a. Gaussiane $\mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \dots, n$ indipendenti, allora $Z = \sum_{i=1}^n X_i \sim \mathcal{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.*

Esempio 8.2.3. *L'altezza della popolazione maschile di 18 anni di una certa popolazione è distribuita secondo una v.a. $X \sim \mathcal{N}(177 \text{ cm}; 9 \text{ cm}^2)$. Ad uno screening viene rilevata l'altezza di un certo campione molto ampio della popolazione maschile di 18 anni, ma lo strumento è mal tarato e commette un errore Gaussiano $W \sim \mathcal{N}(-1 \text{ cm}; 0,4 \text{ cm}^2)$.*

Si stimi la probabilità che un'altezza rilevata risulti almeno di 175 cm.

Soluzione: l'altezza rilevata è data da $X + W$ poiché incorpora l'errore dello strumento e possiamo supporre che le due v.a. siano indipendenti. Per quanto detto prima, sappiamo che $X + W \sim \mathcal{N}(176 \text{ cm}; 9,4 \text{ cm}^2)$, quindi dobbiamo calcolare $P(X + W \geq 175)$. Per farlo standardizziamo la nostra v.a. Gaussiana in modo da ricondurre il calcolo di questa probabilità alla Gaussiana standard. Sappiamo che se $X + W \sim \mathcal{N}(176 \text{ cm}; 9,4 \text{ cm}^2)$, allora

$$\frac{X + W - E(X + W)}{\sqrt{\text{Var}(X + W)}} = \frac{X + W - 176}{\sqrt{9,4}} \sim \mathcal{N}(0; 1),$$

dunque

$$P(X + W \geq 175) = P\left(\frac{X + W - 176}{\sqrt{9,4}} \geq \frac{175 - 176}{\sqrt{9,4}}\right) = 1 - \Phi\left(-\frac{1}{\sqrt{9,4}}\right) = \Phi\left(\frac{1}{\sqrt{9,4}}\right).$$

Anche nel caso delle distribuzioni Gamma, la somma di v.a. presenta una proprietà di invarianza rispetto alla famiglia.

Proposizione 8.2.2. Siano $X_i \sim \Gamma(\alpha_i, \lambda)$ con $\alpha_i, \lambda > 0$ per $i = 1, \dots, n$ v.a. indipendenti, allora

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n \alpha_i, \lambda\right)$$

Dimostrazione: Dimostriamo la proposizione per due v.a. in quanto si può estendere per iterazione. Come al solito applichiamo la formula per la densità della somma di due v.a. indipendenti (8.4)

$$\begin{aligned} f_Z(z) &= \int_0^z \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)} x^{\alpha_1-1} e^{-\lambda x} \frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)} (z-x)^{\alpha_2-1} e^{-\lambda(z-x)} dx \\ &= \frac{1}{\Gamma(\alpha_1)} \frac{1}{\Gamma(\alpha_2)} \lambda^{\alpha_1+\alpha_2} e^{-\lambda z} \int_0^z x^{\alpha_1-1} (z-x)^{\alpha_2-1} dx \end{aligned}$$

usando il cambio di variabile $x = tz$, da cui $dx = zdt$, l'ultimo integrale diventa

$$z^{\alpha_1+\alpha_2-1} \int_0^1 t^{\alpha_1-1} (1-t)^{\alpha_2-1} dt,$$

quindi

$$f_Z(z) = \lambda^{\alpha_1+\alpha_2} z^{\alpha_1+\alpha_2-1} e^{-\lambda z} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^1 t^{\alpha_1-1} (1-t)^{\alpha_2-1} dt$$

cioè è una densità $\Gamma(\alpha_1 + \alpha_2, \lambda)$, da cui deve anche valere

$$\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} = \int_0^1 t^{\alpha_1-1} (1-t)^{\alpha_2-1} dt$$

Il caso generale si ottiene iterando la dimostrazione. □

Alla luce di questa proposizione possiamo rivisitare la somma di esponenziali dicendo che se X_1, \dots, X_n sono v.a. i.i.d. $\exp(\lambda) = \Gamma(1, \lambda)$, allora $X_1 + \dots + X_n \sim \Gamma(n, \lambda)$.

Osservazione 8.2.1. Nella proposizione precedente il fatto che teniamo il secondo parametro fisso nelle distribuzioni Gamma è fondamentale. Non abbiamo un analogo risultato fissando l'altro parametro (come abbiamo già visto nel caso di v.a. esponenziali).

Siano $X \sim \Gamma(\alpha, \lambda)$ e $Y \sim \Gamma(\alpha, \mu)$ due v.a. indipendenti consideriamo $Z = X + Y$, usando la formula (8.4) si ha

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\mu^\alpha}{\Gamma(\alpha)} \int_0^z x^{\alpha-1} (z-x)^{\alpha-1} e^{-\lambda x} e^{-\mu(z-x)} dx \\ &= \frac{(\lambda\mu)^\alpha}{\Gamma(\alpha)^2} e^{-\mu z} \int_0^z x^{\alpha-1} (z-x)^{\alpha-1} e^{-(\lambda-\mu)x} dx \end{aligned}$$

che non è una distribuzione nota, anche se la si può calcolare in alcuni casi specifici.

8.3 Densità e Media condizionata

Finora siamo riusciti a studiare le densità di alcune trasformazioni di vettori aleatori, grazie all'indipendenza delle v.a. considerate.

Quando siamo in presenza di due v.a. X ed Y non necessariamente indipendenti, ma dotate di funzione di densità congiunta $f_{X,Y}(x,y)$, possiamo definire, come nel caso discreto, la densità condizionale di Y dato X come

$$(8.5) \quad f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},$$

per i valori di x tali che $f_X(x) \neq 0$.

Per ogni x fissato, questa è una densità a tutti gli effetti, infatti verifica

1. $f_{Y|X}(y|x) \geq 0$, in quanto rapporto tra un numeratore maggiore od uguale a zero e un denominatore positivo.
2. Per ogni x fissato, si ha

$$\int_{\mathbb{R}} f_{Y|X}(y|x) dy = \int_{\mathbb{R}} \frac{f_{X,Y}(x,y)}{f_X(x)} dy = \frac{1}{f_X(x)} \int_{\mathbb{R}} f_{X,Y}(x,y) dy = \frac{1}{f_X(x)} f_X(x) = 1$$

Leggendo (8.5) al contrario, essa diventa uno strumento per calcolare la densità congiunta e le marginali.

Esempio 8.3.1. Sia Y una v.a. esponenziale di parametro $\lambda > 0$. Se $Y = y$, allora $X|Y = y$ è a v.a. esponenziale di parametro $\frac{1}{y}$.

Possiamo calcolare la densità congiunta

$$f_{X,Y}(x,y) = \lambda e^{-\lambda y} \mathbf{1}_{(0,+\infty)}(y) \frac{1}{y} e^{-\frac{1}{y}x} \mathbf{1}_{(0,+\infty)}(x),$$

dalla quale possiamo calcolare l'altra densità marginale

$$f_X(x) = \int_0^{+\infty} \lambda e^{-\lambda y} \mathbf{1}_{(0,+\infty)}(y) \frac{1}{y} e^{-\frac{1}{y}x} \mathbf{1}_{(0,+\infty)}(x) dy.$$

Conseguentemente, possiamo pensare di definire una media rispetto alla densità condizionale.

Definizione 8.3.1. Diciamo media di Y condizionata ad $X = x$ e si scrive $\mathbb{E}(Y|X = x)$, la quantità

$$(8.6) \quad \mathbb{E}(Y|X = x) = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy$$

Per ogni x fissato, la relazione (8.6) definisce un valore reale, ovvero si sta definendo una funzione che indichiamo con $\phi(x)$. Se applichiamo tale funzione alla v.a. X otterremo una nuova v.a. $Z = \phi(X)$ che chiamiamo media condizionata di Y data X e che denotiamo con $\mathbb{E}(Y|X)$.

La media condizionata è dunque una v.a. essa stessa ed è particolarmente utile per poter calcolare delle medie senza dover ricorrere al calcolo delle densità marginali a partire dalla densità congiunta.

Vale infatti la seguente proposizione che si può riassumere dicendo che la media della media condizionata è la media e si scrive $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$.

Proposizione 8.3.1. Siano X ed Y due v.a. dotate di funzione di densità congiunta $f_{X,Y}(x, y)$ ed entrambe di media finita. Allora

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \mathbb{E}(Y|X = x) f_X(x) dx$$

Dimostrazione: (M) La dimostrazione è abbastanza semplice poiché

$$\begin{aligned} \mathbb{E}(Y) &= \int_{\mathbb{R}} y f_Y(y) dy = \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{X,Y}(x, y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{X,Y}(x, y) dy dx = \int_{\mathbb{R}} \int_{\mathbb{R}} y f_{Y|X}(y|x) dy f_X(x) dx \\ &= \int_{\mathbb{R}} \mathbb{E}(Y|X = x) f_X(x) dx \quad \square \end{aligned}$$

La precedente proposizione dà modo di calcolare la media per passi.

Esempio 8.3.2. A seconda delle condizioni che si verificano, rappresentate da una v.a. $\Theta \sim \Gamma(\alpha, \lambda)$, una certa fila si forma secondo una v.a. di Poisson, cioè se $\Theta = \theta > 0$ allora $N \sim \text{Poisson}(\theta)$, vogliamo calcolare quanto sarà la lunghezza media della fila.

In questo caso per arrivare al calcolo della media, dovremmo prima ottenere la densità congiunta a partire dalla condizionata e dalla marginale, cioè

$$\begin{aligned} f_{N|\Theta}(k, \theta) &= \frac{\theta^k}{k!} e^{-\theta}, \quad f_{\Theta}(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} \mathbf{1}_{\{\theta>0\}} \\ f_N(k) &= \int_{\mathbb{R}} \frac{\theta^k}{k!} e^{-\theta} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta} \mathbf{1}_{\{\theta>0\}} d\theta \Rightarrow \mathbb{E}(N) = \sum_{k=0}^{+\infty} k f_N(k). \end{aligned}$$

Viceversa diventa molto più semplice applicare la precedente proposizione, infatti sappiamo che dato $\Theta = \theta$, poiché $N|\Theta = \theta \sim \text{Poisson}(\theta)$, si ha che $\mathbb{E}(N|\Theta = \theta) = \theta$.

Questo vuol dire che la media condizionata $\mathbb{E}(N|\Theta = \theta)$ ha definito la funzione $\phi(\theta) = \theta$ ovvero l'identità, che, applicata alla v.a., implica $\mathbb{E}(N|\Theta) = \Theta$ e dalla proposizione si ottiene

$$\mathbb{E}(N) = \mathbb{E}(\mathbb{E}(N|\Theta)) = \mathbb{E}(\Theta) = \frac{\alpha}{\lambda}$$

Analoga formula si può ottenere per la varianza.

Avendo notato che $\mathbb{E}(Y|X)$ è una v.a. (come funzione di X) possiamo calcolare, se finita, la sua varianza

$$\text{Var}(\mathbb{E}(Y|X)) = \mathbb{E}[(\mathbb{E}(Y|X))^2] - [\mathbb{E}(\mathbb{E}(Y|X))]^2 = \mathbb{E}[(\mathbb{E}(Y|X))^2] - [\mathbb{E}(Y)]^2$$

dove l'ultima uguaglianza è ottenuta grazie alla proposizione precedente. Questa è quindi **la varianza della media condizionata**.

Ricordiamo che

$$\mathbb{E}[(\mathbb{E}(Y|X))^2] = \int_{\mathbb{R}} (\mathbb{E}(Y|X=x))^2 f_X(x) dx$$

Possiamo anche definire la varianza condizionata di Y data X come la varianza calcolata rispetto alla densità condizionale, cioè

$$\begin{aligned} \text{Var}(Y|X=x) &= \mathbb{E}(Y^2|X=x) - [\mathbb{E}(Y|X=x)]^2 \\ &= \int_{\mathbb{R}} y^2 f_{Y|X}(y|x) dy - \left[\int_{\mathbb{R}} y f_{Y|X}(y|x) dy \right]^2 \end{aligned}$$

la varianza di Y condizionato a $X = x$ e che è una funzione al variare di x , che indichiamo con $\psi(x)$.

Applicando questa funzione alla v.a. X , possiamo definire **la varianza di Y data X** come

$$\text{Var}(Y|X) := \psi(X) = \mathbb{E}(Y^2|X) - [\mathbb{E}(Y|X)]^2.$$

Vale allora la seguente

Proposizione 8.3.2. *Siano X ed Y due v.a. dotate entrambe di momento secondo finito e con densità congiunta $f_{X,Y}(x,y)$, allora vale*

$$(8.7) \quad \text{Var}(Y) = \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X))$$

Dimostrazione: Poniamo $\mathbb{E}(Y) = \mu$, vale la seguente

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}(Y^2 - \mu^2) = \mathbb{E}(Y^2 - [\mathbb{E}(Y|X)]^2 + [\mathbb{E}(Y|X)]^2 - \mu^2) \\ &= \mathbb{E}(Y^2 - [\mathbb{E}(Y|X)]^2) + \mathbb{E}([\mathbb{E}(Y|X)]^2 - \mu^2) \\ &= \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X)) \end{aligned}$$

poiché $\mathbb{E}(\mathbb{E}(Y|X)) = \mu$. □

Esempio 8.3.3. *Riprendendo l'esempio precedente possiamo calcolare facilmente anche la varianza di N , infatti applicando quanto appena detto si ha che $\text{Var}(N|\Theta = \theta) = \theta$ da cui anche la varianza condizionale definisce la funzione $\psi(\theta) = \theta$ da cui $\psi(\Theta) = \Theta$ e si ha*

$$\text{Var}(Y) = \mathbb{E}(\text{Var}(Y|X)) + \text{Var}(\mathbb{E}(Y|X)) = \mathbb{E}(\Theta) + \text{Var}(\Theta) = \frac{\alpha}{\lambda} + \frac{\alpha}{\lambda^2}$$

8.4 Revisione di Teoria

1. Siano X ed Y due v.a. continue indipendenti con densità marginali f_X e f_Y . Calcolare la densità di $Z = \max(X; Y)$.
2. Siano X ed Y due v.a. continue indipendenti con densità marginali f_X e f_Y . Calcolare la densità di $Z = \min(X; Y)$.
3. Siano X ed Y due v.a. esponenziali indipendenti di parametri rispettivamente μ e λ , qual è la varianza di $Z = \min(X, Y)$
4. Dare la definizione di densità condizionale di X data Y , sapendo che le due v.a. hanno densità congiunta $f_{X,Y}(x, y)$.
5. Se X_1, \dots, X_n sono v.a. i.i.d. secondo una $\Gamma(\alpha, \beta)$, qual è la distribuzione di $X_1 + \dots + X_n$?
6. Se X_1, \dots, X_n sono v.a. indipendenti tutte esponenziali di parametro λ , qual è la densità di $X_1 + \dots + X_n$?

8.5 Esercizi

1. Si considerino due variabili aleatorie indipendenti uniformi in $[0, 1]$, calcolare
 - (a) la densità congiunta di X ed $X - Y$;
 - (b) la densità di $X - Y$;
 - (c) $\text{Var}(X - Y)$.
2. Siano $X \sim \exp(\frac{1}{9})$ ed $Y \sim \Gamma(5.4, 3)$ tra loro indipendenti. Siano $Z = X - Y$, $W = \frac{1}{2}XY$
 - (a) Calcolare $\text{Var}(Z)$ e $\text{Var}(W)$.
 - (b) Calcolare $\text{cov}(Z, W)$.
3. Si consideri una v.a. bidimensionale (X, Y) con densità

$$f_{XY}(x, y) = \frac{1}{8}(6 - x - y)\mathbf{1}_{(0,2)}(x)\mathbf{1}_{(2,4)}(y)$$

calcolare

- (a) $\mathbb{E}(Y|X = x)$
 - (b) $\mathbb{E}(Y^2|X = x)$ e $\text{Var}(Y|X = x)$
 - (c) $\mathbb{E}(Y)$
4. La funzione di densità congiunta di X ed Y è data da $f_{X,Y}(x; y) = 8xy$ per $0 < x < y < 1$; trovare $\mathbb{E}(Y|X = x)$ e $\text{Var}(Y|X = x)$.
5. Sia Y una v.a. esponenziale di parametro $\lambda > 0$. Se $Y = y$ allora $X|Y = y$ è una v.a. esponenziale di parametro $\frac{1}{y}$.
Calcolare $\mathbb{E}(X)$ e $\text{Var}(X)$.
6. Per $n \in \mathbb{N}$ fissato, siano X_1, \dots, X_n v.a. indipendenti ed identicamente distribuite tali che $X_k \sim \exp(\frac{\lambda}{n})$ per $\lambda > 0$, $k = 1, \dots, n$. Individuare le leggi di

$$Y_n = \max(X_1, \dots, X_n) \quad \text{e} \quad Z_n = \min(X_1, \dots, X_n).$$

7. Siano X ed Y due v.a. con densità congiunta data da

$$f_{X,Y}(x, y) = \frac{1}{y} e^{-y - \frac{x}{y}} \mathbf{1}_{\{x > 0, y > 0\}}$$

Calcolare le densità marginali, $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\text{cov}(X, Y)$, $\mathbb{E}(X^3|Y)$.

8. La temperatura massima di una certa zona a ferragosto è data da una v.a. $T \sim \Gamma(99, 3)$. Se T prende il valore t , allora il fabbisogno energetico in KW x 1000 di quella zona è distribuito secondo una v.a. $\Gamma(t^2, 4t)$.
- (a) Indicando con X il fabbisogno energetico, scrivere la densità congiunta delle v.a. T, X .
 - (b) Calcolare la media e la varianza del fabbisogno energetico X .

Capitolo 9

SUCCESSIONI DI V.A., FUNZIONI GENERATRICI, LGN, TLC

In questa sezione vediamo come affrontare i problemi che coinvolgono un grande numero di v.a. Matematicamente parlando, vogliamo capire come trattare le successioni di v.a. $\{X_n\}_{n \in \mathbb{N}}$. È quindi chiaro che dovremo anche introdurre dei concetti di convergenza. Due teoremi riassumono i risultati base per la convergenza delle successioni di v.a.: la Legge dei Grandi Numeri (LGN) ed il Teorema del Limite Centrale (TLC). Questi due teoremi sono fondamentali nella Teoria della Probabilità e fondanti della Statistica Matematica.

9.1 Nozioni di convergenza (M)

Quando si trattano successioni di v.a., un punto fondamentale è riuscire a capirne il comportamento asintotico. Per farlo occorre introdurre una nozione di convergenza e in questo ambito si possono effettuare scelte differenti.

Se guardiamo le v.a. come funzioni, vuol dire che siamo interessati a studiare il loro comportamento puntuale e quindi dobbiamo usare un qualche concetto di convergenza di funzioni.

D'altro canto spesso non siamo interessati al comportamento puntuale delle v.a., quanto a valutare la probabilità di eventi definiti dalle stesse. In questo caso, piuttosto che sfruttare una convergenza che osserva il comportamento puntuale delle v.a., può essere utile disporre di un concetto di convergenza che studia il comportamento asintotico di questi eventi.

Infine, in moltissime situazioni tendiamo ad identificare le v.a. mediante le loro funzioni di ripartizione, in questo caso non partiamo più da uno spazio di probabilità definito a priori, stiamo guardando soltanto lo spazio immagine determinato dalla v.a. Quando vediamo i fenomeni aleatori da questa ottica, abbiamo bisogno di una nozione di convergenza che non necessiti più di definire uno spazio di probabilità a priori, ma che utilizzi direttamente il comportamento asintotico delle funzioni di ripartizione.

La prima nozione di convergenza che introduciamo è quindi quella che vede le v.a. come funzioni ed è data da

Definizione 9.1.1. Siano $\{X_n\}_n$ e X v.a. definite tutte sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) , allora diciamo che $X_n \rightarrow X$ con probabilità 1 o quasi ovunque se

$$P(\{\omega : \lim_{n \rightarrow +\infty} X_n(\omega) \neq X(\omega)\}) = 0.$$

Le notazioni correnti per questo tipo di convergenza sono

$$X_n \rightarrow X \quad P = 1, \quad \text{oppure} \quad X_n \rightarrow X \quad P - q.o. \text{ per } n \rightarrow +\infty.$$

La nozione di convergenza quasi ovunque coincide con la convergenza puntuale di funzioni ma su un insieme di probabilità 1.

Esempio 9.1.1. Immaginiamo di prendere lo spazio di probabilità costituito dall'intervallo unitario $[0, 1]$, con la σ -algebra dei Boreliani (ovvero generata da tutti gli intervalli aperti di $[0, 1]$) e come probabilità prendiamo semplicemente quella che associa ad ogni intervallo la sua lunghezza. Su questo spazio di probabilità consideriamo la successione di v.a. data dalle funzioni

$$X_n(\omega) = \mathbf{1}_{[0, 0.5 - \frac{1}{n}] \cup [0.5 + \frac{1}{n}, 1]}$$

queste v.a. convergono quasi ovunque alla funzione costante pari ad 1. Infatti la convergenza è realizzata per tutti i punti dell'intervallo, tranne che per $\omega = 0.5$, dove tutta la successione prende sempre valore pari a zero. Ma la probabilità dell'insieme $\{0.5\}$ (costituito da un singolo punto) è zero.

Un secondo tipo di convergenza è dato dalla

Definizione 9.1.2. Siano $\{X_n\}_n$ e X v.a. definite tutte sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) , allora diciamo che $X_n \rightarrow X$ in probabilità se $\forall \eta, \epsilon > 0$ esiste un $n(\epsilon, \eta) \in \mathbb{N}$ tale che

$$P(|X_n - X| > \epsilon) \leq \eta, \quad \text{per } n > n(\epsilon, \eta)$$

o equivalentemente se per ogni $\epsilon > 0$

$$\lim_{n \rightarrow +\infty} P(|X_n - X| > \epsilon) = 0.$$

e scriviamo per brevità $X_n \rightarrow_P X$.

La disuguaglianza di Chebyshev implica questo tipo di convergenza

Esempio 9.1.2. Si consideri la successione di v.a.

$$X_n \sim \mathcal{N}(5, \frac{1}{n})$$

dalla disuguaglianza di Chebyshev sappiamo che per qualsiasi scelta di $\epsilon > 0$, si ha

$$P(|X_n - \mathbb{E}(X_n)| > \epsilon) \leq \frac{\text{Var}(X_n)}{\epsilon^2} \rightarrow 0, \quad \text{per } n \rightarrow +\infty$$

ovvero

$$P(|X_n - 5| > \epsilon) \leq \frac{1}{\epsilon^2}$$

per qualsiasi valore di ϵ , che è come dire che $X_n \rightarrow_P 5$.

Poiché spesso le v.a. sono individuate soltanto attraverso le loro funzioni di ripartizione o di densità, è utile avere una nozione di convergenza che coinvolga direttamente le funzioni di ripartizione di una successione di v.a.

Definizione 9.1.3. Siano $\{X_n\}_n$ e X v.a. le cui funzioni di ripartizione associate sono rispettivamente $\{F_n\}_n$ e F . Diciamo che X_n converge in distribuzione (o in legge) ad X se

$$\lim_{n \rightarrow +\infty} F_n(x) = F(x)$$

per ogni punto di continuità di F .

Questo tipo di convergenza viene denotata

$$X_n \Rightarrow X, \quad \text{oppure} \quad X_n \rightarrow_{\mathcal{L}} X, \quad \text{oppure} \quad X_n \rightarrow_d X.$$

Le tre nozioni di convergenza date sono una più forte dell'altra, intendendo dire che la convergenza P -q.o. implica la convergenza in probabilità che implica quella in legge.

Esempio 9.1.3. Consideriamo la seguente successione di v.a. indipendenti sullo stesso spazio di probabilità

$$X_n = \begin{cases} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{cases} \quad Y = \begin{cases} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{cases},$$

poiché la distribuzione è sempre la stessa (Bernoulliana), necessariamente $X_n \Rightarrow Y$, ma questa successione non converge in probabilità ad Y , in quanto per ogni $1 > \epsilon > 0$ e per ogni $n \in \mathbb{N}$ si ha

$$P(|X_n - Y| > \epsilon) = P(X_n = 0, Y = 1) + P(X_n = 1, Y = 0) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Infine diamo l'esempio di una successione di v.a. che non converge in distribuzione. Se scegliamo le v.a. $X_n \equiv n$ abbiamo che le funzioni di distribuzione associate

$$F_n(x) = \mathbf{1}_{[n, +\infty)}$$

che convergono alla funzione identicamente uguale a 0, la quale non è una funzione di distribuzione.

9.2 La Legge dei Grandi Numeri

Come abbiamo visto, molti fenomeni aleatori coinvolgono la somma di v.a. nella loro descrizione. È facile immaginare che questa somma possa essere infinita, per esempio per descrivere il numero di teste uscite in una successione infinita di lanci di una moneta equa. Quindi è importante riuscire a capire quale sia il comportamento asintotico di queste serie di v.a.

I teoremi della probabilità importanti in questo ambito sono due, qui presentiamo il primo, che si chiama la **Legge dei Grandi Numeri**.

Teorema 9.2.1. (*Legge dei Grandi Numeri*) Sia $\{X_n\}_n$ una successione di v.a. i.i.d. con media $\mathbb{E}(X_1) = \mu$ e varianza $\text{Var}(X_1) = \sigma^2 < +\infty$. Sia

$$(9.1) \quad \bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Allora per ogni $\epsilon > 0$ vale

$$(9.2) \quad \lim_{n \rightarrow +\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$

Dimostrazione: La dimostrazione si basa sulla disuguaglianza di Chebyshev applicata alla v.a. \bar{X}_n , notando che

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n \mathbb{E}(X_k) = \frac{1}{n} n\mu = \mu$$

e, grazie all'indipendenza delle v.a.,

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Applicando la disuguaglianza di Chebyshev si ha allora

$$P(|\bar{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \quad \text{per } n \rightarrow +\infty \quad \forall \epsilon > 0. \quad \square$$

La successione delle v.a. $\{\bar{X}_n\}_n$ è detta la successione delle medie **empiriche** delle v.a. Sottolineiamo che la media empirica \bar{X}_n è una v.a.

La Legge dei Grandi Numeri, secondo la definizione della sezione precedente, è quindi un esempio di convergenza in probabilità.

Nel teorema dobbiamo pensare la quantità ϵ come piccola, tendente a 0. Per cui il teorema dice che se abbiamo una successione di esperimenti che si ripetono in maniera indipendente e sempre nelle medesime condizioni un numero sufficientemente grande di volte, quasi con certezza possiamo affermare che la distanza tra qualsiasi realizzazione della media empirica e la media effettiva sarà molto piccola. Questo vuol dire che anche qualora noi ignorassimo la distribuzione e la media di una v.a. che rappresenta un certo fenomeno, ripetendo l'esperimento in maniera indipendente e facendo la media aritmetica dei risultati arriveremmo ad avere un'approssimazione della media a noi ignota con qualsiasi precisione desideriamo.

Una v.a. che approssima un parametro reale è detta **stimatore** del parametro. Quindi possiamo dire che la media empirica \bar{X}_n è uno stimatore della media μ . Più precisamente abbiamo le seguenti

Definizione 9.2.1. Chiamiamo *modello statistico* una famiglia di spazi di probabilità $\{\mathcal{M}_\theta = (\Omega, \mathcal{F}, P_\theta)\}_{\theta \in \Theta}$, dove la probabilità dipende da un parametro θ che varia in uno spazio $\Theta \subseteq \mathbb{R}^k$, per qualche $k \in \mathbb{N}$.

Definizione 9.2.2. Dato un modello statistico \mathcal{M}_θ , per ogni valore fissato del parametro, si consideri una successione di v.a. X_1, \dots, X_n, \dots i.i.d. (le osservazioni) con funzione di distribuzione F , determinata da $h(\theta)$ per qualche funzione del parametro $h : \Theta \rightarrow \mathbb{R}$.

Preso un'altra funzione del parametro $f : \Theta \rightarrow \mathbb{R}$, diciamo **stimatore** di $f(\theta)$ una v.a. definita mediante una successione di funzioni delle osservazioni

$$T_n = t_n(X_1, \dots, X_n)$$

che risulta “vicina” a $f(\theta)$. Uno stimatore si dice **non distorto** se $\mathbb{E}(T_n) = f(\theta)$.

La finalità di introdurre questo concetto dovrebbe essere quella di essere in grado di stimare un qualche parametro sulla base delle osservazioni fornite dal fenomeno e quindi occorre specificare una qualche nozione di vicinanza tra lo stimatore ed il parametro. La LGN ci dice quindi che la media campionaria

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

è uno stimatore (non distorto) della media μ di una v.a. se abbiamo scelto le v.a. i.i.d. con quella media, dove $t_n(x_1, \dots, x_n) = \frac{1}{n}(x_1 + \dots + x_n)$ e l'approssimazione del parametro da parte dello stimatore è data dalla nozione di convergenza in probabilità. Come abbiamo visto in precedenza, le medie sono quasi sempre collegate ai parametri che caratterizzano le distribuzioni di probabilità ed il seguente esempio ci fa capire perché in molti fenomeni prendiamo le frequenze relative come le probabilità dei risultati.

Esempio 9.2.1. Supponiamo di voler stabilire se una moneta che stiamo usando è equa oppure no, allora possiamo procedere nella seguente maniera. Si effettuano n lanci ripetuti della moneta, questo vuol dire che stiamo generando una successione di v.a. Bernoulli

$$X_i = \begin{cases} 1 & p \\ 0 & 1 - p, \end{cases}$$

dove il parametro p che rappresenta la probabilità di successo (Testa o Croce secondo i gusti) è incognito. Ma noi sappiamo che $\mathbb{E}(X_i) = p$ per ogni i e quindi la LGN afferma che \bar{X}_n , che conta il numero di volte che si è verificato il successo sul numero totale di prove, è tale che $\bar{X}_n \rightarrow \mathbb{E}(X_1) = p$, quindi possiamo concludere

$$\frac{\# \text{ teste in } n \text{ lanci}}{\# \text{ lanci totali}} = \bar{X}_n \rightarrow p.$$

In altre parole la frequenza relativa delle teste (cioè il rapporto tra il numero di teste e il numero totale di casi) dà un'approssimazione del parametro p che esprime anche l'equità della moneta.

Infatti, se dopo un congruo numero di lanci, ci rendiamo conto che \bar{X}_n si stabilizza (con la precisione che desideriamo noi) intorno al valore $\frac{1}{2}$ potremo concludere che la moneta è equa, altrimenti no.

Alla luce dell'esempio precedente si apre un nuovo problema. Qual è il numero opportuno di ripetizioni dell'esperimento da effettuare per ottenere una precisione desiderata?

Poiché la LGN è dimostrata sulla base della disuguaglianza di Chebyshev per ottenere una stima dell'errore commesso è necessario riuscire a controllare la varianza.

Esempio 9.2.2. *Riprendendo il caso precedente supponiamo di voler commettere un errore al più dell' 1% con probabilità pari a $\frac{1}{100}$. Dalla disuguaglianza di Chebyshev abbiamo*

$$P(|\bar{X}_n - p| > \frac{1}{100}) \leq \frac{\text{Var}(\bar{X}_n)}{(\frac{1}{100})^2} = \frac{p(1-p)}{n} 100^2$$

e vogliamo imporre che questa quantità sia minore di $\frac{1}{100}$. Ma la stima dipende dalla varianza che espressa mediante il parametro incognito p che stiamo stimando. Notiamo però che $p(1-p) \leq \frac{1}{4}$ per ogni $0 < p < 1$, possiamo dunque imporre che

$$\frac{p(1-p)}{n} 100^2 \leq \frac{2500}{n} \leq \frac{1}{100}$$

da cui consegue che $n \geq 250000$.

In realtà sia le prove empiriche che altre stime più precise danno un intero n di gran lunga più basso.

Il nome completo del precedente teorema è Legge **debole** dei Grandi Numeri, contrapposta ad una versione forte che non affronteremo in questo corso. L'aggettivo debole si riferisce al tipo di convergenza che la successione realizza. In realtà sotto le ipotesi da noi presentate (peraltro classiche) vale un risultato più forte.

Teorema 9.2.2. *(Legge Forte dei Grandi Numeri) Sia $\{X_n\}_n$ una successione di v.a. i.i.d. con media $\mathbb{E}(X_1) = \mu$ e varianza $\text{Var}(X_1) = \sigma^2 < +\infty$. Sia*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Allora

$$(9.3) \quad \lim_{n \rightarrow +\infty} \bar{X}_n = \mu, \quad P - q.o.$$

La LGN forte è importante perché dice che, quale che sia la successione di osservazioni realizzate, la media empirica sarà sempre vicina alla media effettiva μ , se il numero di osservazioni è sufficientemente grande. Questo rende il nostro stimatore asintoticamente "esatto", dunque da un punto di vista pratico si può individuare una media μ incognita con qualsiasi precisione si voglia, ripetendo più volte l'esperimento ed osservando quando le cifre decimali della media empirica calcolata sulla base delle osservazioni si stabilizzano alla precisione desiderata.

Invece l'importanza della Legge debole dei Grandi Numeri risiede nel fatto che essa ci fornisce anche una stima dell'errore che si commette approssimando la media con la media

empirica. La LGN è uno strumento molto generale che non richiede particolari specifiche della distribuzione delle v.a. al di là dell'esistenza dei momenti primo e secondo. Quindi abbiamo che la media empirica \bar{X}_n è uno stimatore del parametro p della $\text{Bin}(1, p)$, del parametro λ della $\text{Poisson}(\lambda)$, del parametro μ della $\mathcal{N}(\mu, \sigma^2)$.

Notiamo inoltre che la convergenza in probabilità e la convergenza P -q.o. verificano la seguente importante proprietà.

Se $X_n \rightarrow_P X$ per $n \rightarrow +\infty$ ed f è una funzione a valori reali continua, allora

$$(9.4) \quad f(X_n) \rightarrow_P f(X), \quad n \rightarrow +\infty.$$

Questa proprietà permette di costruire stimatori per i parametri di molte altre distribuzioni, per tutti i momenti di una v.a. e i cosiddetti metodi Monte Carlo per il calcolo di integrali e di medie complesse. Infatti la media di una v.a. è sempre collegata ad almeno uno dei parametri che caratterizzano la sua distribuzione attraverso funzioni continue, quindi applicando la proprietà (9.4), possiamo concludere che $(\bar{X}_n)^{-1}$ rappresenta uno stimatore naturale per il parametro λ di una distribuzione $\exp(\lambda)$, o per il parametro p della distribuzione $\text{geom}(p)$, perché la funzione $\frac{1}{x}$ è continua per $x \neq 0$.

Molte distribuzioni sono caratterizzate da una coppia di parametri, legati ad entrambe la media e la varianza della v.a. (si pensi alle v.a. $\mathcal{N}(\mu, \sigma^2)$ e le $\Gamma(\alpha, \lambda)$ e sarebbe quindi desiderabile riuscire ad avere uno stimatore anche per la varianza di una v.a. Di nuovo questo è possibile grazie ad un'applicazione della LGN.

Supponiamo di avere una successione di v.a., $\{X_n\}_n$ i.i.d. con momenti finiti fino al quarto e con media μ e varianza σ^2 incognite. Se conoscessimo μ , allora anche le v.a. $\{(X_i - \mu)^2\}_i$ sarebbero integrabili con varianza finita ed i.i.d., quindi per la LGN si avrebbe

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 \rightarrow \sigma^2,$$

ma questo stimatore non può essere usato direttamente se la media è incognita. Anche in quest'ultimo caso, si può però costruire uno stimatore (non distorto) per la varianza, prendendo

$$(9.5) \quad s_n^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

sfruttando sempre lo stesso campione.

Dimostriamo che s_n^2 è effettivamente uno stimatore di σ^2 .

$$\begin{aligned}
s_n^2 &= \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu + \mu - \bar{X}_n)^2 \\
&= \frac{1}{n-1} \sum_{k=1}^n [(X_k - \mu)^2 + 2(X_k - \mu)(\mu - \bar{X}_n) + (\mu - \bar{X}_n)^2] \\
&= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 + \frac{2}{n-1} \sum_{k=1}^n (X_k - \mu)(\mu - \bar{X}_n) + \frac{n}{n-1} (\mu - \bar{X}_n)^2 \\
&= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 + 2 \frac{n}{n-1} [2\mu\bar{X}_n - \mu^2 - \bar{X}_n^2] + \frac{n}{n-1} (\mu - \bar{X}_n)^2 \\
&= \frac{1}{n-1} \sum_{k=1}^n (X_k - \mu)^2 - \frac{n}{n-1} (\mu - \bar{X}_n)^2 \\
&= \frac{n}{n-1} \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2 - \frac{n}{n-1} (\mu - \bar{X}_n)^2
\end{aligned}$$

ed il primo termine tende a σ^2 come dimostrato prima, mentre il secondo tende a 0 per la continuità del quadrato anche rispetto alla convergenza in probabilità.

Infine la LGN può essere utilizzata per stimare la distribuzione di una v.a. discreta. Supponiamo di avere una v.a. che prende valori in un insieme finito $\{x_1, \dots, x_m\}$, ma di cui non conosciamo la densità. Come nel caso Bernoulliano, indichiamo con $p_i = P(X = x_i)$ la densità incognita e prendiamo n copie indipendenti (ovvero ripetiamo l'esperimento sempre nelle stesse condizioni) della nostra v.a. X, X_1, \dots, X_n , ed andiamo a contare il numero di volte che è uscito il valore x_i sul numero totale, ovvero

$$\frac{\# \text{ volte } x_i \text{ è stato osservato}}{n} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k = x_i\}}$$

cosiddetta **frequenza empirica**. La LGN implica che questa somma deve convergere al valore incognito p_i essendo le v.a. $\mathbf{1}_{\{X_k = x_i\}} \sim \text{Bin}(1, p_i)$ indipendenti. La precedente osservazione spiega anche perché spesso si attribuisce la probabilità mediante la frequenza empirica.

Per un'applicazione specifica alla distribuzione binomiale vedere l'esempio 2.54 del libro di Baldi [1].

Concludiamo con un importante esempio che rappresenta l'incipit di un grosso settore della Probabilità.

Esempio 9.2.3. *Metodo Monte Carlo.*

Vogliamo calcolare

$$\int_0^1 g(x) dx,$$

che non si calcola con metodi elementari.

Per farlo possiamo sfruttare la LGN, prendendo una successione di v.a. i.i.d. $\{X_n\}_n$ uniformi su $[0, 1]$. Allora anche la successione di v.a.

$$Y_n = g(X_n)$$

è una successione di v.a. i.i.d. con media

$$E(g(X_1)) = \int_0^1 g(x)dx$$

e quindi dalla LGN si ha

$$\frac{1}{n} \sum_{k=1}^n g(X_k) \longrightarrow \int_0^1 g(x)dx$$

Estendendo questa idea a distribuzioni diverse dall'uniforme, si possono calcolare molti tipi di integrali.

9.3 Funzioni Generatrici

In generale in probabilità esistono molte funzioni generatrici che sono funzioni legate alle funzioni di ripartizione mediante un' applicazione biunivoca. Ne consideriamo una in particolare

Definizione 9.3.1. Sia X una v.a. (od un vettore aleatorio $\mathbf{X} = (X_1, \dots, X_d)$), diciamo **funzione generatrice dei momenti** (o trasformata di Laplace) di X

$$M_X(t) = E(e^{\mathbf{t} \cdot \mathbf{X}}), \quad \mathbf{t} \in \mathbb{R}^d,$$

dove con $\mathbf{t} \cdot \mathbf{X}$ abbiamo indicato il prodotto scalare di due vettori in \mathbb{R}^d .

È sottinteso nella definizione che la funzione M_X ha senso soltanto per quei valori \mathbf{t} che rendono la media considerata finita.

D'ora in poi ci restringiamo, per comodità, a considerare soltanto il caso unidimensionale delle v.a. Vale sempre $M_X(0) = 1$ per qualsiasi v.a. inoltre se X è concentrata su valori positivi reali, allora sappiamo $M_X(t) \leq 1$ per $t \leq 0$. Più in generale l'integrale rappresentato dalla funzione generatrice dei momenti è finito in un intorno dello 0. Possiamo dire che $M_X(t)$ è definita in una certa regione di convergenza.

L'importanza delle funzioni generatrici dei momenti risiede nella seguente proposizione.

Proposizione 9.3.1. Siano X ed Y due v.a. con funzioni generatrici dei momenti M_X ed M_Y tali che $M_X(t) = M_Y(t)$ per $t \in (a, b)$ per qualche intervallo aperto. Allora X ed Y hanno la stessa funzione di ripartizione.

Non dimostreremo la precedente proposizione in quanto esula dalle finalità di questo corso e fondamentalmente consiste nel teorema di inversione della trasformata di Laplace, ma sottolineiamo che essa dice che la funzione generatrice dei momenti individua in maniera univoca la funzione di ripartizione.

Questo è un fatto importante perché spesso la funzione generatrice dei momenti, grazie alle proprietà che vedremo, si calcola più facilmente della funzione di ripartizione; quindi l'idea è di procedere per questo calcolo e di sfruttare la proposizione 9.3.1 per dedurre la funzione di ripartizione.

Proprietà della funzione generatrice dei momenti

1. $M_X(0) = 1$ per ogni v.a. X .
2. Se X è una v.a. con funzione generatrice dei momenti M_X definita in un intorno dello 0, allora per ogni $a, b \in \mathbb{R}$ si ha

$$M_{aX+b}(t) = E(e^{t(aX+b)}) = E(e^{taX} e^{tb}) = e^{tb} E(e^{(ta)X}) = e^{tb} M_X(at).$$

3. Se X ed Y sono due v.a. indipendenti con rispettive funzioni generatrici dei momenti M_X ed M_Y definite su un comune intorno dello 0, allora si ha (poiché funzioni di v.a. indipendenti rimangono indipendenti)

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX} e^{tY}) = E(e^{tX}) E(e^{tY}) = M_X(t) M_Y(t).$$

4. Se X è una v.a. con funzione generatrice dei momenti M_X che risulta derivabile per n volte con continuità in un intervallo dello 0, allora vale

$$M_X^{(k)}(0) = E(X^k), \quad \forall k \leq n,$$

infatti si ha

$$M_X^{(k)}(t) = \frac{d^k}{dt^k} E(e^{tX}) = E\left(\frac{d^k}{dt^k} e^{tX}\right) = E(X^k e^{tX})$$

e valutando a 0 si ottiene l'uguaglianza precedente.

L'ultima proprietà chiarisce il perché del nome della funzione.

Può essere utile calcolare le funzioni generatrici dei momenti delle principali distribuzioni.

1. $X \sim \text{Bin}(1, p)$, allora

$$M_X(t) = e^t p + 1 - p = p(e^t - 1) + 1, \quad t \in \mathbb{R}.$$

2. $X \sim \text{Bin}(n, p)$, allora $X = X_1 + \dots + X_n$ con indipendenti $X_i \sim \text{Bin}(1, p)$, dunque

$$M_X(t) = M_{X_1}(t) \dots M_{X_n}(t) = (e^t p + 1 - p)^n, \quad t \in \mathbb{R}.$$

3. $X \sim \text{Poisson}(\lambda)$, allora

$$M_X(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda(1-e^t)}, \quad t \in \mathbb{R}.$$

4. $X \sim \text{geom}(p)$, allora

$$M_X(t) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = pe^t \sum_{k=1}^{\infty} [e^t(1-p)]^{k-1} = \frac{pe^t}{1 - e^t(1-p)}, \quad t < \ln\left(\frac{1}{1-p}\right).$$

5. $X \sim \exp(\lambda)$, allora

$$M_X(t) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

6. $X \sim \mathcal{N}(0; 1)$. Ancora una volta, questo integrale si ottiene completando i quadrati

$$\begin{aligned} M_X(t) &= \int_{\mathbb{R}} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - 2tx}{2}} dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - 2tx + t^2}{2}} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - 2tx + t^2}{2}} e^{\frac{t^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}. \end{aligned}$$

Quindi siamo in grado di trovare la fgm di una qualsiasi $Y \sim \mathcal{N}(\mu; \sigma^2)$, grazie alla proprietà 2., ottenendo che la sua funzione generatrice dei momenti è data da

$$M_Y(t) = e^{t\mu} e^{\sigma^2 \frac{t^2}{2}} \quad t \in \mathbb{R},$$

7. $X \sim \Gamma(\alpha, \lambda)$, allora

$$\begin{aligned} M_X(t) &= \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \int_0^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx \\ &= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^{\infty} \frac{(\lambda-t)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\lambda-t)x} dx = \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \quad t < \lambda. \end{aligned}$$

Nel prossimo capitolo chiariremo l'utilità della funzione generatrice dei momenti.

9.4 Il Teorema del limite centrale

Abbiamo introdotto la funzione generatrice dei momenti. Il seguente teorema ci dice che questa è un importante strumento per verificare la convergenza delle funzioni di ripartizione di una successione di v.a.

Teorema 9.4.1. (Teorema di continuità) Siano $\{X_n\}_n$ e X v.a. Allora per $n \rightarrow +\infty$, $X_n \Rightarrow X$ se e soltanto se $M_{X_n}(t) \rightarrow M_X(t)$ per ogni t in un opportuno intorno dello 0.

Insieme con il fatto che la funzione generatrice dei momenti è in corrispondenza biunivoca con la funzione di ripartizione, riusciamo ad individuare il limite nel senso della convergenza in legge.

Siamo in condizione di enunciare e dimostrare un altro importante risultato di approssimazione che permette di stimare con maggiore precisione (rispetto alla LGN) la distribuzione della somma di v.a. mediante la distribuzione Gaussiana standard.

Teorema 9.4.2. (Teorema del limite centrale)

Sia X_n una successione di v.a. i.i.d. con media μ e varianza σ^2 finite.

Se indichiamo con $S_n = X_1 + \dots + X_n$, allora

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Rightarrow X \sim \mathcal{N}(0; 1), \quad \text{per } n \rightarrow +\infty$$

Dimostrazione: SPDG andiamo a considerare $\mu = 0$ (altrimenti possiamo ripetere la dimostrazione per la successione di v.a. $\{X_n - \mu\}$).

La funzione generatrice dei momenti di $\frac{S_n}{\sigma\sqrt{n}}$ è

$$M_{\frac{S_n}{\sigma\sqrt{n}}}(t) = M_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = M_{X_1+\dots+X_n}\left(\frac{t}{\sigma\sqrt{n}}\right) \stackrel{\text{ind.}}{=} \prod_{i=1}^n M_{X_i}\left(\frac{t}{\sigma\sqrt{n}}\right) \stackrel{\text{i.d.}}{=} \left(M_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

Ma la v.a. X_1 possiede primo e secondo momento finiti, dunque la sua funzione generatrice dei momenti è derivabile due volte con continuità e possiamo approssimarla mediante lo sviluppo di Taylor intorno a 0 al secondo ordine e per la corrispondenza tra momenti e derivate della funzione generatrice dei momenti, abbiamo

$$\begin{aligned} M_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) &= M_{X_1}(0) + M'_{X_1}(0)\frac{t}{\sigma\sqrt{n}} + M''_{X_1}(0)\frac{t^2}{2\sigma^2 n} + o\left(\frac{1}{n}\right) \\ &= 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \end{aligned}$$

e possiamo concludere

$$\left(M_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n = \left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \rightarrow e^{\frac{t^2}{2}}$$

che è esattamente la funzione generatrice dei momenti di una Gaussiana standard. \square

Osservazione 9.4.1. Il TLC potrebbe essere enunciato direttamente per la media campionaria $\bar{X}_n = \frac{S_n}{n}$ piuttosto che per la somma di v.a. Notiamo infatti che se per ogni i , abbiamo $E(X_i) = \mu$ e $\text{Var}(X_i) = \sigma^2$, allora

$$E(S_n) = n\mu, \quad \text{Var}(S_n) = n\sigma^2, \quad E(\bar{X}_n) = \mu, \quad \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

da cui

$$\frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{\sigma^2 n}} = n \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2 n}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}}$$

Il TLC fornisce delle stime molto più accurate della LGN già usando un campione con $n = 30$ o 50 .

Esempi 9.4.1.

1. Pensiamo di voler valutare se una moneta è equa o meno. Un criterio ragionevole sarebbe voler vedere qual è la probabilità che la media empirica si discosti da 0.5 per più della seconda cifra decimale, ovvero

$$P(|\bar{X}_n - p| > 0.01)$$

stabilendo quanto piccola deve essere questa probabilità per ritenere accettabile l'affermazione che la moneta è equa. In questo esempio decidiamo che la soglia per questa probabilità è 0.003.

Applicando la LGN, sapendo che $p(1-p) \leq \frac{1}{4}$ per ogni $0 < p < 1$, abbiamo visto che

$$P(|\bar{X}_n - p| > 0.01) \leq \frac{10000}{4n} \leq \frac{3}{1000}$$

soltanto se $n > 833334$. Quindi dovremmo chiedere di fare circa 800000 lanci prima di stabilire se per noi una media empirica di 0.49 è accettabile.

Applicando invece il TLC si avrebbe

$$\begin{aligned} P(|\bar{X}_n - p| > 0.01) &= P(\bar{X}_n - p > 0.01) + P(\bar{X}_n - p < -0.01) \\ &= P\left(\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} > \frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) + P\left(\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} < \frac{-0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &\simeq (1 - \Phi\left(\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right)) + \Phi\left(\frac{-0.01\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &= 2(1 - \Phi\left(\frac{0.01\sqrt{n}}{\sqrt{p(1-p)}}\right)) \leq 2(1 - \Phi(2 \cdot 0.01\sqrt{n})). \end{aligned}$$

Dalle proprietà della distribuzione Gaussiana Standard sappiamo che questa probabilità è minore di $2 \cdot 0.0014 = 0.0028$ non appena

$$2 \cdot 0.01\sqrt{n} \geq 3 \quad \Rightarrow n \geq 22500$$

2. La concentrazione di un'inquinante in un campione d'acqua rilevato in una certa zona segue una distribuzione esponenziale di parametro 2. Siano X_1, \dots, X_{100} 100 rilevamenti indipendenti: Calcolare

- (a) la probabilità che una rilevazione superi il livello 1;
- (b) approssimativamente la probabilità che il numero totale di rilevazioni che hanno superato il livello 1 sia maggiore di 20

La prima domanda è di facile risoluzione, poiché $P(X_1 > 1) = e^{-2} \simeq 0.14$

Per la seconda possiamo usare direttamente l'approssimazione Normale, infatti abbiamo delle nuove v.a. Bernoulliane

$$Y_i = \begin{cases} 1 & \text{se } X_i > 1 \\ 0 & \text{se } X_i \leq 1 \end{cases} \quad i = 1, \dots, 100$$

anche queste indipendenti, che segnano 1 ogni volta che il livello di guardia viene superato.

Quindi $S_{100} = \sum_{i=1}^{100} Y_i$ è il numero di volte che si è superata la soglia durante le 100 osservazioni e questa v.a. segue una $\text{Bin}(100, 0.14)$, per cui sappiamo che $E(S_{100}) = 14$ e $\text{Var}(S_{100}) = 11.7$, quindi applicando il TLC abbiamo

$$P(S_{100} > 20) = P\left(\frac{S_{100} - 14}{\sqrt{11.7}} > \frac{20 - 14}{\sqrt{11.7}}\right) \simeq 1 - \Phi(1.75) = 1 - 0.95994 = 0.04006.$$

9.5 Revisione di Teoria

1. Dare la definizione di funzione generatrice dei momenti ed dimostrarne le sue proprietà.
2. Enunciare e dimostrare il Teorema del Limite Centrale.
3. Enunciare e dimostrare la Legge dei Grandi Numeri.
4. Se $X \sim \mathcal{N}(3; 2)$ e $Y \sim \exp(2)$, qual è la funzione generatrice dei momenti di $2X + 3Y$?
5. Siano X_1, X_2 , due v.a. indipendenti rispettivamente $\Gamma(\frac{3}{2}, \frac{3}{2})$ e Gaussiana $\mathcal{N}(-1, 4)$, calcolare la funzione generatrice dei momenti di $2X_1 - 3X_2$.
6. La funzione generatrice dei momenti di due v.a. indipendenti identicamente distribuite, X_1, X_2 , è

$$M(t) = \left(\frac{0.5}{0.5 - t}\right)^2, \quad t < \frac{1}{2},$$

qual è la funzione generatrice dei momenti di $Y = -X_1 + 4X_2 + 1$ e dire dove è definita.

7. Dare la definizione di convergenza in legge. (M)
8. Dare un esempio di v.a. che non convergono in legge. (M)
9. Dare un esempio di v.a. che convergono in legge.
10. Se una sequenza di v.a. indipendenti X_1, \dots, X_n hanno tutte funzione generatrice dei momenti data da

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha$$

per parametri λ, α , qual è la funzione generatrice dei momenti di $X_1 + \dots + X_n$?

9.6 Esercizi

1. Sia X una v.a. con legge $f_X(x) = \frac{1}{2}e^{-|x|}$, calcolarne la funzione generatrice dei momenti e dire dove è definita.
2. In un processo produttivo che impiega certi gas nocivi, si ha un sistema di allarme che scatta ogni qual volta rileva la concentrazione di questi gas oltre una certa soglia. Il sistema di allarme è difettoso e con probabilità $\frac{5}{6}$ scatta ad un effettivo superamento della soglia, mentre con probabilità $\frac{1}{8}$ scatta anche se la soglia non è stata superata. La soglia di pericolo viene superata ogni giorno con probabilità $\frac{1}{25}$.
 - (a) Calcolare la probabilità che in un dato giorno l'allarme scatti.
 - (b) Ogni giorno si ripete indipendentemente dagli altri, calcolare approssimativamente la probabilità che l'allarme scatti 20 volte in 100 giorni.
3. Data una v.a. $\mathcal{N}(\mu; \sigma^2)$ calcolarne il momento quarto.
4. Sapendo che il momento quarto di qualsiasi v.a. $\mathcal{N}(\mu; \sigma^2)$ è finito, considerare una successione di v.a. indipendenti $\{X_n\}_n$ tutte $\mathcal{N}(0; 4)$ e calcolare il limite di

$$\frac{X_1^2 + \dots + X_n^2}{n}$$

5. Sia $T \sim \exp(0.2)$
 - (a) Calcolare la densità, la media e la varianza di T^2 .
 - (b) Supponiamo di avere una successione $\{X_n\}$ di v.a. indipendenti tutte con la stessa distribuzione di T^2 . Calcolare approssimativamente per $S_n = X : 1 + \dots + X_n$

$$P\left(\frac{S_{100}}{10} \leq 2\right).$$

6. La statura X dei ragazzi che si presentano alla visita di leva è una v.a. di legge Gaussiana $\mathcal{N}(178\text{cm}, 36\text{cm}^2)$. Un addetto misura le altezze con uno strumento mal tarato, commettendo un errore ulteriormente Gaussiano $W \sim \mathcal{N}(-2\text{cm}, 9\text{cm}^2)$. I ragazzi si presentano indipendentemente alla visita.
 - (a) Calcolare la probabilità che l'altezza misurata sia superiore a 176 cm.
 - (b) Si ripete la misura 100 volte, stimare la probabilità che la media empirica sia maggiore di 178cm.
7. La concentrazione di un'inquinante in un campione d'acqua rilevato in una certa zona segue una distribuzione esponenziale di parametro 0.5. Siano X_1, \dots, X_{100} 100 rilevamenti indipendenti: Calcolare
 - (a) la probabilità che una rilevazione superi il livello 1;

- (b) approssimativamente la probabilità che il numero totale di rilevazioni che hanno superato il livello 1 sia maggiore di 30

8. Due v.a. X ed Y indipendenti sono distribuite secondo una $\Gamma(2; 8)$.

- (a) Calcolare la legge di $X + Y$.
 (b) Siano $(X_1, Y_1), (X_2, Y_2) \dots$ copie indipendenti delle precedenti v.a. e sia

$$Z_n = \sum_{k=1}^n (X_k + Y_k)$$

calcolare approssimativamente $P(Z_{100} - 40 \geq 0)$.

9. Sia $\{X_n\}_{n \in \mathbb{N}}$ una successione di v.a. i.i.d. esponenziali di parametro $\lambda > 0$ e si ponga $Y_n = \frac{X_n}{\log n}$, $n \geq 2$.

- (a) Individuare la legge di Y_n .
 (b) Dire se Y_n converge in legge.

10. Siano X_1, \dots, X_{500} v.a. indipendenti tutte geometriche modificate di parametro $\frac{1}{3}$.

Dare una stima della probabilità che la loro media empirica $\bar{X}_{500} = \sum_{i=1}^{500} X_i$ si discosti dalla sua media più di $\frac{1}{10}$.

11. Il prezzo X di un certo bene è una v.a. di media 10 e varianza 2. Quando questo prezzo viene rilevato viene commesso un errore di arrotondamento rappresentato da una v.a. Y indipendente da X e che segue una legge Gaussiana $\mathcal{N}(0, 0.5)$. Il prezzo viene rilevato in maniera indipendente per 36 giorni lavorativi

Stimare la probabilità che la media empirica dei prezzi risulti inferiore a 9.

12. Sia X una Gaussiana $\mathcal{N}(0, 1)$ e $Y = \exp(X)$.

- (a) Calcolare $E(Y)$ e $\text{Var}(Y)$.
 (b) Siano Y_1, \dots, Y_{40} v.a. indipendenti tutte con la stessa distribuzione di Y . Calcolare APPROSSIMATIVAMENTE $P(\bar{Y}_{40} \leq 1)$, dove

$$\bar{Y}_{40} = \frac{Y_1 + \dots + Y_{40}}{40}.$$

Capitolo 10

ESERCIZI DI RIEPILOGO

1. Dimostrare che per due v.a. X ed Y indipendenti assolutamente continue, vale

$$E(XY) = \mathbb{E}(X)E(Y).$$

2. Se $X \sim \exp(6)$ e $Y \sim \mathcal{N}(-3, 4)$ sono due v.a. indipendenti, qual è la funzione generatrice dei momenti della v.a. $Z = -2X + 4Y$?
3. Siano X_1, \dots, X_{20} v.a. indipendenti $\Gamma(0.5, 2)$, qual è la media di $X_1 + \dots + X_{20}$?
4. Due servizi A, B sono gestiti ognuno da due individui, A_1, A_2, B_1, B_2 con rispettivi tempi di servizio $T_{A_1} \sim \exp(\frac{1}{5})$, $T_{A_2} \sim \exp(\frac{1}{10})$, $T_{B_1} \sim \exp(\frac{1}{6})$, $T_{B_2} \sim \exp(\frac{1}{8})$ tra loro tutti indipendenti.

Il primo dei due operatori che si libera al servizio A , serve il cliente successivo. Lo stesso succede per il servizio B .

- (a) Un cliente arriva al servizio A . In media, quanto tempo dovrà aspettare per essere servito? E per il servizio B ?
 - (b) Un cliente arriva e, dovendo usufruire di entrambi i servizi, sceglie a caso a quale mettersi in fila. Con che probabilità dovrà aspettare meno di 7 minuti prima di essere servito?
 - (c) Se il cliente ha atteso meno di 7 minuti, qual è la probabilità che abbia scelto il servizio B ?
5. Il tempo (in secondi) di un certo gruppo di nuotatori in una specialità, T , è distribuito secondo una $\Gamma(320, 8)$.

Calcolare approssimativamente la probabilità che T sia maggiore od uguale a 45.

Supponendo di avere un campione di 32 nuotatori e che i risultati siano tutti indipendenti l'uno dall'altro, calcolare approssimativamente la probabilità che la media empirica dei 32 risultati risulti minore od uguale a 44.

6. Un test a risposte multiple contiene 40 domande ognuna con 4 opzioni. Ogni domanda ha una singola risposta esatta. Per passare il test occorre rispondere correttamente ad almeno 24 domande.
- (a) Rispondendo a caso a tutte e 40 le domande, in media quante risposte corrette può dare uno studente?
 - (b) **APPROSSIMATIVAMENTE**, rispondendo a caso a tutte le domande, uno studente che probabilità ha di passare il test?
7. In un gioco elettronico passa il turno chi supera un certo livello. Vi sono due giocatori A e B in competizione tra loro. Ad ogni partita, indipendentemente dalle altre, A ha probabilità 70% di superare il livello, B ha invece una probabilità del 80% di superarlo. Ognuno di loro riprova finchè non supera il livello. Ci si ferma quando uno dei due lo supera.
- (a) Se T indica il numero di partite necessarie perchè uno dei due giocatori passi il turno, qual è la sua distribuzione?
 - (b) Si gioca un torneo tra 40 coppie di giocatori, (A_1, B_1) , ..., (A_{40}, B_{40}) , tutte indipendenti e che si comportano nella stessa maniera dei giocatori A, B descritti prima.
Siano T_1, \dots, T_{40} il numero di partite giocate dalle rispettive coppie per avere il giocatore che passa al turno successivo. Calcolare **APPROSSIMATIVAMENTE** la probabilità che il numero totale di partite giocate per avere i 40 componenti del turno successivo sia inferiore a 100.
8. Vi sono tre casse del supermercato A, B, C , che servono ogni cliente con rispettivi tempi indipendenti $\tau_A \sim \exp(\frac{1}{5})$, $\tau_B \sim \exp(\frac{1}{7})$, $\tau_C \sim \exp(\frac{1}{8})$. Il tempo è calcolato in minuti.
- Arrivano alle casse 4 clienti. I primi tre si distribuiscono sulle tre casse contemporaneamente.
- (a) Quanto dovrà aspettare in media il quarto cliente in fila, se la fila è unica?
 - (b) Quanto dovrà aspettare il quarto cliente, se la fila non è unica e sceglie a caso dove mettersi in fila?
9. Il tempo di vita di un certo tipo di lampadine è rappresentato da una v.a. $T \sim \Gamma(28000, 10)$. Il tempo è calcolato in ore.
- (a) Calcolare la media e la varianza di T .
 - (b) Calcolare **approssimativamente** la probabilità che la lampadina funzioni ancora dopo 2850 ore.
10. Siano $X \sim \mathcal{N}(2, 2)$ e $Y \sim \exp(\frac{1}{4})$ tra loro indipendenti. Calcolare la retta di regressione lineare di $Z = XY$ su $W = X - Y$.

11. Il tempo in secondi di un atleta sui 400 m piani è rappresentato da una v.a. con distribuzione $\Gamma(460, 11.5)$. Il cronometro che usa per gli allenamenti presenta un errore di misurazione distribuito secondo una v.a. $\mathcal{N}(0, 2)$.
- (a) Indichiamo con T il tempo dell'atleta misurato dal cronometro. Calcolare la sua media e la sua varianza.
 - (b) L'atleta si allena in 6 mesi per 100 volte, sempre nelle stesse condizioni e in maniera indipendente un giorno dall'altro. Calcolare approssimativamente la probabilità che la media empirica dei tempi misurati nei 100 allenamenti sia maggiore della media effettiva dell'atleta di più di un decimo di secondo?

Indice

1	INTRODUZIONE	1
2	SPAZI DI PROBABILITÀ	2
2.1	Definizioni	2
2.2	Spazi di probabilità equiprobabili	8
2.2.1	Pillole di combinatoria	9
2.3	Revisione di teoria	13
2.4	Esercizi	14
3	PROBABILITÀ CONDIZIONATA ED INDIPENDENZA	15
3.1	Probabilità condizionata	15
3.2	Indipendenza di eventi	18
3.3	Revisione di Teoria	21
3.4	Esercizi	21
4	VARIABILI ALEATORIE DISCRETE	24
4.1	Definizione e densità di probabilità	24
4.2	Modelli estrattivi	29
4.2.1	Estrazioni con reinserimento	30
4.2.2	Estrazioni senza reinserimento	33
4.3	Altre distribuzioni	37
4.4	Densità condizionate e indipendenza	40
4.5	Calcoli con densità congiunte	42
4.6	Revisione di Teoria	45
4.7	Esercizi	46
5	MEDIA E MOMENTI	47
5.1	La Media Matematica	47
5.1.1	Proprietà della media	48
5.1.2	Calcolo delle medie delle principali distribuzioni	49
5.2	Momenti, Varianza e Covarianza	51
5.2.1	Calcolo delle varianze delle principali distribuzioni	56
5.3	Retta di regressione lineare	58
5.4	Revisione di Teoria	59

5.5	Esercizi	60
6	ESERCIZI DI RIEPILOGO	63
7	VARIABILI ALEATORIE CONTINUE	67
7.1	Funzioni di ripartizione e di densità	67
7.2	Trasformazioni di variabili aleatorie	75
7.3	Media e momenti	78
7.4	Le principali densità assolutamente continue e i loro momenti	80
7.4.1	La densità esponenziale	80
7.4.2	La densità Gaussiana	81
7.4.3	La densità Gamma	83
7.5	Applicazioni	85
7.6	Revisione di Teoria	87
7.7	Esercizi	88
8	VARIABILI ALEATORIE A PIÙ DIMENSIONI	90
8.1	Alcuni esempi	90
8.2	Trasformazioni bidimensionali: max, min e somma	93
8.3	Densità e Media condizionata	98
8.4	Revisione di Teoria	101
8.5	Esercizi	101
9	SUCCESSIONI DI V.A., FUNZIONI GENERATRICI, LGN, TLC	103
9.1	Nozioni di convergenza (M)	103
9.2	La Legge dei Grandi Numeri	105
9.3	Funzioni Generatrici	111
9.4	Il Teorema del limite centrale	113
9.5	Revisione di Teoria	116
9.6	Esercizi	117
10	ESERCIZI DI RIEPILOGO	119
11	APPENDICE A - TEORIA DEGLI INSIEMI	124

Capitolo 11

APPENDICE A - TEORIA DEGLI INSIEMI

Insiemi e operazioni su insiemi

Un insieme è una collezione di oggetti ed è usualmente denotato con una lettera maiuscola, A, B, C, \dots , gli oggetti nell'insieme sono detti **elementi** e sono usualmente denotati con le lettere minuscole a, b, c, \dots . Diciamo che l'elemento a appartiene all'insieme A e si scrive $a \in A$. Scriviamo invece $a \notin A$ se a non si trova in A . L'insieme senza elementi è detto **insieme vuoto** ed è indicato con il simbolo \emptyset .

Gli insiemi possono essere descritti:

- elencando gli elementi che vi appartengono $A = \{a, b, c, \dots\}$;
- o attraverso una proprietà

$$M = \{\text{studenti Univaq che frequentano il corso di probabilità}\}.$$

Il numero di elementi di un insieme è detto **cardinalità** dell'insieme.

Definizione 11.0.1.

1. Se tutti gli elementi di A appartengono anche a B , diciamo che A è un **sottoinsieme** di B e scriviamo $A \subseteq B$.

$$A \subseteq B, \quad \text{se } \forall x \in A \Rightarrow x \in B$$

2. Diciamo **intersezione** di due insiemi, $A \cap B$, l'insieme degli elementi che appartengono ad entrambi

$$A \cap B = \{x \text{ tali che } x \in A \text{ e } x \in B\},$$

due insiemi si dicono **disgiunti** se $A \cap B = \emptyset$ (non hanno elementi in comune).

3. Diciamo **unione** di due insiemi, $A \cup B$, l'insieme di tutti gli elementi che sono in A o in B (o in entrambi)

$$A \cup B = \{x \text{ tali che } x \in A \text{ e/o } x \in B\},$$

4. Se abbiamo un insieme universale, Ω , dove si trovano tutti gli insiemi allora, dato un insieme $A \subseteq \Omega$, definiamo il **complementare** di A , come l'insieme costituito da tutti gli elementi che non sono in A

$$A^c = \{x \in \Omega, x \notin A\}.$$

In generale possiamo definire la differenza tra A e B , $A \setminus B$, come l'insieme degli elementi di A che non sono in B .

5. Dati due insiemi A e B , diciamo **Prodotto Cartesiano**

$$A \times B = \{(a, b) : a \in A, b \in B\},$$

l'insieme delle coppie ordinate in cui il primo elemento appartiene ad A ed il secondo a B . Il prodotto Cartesiano può essere esteso a qualsiasi numero di insiemi. $A_1 \times A_2 \times \dots \times A_n$ è l'insieme costituito dalle n -ple ordinate (a_1, a_2, \dots, a_n) , dove il primo elemento $a_1 \in A_1$... fino all'ultimo $a_n \in A_n$.

Esempi 11.0.1.

1. L'insieme precedente, M , è un sottoinsieme di $L = \{\text{studenti di UNIVAQ}\}$.

2. Se $N = \{\text{studenti Univaq che frequentano il corso di fisica}\}$, allora

$$M \cap N = \{\text{studenti Univaq che frequentano il corso di probabilità E il corso di fisica}\}$$

$$M \cup N = \{\text{studenti Univaq che frequentano il corso di probabilità E/O il corso di fisica}\}.$$

3. Naturalmente abbiamo per qualsiasi coppia di insiemi

$$A \cap B \subseteq A, B \subseteq A \cup B.$$

4. Con la notazione precedente

$$M^c = \{\text{studenti Univaq che NON frequentano il corso di probabilità}\}$$

$$(M \cap N)^c = \{\text{studenti Univaq che NON frequentano probabilità E/O fisica}\}$$

5. Se abbiamo i tre insiemi

$$A = \{\text{valori percentuali del tasso di inflazione nazionale}\},$$

$$B = \{\text{valori percentuali del tasso di disoccupazione nazionale}\},$$

$$C = \{\text{valori percentuali del rapporto tra debito pubblico e Prodotto interno lordo}\},$$

allora la tripla $(3, 10, 1.2)$ potrebbe essere la descrizione dello stato economico di un paese.

Proprietà delle operazioni su insiemi

1. Intersezione

$$A \cap A = A$$

$$A \cap B = B \cap A$$

$$A \cap B \subseteq A; A \cap B \subseteq B$$

$$A \cap \emptyset = \emptyset$$

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

2. Unione

$$A \cup A = A$$

$$A \cup B = B \cup A$$

$$A \subseteq A \cup B; B \subseteq A \cup B$$

$$A \cup \emptyset = A$$

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

3. Proprietà distributive

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

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

4. Complementare

$$A \cup A^c = \Omega, \quad A \cap A^c = \emptyset$$

$$(A^c)^c = A, \quad A \subseteq B \Rightarrow B^c \subseteq A^c$$

$$(A \cup B)^c = A^c \cap B^c, \quad (A \cap B)^c = A^c \cup B^c$$

Dimostrazione:

$$(A \cup B)^c = \{x \in \Omega : x \notin A \cup B\} = \{x \in \Omega : x \notin A \text{ e } x \notin B\}$$

$$A^c \cap B^c = \{x \in \Omega : x \in A^c \text{ e } x \in B^c\} = \{x \in \Omega : x \notin A \text{ e } x \notin B\}$$

$$(A \cap B)^c = \{x \in \Omega : x \notin A \cap B\} = \{x \in \Omega : x \notin A \text{ e/o } x \notin B\},$$

$$A^c \cup B^c = \{x \in \Omega : x \notin A \text{ e/o } x \notin B\}$$

Esempio 11.0.1. : Siano $A = \{1, 2, 3\}$ e $B = \{2, 3, 5\}$, allora

$$A \cap B = \{2, 3\}, \quad A \cup B = \{1, 2, 3, 5\}, \quad A \setminus B = \{1\}.$$

Bibliografia

- [1] P. Baldi: Calcolo delle probabilità , Ed. McGraw-Hill Education (2011)