

Appunti di Probabilità e Statistica

Riccardo Agatea

15 aprile 2019

Indice

1	Introduzione	1
2	Spazio di Probabilità	1
2.1	Definizioni	1
2.2	Operazioni sugli Insiemi	3
2.3	σ -algebra \mathcal{F}	5
2.4	Misura di probabilità	7
2.4.1	Misura Empirica	9
2.4.2	Ω discreto	9
2.4.3	Probabilità Uniforme	10
3	Approfondimenti sulla Probabilità	10
3.1	Principio Fondamentale del Conteggio	10
3.2	Probabilità Condizionata	12
3.3	Indipendenza	16
3.3.1	Schema di Bernoulli	18
4	Variabili Aleatorie Discrete	20
4.1	Concetti Base per le Variabili Aleatorie	20
4.2	Valor Medio	23
4.2.1	Valor Medio di una Funzione di una Variabile Aleatoria	26
4.3	Varianza	26
4.4	Variabili Aleatorie Discrete Notevoli	29
4.4.1	Alfabeto Finito	29
4.4.2	Alfabeto Infinito	33

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

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

Per ricavare una stima di I è possibile considerare un insieme $X = \{x_1, x_2, \dots, x_n\}$ 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 $n \rightarrow +\infty$:

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

2 Spazio di Probabilità

2.1 Definizioni

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

$$P : \mathcal{F} \rightarrow [0, 1]$$

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

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

- 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.
- \mathcal{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 $E \in \mathcal{F}$, 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 , \mathcal{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 affinché esca testa la prima volta.

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

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 (Ω, \mathcal{F}, P) si dice discreto se Ω è finito o numerabile:

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

Definizione 2.3. Uno spazio di probabilità (Ω, \mathcal{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 $E \in \mathcal{F}$ 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:

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

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 \mathcal{F} sono \emptyset e Ω :

- \emptyset rappresenta il non presentarsi di alcuno degli eventi elementari in Ω . Chiaramente almeno uno degli eventi elementari si deve presentare, quindi \emptyset 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 (Ω, \mathcal{F}, P) , e due eventi $E, F \in \mathcal{F}$. Essendo \mathcal{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 \mathcal{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
- $E \setminus F$ 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:

$$E \cap F = \emptyset$$

Chiaramente, per gli elementi di \mathcal{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

$$\bigcup_i A_i = A$$

$$A_i \cap A_j = \emptyset \quad \forall i \neq j$$

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:

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

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 $\{E_n\}_n$:

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

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 σ -algebra \mathcal{F}

Abbiamo detto che \mathcal{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 $E \in \mathcal{F}$.

Il fatto che \mathcal{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 \mathcal{F} è una σ -algebra:

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

1. $\mathcal{A} \neq \emptyset$
2. $\forall B \subseteq A \quad B \in \mathcal{A} \Rightarrow B^C \in \mathcal{A}$
3. Presa comunque una famiglia numerabile $(B_n)_n$ di sottoinsiemi di A , vale

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

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 -algre su un dato spazio degli eventi elementari Ω sono:

Esempio 2.7.

$$\mathcal{F} = \{\emptyset, \Omega\}$$

Questa è una σ -algebra perchè:

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

Esempio 2.8.

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

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

$$\mathcal{F} = \{\emptyset, \Omega, E, E^C\}$$

Si dimostra che è una σ -algebra:

Dimostrazione. 1. Trivialmente, $\mathcal{F} \neq \emptyset$

2.

$$\begin{aligned}\emptyset^C &= \Omega \in \mathcal{F} \\ \Omega^C &= \emptyset \in \mathcal{F} \\ E^C &\in \mathcal{F} \\ (E^C)^C &= E \in \mathcal{F}\end{aligned}$$

3. Per dimostrare che \mathcal{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 \emptyset , perchè esso è l'elemento neutro dell'unione, e quindi $\emptyset \cup x \in \mathcal{F} \Leftrightarrow x \in \mathcal{F} \ \forall x$. Resta da dimostrare che $E \cup E^C \in \mathcal{F}$, che è triviale in quanto $E \cup E^C = \Omega$.

◁

Le proprietà di una σ -algebra sono:

- $\emptyset, \Omega \in \mathcal{F} \quad \forall \mathcal{F}$

Dimostrazione.

$$\begin{aligned}\mathcal{F} &\neq \emptyset \\ \Rightarrow \exists G &\in \mathcal{F} \\ \Rightarrow G^C &\in \mathcal{F} \\ \Rightarrow G \cup G^C &\in \mathcal{F} \\ \Rightarrow \Omega &\in \mathcal{F} \\ \Rightarrow \Omega^C &= \emptyset \in \mathcal{F}\end{aligned}$$

◁

- \mathcal{F} è chiusa rispetto le unioni finite

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

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

◁

- \mathcal{F} è chiusa rispetto le intersezioni

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

$$\begin{aligned}
& \forall i \ E_i \in \mathcal{F} \\
& \Rightarrow \forall i \ E_i^C \in \mathcal{F} \\
& \Rightarrow \bigcup_i E_i^C \in \mathcal{F} \\
& \Rightarrow \left(\bigcup_i E_i^C\right)^C \in \mathcal{F} \\
& \Rightarrow \bigcap_i E_i \in \mathcal{F} \quad \quad \quad \hookrightarrow
\end{aligned}$$

Se Ω è discreto, allora si può sempre prendere $\mathcal{F} = \mathbb{P}(\Omega)$, 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

$$P: \mathcal{F} \rightarrow [0, 1]$$

si dice misura di probabilità se

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

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

Le proprietà di P sono:

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

Dimostrazione.

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

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

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

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

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

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

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) \hookrightarrow

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

$$P\left(\bigcup_n E_n\right) \leq \sum_n P(E_n)$$

Dimostrazione. (ecc) \hookrightarrow

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

Dimostrazione. (per casa) \hookrightarrow

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

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

2.4.1 Misura Empirica

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

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

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

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

Definizione 2.11. Si dice misura di probabilità empirica la funzione

$$\begin{aligned} \hat{P}_n : \mathcal{F} &\rightarrow [0, 1] \\ E &\mapsto \frac{n_E}{n} \end{aligned}$$

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 Ω discreto

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

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

Segue dalla σ -additività di P che:

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

Si ricava che se Ω è finito

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

identificare P equivale ad identificare una famiglia

$$\{p_1, \dots, p_n\}$$

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:

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

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

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

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. $\forall r \in \{1, \dots, R\}$ la scelta va eseguita fra n_r elementi
2. $\forall r \in \{1, \dots, R\}$ 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 $k \leq n$)

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:

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

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:

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

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:

$$\begin{aligned} 2^n &= \sum_{k=0}^n \binom{n}{k} \\ \binom{n}{k} &= \binom{n}{n-k} \\ \binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1} \end{aligned}$$

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:

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

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 (Ω, \mathcal{F}, P) uno spazio di probabilità, sia $F \in \mathcal{F}$ tale che $P(F) > 0$; si definisce misura di probabilità condizionata a F la funzione

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

dove:

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

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 :

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

□

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

Alcune proprietà:

- $P(E^C|F) = 1 - P(E|F) \quad \forall E \in \mathcal{F}$
- $E \cap F = \emptyset \Rightarrow P(E|F) = 0$
- $F \subseteq E \Rightarrow P(E|F) = 1$
- $P(E) = 0 \Rightarrow P(E|F) = 0$

- $P(E) = 1 \Rightarrow P(E|F) = 1$
- *Formula di moltiplicazione:* sia $\{E_n\}_n$ una famiglia di eventi non necessariamente disgiunti, allora vale

$$\begin{aligned} P\left(\bigcap_{i=1}^n E_i\right) &= P\left(E_n \mid \bigcap_{i=1}^{n-1} E_i\right) \cdot \dots \cdot P(E_1) \\ &= \prod_{k=2}^n P\left(E_k \mid \bigcap_{i=1}^{k-1} E_i\right) \cdot P(E_1) \end{aligned}$$

- 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:

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

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

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

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

$$\begin{aligned} \bigcup_i F_i &= \Omega \\ F_i \cap F_j &= \emptyset \quad \forall i \neq j \end{aligned}$$

Inoltre, imponiamo che

$$P(F_i) > 0 \quad \forall i$$

Allora, dato un evento $E \in \mathcal{F}$ 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:

$$\begin{aligned}
 E &= \bigcup_{i=1}^n (E \cap F_i) \\
 \Rightarrow P(E) &= P\left(\bigcup_{i=1}^n (E \cap F_i)\right) \\
 &= \sum_{i=1}^n P(E \cap F_i) \\
 &= \sum_{i=1}^n \frac{P(E \cap F_i)}{P(F_i)} P(F_i) \\
 &= \sum_{i=1}^n P(E|F_i) P(F_i) \quad \quad \quad \hookrightarrow
 \end{aligned}$$

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 $E, F \in \mathcal{F}$, 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:

- $E \perp\!\!\!\perp \emptyset$
- $E \perp\!\!\!\perp \Omega$
- $E \perp\!\!\!\perp E \Leftrightarrow E = \Omega \vee E = \emptyset$
- 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 $E, F, G \in \mathcal{F}$, 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 $\{E_n\}_n \subseteq \mathcal{F}$, si dice famiglia di eventi indipendenti se e solo se $\forall r \in \{2, 3, \dots, n\}$ 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 $\{E_n\}_n \subseteq \mathcal{F}$, 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 = "l' i -esimo studente ha risposto correttamente"

E_i = "ci sono state i 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:

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

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 $\Omega = \{0, 1\}^n$, $\mathcal{F} = \mathbb{P}(\Omega)$, e da una misura di probabilità P che rispetti le condizioni:

- Gli eventi

$$B_j = \{(b_1, \dots, b_n) | b_j = 1\}$$

sono equiprobabili, cioè

$$P(B_j) = p \quad \forall j \in \{1, \dots, n\} \text{ con } p \in [0, 1]$$

- La famiglia

$$\{B_1, \dots, B_n\}$$

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 $\{\text{"funziona"}, \text{"non funziona"}\}$ come

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

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

Per il sistema a 3 componenti vale:

$$\begin{aligned}P(\text{"sistema funziona"}) &= P(\text{"almeno due componenti su tre funzionano"}) \\ &= P(\{\text{"esattamente 2 componenti funzionano"}\} \cup \{\text{"esattamente 3 componenti funzionano"}\}) \\ &= P(\{\text{"2 componenti su 3 funzionano"}\}) + P(\{\text{"tutti e 3 i componenti funzionano"}\}) \\ &= P\left(\bigcup_{I \subset \{1,2,3\} \wedge |I|=2} \{(b_1, b_2, b_3) \in \{0, 1\}^3 \mid b_j = 1 \Leftrightarrow j \in I\}\right) + P(\{(1, 1, 1)\})\end{aligned}$$

essendo gli eventi nell'unione disgiunti:

$$\begin{aligned}&= \sum_{I \subset \{1,2,3\} \wedge |I|=2} P(\{(b_1, b_2, b_3) \in \{0, 1\}^3 \mid b_j = 1 \Leftrightarrow j \in I\}) + P(\{(1, 1, 1)\}) \\ &= \sum_{I \subset \{1,2,3\} \wedge |I|=2} \prod_{j \in I} P(B_j) \cdot \prod_{j \notin I} P(B_j^C) + P(\{(1, 1, 1)\})\end{aligned}$$

siccome $|I| = 2$:


$$\begin{aligned}&= \sum_{I \subset \{1,2,3\} \wedge |I|=2} p^2(1-p) + p^3 \\ &= \binom{3}{2} p^2(1-p) + p^3 \\ &= 3 \cdot p^2(1-p) + p^3\end{aligned}$$

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)

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

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, assegnando 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 ogni $\omega \in \Omega$ viene mappato in uno e un solo $x_k \in \mathcal{X}$), 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

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

$X^{-1}(x_k)$ 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

$$p_X : \mathcal{X} \rightarrow [0, 1] \\ x_k \mapsto p_X(x_k) = P^X(\{x_k\})$$

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

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

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€ per 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:

$$\begin{aligned} p_X(0) &= P(X^{-1}(0)) = P(\text{"non sono estratte carte di picche"}) \\ &= \frac{\binom{39}{3}}{\binom{52}{3}} = \frac{703}{1700} \end{aligned}$$

$$\begin{aligned} p_X(1) &= P(X^{-1}(1)) = P(\text{"è estratta una sola carta di picche"}) \\ &= \frac{13 \cdot \binom{39}{2}}{\binom{52}{3}} = \frac{741}{1700} \end{aligned}$$

$$\begin{aligned} p_X(2) &= P(X^{-1}(2)) = P(\text{"sono estratte due carte di picche"}) \\ &= \frac{\binom{13}{2} \cdot 39}{\binom{52}{3}} = \frac{234}{1700} \end{aligned}$$

$$\begin{aligned} p_X(3) &= P(X^{-1}(3)) = P(\text{"sono estratte tre carte di picche"}) \\ &= \frac{\binom{13}{3}}{\binom{52}{3}} = \frac{22}{1700} \end{aligned}$$



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

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

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 $X : \Omega \rightarrow \mathbb{R}$ variabile aleatoria e $g : \mathbb{R} \rightarrow \mathbb{R}$ 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 esista; 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 $X : \Omega \rightarrow \mathbb{R}$ variabile aleatoria discreta. Allora si ha*

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

Dimostrazione.

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

Siccome al variare di k gli insiemi $\{\omega \mid X(\omega) = x_k\}$ 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:

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

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

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

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

Dimostrazione.

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

Per il teorema fondamentale:

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

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

Dimostrazione.

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

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 $X \geq 0$, cioè tale che $\mathcal{X} \subseteq \mathbb{R}_+$, allora anche $E(X)$ è non negativo

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

Dimostrazione. Per definizione:

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

Per definizione:

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

Per H_p :

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

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

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

Dimostrazione. Segue dalle altre proprietà:

$$X \geq Y \Leftrightarrow X - Y \geq 0$$

$$\text{Per positività: } \Leftrightarrow E(X - Y) \geq 0$$

$$\text{Per linearità: } \Leftrightarrow E(X) - E(Y) \geq 0$$

$$\Leftrightarrow E(X) \geq E(Y) \quad \hookrightarrow$$

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

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

$$\bar{x} = \sup(\mathcal{X})$$

Allora:

$$\underline{x} \leq E(X) \leq \bar{x}$$

Dimostrazione. Per definizione:

$$\underline{x} \leq x_k \leq \bar{x} \quad \forall x_k \in \mathcal{X}$$

Segue:

$$\begin{aligned} \sum_{x_k \in \mathcal{X}} \underline{x} \cdot p_X(x_k) &\leq \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leq \sum_{x_k \in \mathcal{X}} \bar{x} \cdot p_X(x_k) \\ \underline{x} \sum_{x_k \in \mathcal{X}} p_X(x_k) &\leq \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leq \bar{x} \sum_{x_k \in \mathcal{X}} p_X(x_k) \\ \underline{x} \cdot 1 &\leq \sum_{x_k \in \mathcal{X}} x_k \cdot p_X(x_k) \leq \bar{x} \cdot 1 \end{aligned}$$

Segue la tesi. \square

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 $X(\omega) = a \quad \forall \omega \in \Omega$, e quindi vale

$$\begin{aligned} \mathcal{X} &= \{a\} \\ p_X(a) &= 1 \end{aligned}$$

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:

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

Notiamo che

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

Ma

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

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:

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

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

1. $Var(X) \geq 0$, e in particolare

$$Var(X) = 0 \Leftrightarrow X = k \in \mathbb{R}$$

Dimostrazione. Segue immediatamente dalla definizione. □

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

Dimostrazione.

$$\begin{aligned} \text{Var}(aX) &= E((aX - E(aX))^2) \\ &= E((aX - aE(X))^2) \\ &= E(a^2(X - E(X))^2) \\ &= a^2 E((X - E(X))^2) \end{aligned} \quad \mathcal{Q}$$

$$3. \text{Var}(X + a) = \text{Var}(X)$$

Dimostrazione.

$$\begin{aligned} \text{Var}(X + a) &= E((X + a - E(X + a))^2) \\ &= E((X + a - E(X) + E(a))^2) \\ &= E((X + a - E(X) + a)^2) \\ &= E((X - E(X))^2) \\ &= \text{Var}(X) \end{aligned} \quad \mathcal{Q}$$

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:

E_i = "Estrazione dall'urna 2 una pallina di colore i "

T_i = "Trasferimento dall'urna 1 all'urna 2 una pallina di colore i "

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:

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

4.4 Variabili Aleatorie Discrete Notevoli

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

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

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 variabile 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.

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

La sommatoria corrisponde alla somma dei valori della densità discreta di una variabile aleatoria $Y \sim \text{Bin}(n-1, p)$ 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:

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

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

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

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

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

- 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:


$$Y_i \sim \text{Be}\left(\frac{1}{36}\right)$$

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

$$Y = \sum_{i=1}^{24} Y_i \sim \text{Bin}\left(24, \frac{1}{36}\right)$$

Y è il numero di volte in cui abbiamo ottenuto un doppio 6; analogamente a prima:

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y < 1) \\ &= 1 - P(Y = 0) \\ &= 1 - \binom{24}{0} \left(\frac{1}{36}\right)^0 \left(1 - \frac{1}{36}\right)^{24} \\ &= 1 - \left(\frac{35}{36}\right)^{24} \approx 0.49 \end{aligned}$$

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 Una variabile X si dice geometrica, e si indica

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

NB. Indichiamo:

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

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:

- $p_X(k) \geq 0 \quad \forall k \in \mathbb{N}$, 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:

$$\begin{aligned}
 E(X) &= \sum_{k \in \mathbb{N}} k \cdot p_X(k) \\
 &= \sum_{k \in \mathbb{N}} k \cdot (1-p)^{k-1} \cdot p \\
 &= p \sum_{k \in \mathbb{N}} k \cdot (1-p)^{k-1}
 \end{aligned}$$

ricordiamo che:

$$k(1-p)^{k-1} = -\frac{d}{dp}[(1-p)^k]$$

segue:

$$= -p \sum_{k \in \mathbb{N}} \frac{d}{dp}[(1-p)^k]$$

per la linearità della derivata:

$$= -p \frac{d}{dp} \left[\sum_{k \in \mathbb{N}} (1-p)^k \right]$$

per $k = 1$ risulta un termine 1 della somma,

che quindi è irrilevante per la derivata,

e quindi si può far partire la sommatoria da 0

$$\begin{aligned}
 &= -p \frac{d}{dp} \left[\sum_{k \in \mathbb{N}_0} (1-p)^k \right] \\
 &= -p \frac{d}{dp} \left[\frac{1}{p} \right] \\
 &= -p \left(-\frac{1}{p^2} \right) = \frac{1}{p}
 \end{aligned}$$

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:

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

2. La probabilità di vincere alla sesta giocata è:

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

