Appunti di Probabilità e Statistica

Riccardo Agatea

15 maggio 2019

Indice

1 Introduzione

La probabilità si prefigge di studiare eventi casuali in modo matematico. Per evento casuale non si intende solo un evento prettamente aleatorio, come il lancio di un dado, ma anche un evento essenzialmente deterministico che richiederebbe, per essere previsto, lo studio di un sistema complesso, influenzato da un numero così alto di variabili, che risulta più semplice ed efficiente studiarlo come se fosse casuale. Inoltre, in alcune situazioni la probabilità può intervenire nonostante il caso sia completamente assente: uno degli esempi più importanti è l'integrazione con il metodo Monte Carlo. Considerato un integrale

$$I = \int_0^1 f(x)dx$$

difficilmente risolvibile individuando un'antiderivata di $f(x)$, dal teorema della media integrale è noto che

$$\exists \xi \in [0,1]\colon \int_0^1 f(x)dx = f(\xi)$$

Per ricavare una stima di I è possibile considerare un insieme $X = \{x_1, x_2, ..., x_n\}s$ di valori casuali tale che $X \subseteq [0,1]$. La media dei valori assunti dalla funzione in questi punti, per le leggi dei grandi numeri, tende a I per $n \to +\infty$:

$$\frac{1}{n} * \sum_{i=1}^{n} f(x_i) \xrightarrow[n\to+\infty]{} I$$

2 Spazio di Probabilità

2.1 Definizioni

Definizione 2.1. Sia Ω un insieme non vuoto, sia $\mathscr{F}$ una famiglia di sottoinsiemi di Ω (cioè un insieme di sottoinsiemi di Ω), e sia

$$P : \mathscr{F} \to [0,1]$$

una funzione tale che $P(\Omega) = 1$. Si chiama spazio di probabilità la terna ordinata

$$(\Omega, \mathscr{F}, P)$$

- Ω si chiama spazio degli eventi elementari, contiene tutti i risultati possibili di un dato esperimento. I suoi elementi si dicono eventi elementari o esiti specifici.
- $\mathscr{F}$ ha la struttura di una σ-algebra (concetto che verrà studiato più avanti); intuitivamente contiene tutti gli eventi che si vogliono studiare in senso probabilistico.
- P si dice misura di probabilità. Se $E \in \mathscr{F}$, e quindi E è un evento, $P(E)$ è la probabilità che E si realizzi. Si intende che se $P(E) = 0$ allora E non si verifica mai, se $P(E) = 1$ allora E si verifica sempre.

NB. Per uno stesso esperimento Ω, $\mathscr{F}$ e P non sono necessariamente unici. Generalmente si cerca di scegliere Ω in modo che sia il più ridotto possibile, per cercare di minimizzare la complessità degli eventi da studiare.

Sono dati alcuni esempi di esperimenti, e dei relativi spazi degli eventi elementari.

Esempio 2.1. Esperimento: lancio di una moneta.

$$\Omega = \{T, C\}$$

Esempio 2.2. Esperimento: 3 lanci consecutivi di una moneta.

$$\Omega = \{TTT, TTC, TCT, TCC, CTT, CTC, CCT, CCC\}$$

Esempio 2.3. Esperimento: conteggio del numero di volte in cui esce testa su 3 lanci di una moneta.

$$\Omega = \{0, 1, 2, 3\}$$

Esempio 2.4. Esperimento: conteggio del numero di lanci necessari affinchè esca testa la prima volta.

$$\Omega = \{1, 2, 3, ...\} = \mathbb{N}$$

Esempio 2.5. Esperimento: misura della durata di vita di una lampadina (supponendo che la sensibilità del sistema di misura sia essenzialmente nulla, cioè che le misure siano corrette fino all'infinitesima cifra significativa).

$$\Omega = \mathbb{R}_+ \quad \text{oppure} \quad \Omega = [0, a] \subsetneq \mathbb{R}_+$$

Introduciamo alcuni concetti:

Definizione 2.2. Uno spazio di probabilità $(\Omega, \mathscr{F}, P)$ si dice discreto se Ω è finito o numerabile:

$$|\Omega| \leqslant |\mathbb{N}|$$

Definizione 2.3. Uno spazio di probabilità $(\Omega, \mathscr{F}, P)$ si dice continuo se Ω è non numerabile:

$$|\Omega| = |\mathbb{R}|$$

Definizione 2.4. Sia $\omega \in \Omega$ il risultato osservato dall'esecuzione di un dato esperimento, e sia $E \in \mathscr{F}$ un evento. Si dice che E si è verificato se e solo se

$$\omega \in E$$

Esempio 2.6. Esperimento: 3 lanci consecutivi di una moneta. Consideriamo:

$$\Omega = \{TTT, TTC, TCT, TCC, CTT, CTC, CCT, CCC\}$$
$$E = \{\omega \in \Omega \mid \text{testa e croce si alternano}\}$$

Dall'esecuzione dell'esperimento, si osserva $\omega = TTC$.

$$\omega \notin E \Rightarrow E \text{ non si è verificato}$$

Ripetendo l'esperimento si osserva $\omega = TCT$.

$$\omega \in E \Rightarrow E \text{ si è verificato}$$

Due particolari eventi sempre presenti in $\mathscr{F}$ sono $\varnothing$ e Ω:

- $\varnothing$ rappresenta il non presentarsi di alcuno degli eventi elementari in Ω. Chiaramente almeno uno degli eventi elementari si deve presentare, quindi $\varnothing$ non si verifica mai.
- Ω rappresenta il presentarsi di uno degli eventi elementari, senza limitazioni, e quindi si verifica sempre.

2.2 Operazioni sugli Insiemi

Consideriamo uno spazio di probabilità $(\Omega, \mathscr{F}, P)$, e due eventi $E, F \in \mathscr{F}$. Essendo $\mathscr{F}$ una famiglia di sottoinsiemi di Ω, E e F sono insiemi, e quindi si possono applicare ad essi le comuni operazioni sugli insiemi. Quelli che si ottengono sono ancora eventi di $\mathscr{F}$, di cui si può studiare il verificarsi in funzione del verificarsi di E e F.

- $E \cap F$ si verifica $\Leftrightarrow$ sia E che F si verificano
- $E \cup F$ si verifica $\Leftrightarrow$ almeno uno fra E e F si verifica
- E^C si verifica $\Leftrightarrow$ E non si verifica
- $E \smallsetminus F$ si verifica $\Leftrightarrow$ E si verifica e F non si verifica
- $E \Delta F$ si verifica $\Leftrightarrow$ solo uno fra E ed F si verifica

Definizione 2.5. E ed F si dicono incompatibili se sono disgiunti:

$$E \cap F = \varnothing$$

Chiaramente, per gli elementi di $\mathscr{F}$ continuano a valere la proprietà distributiva dell'intersezione sull'unione e dell'unione sull'intersezione, e le due leggi di De Morgan. Usando le leggi di De Morgan, è sempre possibile esprimere un'espressione in funzione del complemento e di una fra unione e intersezione.

Partizioni Ricordiamo la definizione di partizione di un insieme:

Definizione 2.6. Sia A un insieme, sia $(A_i)_i$ una famiglia di suoi sottoinsiemi. $(A_i)_i$ si dice partizione di A se e solo se

$$\bigcup_i A_i = A$$
$$A_i \cap A_j = \varnothing \quad \forall i \neq j$$

Se l'insieme considerato è uno spazio degli eventi Ω, allora una sua partizione rappresenta una famiglia di eventi incompatibili, ma per i quali c'è la garanzia che uno e uno solo di essi si verifichi. Dato un evento $F \subseteq \Omega$ e una partizione $(E_i)_i$ di Ω, è sempre possibile *scomporre* F rispetto $(E_i)_i$, cioè si può sempre rappresentare F come l'unione delle intersezioni di F con gli insiemi di $(E_i)_i$:

$$F = \bigcup_i (F \cap E_i)$$

Chiaramente, siccome $(E_i)_i$ è una partizione di Ω, $(F \cap E_i)_i$ è una partizione di F. In modo simile, è sempre possibile scomporre l'unione di due eventi in una sua partizione: se $E, F \subseteq \Omega$ allora

$$E \cup F = (E \cap F^C) \cup (E^C \cap F) \cup (E \cap F)$$

Concettualmente, significa che "o si verifica solo E, o si verifica solo F, oppure si verificano entrambi". Un'altra scomposizione, più generalizzabile è:

$$E \cup F = E \cup (F \smallsetminus E)$$

In questo caso, il significato è "o si verifica E, e non ci interessiamo di F, oppure se E non si verifica, allora si verifica F". Questa scomposizione si può generalizzare alla generica famiglia di eventi $\{E_n\}_n$:

$$\bigcup_{k=1}^{n} E_k = E_1 \cup \left(\bigcup_{k=2}^{n} \left(E_k \smallsetminus \bigcup_{i=1}^{k-1} E_j \right) \right)$$

Esercizio 2.1. Dati tre eventi E, F, G, trovare l'evento corrispondente a

- si verifica solo E
- si verificano E e F, ma non G
- si verifica esattamente uno degli eventi E, F, o G
- si verifica almeno uno dei tre
- si verificano esattamente due degli eventi

2.3 Spazio degli Eventi

Abbiamo detto che $\mathscr{F}$ è una famiglia di sottoinsiemi di Ω che contiene tutti gli eventi che ci interessa analizzare. Diamo una definizione formale di evento:

Definizione 2.7. Un sottoinsieme E di Ω si dice evento se e solo se $E \in \mathscr{F}$.

Il fatto che $\mathscr{F}$ contenga gli eventi che ci interessano ha delle conseguenze, nello specifico significa che se un evento è presente devono essere presenti anche tutti gli eventi "equivalenti". Ad esempio, se l'evento "è uscita sempre testa" è presente, anche "non è uscita mai croce" deve essere presente, perchè danno la stessa informazione. Per questo $\mathscr{F}$ è una σ-algebra:

Definizione 2.8. Dato un insieme A e una famiglia di suoi sottoinsiemi $\mathscr{A} \subseteq \mathbb{P}(A)$, $\mathscr{A}$ è una σ-algebra se e solo se:

1. $\mathscr{A} \neq \varnothing$
2. $\forall B \subseteq A \quad B \in \mathscr{A} \Rightarrow B^C \in \mathscr{A}$

3. Presa comunque una famiglia numerabile $(B_n)_n$ di sottoinsiemi di A, vale

$$\forall n \; B_n \in \mathscr{A} \Rightarrow \bigcup_i B_i \in \mathscr{A}$$

Bastano queste condizioni perchè, come abbiamo detto, tutte le operazioni fra insiemi possono essere ridotte a combinazioni di complemento e unione. Due esempi particolarmente importanti di σ-algebre su un dato spazio degli eventi elementari Ω sono:

Esempio 2.7.

$$\mathscr{F} = \{\varnothing, \Omega\}$$

Questa è una σ-algebra perchè:

1. Chiaramente non è vuota
2. $\varnothing^C = \Omega \in \mathscr{F}$ e $\Omega^C = \varnothing \in \mathscr{F}$
3. $\varnothing \cup \Omega = \Omega \in \mathscr{F}$

Esempio 2.8.

$$\mathscr{F} = \mathbb{P}(\Omega)$$

Con $\mathbb{P}(\Omega)$ indichiamo l'insieme delle parti di Ω. La dimostrazione è triviale.

Definizione 2.9. Sia $E \subseteq \Omega$, la σ-algebra generata da E è

$$\mathscr{F} = \{\varnothing, \Omega, E, E^C\}$$

Si dimostra che è una σ-algebra:

Dimostrazione. 1. Trivialmente, $\mathscr{F} \neq \varnothing$

2.
$$\begin{aligned} \varnothing^C &= \Omega \in \mathscr{F} \\ \Omega^C &= \varnothing \in \mathscr{F} \\ E^C &\in \mathscr{F} \\ (E^C)^C &= E \in \mathscr{F} \end{aligned}$$

3. Per dimostrare che $\mathscr{F}$ è chiusa rispetto l'unione non è necessario considerare le unioni con Ω, perchè $\Omega \cup x = \Omega \; \forall x$. Non è necessario considerare neanche $\varnothing$, perchè esso è l'elemento neutro dell'unione, e quindi $\varnothing \cup x \in \mathscr{F} \Leftrightarrow x \in \mathscr{F} \; \forall x$. Resta da dimostrare che $E \cup E^C \in \mathscr{F}$, che è triviale in quanto $E \cup E^C = \Omega$.

Le proprietà di una σ-algebra sono:

- $\varnothing, \Omega \in \mathscr{F} \quad \forall \mathscr{F}$

  Dimostrazione.

$$\begin{aligned} &\mathscr{F} \neq \varnothing \\ \Rightarrow &\exists G \in \mathscr{F} \\ \Rightarrow &G^C \in \mathscr{F} \\ \Rightarrow &G \cup G^c \in \mathscr{F} \\ \Rightarrow &\Omega \in \mathscr{F} \\ \Rightarrow &\Omega^C = \varnothing \in \mathscr{F} \end{aligned}$$

- $\mathscr{F}$ è chiusa rispetto le unioni finite

Dimostrazione. Sia $\{E_1, ..., E_n\} \subseteq \mathscr{F}$ una famiglia finita di sottoinsiemi di Ω. Questa può essere estesa definendo $E_i = \varnothing \ \forall i > n$; segue:

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i \in \mathbb{N}} B_i \in \mathscr{F}$$

- $\mathscr{F}$ è chiusa rispetto le intersezioni

Dimostrazione. Sia $(E_n)_n \subseteq \mathscr{F}$ una famiglia numerabile di sottoinsiemi di Ω.

$$\begin{aligned}
&\forall i \ E_i \in \mathscr{F} \\
&\Rightarrow \forall i \ E_i^C \in \mathscr{F} \\
&\Rightarrow \bigcup_i E_i^C \in \mathscr{F} \\
&\Rightarrow (\bigcup_i E_i^C)^C \in \mathscr{F} \\
&\Rightarrow \bigcap_i E_i \in \mathscr{F}
\end{aligned}$$

Se Ω è discreto, allora si può sempre prendere $\mathscr{F} = \mathbb{P}(\Omega)$, ed è quello che si fa di solito.

2.4 Misura di probabilità

Diamo una definizione di misura di probabilità:

Definizione 2.10. Una funzione

$$P : \mathscr{F} \to [0, 1]$$

si dice misura di probabilità se

1. $P(\Omega) = 1$
2. P è σ-additiva, cioè: sia $(E_i)_i \subseteq \mathscr{F}$ una famiglia numerabile di eventi mutualmente incompatibili, allora

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

Le proprietà di P sono:

- $\forall E \in \mathscr{F} \quad P(E^C) = 1 - P(E)$

Dimostrazione.

$$\begin{aligned}
&E \cup E^C = \Omega \\
&\Rightarrow P(E \cup E^C) = P(\Omega) \\
&\Rightarrow P(E) + P(E^C) = 1 \\
&\Rightarrow P(E^C) = 1 - P(E)
\end{aligned}$$

- $\forall E, F \in \mathscr{F} \quad E \subseteq F \Rightarrow P(E) \leqslant P(F)$

Dimostrazione.

$$\begin{aligned}
F &= (F \cap E) \cup (F \cap E^C) \\
&= E \cup (F \cap E^C) \\
\Rightarrow P(F) &= P(E) + P(F \cap E^C)
\end{aligned}$$

Essendo $P(F \cap E^C) \geqslant 0$, segue la tesi.

- $\forall E, F \in \mathscr{F} \quad P(E \cup F) = P(E) + P(F) - P(E \cap F)$

  Dimostrazione. Per definizione di misura di probabilità, P è additiva su unioni disgiunte, quindi:

  $$E \cup F = E \cup (F \setminus E)$$
  $$\Rightarrow P(E \cup F) = P(E) + P(F \setminus E)$$

  Ricordiamo che, per definizione:

  $$F \setminus E = F \cap E^C$$

  Inoltre:

  $$F = (F \cap E) \cup (F \cap E^C)$$

  Sostituendo:

  $$F \setminus E = F \cap E^C$$

  (ecc)

- La precedente si può generalizzare a una famiglia $\{E_n\}_n$ di eventi non necessariamente disgiunti:

  $$P(\bigcup_n E_n) \leqslant \sum_n P(E_n)$$

  Dimostrazione. (ecc)

- $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$

  Dimostrazione. (per casa)

- $P(E \Delta F) = P(E) + P(F) - 2P(E \cap F)$

Chiaramente, P non è determinata univocamente da Ω e $\mathscr{F}$, però ci sono dei metodi "standard" per costruirla, soprattutto nelle applicazioni pratiche.

2.4.1 Misura Empirica

Consideriamo una coppia $(\Omega, \mathscr{F})$ fissata in modo da rappresentare un esperimento ripetibile. Eseguendo l'esperimento n volte si otterrà una sequenza di n risultati

$$(\omega_1, ..., \omega_n) \quad \omega_i \in \Omega \ \forall i \in \{1, ..., n\}$$

Sia $E \in \mathscr{F}$, definiamo una quantità n_E che rappresenti il numero di volte in cui E si è verificato

$$n_E := |\{i \in \{1, ..., n\} : \omega_i \in E\}|$$

Definizione 2.11. Si dice misura di probabilità empirica la funzione

$$\hat{P}_n : \mathscr{F} \to [0, 1]$$
$$E \mapsto \frac{n_E}{n}$$

Questa misura in modo empirico il numero di volte in cui un dato evento si è verificato. Il limite più importante è che l'esperimento deve essere ripetibile.

2.4.2 Spazio campionario discreto

Nel caso in cui Ω sia discreto e $\mathscr{F} = \mathbb{P}(\Omega)$, si può dimostrare che per identificare univocamente P è sufficiente identificare i valori p_ω tali che

$$p_\omega = P(\{\omega\}) \quad \text{con} \sum_{\omega\in\Omega} p_\omega = 1$$

Segue dalla σ-additività di P che:

$$P(E) = \sum_{\omega\in E} p_\omega \quad \forall E \in \mathscr{F}$$

Si ricava che se Ω è finito

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

identificare P equivale ad identificare una famiglia

$$\{p_1, ..., p_n\}$$

tale che

$$\sum_{i=1}^{n} p_i = 1$$

Osservazione 2.1. In realtà sono sufficienti $n-1$ valori p_i, perchè vale necessariamente:

$$p_n = 1 - \sum_{i=1}^{n-1} p_i$$

Se invece Ω è numerabile, identificare P equivale ad identificare una successione $(p_n)_n$ tale che

$$\sum_{i=1}^{+\infty} p_i = 1$$

2.4.3 Probabilità Uniforme

Un caso particolare del punto precedente si ha se Ω è finito e tutti gli eventi elementari hanno la stessa probabilità:

Definizione 2.12. Sia Ω spazio di probabilità finito, la funzione di probabilità uniforme è la misura di probabilità identificata dalla famiglia di valori p_i tali che:

$$p_i = \frac{1}{n} \quad \forall i \in \{1, ..., n\}$$

In questo caso, vale che dato un evento $E \in \mathscr{F}$:

$$\begin{aligned} P(E) &= \sum_{\omega\in E} P(\{\omega\}) \\ &= \sum_{\omega\in E} p_\omega \\ &= \sum_{\omega\in E} \frac{1}{n} \\ &= \frac{1}{n} * \sum_{\omega\in E} 1 \\ &= \frac{1}{n} * |E| \\ &= \frac{|E|}{|\Omega|} \end{aligned}$$

In altre parole, la probabilità di un evento E è pari al numero di risultati favorevoli ad E diviso per il numero totale di risultati.

3 Approfondimenti sulla Probabilità

3.1 Principio Fondamentale del Conteggio

Principio (Fondamentale del Conteggio)**.** *Dovendo eseguire R scelte consecutive tali che*

1. $\forall r \in \{1, ..., R\}$ *la scelta va eseguita fra n_r elementi*
2. $\forall r \in \{1, ..., R\}$ *la scelta non dipende dalle precedenti*
3. *da due successioni distinte di R scelte si giunge a risultati diversi*

Il numero totale di risultati possibili è

$$\prod_{i=1}^{R} n_i$$

In generale, i problemi in cui interviene il principio fondamentale del conteggio riguardano il campionamento, secondo uno schema riconducibile all'estrazione di k palline da un'urna che ne contiene n. Ci sono due possibilità riguardo le palline, ed altrettante riguardo l'estrazione:

- Per quanto riguarda le palline, possono essere:
  - tutte diverse (e quindi, virtualmente numerabili da 1 a n)
  - diverse a gruppi (con $m < n$ gruppi)
- Per quanto riguarda l'estrazione, può essere:
  - con reinserimento (con la possibilità di estrarre $k \in \mathbb{N}$ palline)
  - senza reinserimento (limitando il numero di estrazioni a $k \leqslant n$)

Inoltre è possibile distinguere fra due possibilità sugli aspetti interessanti delle palline estratte:

- ci interessa la disposizione delle palline estratte (cioè ci interessa quali palline sono estratte e in quale ordine)
- ci interessa la combinazione delle palline estratte (cioè ci interessa solo quali palline sono estratte, ma non in quale ordine, e quindi ci interessa quante palline di ciascun tipo sono state estratte)

I problemi si possono quindi analizzare in base alle loro caratteristiche attraverso il calcolo combinatorio.

Disposizione con reinserimento In questo caso, assumendo che le palline siano tutte diverse, ad ogni estrazione ci sono n possibilità diverse di estrarre la pallina, e quindi le possibili disposizioni sono:

$$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ volte}} = n^k$$

Disposizione senza reinserimento Problema molto simile al precedente, con la differenza che le palline non vengono reinserite nell'urna. Ad ogni estrazione il numero di palline che è possibile estrarre diminuisce di 1. Le possibili disposizioni sono:

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Un caso particolare si ha quando $k = n$, cioè se vengono estratte tutte le palline dall'urna. In questo caso, le disposizioni sono:

$$\frac{n!}{(n-k)!} = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

Queste si dicono permutazioni, in quanto sono modi diversi di disporre gli stessi elementi.

Combinazioni senza reinserimento Anche in questo caso ad ogni estrazione il numero di palline che è possibile estrarre diminuisce di 1; inoltre, data una sequenza di palline estratte, questa è equivalente a tutte le sue permutazioni. Il numero di combinazioni si può quindi ottenere dividendo il numero di disposizioni ottenibili in questa situazione per il numero di permutazioni possibili della sequenza estratta:

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

Per convenzione si pone:

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

Un'applicazione importante del coefficiente binomiale è la formula del binomio di Newton:

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

Alcune proprietà sono:

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

$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

Permutazioni con palline uguali Nel caso i colori siano $m < n$, e ci siano n_i palline dell'i-esimo colore, il numero di modi di estrarle tutte si può ottenere dividendo il numero di modi di estrarre tutte le palline se fossero tutte diverse per il numero di modi di scambiare fra di loro le palline uguali:

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!}$$

Questo caso corrisponde al trovare il numero di anagrammi di una parola con lettere ripetute.

3.2 Probabilità Condizionata

Finora abbiamo visto situazioni in cui la probabilità di un dato evento E dipende esclusivamente da E, cioè in cui la misura di quanto verosimile sia E è basata solo sulle caratteristiche "di partenza" dello spazio di probabilità. In alcuni casi, è utile valutare come varia la probabilità di E se sono fornite informazioni aggiuntive; l'obiettivo è formalizzare il concetto molto intuitivo che, ad esempio, la probabilità che piova quando è sereno è diversa dalla probabilità che piova quando è nuvoloso. In particolare, l'informazione relativa al clima deve avere un effetto sulla probabilità che piova. Questo viene formalizzato introducendo la misura di probabilità condizionata:

Definizione 3.1. Sia $(\Omega, \mathscr{F}, P)$ uno spazio di probabilità, sia $F \in \mathscr{F}$ tale che $P(F) > 0$; si definisce misura di probabilità condizionata a F la funzione

$$\begin{aligned} P(\cdot|F) : \mathscr{F} &\to [0,1] \\ E &\mapsto P(E|F) \end{aligned}$$

dove:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \quad \forall E \in \mathscr{F}$$

La probabilità $P(E|F)$ misura la probabilità che se si verifica F, si verifichi anche E. Si dimostra che la misura di probabilità condizionata $P(\cdot|F)$è effettivamente una misura di probabilità:

Dimostrazione. 1.

$$\begin{aligned} P(\Omega|F) &= \frac{P(\Omega \cap F)}{P(F)} \\ &= \frac{P(F)}{P(F)} = 1 \end{aligned}$$

2. Considerando una famiglia $\{E_n\}_n$ di sottoinsiemi di Ω:

$$P\left(\bigcup_n E_n \middle| F\right) = \frac{P\left(\left(\bigcup_n E_n\right) \cap F\right)}{P(F)} = \frac{P\left(\bigcup_n (E_n \cap F)\right)}{P(F)} = \sum_n \frac{P(E_n \cap F)}{P(F)}$$

Nota. Chiaramente, ha senso considerare la probabilità di E condizionata a F solo se $P(F) > 0$.

Alcune proprietà:

- $P(E^C|F) = 1 - P(E|F) \quad \forall E \in \mathscr{F}$
- $E \cap F = \varnothing \Rightarrow P(E|F) = 0$
- $F \subseteq E \Rightarrow P(E|F) = 1$
- $P(E) = 0 \Rightarrow P(E|F) = 0$
- $P(E) = 1 \Rightarrow P(E|F) = 1$
- *Formula di moltiplicazione*: sia $\{E_n\}_n$ una famiglia di eventi non necessariamente disgiunti, allora vale

$$P\left(\bigcap_{i=1}^{n} E_i\right) = P\left(E_n | \bigcap_{i=1}^{n-1} E_i\right) \cdot \ldots \cdot P(E_1) = \prod_{k=2}^{n} P\left(E_k | \bigcap_{i=1}^{k-1} E_i\right) \cdot P(E_1)$$

- Dalla precedente, per $n = 2$, si ricava

$$P(E \cap F) = P(E|F) \cdot P(F)$$

Nota. Nel contesto della probabilità condizionata, $P(E)$ si dice anche probabilità a priori di E, mentre $P(E|F)$ è detta, per contrasto, probabilità a posteriori.

Vediamo adesso alcune formule legate alla probabilità condizionata.

Formula di Probabilità Totale Partendo dalla formula di moltiplicazione:

$$P(E \cap F) = P(E|F) \cdot P(F)$$
$$P(E \cap F^C) = P(E|F^C) \cdot P(F^C)$$

$E \cap F$ e $E \cap F^C$ sono eventi disgiunti, e vale $(E \cap F) \cup (E \cap F^C) = E$. Segue:

$$P(E) = P(E \cap F) + P(E \cap F^C) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

Questa formula si può generalizzare: consideriamo una partizione $\{F_n\}_n$ di Ω, cioè una famiglia di sottoinsiemi di Ω tale che

$$\bigcup_i F_i = \Omega$$
$$F_i \cap F_j = \varnothing \quad \forall i \neq j$$

Inoltre, imponiamo che
$$P(F_i) > 0 \quad \forall i$$
Allora, dato un evento $E \in \mathscr{F}$ si può scrivere:
$$P(E) = \sum_{i=1}^{n} P(E|F_i) \cdot P(F_i)$$

Dimostrazione. Essendo $\{F_n\}_n$ una partizione di Ω, vale:
$$\begin{aligned}
E &= \bigcup_{i=1}^{n}(E \cap F_i) \\
\Rightarrow P(E) &= P\left(\bigcup_{i=1}^{n}(E \cap F_i)\right) \\
&= \sum_{i=1}^{n} P(E \cap F_i) \\
&= \sum_{i=1}^{n} \frac{P(E \cap F_i)}{P(F_i)} P(F_i) \\
&= \sum_{i=1}^{n} P(E|F_i)P(F_i)
\end{aligned}$$

Formula di Bayes La formula di Bayes segue da un'interpretazione della formula di probabilità totale: ciascuno degli eventi F_i si può interpretare come una possibile causa di E. Ciascuno degli addendi $P(E|F_i)P(F_i)$ rappresenta la probabilità che E si verifichi dopo che si è verificato F_i. Ci chiediamo ora quale sia la probabilità che, avendo osservato che E si verifica, si sia verificato F_i. Applicando le formule di moltiplicazione e di probabilità totale, considerando uno specifico evento F_k della partizione, segue:
$$\begin{aligned}
P(F_k|E) &= \frac{P(F_k \cap E)}{P(E)} \\
&= \frac{P(E|F_k)P(F_k)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}
\end{aligned}$$

Esempio 3.1. Un'applicazione della formula di Bayes si ha in molti casi in cui si cerca di studiare un dato fenomeno non direttamente osservabile, e che quindi porta alla necessità di effettuare un test indiretto per valutare il presentarsi o meno del fenomeno. In modo semplificato, possiamo considerare che per l'evento ha $\{D, d\}$ come possibili outcome, rappresentanti rispettivamente il verificarsi o meno dell'evento, e il test ha $\{T, t\}$ come possibili risultati, che rappresentano rispettivamente il rilevare il verificarsi o il non verificarsi del fenomeno. Gli outcome dell'esperimento sono quindi:

- TD se il test ha correttamente rilevato il verificarsi dell'esperimento
- td se il test ha correttamente rilevato il non verificarsi dell'esperimento
- Td se il test ha rilevato un falso positivo, detto anche errore di prima specie
- tD se il test ha rilevato un falso negativo, detto anche errore di seconda specie

Due probabilità che caratterizzano l'esperimento sono:

- $P(T|D)$, detta sensibilità
- $P(t|d)$, detta specificità

In particolare, date queste probabilità, è possibile ricavare quanto probabile sia che il fenomeno si sia verificato dopo aver ricevuto un risultato positivo al test:
$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|d)P(d)}$$

3.3 Indipendenza

Definizione 3.2. Dati due eventi $E, F \in \mathscr{F}$, E e F si dicono indipendenti, e si indica

$$E \perp\!\!\!\perp F$$

se e solo se

$$P(E \cap F) = P(E)P(F)$$

Il concetto di eventi indipendenti formalizza l'idea che se due eventi sono completamente slegati, allora la probabilità che uno si verifichi non ha nessuna influenza sulla probabilità che si verifichi anche l'altro. Alcune proprietà:

- $E \perp\!\!\!\perp \varnothing$
- $E \perp\!\!\!\perp \Omega$
- $E \perp\!\!\!\perp E \Leftrightarrow E = \Omega \vee E = \varnothing$
- Se $P(E) > 0$ e $P(F) > 0$, allora vale

$$E \perp\!\!\!\perp F \Leftrightarrow P(E|F) = P(E) \Leftrightarrow P(F|E) = P(F)$$

- $E \perp\!\!\!\perp F \Leftrightarrow E^C \perp\!\!\!\perp F \Leftrightarrow E \perp\!\!\!\perp F^C \Leftrightarrow E^C \perp\!\!\!\perp F^C$

Per tre eventi, la definizione di indipendenza è lievemente più complessa:

Definizione 3.3. Dati tre eventi $E, F, G \in \mathscr{F}$, questi si dicono indipendenti se e solo se

1. $$E \perp\!\!\!\perp F \wedge E \perp\!\!\!\perp G \wedge F \perp\!\!\!\perp G$$

2. $$P(E \cap F \cap G) = P(E)P(F)P(G)$$

NB. Le due condizioni sono slegate: ci sono casi in cui gli eventi sono indipendenti a due a due, ma la loro intersezione non ha probabilità pari al prodotto delle singole probabilità, o viceversa la probabilità dell'intersezione coincide con il prodotto delle probabilità, ma gli eventi non sono indipendenti a due a due.

Il concetto di indipendenza si estende a più eventi:

Definizione 3.4. Data una famiglia finita di eventi $\{E_n\}_n \subseteq \mathscr{F}$, si dice famiglia di eventi indipendenti se e solo se $\forall r \in \{2, 3, ..., n\}$ ogni sottofamiglia di $\{E_n\}_n$ composta da r eventi è indipendente.

Le due condizioni continuano ad essere slegate. Si nota che il numero di casi da considerare aumenta quadraticamente al crescere del numero di eventi.

Definizione 3.5. Data una famiglia numerabile di eventi $\{E_n\}_n \subseteq \mathscr{F}$, si dice famiglia di eventi indipendenti se $\forall \{F_k\}_k \subseteq \{E_n\}_n$ con $\{F_k\}_k$ famiglia finita, $\{F_k\}_k$ è famiglia di eventi indipendenti.

Osservazione 3.1. Se $\{A_n\}_n$ è una famiglia di eventi indipendenti (non importa quanti), allora anche la famiglia formata sostituendo alcuni (non necessariamente tutti) eventi con i rispettivi complementari è indipendente.

Esercizio 3.1. Dimostrare che se $\{A, B\}$ sono indipendenti, anche $\{A^C, B^C\}$ sono indipendenti

Soluzione.

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

Esercizio 3.2. A tre studenti viene posta la stessa domanda. Sappiamo che il primo risponde correttamente con probabilità $\frac{1}{2}$, il secondo con probabilità $\frac{2}{3}$, il terzo con probabilità $\frac{1}{3}$. Gli studenti non possono comunicare fra loro. Vogliamo calcolare la probabilità che, se un solo studente ha risposto correttamente, sia stato il secondo.

Soluzione. Consideriamo gli eventi relativi ai possibili outcome di cui ci può interessare:

$$S_i = \text{"l'}i\text{-esimo studente ha risposto correttamente"}$$
$$E_i = \text{"ci sono state } i \text{ risposte esatte"}$$

Siccome è noto che solo uno studente ha risposto correttamente, la probabilità da calcolare è la probabilità di S_2 condizionato a E_1:

$$P(S_2|E_1) = \frac{P(S_2 \cap E_1)}{P(E_1)}$$

Concettualmente, possiamo riscrivere $S_2 \cap E_1$ ed E_1 come unioni ed intersezioni di eventi più semplici:

$$P(S_2|E_1) = \frac{P(S_2 \cap E_1)}{P(E_1)} = \frac{P(S_1^C \cap S_2 \cap S_3^C)}{P((S_1 \cap S_2^C \cap S_3^C) \cup (S_1^C \cap S_2 \cap S_3^C) \cup (S_1^C \cap S_2^C \cap S_3))}$$

Siccome gli eventi parte dell'unione sono disgiunti, la probabilità si può distribuire:

$$P(S_2|E_1) = \frac{P(S_2 \cap E_1)}{P(E_1)} = \frac{P(S_1^C \cap S_2 \cap S_3^C)}{P(S_1 \cap S_2^C \cap S_3^C) + P(S_1^C \cap S_2 \cap S_3^C) + P(S_1^C \cap S_2^C \cap S_3)}$$

Siccome gli studenti non possono comunicare, la risposta di ciascuno non influenza la risposta degli altri, e quindi possiamo assumere che i tre eventi siano indipendenti:

$$\begin{aligned}
&\frac{P(S_1^C \cap S_2 \cap S_3^C)}{P(S_1 \cap S_2^C \cap S_3^C) + P(S_1^C \cap S_2 \cap S_3^C) + P(S_1^C \cap S_2^C \cap S_3)} \\
&= \frac{P(S_1^C)P(S_2)P(S_3^C)}{P(S_1)P(S_2^C)P(S_3^C) + P(S_1^C)P(S_2)P(S_3^C) + P(S_1^C)P(S_2^C)P(S_3)} \\
&= \frac{P(\frac{1}{2})P(\frac{2}{3})P(\frac{2}{3})}{P(\frac{1}{2})P(\frac{1}{3})P(\frac{2}{3}) + P(\frac{1}{2})P(\frac{2}{3})P(\frac{2}{3}) + P(\frac{1}{2})P(\frac{1}{3})P(S_3)}\ 0.57
\end{aligned}$$

3.3.1 Schema di Bernoulli

Detto anche *Modello Binomiale per Eventi.* Consideriamo un esperimento con queste caratteristiche:

- Vengano eseguite n prove identiche, effettuate in sequenza
- Ciascuna prova ha solo due possibili risultati, che generalmente si indicano con $\{0, 1\}$
- Il risultato di ciascuna prova non influenza il risultato di nessuna delle altre

L'esperimento si può descrivere con un modello di probabilità dato da $\Omega = \{0, 1\}^n$, $\mathscr{F} = \mathbb{P}(\Omega)$, e da una misura di probabilità P che rispetti le condizioni:

- Gli eventi
$$B_j = \{(b_1, ..., b_n)|b_j = 1\}$$
sono equiprobabili, cioè
$$P(B_j) = p \quad \forall j \in \{1, ..., n\} \text{ con } p \in [0, 1]$$
- La famiglia
$$\{B_1, ..., B_n\}$$
è una famiglia di eventi indipendenti.

Esempio 3.2. Consideriamo un sistema di comunicazione formato da n componenti, ciascuno dei quali ha probabilità p di funzionare. Nella sua totalità, il sistema funziona se almeno metà delle componenti funzionano. Ci chiediamo per quali valori di p un sistema a tre componenti ha maggiore probabilità di funzionare rispetto a un sistema formato da un solo componente.

Soluzione. L'idea è di rappresentare il problema attraverso lo schema di Bernoulli. Identifichiamo:

- una prova con il test di un componente

- il numero di prove ripetute con il numero di componenti
- gli esiti $\{0,1\}$ con $\{"funziona","non\ funziona"\}$ come

$$\begin{aligned} "funziona" &\mapsto 1 \\ "non\ funziona" &\mapsto 0 \end{aligned}$$

- la probabilità di successo con $P(B_j) = p$

Per il sistema a 3 componenti vale:

$$\begin{aligned}
&P(\text{"sistema funziona"}) \\
&= P(\text{"almeno due componenti su tre funzionano"}) \\
&= P(\{\text{"esattamente 2 componenti funzionano"}\} \cup \{\text{"esattamente 3 componenti funzionano"}\}) \\
&= P(\{\text{"2 componenti su 3 funzionano"}\}) + P(\{\text{"tutti e 3 i componenti funzionano"}\}) \\
&= P(\bigcup_{I\subset\{1,2,3\}\wedge|I|=2} \{(b_1,b_2,b_3) \in \{0,1\}^3 | b_j = 1 \Leftrightarrow j \in I\}) + P(\{(1,1,1)\}) \\
&\text{essendo gli eventi nell'unione disgiunti:} \\
&= \sum_{I\subset\{1,2,3\}\wedge|I|=2} P(\{(b_1,b_2,b_3) \in \{0,1\}^3 | b_j = 1 \Leftrightarrow j \in I\}) + P(\{(1,1,1)\}) \\
&= \sum_{I\subset\{1,2,3\}\wedge|I|=2} \prod_{j\in I} P(B_j) \cdot \prod_{j\notin I} P(B_j^C) + P(\{(1,1,1)\}) \\
&\text{siccome } |I| = 2\text{:} \\
&= \sum_{I\subset\{1,2,3\}\wedge|I|=2} p^2(1-p) + p^3 \\
&= \binom{3}{2} p^2(1-p) + p^3 \\
&= 3 \cdot p^2(1-p) + p^3
\end{aligned}$$

Per il sistema ad una componente, chiaramente vale:

$$P(\text{"sistema funziona"}) = p$$

Quindi dobbiamo cercare p tale che:

$$\begin{aligned}
&3 \cdot p^2(1-p) + p^3 > p \\
\Leftrightarrow &3p^2 - 2p^3 > p \\
\Leftrightarrow &p > \frac{1}{2}
\end{aligned}$$

Nel complesso, si ottiene che ha senso usare più componenti solo se ciascuna ha più probabilità di funzionare di $\frac{1}{2}$.

4 Variabili Aleatorie Discrete

4.1 Concetti Base per le Variabili Aleatorie

Spesso, negli esperimenti concreti ha senso raccogliere i dati attraverso dei costrutti che si possano analizzare più approfonditamente degli spazi di probabilità.

Definizione 4.1. Sia $(\Omega, \mathbb{P}(\Omega), P)$ spazio di probabilità discreto. Si dice variabile aleatoria una mappa (funzione)

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

Definizione 4.2. Data una variabile aleatoria X si dice alfabeto di X e si indica $\mathcal{X}$ l'immagine di Ω rispetto X:

$$\mathcal{X} = \{x \in \mathbb{R} | \exists\omega \in \Omega : X(\omega) = x\}$$

Nota. Dato uno spazio di probabilità ci sono infinite variabili aleatorie associate.

Per un dato spazio di probabilità $(\Omega, \mathbb{P}(\Omega), P)$ e un evento $E \in \mathbb{P}(\Omega)$, una variabile aleatoria che si può sempre definire è detta variabile aleatoria indicatrice, ed è definita come:

$$\mathbb{1}_E(\omega) = \begin{cases} 1 & \text{se } \omega \in E \\ 0 & \text{se } \omega \notin E \end{cases}$$

Osservazione 4.1. Le variabili aleatorie sono funzioni a valori reali, quindi se sono definite sullo stesso Ω si possono combinare attraverso le consuete operazioni aritmetiche fra reali.

Possiamo assegnare ad un certo alfabeto $\mathcal{X}$ una probabilità, asegnando ai singoletti di $\mathcal{X}$ (cioè ai suoi sottoinsiemi di cardinalità 1) una probabilità P^X che sia compatibile con la misura di probabilità P dello spazio di probabilità su cui è definita X. Per farlo, consideriamo gli eventi definiti da

$$E_k = \{\omega \in \Omega | X(\omega) = x_k\} = X^{-1}(x_k)$$

Siccome X è una funzione, questi eventi sono disgiunti e la loro unione è Ω (perchè per definizione di funzione <u>ogni</u> $\omega \in \Omega$ viene mappato in <u>uno e un solo</u> $x_k \in \mathcal{X}$), quindi formano una partizione di Ω. Possiamo definire la probabilità P^X come P calcolata sugli eventi E_k:

Definizione 4.3. Sia X variabile aleatoria sullo spazio di probabilità $(\Omega, \mathbb{P}(\Omega, P)$. Si dice misura di probabilità indotta la funzione

$$\begin{aligned} P^X &\to [0,1] \\ \{x_k\} &\mapsto P^X(\{x_k\}) = P(X^{-1}(x_k)) \end{aligned}$$

$X^{-1}(x_k)$ è l'insieme degli eventi elementari che X mappa in x_k, quindi $P^X(\{x_k\})$ rappresenta la probabilità che l'esito dell'esperimento venga mappato nel numero x_k da X. In altre parole, $P^X(\{x_k\})$ rappresenta la probabilità dell'evento "si verifica un esito che viene mappato in x_k da X".

Lemma 4.1. *La misura di probabilità indotta è una misura di probabilità, definita sullo spazio campionario $\mathcal{X}$ e sulla σ-algebra $\mathbb{P}(\mathcal{X})$.*

In generale, vorremmo fosse possibile definire una variabile aleatoria indipendentemente dall'esperimento da cui deriva, cioè vorremmo dare una descrizione probabilistica di X che non sia basata sullo spazio di probabilità $(\Omega, \mathbb{P}(\Omega), P)$. È sempre possibile mostrare i valori che X può assumere mostrando il suo alfabeto, ma con gli strumenti introdotti finora non è possibile caratterizzare la probabilità che assuma ciascun valore; abbiamo introdotto la probabilità indotta, che però è ancorata nello spazio di probabilità sottostante a X. Per separare X da Ω introduciamo la densità di probabilità.

Definizione 4.4. Sia X variabile aleatoria con alfabeto $\mathcal{X}$, e sia P^X la sua probabilità indotta. Si dice densità di probabilità di X la funzione

$$\begin{aligned} p_X : X &\to [0,1] \\ x_k &\mapsto p_X(x_k) = P^X(\{x_k\}) \end{aligned}$$

Esercizio 4.1. Lancio una moneta e un dado; se i risultati che ottengo hanno la stessa iniziale vinco 1€, altrimenti perdo 50cents. Sia X la variabile aleatoria che descrive la vincita o la perdita. Determinare alfabeto e densità di X.

Soluzione. L'alfabeto è

$$\mathcal{X} = \{1, -\frac{1}{2}\}$$

Per identificare la densità discreta è necessario identificare lo spazio campionario:

$$\Omega = \{(x,y) | x \in \{T, C\}, y \in \{1,2,3,4,5,6\}\}$$

Si ricava:

$$p_X(1) = P(X^{-1}(1)) = P(\{(T,3),(C,5)\}) = \frac{2}{12}$$

Notando che $X = -\frac{1}{2} \Leftrightarrow X \neq 1$:

$$p_X(-\frac{1}{2}) = P(X^{-1}(-\frac{1}{2})) = P(\{(T,3),(C,5)\}^C) = 1 - \frac{2}{12} = \frac{10}{12}$$

Si nota che alfabeto e densità di una variabile aleatoria non identificano l'esperimento originale. Mentre è sempre possibile, dato lo spazio di probabilità, ricavare la variabile aleatoria e la misura di probabilità indotta, molto più di frequente viene data direttamente la variabile aleatoria, caratterizzandola attraverso il suo alfabeto e la sua densità discreta:

Definizione 4.5. Sia X una variabile aleatoria, questa può essere definita indicando:

- il suo alfabeto $\mathcal{X}$
- la sua densità discreta $p_X(x_k)\ \forall x_k \in \mathcal{X}$

Questa si dice definizione o descrizione probabilistica di X.

Esercizio 4.2. Estraiamo 3 carte da un mazzo di 52. Vinciamo 1€ pero ogni carta di picche estratta. Sia X la variabile aleatoria che descrive la vincita; determinarne alfabeto e densità.

Soluzione. L'alfabeto è:

$$\mathcal{X} = \{0, 1, 2, 3\}$$

La densità si ricava dall'esperimento:

$$\begin{aligned}
p_X(0) = P(X^{-1}(0)) &= P(\text{"non sono estratte carte di picche"}) \\
&= \frac{\binom{39}{3}}{\binom{52}{3}} = \frac{703}{1700} \\
p_X(1) = P(X^{-1}(1)) &= P(\text{"è estratta una sola carta di picche"}) \\
&= \frac{13 \cdot \binom{39}{2}}{\binom{52}{3}} = \frac{741}{1700} \\
p_X(2) = P(X^{-1}(2)) &= P(\text{"sono estratte due carte di picche"}) \\
&= \frac{\binom{13}{2} \cdot 39}{\binom{52}{3}} = \frac{234}{1700} \\
p_X(3) = P(X^{-1}(3)) &= P(\text{"sono estratte tre carte di picche"}) \\
&= \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{22}{1700}
\end{aligned}$$

A volte ha senso comporre una variabile aleatoria con una funzione reale, ottenendo un'altra variabile aleatoria. Vediamo un esempio:

Esempio 4.1. Sia X variabile aleatoria con alfabeto

$$\mathcal{X} = \{-1, 1, 3\}$$

e densità discreta

$$\begin{aligned}
p_X(-1) = p_X(3) &= \frac{1}{4} \\
p_x(1) &= \frac{1}{2}
\end{aligned}$$

Caratterizziamo $Y = g(X) = X^2$, che è data dalla composizione di X e la funzione

$$\begin{aligned}
g : \mathbb{R} &\to \mathbb{R} \\
x &\mapsto x^2
\end{aligned}$$

L'alfabeto di Y è dato dall'applicazione di g a tutti gli elementi di $\mathcal{X}$:

$$\mathcal{Y} = \{1, 9\}$$

Per calcolare la densità discreta di Y bisogna ricondurre il problema alla variabile X:

$$\begin{aligned} p_Y(1) &= P(X^{-1}(-1) \cup X^{-1}(1)) \\ &= P(X^{-1}(-1)) + P(X^{-1}(1)) \\ &= p_X(-1) + p_X(1) = \frac{3}{4} \\ p_Y(9) &= P(X^{-1}(3)) \\ &= p_X(3) = \frac{1}{4} \end{aligned}$$

Questo concetto si formalizza:

Definizione 4.6. Siano $X : \Omega \to \mathbb{R}$ variabile aleatoria e $g : \mathbb{R} \to \mathbb{R}$ funzione reale di variabile reale. Si dice funzione di X la composizione $Y = g \circ X$, la quale è ancora una variabile aleatoria caratterizzata da

- Un alfabeto, il quale è l'immagine di $\mathcal{X}$ attraverso g:

$$\mathcal{Y} = g(\mathcal{X})$$

- Una densità discreta

4.2 Valor Medio

Il valor medio di una variabile aleatoria X rappresenta il valore attorno a cui si distribuiscono i valori assunti da X.

Definizione 4.7. Sia X variabile aleatoria discreta con alfabeto $\mathcal{X}$ e densità di probabilità p_X. Il valor medio di X è definito come:

$$E(X) = \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k)$$

Il valor medio ha anche molti altri nomi, fra cui media, speranza matematica, valore atteso, e molti altri.

Osservazione 4.2. Se X è discreta anche $\mathcal{X}$ lo è, ma non è detto che sia finito. Se è numerabile, il valor medio diventa la somma di una serie, la quale non è detto che esiste; se la serie non converge, il valor medio non esiste.

Il valor medio si può intendere come il baricentro di un sistema formato da una serie di masse disposte su una sbarra, dove x_k è il braccio della k-esima massa e $p_X(x_k)$ è il suo peso.

Teorema 4.1 (Fondamentale del valor medio)**.** *Sia $X : \Omega \to \mathbb{R}$ variabile aleatoria discreta. Allora si ha*

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\})$$

Dimostrazione.

$$\begin{aligned} E(X) &= \sum_{x_k \in \mathcal{X}} x_k P(x_k) \\ &= \sum_{x_k \in \mathcal{X}} x_k P(X^{-1}(x_k)) \\ &= \sum_{x_k \in \mathcal{X}} x_k P(\{\omega \in \Omega | X(\omega) = x_k\}) \\ &= \sum_{x_k \in \mathcal{X}} \sum_{w : X(\omega) = x_k} x_k P(\{\omega\}) \\ &= \sum_{x_k \in \mathcal{X}} \sum_{w : X(\omega) = x_k} X(\omega) P(\{\omega\}) \end{aligned}$$

Siccome al variare di k gli insiemi $\{\omega | X(\omega) = x_k\}$ sono disgiunti, e i valori x_k sono tutti scorsi dalla prima sommatoria, questi formano una partizione di Ω, quindi le due sommatorie corrispondono ad un unica sommatoria su tutto Ω stesso:

$$= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$$

Questo teorema nella pratica non è utilizzato, ma permette di dimostrare alcune proprietà importanti del valor medio: siano $X, Y : \Omega \to \mathbb{R}$ variabili aleatorie, siano $a, b \in \mathbb{R}$, allora valgono

1. Linearità: $E(aX + bY) = aE(X) + bE(Y)$

   (a) $E(aX) = aE(X)$

   Dimostrazione.

   $$\begin{aligned} aX : \Omega &\to \mathbb{R} \\ \omega &\mapsto aX(\omega) \end{aligned}$$

   Per il teorema fondamentale:

   $$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} aX(\omega) \cdot P(\{\omega\}) \\ &= a \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) \\ &= aE(X) \end{aligned}$$

   (b) $E(X + Y) = E(X) + E(Y)$

   Dimostrazione.

   $$\begin{aligned} X + Y : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) + Y(\omega) \end{aligned}$$

   Per il teorema fondamentale:

   $$\begin{aligned} E(X) &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \cdot P(\{\omega\}) \\ &= \sum_{\omega \in \Omega} (X(\omega) \cdot P(\{\omega\}) + Y(\omega) \cdot P(\{\omega\})) \\ &= \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) + \sum_{\omega \in \Omega} Y(\omega) \cdot P(\{\omega\}) \\ &= E(X) + E(Y) \end{aligned}$$

2. Positività: se X è una variabile aleatoria non negativa (cioè tale che $X \geqslant 0$, cioè tale che $\mathcal{X} \subseteq \mathbb{R}_+$, allora anche $E(X)$ è non negativo

   $$X \geqslant 0 \Rightarrow E(X) \geqslant 0$$

   Dimostrazione. Per definizione:

   $$E(X) = \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k)$$

   Per definizione:

   $$p_X(x_k) \geqslant 0 \quad \forall x_k \in \mathcal{X}$$

   Per Hp:

   $$x_k \geqslant 0 \quad \forall x_k \in \mathcal{X}$$

   Segue che i prodotti sono non negativi, e la sommatoria è a sua volta non negativa.

3. Monotonia: $X \geqslant Y \Rightarrow E(X) \geqslant E(Y)$

   Dimostrazione. Segue dalle altre proprietà:

   $$\begin{aligned} X \geqslant Y &\Leftrightarrow X - Y \geqslant 0 \\ \text{Per positività:} \quad &\Leftrightarrow E(X - Y) \geqslant 0 \\ \text{Per linearità:} \quad &\Leftrightarrow E(X) - E(Y) \geqslant 0 \\ &\Leftrightarrow E(X) \geqslant E(Y) \end{aligned}$$

4. Limiti inferiore e superiore: essendo $\mathcal{X}$ discreto, ha limiti superiore e inferiore (che potrebbero essere infiniti); indichiamo

$$\underline{x} = inf(\mathcal{X})$$
$$\overline{x} = sup(\mathcal{X})$$

   Allora:

$$\underline{x} \leqslant E(X) \leqslant \overline{x}$$

   Dimostrazione. Per definizione:

$$\underline{x} \leqslant x_k \leqslant \overline{x} \quad \forall x_k \in \mathcal{X}$$

   Segue:

$$\sum_{x_k \in \mathcal{X}} \underline{x} \cdot p_X(x_k) \leqslant \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leqslant \sum_{x_k \in \mathcal{X}} \overline{x} \cdot p_X(x_k)$$
$$\underline{x} \sum_{x_k \in \mathcal{X}} p_X(x_k) \leqslant \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leqslant \overline{x} \sum_{x_k \in \mathcal{X}} p_X(x_k)$$
$$\underline{x} \cdot 1 \leqslant \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leqslant \overline{x} \cdot 1$$

   Segue la tesi.

Osservazione 4.3. Se $a \in \mathbb{R}$, vale sempre $E(a) = a$, perchè possiamo vedere una costante come un valore assunto da una variabile aleatoria X tale che $X(\omega) = a \ \forall \omega \in \Omega$, e quindi vale

$$\mathcal{X} = \{a\}$$
$$p_X(a) = 1$$

4.2.1 Valor Medio di una Funzione di una Variabile Aleatoria

Abbiamo visto che se X è una variabile aleatoria e g una funzione reale di variabile reale, possiamo definire la variabile aleatoria $Y = g \circ X$. Vogliamo ora calcolare $E(Y)$; per questo introduciamo un teorema (che non dimostriamo).

Teorema 4.2.

$$E(Y) = \sum_{x_k \in \mathcal{X}} g(x_k) \cdot p_X(x_k)$$

Questo teorema ci permette di calcolare $E(Y)$ senza la necessità di caratterizzare completamente Y.

4.3 Varianza

Mentre il valor medio quantifica un valore attorno a cui i valori assunti da una variabile aleatoria si distribuiscono, la varianza quantifica la dispersione dei valori.

Definizione 4.8. Sia X variabile aleatoria di alfabeto $\mathcal{X}$ e densità p_X; la varianza di X si indica con $Var(X)$, ed è definita come

$$Var(X) = \sum_{x_k \in \mathcal{X}} (x_k - E(X))^2 p_X(x_k)$$

Esempio 4.2. Consideriamo le variabili aleatorie:

$$X_1 \in \{-1, \frac{1}{4}, \frac{3}{4}\}$$
$$p_{X_1}(-1) = p_{X_1}(\frac{1}{4}) = p_{X_1}(\frac{3}{4}) = \frac{1}{3}$$
$$X_2 \in \{-10, 10\}$$
$$p_{X_2}(-10) = p_{X_2}(10) = \frac{1}{2}$$

Notiamo che

$$E(X_1) = -1 \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{3} + \frac{3}{4} \cdot \frac{1}{3} = 0$$
$$E(X_2) = -10 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 0$$

Ma

$$Var(X_1) = (-1)^2 \cdot \frac{1}{3} + (\frac{1}{4})^2 \cdot \frac{1}{3} + (\frac{3}{4})^2 \cdot \frac{1}{3} \approx 0.524$$
$$Var(X_2) = (-10)^2 \cdot \frac{1}{2} + (10)^2 \cdot \frac{1}{2} = 100$$

Osservazione 4.4. La varianza si può definire in funzione solo della media:

$$Var(X) = E((X - E(X))^2)$$

Si può vedere considerando $Var(X) = E(g(X))$ con $g(x) = (x - E(X))^2$.

Dall'osservazione possiamo ricavare alcuni risultati interessanti:

$$\begin{aligned} Var(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2 \cdot X \cdot E(X) + (E(X))^2) \\ &\text{ricordando le proprietà del valor} \\ &\text{medio, e ricordando che } E(a) = a\text{:} \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

In molti testi questa è data come definizione di varianza. Dalla definizione si ricava che $Var(X) \geqslant 0$, quindi deve sempre valere $E(X^2) - (E(X))^2$. Vediamo alcune proprietà della varianza:

1. $Var(X) \geqslant 0$, e in particolare
$$Var(X) = 0 \Leftrightarrow X = k \in \mathbb{R}$$

   Dimostrazione. Segue immediatamente dalla difinizione. ∎

2. $Var(aX) = a^2 \cdot Var(X)$

   Dimostrazione.
$$\begin{aligned} Var(aX) &= E((aX - E(aX))^2) \\ &= E((aX - aE(X))^2) \\ &= E(a^2(X - E(X))^2) \\ &= a^2 E((X - E(X))^2) \end{aligned}$$
∎

3. $Var(X + a) = Var(X)$

   Dimostrazione.
$$\begin{aligned} Var(X + a) &= E((X + a - E(X + a))^2) \\ &= E((X + a - E(X) + E(a))^2) \\ &= E((X + a - E(X) + a)^2) \\ &= E((X - E(X))^2) \\ &= Var(X) \end{aligned}$$
∎

Esercizio 4.3. Abbiamo due urne

1. La prima urna contiene 1 palline dorata, 4 palline verdi e 15 palline bianche
2. La seconda urna contiene 4 palline verdi e 25 palline bianche

Viene spostata una pallina scelta a caso dall'urna 1 all'urna 2, e successivamente estraiamo una pallina dall'urna 2.

- Se estraiamo la pallina dorata, vinciamo 50€
- Se estraiamo una pallina verde, perdiamo 1€

Sia X la variabile aleatoria che rappresenta la vincita, si calcoli la sua varianza.

Soluzione. I valori che può assumere X sono

$$\mathcal{X} = \{-1, 0, 50\}$$

Per comodità di notazione, definiamo i seguenti eventi:

$$\begin{aligned} E_i &= \text{"Estrazione dall'urna 2 una pallina di colore } i\text{"} \\ T_i &= \text{"Trasferimento dall'urna 1 all'urna 2 una pallina di colore } i\text{"} \end{aligned}$$

Ricordando la formula di probabilità totale, la densità discreta di X è:

$$\begin{aligned} p_X(-1) &= P(E_V) \\ &= P(E_V|T_D) \cdot P(T_D) + P(E_V|T_V) \cdot P(T_V) + P(E_V|T_B) \cdot P(T_B) \\ &= \frac{4}{30} \cdot \frac{1}{20} + \frac{5}{30} \cdot \frac{4}{20} + \frac{4}{30} \cdot \frac{15}{20} = \frac{7}{50} \\ p_X(0) &= P(E_B) \\ &= P(E_B|T_D) \cdot P(T_D) + P(E_B|T_V) \cdot P(T_V) + P(E_B|T_B) \cdot P(T_B) \\ &= \frac{25}{30} \cdot \frac{1}{20} + \frac{25}{30} \cdot \frac{4}{20} + \frac{26}{30} \cdot \frac{15}{20} = \frac{103}{120} \\ p_X(50) &= P(E_D) \\ &= P(E_D|T_D) \cdot P(T_D) + P(E_D|T_V) \cdot P(T_V) + P(E_D|T_B) \cdot P(T_B) \\ &= \frac{1}{30} \cdot \frac{1}{20} + \frac{0}{30} \cdot \frac{4}{20} + \frac{0}{30} \cdot \frac{15}{20} = \frac{1}{600} \end{aligned}$$

Calcoliamo ora la varianza:

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= \left[(-1)^2 \cdot \frac{7}{50} + (0)^2 \cdot \frac{103}{120} + (50)^2 \cdot \frac{1}{600}\right] \\ &- \left[(-1) \cdot \frac{7}{50} + (0) \cdot \frac{103}{120} + (50) \cdot \frac{1}{600}\right]^2 \approx 4.3 \end{aligned}$$

4.4 Variabili Aleatorie Discrete Notevoli

Definizione 4.9. Siano X e Y due varibili aleatorie con lo stesso alfabeto, e con stessa densità di probabilità sugli stessi valori:

$$\begin{gathered} \mathcal{X} = \mathcal{Y} \\ p_X(x_k) = p_Y(y_j) \quad \forall x_k = y_j \end{gathered}$$

Allora X e Y si dicono probabilisticamente equivalenti e si indica

$$X \sim Y$$

NB. È sbagliato dire che $X = Y$, perchè le due variabili potrebbero essere definite su spazi di probabilità completamente diversi.

4.4.1 Alfabeto Finito

Variabili Aleatorie di Bernoulli Le variabili aleatorie di Bernoulli rappresentano il risultato di una singola prova in uno schema di Bernoulli. Una variabile X è variabile di Bernoulli, e si indica con

$$X \sim Be(p) \quad \text{con } p \in [0,1]$$

se è caratterizzata da:

$$\begin{aligned} \mathcal{X} &= \{0,1\} \\ p_X(1) &= p \\ p_X(0) &= 1-p \\ E(X) &= p \\ Var(X) &= 1-p \end{aligned}$$

Esempio 4.3. La variabile $\mathbb{1}_E$ è una varibile di Bernoulli:

$$\mathbb{1}_E \sim Be(P(E))$$

Variabili Aleatorie Binomiali Una variabile X è variabile binomiale, e si indica con

$$X \sim Bin(n,p) \quad \text{con } p \in [0,1], n \in \mathbb{N} \setminus \{0\}$$

se è caratterizzata da:

$$\begin{aligned} \mathcal{X} &= \{0,1,\dots,n\} \\ p_X(k) &= \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k \in \mathcal{X} \\ E(X) &= np \\ Var(X) &= np(1-p) \end{aligned}$$

Le variabili di questa famiglia sono quelle associate agli schemi di Bernoulli.

Osservazione 4.5. Se all'i-esima prova viene associata una variabile di Bernoulli $X_i \sim Be(p)$, la variabile $X \sim Bin(n,p)$ è la somma delle X_i:

$$X = \sum_{i=1}^{n} X_i \sim Bin(n,p)$$

Infatti, siccome X_i vale 1 se e solo se l'i-esima prova ha avuto successo, la somma degli X_i è il numero di prove che hanno avuto successo.

Dimostriamo solo il calcolo del valor medio:

Dimostrazione.

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \cdot p_X(k) \\
&= \sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= 0 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} + \sum_{k=1}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \frac{n!}{(n-k)!k!} \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \cdot \frac{n!}{(n-k)!(k-1)!} \cdot p^k \cdot (1-p)^{n-k} \\
&\text{indico } h = k-1 \\
&= \sum_{h=0}^{n-1} \cdot \frac{n!}{(n-1-h)!h!} \cdot p^{h+1} \cdot (1-p)^{n-1-h} \\
&= n \cdot p \cdot \sum_{h=0}^{n-1} \cdot \frac{(n-1)!}{(n-1-h)!h!} \cdot p^h \cdot (1-p)^{n-1-h}
\end{aligned}
$$

La sommatoria corrisponde alla somma dei valori della densità discreta di una variabile aleatoria $Y \sim Bin(n-1,p)$ sui valori del suo alfabeto $\mathcal{Y} = \{0,\dots,n-1\}$; per definizione di p_Y, questa somma è 1

$$= n \cdot p \cdot \sum_{h=0}^{n-1} p_Y(h) = n \cdot p$$

Esercizio 4.4. È più facile ottenere almeno un 6 lanciando 4 volte un dado equilibrato, oppure ottenere un doppio 6 lanciando una coppia di dadi 24 volte?

Soluzione. L'esercizio riguarda due esperimenti, entrambi consistenti di prove ripetute, e quindi li studiamo con lo schema di Bernoulli.

- Lanciamo un dado 4 volte, e vediamo se esce almeno un 6. Ciascun lancio si può modellare con una variabile

  $$X_i = \begin{cases} 1 & \text{se all'}i\text{-esimo lancio esce 6} \\ 0 & \text{altrimenti} \end{cases}$$

  4 lanci $\Rightarrow i \in \{1,2,3,4\}$. Chiaramente:

  $$X_i \sim Be\left(\frac{1}{6}\right)$$

  Il risultato dell'esperimento si può studiare con una variabile aleatoria:

  $$X = \sum_{i=1}^{4} X_i \sim Bin\left(4, \frac{1}{6}\right)$$

  X è il numero di volte in cui abbiamo ottenuto 6; calcoliamo:

  $$
  \begin{aligned}
  P(X \geqslant 1) &= 1 - P(X < 1) \\
  &= 1 - P(X = 0) \\
  &= 1 - \binom{4}{0}\left(\frac{1}{6}\right)^0\left(1 - \frac{1}{6}\right)^4 \\
  &= 1 - \left(\frac{5}{6}\right)^4 \approx 0.518
  \end{aligned}
  $$

- Lanciamo due dadi 24 volte, e vediamo se esce almeno una volta un doppio 6. Ciascun lancio si può modellare con una variabile
$$Y_i = \begin{cases} 1 & \text{se all'}i\text{-esimo lancio esce un doppio 6} \\ 0 & \text{altrimenti} \end{cases}$$
24 lanci $\Rightarrow i \in \{1, \dots, 24\}$. Chiaramente:
$$Y_i \sim Be\left(\frac{1}{36}\right)$$
Il risultato dell'esperimento si può studiare con una variabile aleatoria:
$$Y = \sum_{i=1}^{4} Y_i \sim Bin\left(24, \frac{1}{36}\right)$$
Y è il numero di volte in cui abbiamo ottenuto un doppio 6; analogamente a prima:
$$\begin{aligned} P(Y \geqslant 1) &= 1 - P(Y < 1) \\ &= 1 - P(Y = 0) \\ &= 1 - \binom{24}{0}\left(\frac{1}{36}\right)^0\left(1 - \frac{1}{36}\right)^{24} \\ &= 1 - \left(\frac{35}{36}\right)^{24} \approx 0.49 \end{aligned}$$

Quindi è più probabile ottenere un 6 con un dado che ottenere una coppia di 6 con due dadi, anche con molti più lanci. Notevole è il fatto che le due variabili X e Y hanno stesso valor medio e stessa varianza, ma ciò non significa che le due variabili si comportano allo stesso modo.

4.4.2 Alfabeto Infinito

Variabili Aleatorie Geometriche

NB. Indichiamo:
$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, \dots\} \\ \mathbb{N}_0 &= \{0, 1, 2, 3, \dots\} \end{aligned}$$

Una variabile X si dice geometrica, e si indica
$$X \sim Ge(p) \quad \text{con } p \in [0, 1]$$
se è caratterizzata da:
$$\begin{aligned} \mathcal{X} &= \mathbb{N} \\ p_X(k) &= (1-p)^{k-1}p \\ E(X) &= \frac{1}{p} \\ Var(X) &= \frac{1-p}{p^2} \end{aligned}$$
Una variabile aleatoria geometrica rappresenta, in uno schema di Bernoulli, l'indice della prima prova che ha successo. Se consideriamo le variabili $X_i \sim Be(p)$, con p probabilità che una prova abbia successo e $i \in \mathbb{N}$ allora:
$$X = min\{i \in \mathbb{N} | X_i = 1\} \sim Ge(p)$$
La densità discreta di X segue da questa osservazione: se $X = k$, allora $X_i = 0\ \forall i < k$ e $X_k = 1$. Segue:
$$\begin{aligned} p_X(k) &= P(X = k) \\ &= P(X_1 = 0, X_2 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &\text{per l'indipendenza e l'equiprobabilità delle prove:} \\ &= \prod_{j=1}^{k-1} P(X_j = 0) P(X_k = 1) \\ &= (1-p)^{k-1}p \end{aligned}$$

Osservazione 4.6. La funzione
$$p_X(k) = (1-p)^{k-1}p \quad k \in \mathbb{N}$$
è una densità discreta di probabilità. Infatti:

- $p_X(k) \geqslant 0 \ \forall k \in \mathbb{N}$, trivialmente
- $\sum_{k\in\mathbb{N}} p_X(k) = 1$, perchè

$$\begin{aligned}
\sum_{k\in\mathbb{N}} p_X(k) &= \sum_{k\in\mathbb{N}} (1-p)^{k-1}p \\
&= p\sum_{k\in\mathbb{N}} (1-p)^{k-1} \\
&\text{indicando } h = k-1 \\
&= p\sum_{h\in\mathbb{N}_0} (1-p)^{h} \\
&\text{per la convergenza della serie geometrica:} \\
&= p\cdot\frac{1}{1-(1-p)} = 1
\end{aligned}$$

Dimostriamo il calcolo del valor medio:

$$\begin{aligned}
E(X) &= \sum_{k\in\mathbb{N}} k\cdot p_X(k) \\
&= \sum_{k\in\mathbb{N}} k\cdot(1-p)^{k-1}\cdot p \\
&= p\sum_{k\in\mathbb{N}} k\cdot(1-p)^{k-1} \\
&\text{ricordiamo che:} \\
&k(1-p)^{k-1} = -\frac{d}{dp}[(1-p)^k] \\
&\text{segue:} \\
&= -p\sum_{k\in\mathbb{N}} \frac{d}{dp}[(1-p)^k] \\
&\text{per la linearità della derivata:} \\
&= -p\frac{d}{dp}\left[\sum_{k\in\mathbb{N}}(1-p)^k\right] \\
&\text{per } k = 0 \text{ risulta un termine 1 della somma,} \\
&\text{che quindi è irrilevante per la derivata,} \\
&\text{e quindi si può far partire la sommatoria da 0} \\
&= -p\frac{d}{dp}\left[\sum_{k\in\mathbb{N}_0}(1-p)^k\right] \\
&= -p\frac{d}{dp}\left[\frac{1}{p}\right] \\
&= -p\left(-\frac{1}{p^2}\right) = \frac{1}{p}
\end{aligned}$$

Per il calcolo della varianza si utilizza lo stesso "trucco", ma utilizzando la derivata seconda, perchè invece di k c'è k^2.

Nota. In alcuni libri, la variabile geometrica è definita come il numero di insuccessi prima del primo successo, invece dell'indice del primo successo stesso. Risulta che il valore di questa variabile è traslato di -1 rispetto alla variabile geometrica che usiamo noi. Le differenze sono poche: l'alfabeto è $\mathbb{N}_0$ e la densità discreta ha k invece di $k-1$ all'esponente; valor medio e varianza si possono ottenere considerando $E(X-1)$ e $Var(X-1)$.

Esercizio 4.5. Una probabilità notevole da calcolare relativa alle variabili aleatorie geometriche è detta probabilità di lunga attesa, cioè la probabilità che il primo successo avvenga dopo una certa prova k:

$$P(X > k)$$

Soluzione.

$$\begin{aligned}
P(X > k) &= \sum_{j=k+1}^{+\infty} P(X = j) \\
&= \sum_{j=k+1}^{+\infty} (1-p)^{j-1} p \\
&= p \sum_{j=k+1}^{+\infty} (1-p)^{j-1} \\
&\text{indico } h = j - k - 1 \\
&= p \sum_{h=0}^{+\infty} (1-p)^{h+k} \\
&= p(1-p)^k \sum_{h=0}^{+\infty} (1-p)^h \\
&= p(1-p)^k \frac{1}{p} = (1-p)^k
\end{aligned}$$

Esercizio 4.6. Alla roulette (numeri da 0 a 36) si scommette ripetutamente su un numero tra 1 e 12 (compresi). Si calcoli:

1. La probabilità di perdere nelle prime 5 giocate
2. La probabilità di vincere alla sesta giocata

Soluzione. La situazione si può inserire in uno schema di Bernoulli: ciascuna prova corrisponde ad una giocata, e una prova ha successo se in quella giocata esce un numro compreso in senso largo fra 1 e 12. Ciascuna prova ha successo con probabilità $\frac{12}{37}$, e a priori non sappiamo quante prove vengono eseguite. Definiamo una variabile geometrica associata all'esperimento:

$$X \sim Ge\left(\frac{12}{37}\right)$$

Possiamo risolvere i due quesiti:

1. La probabilità di perdere nelle prime 5 giocate è data dalla probabilità di lunga attesa:

$$P(X \geqslant 5) = \left(1 - \frac{12}{37}\right)^5 \approx 0.14$$

2. La probabilità di vincere alla sesta giocata è:

$$P(X = 6) = p_X(6) = \left(1 - \frac{12}{37}\right)^5 \cdot \frac{12}{37} \approx 0.046$$

Variabili Aleatorie di Poisson Una variabile aleatoria X si dice di Poisson, o con distribuzione come Poisson, e si indica

$$X \sim Po(\lambda)$$

se è caratterizzata da

$$\begin{aligned}
\mathcal{X} &= \mathbb{N}_0 \\
p_X(k) &= e^{-\lambda} \frac{\lambda^k}{k!} \\
E(X) &= \lambda \\
Var(X) &= \lambda
\end{aligned}$$

$p_X(k)$ è una densità discreta perchè lo sviluppo in serie di Taylor di e^x è:

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}$$

quindi:

$$\begin{aligned}
\sum_{k=0}^{+\infty} p_X(k) &= \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \cdot e^{\lambda} = 1
\end{aligned}$$

Dimostriamo $E(X)$:

Dimostrazione.

$$\begin{aligned}
E(X) &= \sum_{k \in \mathbb{N}_0} k \cdot p_X(k) \\
&= \sum_{k=0}^{+\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} \\
&\text{per } k = 0 \text{ si ottiene un addendo pari a } 0 \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} k \cdot \frac{\lambda^k}{k!} \\
&k \neq 0 \Rightarrow \text{si semplifica:} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} \\
&\text{indico } h = k - 1 \\
&= e^{-\lambda} \sum_{h=0}^{+\infty} \frac{\lambda^{h+1}}{h!} \\
&= e^{-\lambda} \lambda \sum_{h=0}^{+\infty} \frac{\lambda^h}{h!} \\
&= e^{-\lambda} \lambda e^{\lambda} = \lambda
\end{aligned}$$

Per la varianza il ragionamento è molto simile. Le variabili di Poisson si usano per studiare eventi che si verificano molto poco probabilmente, ma per cui vengono eseguite molte prove. Si possono vedere come un limite per $n \to +\infty$ di $Bin(n, p)$; in particolare, il valore di una variabile aleatoria $Po(\lambda)$ è, come per $Bin(n, p)$, il numero di volte in cui una prova ha successo in uno schema di Bernoulli, in cui vengano eseguite molte prove, con una p molto piccola. Formalizzeremo questo concetto.

Esercizio 4.7. Il numero di meteoriti che colpiscono un satellite durante la sua orbita si distribuisce come una variabile aleatoria di Poisson di parametro λ. Il satellite compie un'orbita in un giorno, ed in media è colpito da 3 meteoriti. Si calcoli la probabilità che nel percorrere 5 orbite il numero di meteoriti che colpiscono il satellite sia $\leqslant 3$.

Soluzione. Dal testo, il numero di meteoriti che colpiscono un satellite durante la sua orbita è $X \sim Po(\lambda)$. Siccome la domanda si riferisce a 5 orbite, studieremo questo caso specifico. In media il satellite viene colpito da 3 meteoriti per orbita, quindi in media in 5 orbite verrà colpito da 15 meteoriti. Quindi:

$$E(X) = 15 \wedge E(X) = \lambda \Rightarrow \lambda = 15$$

Segue:

$$p_X(k) = e^{-15} \frac{15^k}{k!}$$

Calcoliamo:

$$\begin{aligned}P(X \leqslant 3) &= p_X(0) + p_X(1) + p_X(2) + p_X(3)\\ &= e^{-15}\frac{15^0}{0!} + e^{-15}\frac{15^1}{1!} + e^{-15}\frac{15^2}{2!} + e^{-15}\frac{15^3}{3!} \approx 0.0002\end{aligned}$$

Formaliziamo l'intuizione che $Po(\lambda)$ sia il limite di $Bin(n, p)$ per n grandi attraverso un teorema:

Teorema 4.3 (Limite di Poisson)**.** *Siano X_n e Y due variabili aleatorie tali che*

$$\begin{aligned}X_n &\sim Bin\left(n, \frac{\lambda}{n}\right)\\ Y &\sim Po(\lambda)\end{aligned}$$

Allora

$$\forall k \in \mathbb{N}_0 \qquad \lim_{n\to+\infty} p_{X_n}(k) = p_Y(k)$$

Dimostrazione. Dalle caratteristiche di X_n vale:

$$\begin{aligned}p_{Xn}(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{n!}{(n-k)!k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}\end{aligned}$$

Portando al limite:

$$\lim_{n\to+\infty}\frac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

Per le proprietà dei limiti:

$$\begin{aligned}&\lim_{n\to+\infty}\frac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{n^k} = \lim_{n\to+\infty}\frac{n^k + o(n^k)}{n^k} = 1\\ &\lim_{n\to+\infty}\frac{\lambda^k}{k!} = \frac{\lambda^k}{k!}\\ &\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}\\ &\lim_{n\to+\infty}\left(1-\frac{\lambda}{n}\right)^k = 1\end{aligned}$$

Segue:

$$\begin{aligned}\lim_{n\to+\infty} p_{X_n}(k) &= \lim_{n\to+\infty}\frac{n\cdot(n-1)\cdot\dots\cdot(n-k+1)}{n^k}\cdot\frac{\lambda^k}{k!}\cdot\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}\\ &= 1\cdot\frac{\lambda^k}{k!}\cdot\frac{e^{-\lambda}}{1}\\ &= e^{-\lambda}\frac{\lambda^k}{k!} = p_Y(k)\end{aligned}$$

Euristicamente si verifica che $Bin(n, p)$ si può approssimare con $Po(np)$ se

$$\begin{aligned}n &> 100\\ p &< 0.01\\ np &\leqslant 20\end{aligned}$$

Si può dimostrare che dal Teorema Limite di Poisson seguono anche

$$\lim_{n \to +\infty} E(X_n) = E(Y)$$
$$\lim_{n \to +\infty} Var(X_n) = Var(Y)$$

Nelle applicazioni reali il teorema si può applicare a situazioni compatibili con lo schema di Bernoulli in cui si verificano un numero $n \gg 1$ di prove ripetute indipendenti, ciascuna con una probabilità $p \ll 1$ di avere successo. Generalmente n e p non sono noti (ad esempio, il numero di terremoti di intensità più alta di una certa soglia in una zona sismica non è misurabile, non si possono contare uno per uno, e soprattutto non si può calcolare con una certa accuratezza la probabilità che se ne verifichi uno), ma si può calcolare la media (eseguendo misurazioni ripetute nel tempo), che per le variabili di famiglia $Bin(n, p)$ è np. Non essendo noti n e p non si può usare una variabile binomiale, e quindi si approssima con una variabile di Poisson $Po(\lambda) = Po(np)$. Oltretutto, per n grandi il calcolo dei coefficienti binomiali nella densità discreta di $Bin(n, p)$ diventa molto difficoltoso, mentre la densità discreta di $Po(\lambda)$ è sempre molto calcolabile. Negli esercizi, le possibilità di applicazione del Teorema Limite sono due:

- Può essere che siano dati un numero di prove e una probabilità di successo che rispettano l'euristica, e per cui i calcoli con $B(n, p)$ sono difficili
- Può essere che non siano dati n e p ma direttamente il valore medio np, e quindi si usa direttamente $Po(np)$

Esercizio 4.8. La probabilità che in una partita di poker venga servito un full è 0.0014. Vengono giocate 1000 partite. Calcolare la probabilità che vengano serviti alpiù 3 full.

Soluzione. L'esercizio si può studiare attraverso uno schema di Bernoulli in cui:

- La prova ripetuta è una partita di poker, e il successo consiste nel servizio di un full.
- Il numero di prove ripetute è $n = 1000$
- la probabilità di successo è $p = 0.0014$

Il numero di successi su n prove si può rappresentare con una variabile aleatoria X tale che:

$$X \sim Bin(n, p) = Bin(1000, 0.0014)$$

Valgono

$$\begin{aligned} n &= 1000 > 100 \\ p &= 0.0014 < 0.01 \\ np &= 1.4 \leqslant 20 \end{aligned}$$

Possiamo quindi utilizzare l'approssimazione di Poisson: consideriamo una variabile aleatoria Y tale che:

$$Y \sim Po(np) = Po(1.4)$$

Calcoliamo la probabilità richiesta:

$$\begin{aligned} &P(X \leqslant 3) \approx \\ &P(Y \leqslant 3) = \\ &P(Y = 0) + P(Y = 1) + P(Y = 2) + P(Y = 3) = \\ &e^{-1.4}\frac{(1.4)^0}{0!} + e^{-1.4}\frac{(1.4)^1}{1!} + e^{-1.4}\frac{(1.4)^2}{2!} + e^{-1.4}\frac{(1.4)^3}{3!} \approx 0.95 \end{aligned}$$

Esercizio 4.9. Il numero di influenze che una data persona contrae in un anno si distribuisce come $Po(5)$. Viene proposto un vaccino che fa diminuire a 3 il numero di influenze contratte annualmente, ma funziona solo nel 75% dei casi. Se un individuo si vaccina e quell'anno contrae solo 2 influenze, qual è la probabilità che il vaccino sia stato efficace?

Soluzione. Definiamo alcuni eventi di interesse:

$$E = \{\text{"il vaccino è stato efficace"}\}$$
$$F = \{\text{"l'individuo è vaccinato e contrae 2 influenze"}\}$$

Definiamo inoltre le variabili aleatorie

$$X_i = Po(i) \quad i \in \{3, 5\}$$

Dobbiamo calcolare la probabilità $P(E|F)$. Non abbiamo informazioni su questa probabilità condizionata, ma possiamo ricavarne sul contrario, quindi applichiamo il Teorema di Bayes e la formula di probabilità totale:

$$\begin{aligned} P(E|F) &= \frac{P(F|E)P(E)}{P(F)} \\ &= \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)} \end{aligned}$$

Dal testo, $P(E) = 0.75$, e quindi $P(E^C) = 0.25$. Le probabilità condizionate a E e E^C sono date dalle probabilità che le rispettive variabili aleatorie di Poisson valgano 2:

$$\frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)} =$$
$$\frac{P(X_3 = 2)P(E)}{P(X_3 = 2)P(E) + P(X_5 = 2)P(E^C)} =$$
$$\frac{e^{-3}\frac{3^2}{2!} \cdot 0.75}{e^{-3}\frac{3^2}{2!} \cdot 0.75 + e^{-5}\frac{5^2}{2!} \cdot 0.25}$$

5 Vettori Aleatori Discreti

Con gli strumenti che abbiamo non siamo ancora in grado di calcolare alcune cose che idealmente dovrebbero essere semplici, ad esempio $E(XY)$ oppure $Var(X + Y)$, con X e Y variabili aleatorie. Consideriamo in particolare $Var(X + Y)$:

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

Definiamo una grandezza che caratterizza X e Y:

Definizione 5.1. Si definisce la covarianza fra due variabili aleatorie X e Y:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Risulta:

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Dobbiamo analizzare la covarianza più nel dettaglio. Osserviamo, dalla definizione, due cose:

Osservazione 5.1.

$$Y = X \Rightarrow Cov(X, Y) = Cov(X, X) = Var(X)$$

Osservazione 5.2.

$$\begin{aligned}
Cov(X,Y) &= E[(X - E(X))(Y - E(Y))] \\
&= E[XY - YE(X) - XE(Y) + E(X)E(Y))] \\
&= E(XY) - E(X)E(Y) \cancel{- E(X)E(Y)} \cancel{+ E(X)E(Y)} \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

Questa è una definizione alternativa della covarianza.

Per calcolare la covarianza sfruttando l'Osservazione 2, avendo caratterizzato X e Y attraverso i rispettivi alfabeti e le rispettive densità discrete, è piuttosto semplice calcolare $E(X)$ e $E(Y)$; per calcolare $E(XY)$ è necessario caratterizzare la variabile XY, e in particolare calcolare $p_{XY}(x_i \cdot y_j)$ per ogni coppia $(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}$. Questi valori della densità discreta corrispondono alle probabilità $P(X = x_i, Y = y_j)$, che corrispondono alle probabilità delle intersezioni fra i due eventi $\{X = x_i\}$ e $\{Y = y_j\}$; queste probabilità sono calcolabili esplicitamente solo se i due eventi sono indipendenti. Il motivo per cui non sono calcolabili in generale a partire da p_X e p_Y è che racchiudono e codificano tutte le informazioni sul modo in cui X e Y si relazionano, che sono assenti dalle densità discrete prese singolarmente. Dobbiamo quindi definire un costrutto più potente delle variabili aleatorie per trattare questi casi:

Definizione 5.2. Sia $(\Omega, \mathbb{P}(\Omega), P)$ uno spazio di probabilità discreto. Si dice vettore aleatorio n-dimensionale una funzione:

$$\begin{aligned}
\boldsymbol{V} &: \Omega \to R^n \\
&\omega \mapsto \boldsymbol{V}(\omega) = (X_1(\omega), \dots, X_n(\omega))
\end{aligned}$$

Noi ci concentreremo sul caso $n = 2$:

$$\begin{aligned}
\boldsymbol{V} &: \Omega \to R^2 \\
&\omega \mapsto \boldsymbol{V}(\omega) = (X(\omega), Y(\omega))
\end{aligned}$$

X e Y sono variabili aleatorie discrete, come quelle che abbiamo visto fino adesso, con alfabeti $\mathcal{X}$ e $\mathcal{Y}$ e densità discrete p_X e p_Y.

5.1 Caratterizzazione dei Vettori Aleatori

Cerchiamo adesso di caratterizzare i vettori aleatori come abbiamo fatto per le variabili aleatorie.

Definizione 5.3. Sia $\boldsymbol{V}$ un vettore aleatorio discreto bidimensionale. Si dice alfabeto di $\boldsymbol{V}$ l'insieme delle coppie di valori che il vettore può assumere. Si ha

$$\mathcal{V} \subseteq \mathcal{X} \times \mathcal{Y}$$

Generalemente l'inclusione $\mathcal{V} \subseteq \mathcal{X} \times \mathcal{Y}$ è un'inclusione stretta. Si verifica l'uguaglianza quando $\{X = x_i\}$ e $\{Y = y_j\}$ sono eventi indipendenti per ogni coppia $(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}$. Definiamo andesso il corrispettivo della densità discreta:

Definizione 5.4. Si chiama densità discreta congiunta di un vettore aleatorio $\boldsymbol{V} = (X, Y)$, o più comunemente delle variabili X e Y, la mappa

$$\begin{aligned}
p_{XY} &: \mathcal{X} \times \mathcal{Y} \to [0, 1] \\
&(x_i, y_j) \mapsto P(X = x_i, Y = y_j)
\end{aligned}$$

Osservazione 5.3. Si considera $\mathcal{X} \times \mathcal{Y}$ come dominio di p_{XY} per comodità. Per i punti $(x_i, y_j) \notin \mathcal{V}$ vale $p_{XY}(x_i, y_j) = 0$, perchè se $\boldsymbol{V}$ non può assumere valore (x_i, y_j), significa che $\{X = x_i\} \cap \{Y = y_j\} = \varnothing$

La densità congiunta è una misura di probabilità, quindi vale

1. $$p_{XY}(x_i, y_j) \geqslant 0 \quad \forall (x_i, y_j) \in \mathcal{V}$$

2. $$\sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} p_{XY}(x_i, y_j) = 1$$

Lemma 5.1. *Le densità discrete p_X e p_Y di X e Y rispettivamente (dette anche densità marginali) si ottengono da p_{XY} come*

$$p_X(x_i) = \sum_{y_j \in \mathcal{Y}} p_{XY}(y_j) \qquad\qquad p_Y(y_j) = \sum_{x_i \in \mathcal{X}} p_{XY}(x_i)$$

Mentre è possibile calcolare le densità marginali dalla densità congiunta, non è sempre possibile il viceversa. Si possono adesso definire valor medio e varianza di variabili aleatorie ottenute come funzioni di più variabili aleatorie (cioè come funzioni di vettori aleatori):

Definizione 5.5. Sia (X, Y) vettore aleatorio discreto, sia $g : \mathbb{R}^2 \to \mathbb{R}$ funzione sufficientemente regolare. Allora si definisce

$$E[g(X,Y)] = \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} g(x_i, y_j) p_{XY}(x_i, y_j)$$

5.2 Indipendenza fra Variabili Aleatorie

Introduciamo il concetto di indipendenza fra variabili aleatorie per una coppia di variabili:

Definizione 5.6. La coppia (X, Y) si dice indipendente (o analogamente X e Y si dicono indipendenti) se e solo se

$$p_{XY}(x_i, y_j) = p_X(x_i) p_Y(y_j) \quad \forall x_i \in \mathcal{X}, \forall y_j \in \mathcal{Y}$$

La definizione è analoga a quella di eventi indipendenti, con la differenza che in questo caso bisogna dimostrare che tutte le probabilità congiunte si fattorizzano nel prodotto delle marginali, quindi generalmente è un calcolo piuttosto oneroso.

Osservazione 5.4.

$$X, Y \text{ indipendenti} \Leftrightarrow f(X), g(Y) \text{ indipendenti} \quad \forall f, g : \mathbb{R} \to \mathbb{R}$$

Vediamo come questo concetto influenza il valor medio

Teorema 5.1. *Se X e Y sono indipendenti, allora*

$$E(XY) = E(X)E(Y)$$

Dimostrazione.

$$\begin{aligned} E(XY) &= \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} x_i y_j p_{XY}(x_i, y_j) \\ &\text{per l'indipendenza di } X \text{ e } Y\text{:} \\ &= \sum_{x_i \in \mathcal{X}} \sum_{y_j \in \mathcal{Y}} x_i y_j p_X(x_i) p_Y(y_j) \\ &= \sum_{x_i \in \mathcal{X}} x_i p_X(x_i) \sum_{y_j \in \mathcal{Y}} y_j p_Y(y_j) \\ &= E(X)E(Y) \end{aligned}$$

Da questo teorema seguono due conseguenze: se X e Y sono indipendenti

1. $Cov(X, Y) = 0$

   Dimostrazione.

   $$\begin{aligned} Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) = 0 \end{aligned}$$

   Nota. Se $Cov(X, Y) = 0$ allora X e Y si dicono scorrelate. Chiaramente, l'indipendenza implica la scorrelazione; non è vero il viceversa.

2. $Var(X + Y) = Var(X) + Var(Y)$

Dimostrazione. Abbiamo visto che

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Ma dal punto 1. $Cov(X,Y) = 0$, quindi:

$$Var(X+Y) = Var(X) + Var(Y) \cancel{+2Cov(X,Y)} = Var(X) + Var(Y)$$

NB. Dal punto 2. sappiamo che se le variabili sono indipendenti la varianza è additiva, ma non lineare, perchè mentre si distribuisce sulle somme, le costanti continuano ad essere "portate fuori" al quadrato.

5.2.1 Generalizzazione a più Variabili

Definizione 5.7. L'insieme di variabili aleatorie $\{X_1, X_2, \dots, X_n\}$ è detto indipendente se e solo se

$$p_{X_1,\dots,X_n}(x_1,\dots,x_n) = p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n) \quad \forall x_1,\dots,x_n$$

Non vedremo molto sull'indipendenza di più variabili, ma una proprietà importante è

Proposizione 5.1. *Se $X_1,\dots,X_n$ sono indipendenti, allora*

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Esercizio 5.1. Si lancia un dado truccato in modo che

$$P(\text{"ottenere punteggio } i\text{"}) = \begin{cases} p & \text{se } i \text{ è dispari} \\ 2p & \text{se } i \text{ è pari} \end{cases}$$

con $p \in [0,1]$. Indicando con D il punteggio del dado, definiamo le variabili aleatorie

$$X = \begin{cases} 1 & \text{se } D \text{ è pari} \\ 0 & \text{altrimenti} \end{cases}$$

e

$$Y = \begin{cases} 1 & \text{se } D > 3 \\ 0 & \text{altrimenti} \end{cases}$$

Determinare la densità congiunta di X e Y.

Soluzione. Determiniamo p sfruttando la condizione di normalizzazione:

$$3p + 3(2p) = 1 \Leftrightarrow p = \frac{1}{9}$$

Per il punteggio del dado si ottiene:

$$P(D=1) = P(D=3) = P(D=5) = \frac{1}{9}$$
$$P(D=2) = P(D=4) = P(D=6) = \frac{2}{9}$$

Caratterizziamo il vettore (X,Y): l'alfabeto è

$$\mathcal{X} \times \mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1)\}$$

La densità congiunta è data da:

$$
\begin{aligned}
p_{XY}(0,0) &= P(\{D \text{ dispari}\} \cap \{D \leqslant 3\}) \\
&= P(D \in \{1,3\}) \\
&= P(D=1) + P(D=3) = \frac{2}{9} \\
p_{XY}(0,1) &= P(\{D \text{ dispari}\} \cap \{D > 3\}) \\
&= P(D=5) = \frac{1}{9} \\
p_{XY}(1,0) &= P(\{D \text{ pari}\} \cap \{D \geqslant 3\}) \\
&= P(D=2) = \frac{2}{9} \\
p_{XY}(1,1) &= P(\{D \text{ pari}\} \cap \{D > 3\}) \\
&= P(D \in \{4,6\}) \\
&= P(D=4) + P(D=6) = \frac{4}{9}
\end{aligned}
$$

Esercizio 5.2. Sul tavolo ci sono due monete, delle quali una sola è equilibrata. Quella truccata ha una probabilità di ottenere testa di $\frac{1}{3}$. Si sceglie a caso una delle due monete e si lancia due volte. Consideriamo le variabili aleatorie

$$
X_i = \begin{cases} 1 & \text{se all'}i\text{-esimo lancio si ottiene testa} \\ 0 & \text{altrimenti} \end{cases}
$$

per $i = 1, 2$.

1. Calcolare la densità congiunta di (X_1, X_2)
2. Calcolare $E(X_1)$ e $E(X_2)$
3. Dire se X_1 e X_2 sono indipendenti

Soluzione. 1. Definiamo l'evento

$$F = \text{"prendiamo la moneta onesta"}$$

Possiamo adesso calcolare la densità congiunta di (X_1, X_2), considerando che l'alfabeto di (X_1, X_2) è $\{0,1\}^2$, ed utilizzando la formula di probabilità totale:

$$
\begin{aligned}
p_{X_1X_2}(0,0) &= P(X_1 = 0 \cap X_2 = 0) \\
&= P(X_1 = 0 \cap X_2 = 0|F)P(F) + P(X_1 = 0 \cap X_2 = 0|F^C)P(F^C) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{25}{72} \\
p_{X_1X_2}(0,1) &= P(X_1 = 0 \cap X_2 = 1) \\
&= P(X_1 = 0 \cap X_2 = 1|F)P(F) + P(X_1 = 0 \cap X_2 = 1|F^C)P(F^C) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{17}{72} \\
p_{X_1X_2}(1,0) &= P(X_1 = 1 \cap X_2 = 0) \\
&= P(X_1 = 1 \cap X_2 = 0|F)P(F) + P(X_1 = 1 \cap X_2 = 0|F^C)P(F^C) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{2} = \frac{17}{72} \\
p_{X_1X_2}(1,1) &= P(X_1 = 1 \cap X_2 = 1) \\
&= P(X_1 = 1 \cap X_2 = 1|F)P(F) + P(X_1 = 1 \cap X_2 = 1|F^C)P(F^C) \\
&= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{13}{72}
\end{aligned}
$$

2. Notiamo che

$$X_i \sim Be(p_i)$$

Quindi, $E(X_i) = p_i$. È quindi sufficiente calcolare le due densità marginali:

$$\begin{aligned} E(X_1) &= p_{X_1}(1) \\ &= p_{X_1X_2}(1,0) + p_{X_1X_2}(1,1) \\ &= \frac{17}{72} + \frac{13}{72} = \frac{5}{12} \\ E(X_2) &= p_{X_1}(2) \\ &= p_{X_1X_2}(0,1) + p_{X_1X_2}(1,1) \\ &= \frac{17}{72} + \frac{13}{72} = \frac{5}{12} \end{aligned}$$

3. Non sono indipendenti, perchè si ha

$$p_{X_1X_2}(1,1) \neq p_{X_1}(1)p_{X_2}(1)$$

6 Variabili Aleatorie Assolutamente Continue

La definizione essenzialmente è la stessa delle variabili aleatorie discrete, se non che lo spazio campionario sottostante sarà continuo invece che discreto. In realtà c'è un problema nella definizione della misura di probabilità per spazi campionari di questo tipo, che ha svariate sfaccettature, e che noi non vedremo. Noi daremo una descrizione probabilistica delle variabili aleatorie assolutamente continue, lasciando che lo spazio di probabilità rimanga implicito.

Definizione 6.1. Una variabile aleatoria X si dice (assolutamente) continua se esiste una funzione $f_X : \mathbb{R} \to \mathbb{R}$, detta densità di probabilità, tale che per ogni intervallo $[a, b] \subseteq \mathbb{R}$ vale

$$P(a \leqslant X \leqslant b) = \int_a^b f_X(x)dx$$

Nota. 1. Se non ci sono ambiguità, generalmente la X a pedice di f si omette.

2. Esistono anche variabili aleatorie continue non assolutamente, ma noi non le vedremo, quindi spesso ometteremo il termine "assolutamente".

La funzione f deve soddisfare

1. Positività

$$f_X(x) \geqslant 0 \quad \forall x \in \mathbb{R}$$

2. Normalizzazione

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

Nota. La funzione densità di probabilità ha un significato diverso dalla densità discreta, perchè quest'ultima è effettivamente una probabilità, invece la funzione f non lo è, mentre lo è il suo integrale.

Osservazione 6.1.

- La funzione f non è necessariamente continua.
- La probabilità che X appartenga all'intervallo $[a, b]$ è data dall'area sottesa dal grafico di f_X, l'asse x e le due rette $x = a$ e $x = b$.
- Non si può parlare di "probabilità che X assuma valore a", perchè la questa sarebbe data da

$$\begin{aligned} P(X = a) &= \lim_{\varepsilon \to 0} P(a - \varepsilon \leqslant X \leqslant a + \varepsilon) \\ &= \lim_{\varepsilon \to 0} \int_{a-\varepsilon}^{a+\varepsilon} f_X(x)dx = 0 \end{aligned}$$

Questo non va interpretato come "l'evento $X = a$ è impossibile", ma il significato è che la probabilità $P(X = a)$ non ha significato nell'ambito delle variabili aleatorie continue.

- La funzione f_x va intesa come una densità in senso fisico: per un certo valore $a \in \mathbb{R}$, $f_x(a)$ rappresenta non la probabilità di a ma il modo in cui la massa di probabilità di X si distribuisce attorno ad a. In un senso molto approssimato, si può intendere $f_X(a)$ come la probabilità che X assuma un valore in un intorno di a diviso l'ampiezza dell'intorno, quando questa ampiezza tende a 0.

Esercizio 6.1. Sia X variabile aleatoria continua con densità

$$f_X(x) = \begin{cases} c(4x - 2x^2) & \text{se } 0 < x < 2 \\ 0 & \text{altrimenti} \end{cases}$$

Calcoliamo c e $P(X > 1)$.

Soluzione. Siccome f_X è una densità deve valere:

$$\begin{aligned} \int_{\mathbb{R}} f(x)dx &= 1 \\ \int_{-\infty}^{0} 0 + \int_{0}^{2} c(4x - 2x^2)dx + \int_{2}^{+\infty} &= 1 \\ 0 + c\left[2x^2 - \frac{2}{3}x^3\right]_0^2 + 0 &= 1 \\ \frac{8}{3}c &= 1 \\ c &= \frac{3}{8} \end{aligned}$$

Possiamo calcolare $P(X > 1)$:

$$\begin{aligned} P(X > 1) &= \int_{1}^{+\infty} f_X(x)dx \\ &= \frac{3}{8}\int_{1}^{2}(4x - 2x^2) = \frac{1}{2} \end{aligned}$$

In alcuni casi le stesse informazioni fornite dalla funzione di densità vengono date da un'altra funzione, detta di distribuzione o ripartizione, che può essere più comoda. In questo modo si ottiene una caratterizzazione equivalente della variabile aleatoria in esame.

Definizione 6.2. Sia X una variabile aleatoria continua. Si dice funzione di distribuzione o di ripartizione di X la funzione

$$\begin{aligned} F_X : \mathbb{R} &\to [0, 1] \\ x &\mapsto F_X(x) = P(X \leqslant x) \end{aligned}$$

Osservazione 6.2. Se F_X è nota, possiamo calcolare $P(X \in (a, b])$; infatti:

$$\begin{aligned} P(X \in (a, b]) &= P(a < X \leqslant b) \\ &= P(\{X \leqslant b\} \setminus \{X \leqslant a\}) \\ &= P(\{X \leqslant b\}) - P(\{X \leqslant a\}) \\ &= F_X(b) - F_X(a) \end{aligned}$$

Vediamo due proprietà della funzione di distribuzione

1. $$\lim_{x \to -\infty} F_X(x) = 0 \qquad\qquad \lim_{x \to +\infty} F_X(x) = 1$$

2. $$x \leqslant y \Rightarrow F_X(x) \leqslant F_X(y)$$

Fra funzione densità di probabilità e funzione di ripartizione c'è un legame piuttosto stretto, che deriva direttamente dalla definizione di F_X:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$
$$f_X(x) = \frac{d}{dx}F_X(x)$$

La funzione di distribuzione si può utilizzare per calcolare la densità di funzioni invertibili di variabili aleatorie; questo non lo formalizziamo, ma lo vediamo con un esempio.

Esempio 6.1. Supponiamo che X sia una variabile continua con densità f_X e funzione di distribuzione F_X. Definiamo $Y = 2X$, e determiniamo la densità di Y.

Soluzione. La cosa più semplice è caratterizzare la funzione di distribuzione di Y:

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) \\ &= P(2X \leqslant y) \\ &= P\left(X \leqslant \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right) \end{aligned}$$

Possiamo quindi calcolare la densità:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) \\ &= \frac{d}{dy}F_X\left(\frac{y}{2}\right) \\ &= F'_X\left(\frac{y}{2}\right) \cdot \frac{1}{2} = \frac{f_X\left(\frac{y}{2}\right)}{2} \end{aligned}$$

6.1 Valor Medio e Varianza di Variabili Aleatorie Continue

Definizione 6.3. Sia X variabile aleatoria continua con densità f_X; il suo valor medio è dato da

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx$$

se l'integrale esiste finito.

NB. Non tutte le variabili aleatorie continue hanno valor medio. Ad esempio, le variabili aleatorie di Cauchy, cioè con densità

$$f_X(x) = \frac{1}{\pi}\frac{1}{x^2+1}$$

non hanno valor medio.

Teorema 6.1. *Sia X una variabile aleatoria continua, sia $g : \mathbb{R} \to \mathbb{R}$ una funzione sufficientemente regolare. Allora $g(X)$ è ancora una variabile aleatoria continua, ed il suo valor medio è*

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f_X(x)dx$$

se l'integrale esiste finito.

Nota. Chiaramente, potrebbe essere che X abbia valor medio, ma che $g(X)$ non ce l'abbia

Come applicazione del teorema, possiamo scegliere $g(x) = (x - E(X))^2$ ed ottenere la varianza di X come

$$Var(X) = E[(X - E(X))^2] = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x)dx$$

Come per le variabili aleatorie discrete, vale

$$Var(X) = E(X^2) - E(X)^2$$

Chiaramente, anche per calcolare $E(X^2)$ è necessario applicare il teorema.

6.2 Variabili Aleatorie Continue Notevoli

Analogamente alle variabili aleatorie discrete, si possono considerare due variabili equivalenti se hanno la stessa probabilità:

Definizione 6.4. Siano X e Y due varibili aleatorie continue con stessa funzione densità di probabilità:

$$f_X(x) = f_Y(x) \quad \forall x \in \mathbb{R}$$

Allora X e Y si dicono probabilisticamente equivalenti e si indica

$$X \sim Y$$

Vaiabili Aleatorie Uniformi Una variabile X si dice uniforme e si indica

$$X \sim U(a,b) \quad \text{con } a < b$$

se è caratterizzata da

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a,b] \\ 0 & \text{se } x \notin [a,b] \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leqslant x \leqslant b \\ 1 & \text{se } x > b \end{cases}$$

$$E(X) = \frac{b+a}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

Le espressioni di $F_X(x)$, $E(X)$ e $Var(X)$ si ricavano da $f_X(x)$ applicando le definizioni.

Variabili Aleatorie Esponenziali Una variabile X si dice esponenziale e si indica

$$X \sim Exp(\lambda) \quad \text{con } \lambda > 0$$

se è caratterizzata da

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geqslant 0 \\ 0 & \text{se } x < 0 \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-\lambda x} & \text{se } x \geqslant 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Le espressioni di $F_X(x)$, $E(X)$ e $Var(X)$ si ricavano attraverso le definizioni (per calcolare $E(X)$e $Var(X)$ bisogna integrare per parti).

Usi delle Variabili Aleatorie Esponenziali Le variabili esponenziali si utilizzano per modellare tempi di primo arrivo o tempi di attesa, e infatti sono fortemente legate alle variabili di Poisson. Supponiamo ad esempio che N_1 sia la variabile aleatoria che modella le richieste di servizio di un server per unità di tempo. Tipicamente si assume che $N_1 \sim Po(\lambda)$, dove λ è il numero medio di richieste al server per unità di tempo. Il numero di richieste in t unità di tempo sarà quindi $N_t \sim Po(\lambda t)$. Definiamo ora W la variabile aleatoria che rappresenta il tempo di attesa fino all'arrivo della prima richiesta; si ha

$$P(W > t) = P(N_t = 0) = e^{-\lambda t}$$

Segue che
$$P(W \leqslant t) = 1 - P(W > t) = 1 - e^{-\lambda t}$$
Se t è un istante infinitesimo di tempo, questa è la funzione di ripartizione di W:
$$F_X(t) = 1 - e^{-\lambda t}$$
Quindi $W \sim Exp(\lambda)$.

Proprietà di Assenza di Memoria Se $X \sim Exp(\lambda)$, allora X "non ha memoria":
$$P(X \geqslant T + t | X \geqslant T) = P(X \geqslant t)$$
Dimostrazione.
$$\begin{aligned} P(X \geqslant T + t | X \geqslant T) &= \frac{P(X \geqslant T + t, X \geqslant T)}{P(X \geqslant T)} \\ &= \frac{P(X \geqslant T + t)}{P(X \geqslant T)} \\ &= \frac{1 - P(X \leqslant T + t)}{1 - P(X \leqslant T)} \\ &= \frac{1 - F_X(T + t)}{1 - F_X(T)} \\ &= \frac{e^{-\lambda(T+t)}}{e^{-\lambda T}} \\ &= e^{-\lambda t} \\ &= 1 - F_X(t) \\ &= 1 - P(X \leqslant t) \\ &= P(X \geqslant t) \end{aligned}$$

Le variabili esponenziali sono le uniche variabili assolutamente continue ad avere questa proprietà.

Esercizio 6.2. La lunghezza di una telefonata (in minuti) è una variabile aleatoria $X \sim Exp(\frac{1}{10}$. Se qualcuno arriva prima di noi alla cabina telefonica, determinare la probabilità di dover aspettare

1. più di 10 minuti
2. tra 10 e 20 minuti

Soluzione. 1.
$$\begin{aligned} P(X > 10) &= 1 - P(X \leqslant 10) \\ &= 1 - F_X(10) \\ &= 1 - 1 + e^{-10 \cdot \frac{1}{10}} = e^{-1} \approx 0.368 \end{aligned}$$

2.
$$\begin{aligned} P(10 < X < 20) &= F_X(20) - F_X(10) \\ &= 1 - e^{-20 \cdot \frac{1}{10}} - 1 + e^{-10 \cdot \frac{1}{10}} \\ &= e^{-1} - e^{-2} \approx 0.233 \end{aligned}$$

Esercizio 6.3. Il numero di km che un'automo percorre prima che la batteria ceda è una variabile aleatoria esponenziale di media 10000. Se una persona desidera fare un viaggio di 5000km, qual è la probabilità che effettui il viaggio senza cambiare la batteria?

Soluzione. Definiamo X la variabile aleatoria che descrive la durata della batteria misurata in $Mm = 10^6 m$. Quindi, siccome la media di X è 10, vale

$$X \sim Exp(\frac{1}{10})$$

Sia T il tempo di usura attuale della batteria. Vogliamo calcolare

$$P(X > T + 5 | X > T)$$

Per l'assenza di memoria delle variabili esponenziali:

$$P(X > T + 5 | X > T) = P(X > 5)$$

Segue:

$$\begin{aligned} P(X > 5) &= 1 - P(X \leqslant 5) \\ &= 1 - F_X(5) \\ &= e^{-5 \cdot \frac{1}{10}} \\ &= e^{-\frac{1}{2}} \approx 0.604 \end{aligned}$$

Varaibile Aleatorie Gaussiane o Normali Una variabile si dice gaussiana o normale e si indica

$$X \sim N(\mu, \sigma^2) \quad \text{con } \mu \in \mathbb{R}, \sigma^2 > 0$$

se è caratterizzata da

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Osservazione 6.3. Per le variabili normali non è possibile dare la funzione di distribuzione perchè e^{-x^2} non ammette primitiva in modo analitico.

La funzione densità di X ha un grafico detto "a campana". Il grafico è simmetrico rispetto μ, e la larghezza è data dalla varianza. Calcoliamo $E(X)$:

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\text{sommo e sottraggo } \mu \\ &= \int_{-\infty}^{+\infty} \frac{x - \mu + \mu}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{+\infty} \frac{x - \mu}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\text{il secondo integrale è 1, in quanto la funzione integranda è } f_X(x) \\ &\text{cambio variabile nel primo integrale: } y = \frac{x-\mu}{\sqrt{2\sigma^2}},\ dy = dx \\ &= \int_{-\infty}^{+\infty} \frac{y}{\sqrt{\pi}} e^{-y^2} dy + \mu \\ &\text{la funzine integranda è dispari, e il dominio è simmetrico rispetto 0} \\ &\text{quindi l'integrale è nullo} \\ &= 0 + \mu = \mu \end{aligned}$$

Per calcolare la varianza invece di usare la relazione $Var(X) = E(X^2) - (E(X))^2$ in questo caso è più comodo usare la definizione, perchè con lo stesso cambio di variabile di prima e un'integrazione per parti si ricava. Il risultato è $Var(X) = \sigma^2$.

Notazione. σ si dice deviazione standard.

Trasformazioni Affini di Variabili Normali Sia $X \sim N(\mu, \sigma^2)$, definiamo $Y = aX + b$. Allora

$$Y \sim N(a\mu + b, a^2\sigma^2)$$

Quindi non è solo possibile ricavare valor medio e varianza di Y (che sono noti dalle loro proprietà), ma sappiamo anche che Y è ancora normale.

Dimostrazione. Caratterizziamo i parametri:

-
$$\begin{aligned} E(Y) &= E(aX + b) \\ &= aE(X) + b \\ &= a\mu + b \end{aligned}$$

-
$$\begin{aligned} Var(Y) &= Var(aX + b) \\ &= Var(aX) \\ &= a^2 Var(X) \\ &= a^2\sigma^2 \end{aligned}$$

Determiniamo ora la funzione di distribuzione di Y: per un generico $y \in \mathbb{R}$

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) \\ &= P(aX + b \leqslant y) \\ &= P\left(X \leqslant \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Deriviamo:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) \\ &= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\ &= F_X'\left(\frac{y-b}{a}\right) \frac{d}{dy} \frac{y-b}{a} \\ &= f_Y\left(\frac{y-b}{a}\right) \cdot \frac{1}{a} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}} \frac{1}{a} \\ &= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \cdot e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}} \end{aligned}$$

La conseguenza più importante di questa proprietà è che tutte le variabili aleatorie normali si possono ricondurre ad una variabile normale detta centrata standard, che si indica generalmente con Z, che ha media 0 e varianza 1. Se $X \sim N(\mu, \sigma^2)$, allora la normale centrata standard è data da

$$Z = \frac{X - \mu}{\sigma}$$

Questo è molto utile per due motivi:

1. Siccome F_X non si può calcolare analiticamente, va stimata con l'analisi numerica; in tabella è fornita la funzione di ripartizione della centrata standard, e tutte le altre vanno ricavate.

2. Si possono confrontare dati che si distribuiscono con normali molto diverse, riconducendosi alla centrata standard.

Lemma 6.1. *Se $X_1, \dots, X_n$ sono variabili aleatorie normali indipendenti, con*

$$X_i \sim N(\mu_i, \sigma_i^2)$$

Siano $a_1, \dots, a_n \in \mathbb{R}$. Allora

$$\sum_{i=n}^{n} a_i X_i \sim N(\sum_{i=n}^{n} a_i \mu_i, \sum_{i=n}^{n} a_i^2 \sigma_i^2)$$

Notazione. La funzione di distribuzione di una variabile normale centrata standard si indica

$$\Phi(z)$$

Esercizio 6.4. Una macchina confeziona barattoli di caffè del contenuto nominale di $250g$ di caffè. Il peso reale è invece una variabile aleatoria normale di media $250g$.

1. Calcolare la deviazione standard del peso sapendo che il 5% dei barattoli pesa più di $250g$

2. Calcolare la probabilità che un barattolo pesi meno di $245g$

Soluzione. 1. Indichiamo con X il peso dei barattoli, quindi:

$$X \sim N(250, \sigma^2)$$

Vogliamo fissare σ in modo che

$$P(X > 252) = 0.05$$

Ci riportiamo alla normale centrata standard:

$$\begin{aligned} P\left(\frac{X - 250}{\sigma} > \frac{252 - 250}{\sigma}\right) &= P\left(Z > \frac{2}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{2}{\sigma}\right) = 0.05 \end{aligned}$$

Si ha

$$\Phi\left(\frac{2}{\sigma}\right) = 0.95$$

Da tabella si ricava

$$\frac{2}{\sigma} \approx 1.645 \Rightarrow \sigma \approx 1.22$$

2. Normalizziamo e risolviamo:

$$\begin{aligned} P(X < 245) &= P\left(\frac{X - 250}{1.22} < \frac{245 - 250}{1.22}\right) \\ &= P(Z < -4.098) \\ &= \Phi(-4.098) \\ &= 1 - \Phi(4.098) \approx 1 - 1 = 0 \end{aligned}$$