

Schede riassuntive di *Elementi di Probabilità e Statistica*

A cura di Gabriel Antonio Videtta¹
g.videtta1@studenti.unipi.it

A.A. 2023-2024

Ultimo aggiornamento: 19 aprile 2024

¹Basato su un layout di [Luca Lombardo](#) e di [Francesco Sorce](#).

Indice

Notazioni impiegate	2		
Prerequisiti matematici	5		
Lista delle identità sulle sommatorie	7		
1 Spazi di probabilità in generale	8		
1.1 Definizioni preliminari	8		
1.1.1 Esperimento aleatorio, spazi campionari	8		
1.1.2 σ -algebre e spazi misurabili	8		
1.1.3 Insiemi discreti e σ -algebra naturale	8		
1.1.4 Proprietà di una σ -algebra e σ -algebra generata	8		
1.2 Corrispondenze logiche e relazionali tra eventi	8		
1.3 Misure di probabilità	9		
1.3.1 La probabilità P su Ω e spazi di probabilità	9		
1.3.2 Proprietà della probabilità P	9		
1.3.3 Eventi incompatibili, quasi certi e trascurabili, proprietà che accadono q.c.	9		
1.4 Probabilità condizionata	9		
1.4.1 Definizione di $P(\cdot B)$	9		
1.4.2 Regola della catena, formula delle probabilità totali e Teorema di Bayes	9		
1.4.3 Rapporto di influenza, correlazione positiva e negativa	10		
1.5 Indipendenza stocastica tra eventi	10		
2 Probabilità discreta	12		
2.1 Funzione di densità discreta	12		
2.1.1 Definizione per il caso discreto	12		
2.1.2 Range di una probabilità discreta e restrizione	12		
2.1.3 Misure di probabilità discrete su spazi campionari non discreti e discretizzazione	12		
2.2 Variabili aleatorie discrete	13		
2.2.1 Definizione di v.a. discreta e composizione	13		
2.2.2 Legge di una v.a. X	13		
2.2.3 Uguaglianza q.c., medesima legge e stabilità per composizione	13		
2.2.4 Variabile aleatoria multivariata, leggi congiunte e marginali	14		
2.2.5 Indipendenza di variabili aleatorie discrete e stabilità per congiunzione e composizione	14		
2.3 Valore atteso e momenti	15		
2.3.1 Valore atteso su v.a. integrabili e/o non negative	15		
		2.3.2 Proprietà del valore atteso e moltiplicatività per v.a. indipendenti	16
		2.3.3 Valore atteso condizionale	16
		2.3.4 Momenti (assoluti) n -esimi	16
		2.3.5 Disuguaglianza di Markov, di Hölder, di Cauchy-Schwarz e di Jensen	17
		2.4 Altri indici di centralità: moda e mediana	17
		2.5 Indici di dispersione: covarianza, varianza, dev. standard e coeff. di correlazione	17
		2.5.1 Definizioni e covarianza come forma bilineare simmetrica	17
		2.5.2 Identità sulla (co)varianza e disuguaglianza di Chebyshev	18
		2.5.3 Coeff. di correlazione e retta di regressione lineare	18
		2.6 Legge dei grandi numeri (LGN), media campionaria e limite in senso probabilistico	19
		2.6.1 Definizioni ed enunciato	19
		2.6.2 Trasformata di Cramer per l'ottimizzazione della stima	19
		2.7 Teorema centrale del limite (TCL, o TLC)	20
		2.7.1 Intuizione del TCL: <i>zoom-in</i> e <i>scaling</i>	20
		2.7.2 Enunciato del TCL e Teorema di De Moivre-Laplace per la distr. binomiale	20
		2.8 Modelli probabilistici classici	20
		2.8.1 Probabilità uniforme	20
		2.8.2 Sequenze di esperimenti e modello delle prove ripetute di Bernoulli	20
		Tabella e proprietà delle distribuzioni discrete	22
		Tabella e proprietà della f.d.r. $\Phi(x)$ di una normale standard	23

Notazioni impiegate

Algebra lineare

- q_φ – dato uno spazio vettoriale V equipaggiato con un prodotto scalare φ , q_φ è la forma quadratica associatogli, ovvero $q_\varphi(v) = \varphi(v, v)$.
- $\|v\|_\varphi$ – dato uno spazio vettoriale reale V equipaggiato con un prodotto scalare (semi)definito positivo φ , $\|\cdot\|_\varphi$ è la (semi)norma indotta da φ , ovvero $\|v\|_\varphi = \sqrt{q_\varphi(v)} = \sqrt{\varphi(v, v)}$.
- vettore isotropo – vettore che annulla la forma quadratica.
- vettore anisotropo – vettore non isotropo, vettore che non annulla la forma quadratica.
- $\cos_\varphi(v, w)$, $\cos(v, w)$ – dati due vettori anisotropi v, w su uno spazio vettoriale reale V equipaggiato di un prodotto scalare semidefinito positivo φ , si definisce $\cos_\varphi(v, w)$ (o $\cos(v, w)$ se φ è noto dal contesto) in modo tale che:

$$\cos_\varphi(v, w) = \frac{\varphi(v, w)}{\|v\|_\varphi \cdot \|w\|_\varphi}.$$

- vettore v ortogonale a w per φ – Due vettori v, w tali per cui $\varphi(v, w) = 0$.
- V_φ^\perp – Radicale del prodotto scalare (o hermitiano) φ sullo spazio V , ovvero sotto-spazio dei vettori ortogonali ai vettori di tutto lo spazio.
- $CI(\varphi)$ – Sottoinsieme dei vettori di V che annullano q_φ , ossia sottoinsieme dei vettori isotropi.
- $C_\varphi(v, w)$ – coefficiente di Fourier di v rispetto a w , ossia $C(v, w) \stackrel{\text{def}}{=} \frac{\varphi(v, w)}{\varphi(v, v)}$.

Analisi matematica

- $f(A_i) \nearrow x$ – la successione $(f(A_i))_{i \in \mathbb{N}}$ a valori in \mathbb{R} è crescente al crescere di i e ha come limite x .
- $f(A_i) \searrow x$ – la successione $(f(A_i))_{i \in \mathbb{N}}$ a valori in \mathbb{R} è decrescente al crescere di i e ha come limite x .
- esponente coniugato di p – per $p > 1$, l'esponente coniugato p' di p è un numero reale $p' > 1$ tale per cui:

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

- $\|x\|_p$ – norma p -esima del vettore $x \in \mathbb{R}^n$, ovvero:

$$\|x\|_p = \left(\sum_{i \in [n]} |x_i|^p \right)^{\frac{1}{p}}.$$

Per $p = 2$, si scrive semplicemente $\|x\|$, e coincide con la norma indotta dal prodotto scalare canonico di \mathbb{R}^n .

- $f > g$ – Per una funzione f a valori reali, come affermazione corrisponde a dire che per un qualsiasi punto del dominio x , $f(x) > g(x)$. Si estende naturalmente a $<$, \geq , \leq (eventualmente con catene di disuguaglianze). Da non confondersi con l'insieme $f > g$.
- a – Per una costante $a \in \mathbb{R}$ la mappa costante $D \ni d \mapsto a \in \mathbb{R}$; la sua interpretazione dipende dal contesto.
- f^+ – Parte positiva di una mappa f a valori reali, ovvero $f^+(a)$ è uguale a $f(a)$ se $f(a) \geq 0$ e 0 altrimenti.
- f^- – Parte negativa di una mappa f a valori reali, ovvero $f^-(a)$ è uguale a $-f(a)$ se $f(a) \leq 0$ e 0 altrimenti. In questo modo $f = f^+ - f^-$.
- \exp – Funzione esponenziale e^x .
- $\log \equiv \ln = \log_e$ – Logaritmo naturale, ossia logaritmo in base e .

Combinatoria

- $D_{n,k} = \frac{n!}{(n-k)!}$ – numero di disposizioni ottenute prendendo k elementi tra n oggetti.
- $\binom{n}{k} = C_{n,k}$ – il coefficiente binomiale n su k , ovvero il numero di combinazioni possibili prendendo k elementi tra n oggetti; equivale a $\frac{n!}{(n-k)!k!} = D_{n,k}/k!$. Alternativamente, il numero di sottoinsiemi di k elementi in $[n]$.
- $S(I)$ – gruppo simmetrico relativo a I , gruppo delle permutazioni di I .
- S_n – n -esimo gruppo simmetrico, gruppo delle permutazioni di $[n]$.

Teoria degli insiemi

- $\mathcal{P}(\Omega)$ – insieme delle parti di Ω , ossia insieme dei sottoinsiemi di Ω .
- $A \cup B$ – unione disgiunta di A e B , ovverosia $A \cup B$ con l'ipotesi che $A \cap B = \emptyset$ (la notazione si estende naturalmente a una famiglia di insiemi a due a due disgiunti).
- $A \Delta B = A \setminus B \cup B \setminus A$ – differenza simmetrica tra A e B .
- $[n]$ – l'insieme $\{1, \dots, n\}$.
- $\prod_{i \in I} S_i$ con S_i insieme e I ordinato – prodotto cartesiano degli S_i , ordinato secondo I .
- $[[n]]$ – l'insieme $\{0, \dots, n\} = \{0\} \cup [n]$.
- $\#A, |A|$ – numero di elementi di A , o semplicemente la cardinalità di A .
- insieme finito – insieme in biezione con $[n]$ per qualche $n \in \mathbb{N}$.
- insieme numerabile – insieme in biezione con \mathbb{N} .
- $A_i \nearrow A$ – la famiglia $(A_i)_{i \in \mathbb{N}}$ è crescente e ha come limite A , ovverosia $A_i \subseteq A_{i+1}$ per ogni $i \in \mathbb{N}$ e $\bigcup_{i \in \mathbb{N}} A_i = A$.
- $A_i \searrow A$ – la famiglia $(A_i)_{i \in \mathbb{N}}$ è decrescente e ha come limite A , ovverosia $A_i \supseteq A_{i+1}$ per ogni $i \in \mathbb{N}$ e $\bigcap_{i \in \mathbb{N}} A_i = A$.
- ω_i – i -esima coordinata di $\omega \in \Omega$, se Ω è un prodotto cartesiano di finiti termini o di un numero numerabile di termini.
- $A^1 \stackrel{\text{def}}{=} A$ – useremo questa notazione per comodità.
- A^c – il complementare di A riferito a Ω , quindi $\Omega \setminus A$, in modo tale che $\Omega = A \cup A^c$.
- $X^{-1}(A)$ – controimmagine dell'insieme $A \subseteq C$ in riferimento alla funzione $X : D \rightarrow C$, ovverosia $X^{-1}(A) = \{\omega \in D \mid X(\omega) \in A\}$.
- $S_X, \text{im } X$ – immagine della funzione X .
- $\text{supp } X$ – supporto di X , ovverosia sottoinsieme del dominio degli elementi che non annullano X .
- $1_A, I_A$ – funzione indicatrice di A , ovverosia la funzione $1_A : B \rightarrow \{[1]\} \subseteq \mathbb{R}$ riferita ad $A \subseteq B$ tale per cui:

$$1_A(b) = \begin{cases} 1 & \text{se } b \in A, \\ 0 & \text{altrimenti.} \end{cases}$$
- \Rightarrow – simbolo utilizzato al posto \rightarrow quando si elencano più funzioni che condividono o lo stesso dominio o lo stesso codominio (e.g. $f, g : A, B \Rightarrow C$ elenca una funzione $f : A \rightarrow C$ e una $g : B \rightarrow C$; $f, g : A \Rightarrow B, C$ elenca una funzione $f : A \rightarrow B$ e una $g : A \rightarrow C$).

Probabilità e teoria della misura

- Ω – spazio campionario, l'insieme di tutti i possibili esiti dell'esperimento aleatorio considerato.
- $\sigma(\tau)$ – σ -algebra generata dalla famiglia $\tau \subseteq \mathcal{P}(\Omega)$.
- $\sigma\{A_1, \dots, A_n\}$ – σ -algebra generata dalla famiglia $\tau = \{A_1, \dots, A_n\} \subseteq \mathcal{P}(\Omega)$.
- \mathcal{F} – σ -algebra relativa a Ω , ossia l'insieme dei possibili eventi.
- (Ω, \mathcal{F}) – spazio misurabile.
- P – misura di probabilità su uno spazio misurabile.
- (Ω, \mathcal{F}, P) – spazio di probabilità.
- q.c. – quasi certo/quasi certamente.
- p – per Ω discreto, funzione di densità discreta.
- v.a. – variabile aleatoria.
- P^X – legge della v.a. X rispetto a P .
- p_X – densità della legge della v.a. X , rispetto a P .
- $X \in A$ – per una v.a. $X : \Omega \rightarrow S$, $X \in A$ è l'insieme $X^{-1}(A)$. Si estende naturalmente al caso \notin .
- $X = a$ – per una v.a. $X : \Omega \rightarrow S$, $X = a$ è l'insieme $X^{-1}(a)$. Si estende naturalmente al caso \neq .
- $X = Y$ – per due v.a. $X, Y : \Omega \Rightarrow S$ l'insieme $\{\omega \in \Omega \mid X(\omega) = Y(\omega)\}$. Si estende naturalmente al caso \neq e in modo analogo a $>, <, \leq, \geq$.
- $X > a$ – per una v.a. reale $X : \Omega \rightarrow \mathbb{R}$, $X > a$ è l'insieme $X^{-1}((a, \infty))$; per una v.a. discreta $X : \Omega \rightarrow \mathbb{R}$ è l'insieme $X^{-1}(\{m \in \mathbb{N} \mid m > a\})$. Si estende naturalmente ai casi $<, \leq, \geq$ (eventualmente anche con una catena di disuguaglianze). Da non confondersi con l'affermazione $X > a$ per X a valori reali.
- $\varphi(X)$ – per una v.a., la composizione $\varphi \circ X$.
- $\stackrel{(d)}{=} , \sim$ – per due v.a. $X, Y : \Omega_1, \Omega_2 \Rightarrow S$ indica l'uguaglianza di legge, ovverosia $P_{\Omega_1}^X = P_{\Omega_2}^Y$.
- i.d. – identicamente distribuite; utilizzato in relazione a un gruppo di v.a. che condividono la stessa legge (spesso rispetto a uno stesso Ω).
- i.i.d. – indipendenti e identicamente distribuite; utilizzato in relazione a un gruppo di v.a. indipendenti che condividono la stessa legge (spesso rispetto a uno stesso Ω).
- $(X_i)_{i \in I}$ – famiglia di v.a., oppure v.a. congiunta.
- (X_1, \dots, X_n) – per una famiglia $(X_i : \Omega \rightarrow S_i)_{i \in [n]}$ di v.a. indica la v.a. congiunta (multivariata) $(X_1, \dots, X_n) : \Omega \rightarrow \prod_{i \in [n]} S_i$, $\omega \mapsto (X_1(\omega), \dots, X_n(\omega))$. Se la famiglia è composta da due variabili, si dice anche *coppia bivariata*.

- $P(A, B) \stackrel{\text{def}}{=} P(A \cap B)$ – notazione introdotta per scrivere più comodamente $P(X = x, Y = y)$ in luogo di $P((X = x) \cap (Y = y))$. Si generalizza in modo naturale a più eventi.
- $L(A, B) \stackrel{\text{def}}{=} \frac{P(A|B)}{P(A)}$ – rapporto di influenza tra A e B .
- $\otimes_{i \in [n]} P_i = P_1 \otimes \dots \otimes P_n$ – Date P_i probabilità su S_i discreto, $P_1 \otimes \dots \otimes P_n \stackrel{\text{def}}{=} P$ è la misura di probabilità naturale su $\prod_{i \in [n]} S_i$ tale per cui le proiezioni π_i siano v.a. discrete indipendenti e per cui $P(\pi_i = x_i) = p_i(x_i)$ per ogni $x_i \in S_i$, $i \in [n]$.
- $\mathbb{E}[X]$ – valore atteso di X .
- $\mathbb{E}[X | A] \stackrel{\text{def}}{=} \frac{\mathbb{E}[X \cdot 1_A]}{P(A)}$ – valore atteso di X condizionato a A .
- $\text{Cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ – covarianza di X e Y .
- $\text{Var}(X) \stackrel{\text{def}}{=} \text{Cov}(X, X)$ – varianza di X .
- $\sigma(X) \stackrel{\text{def}}{=} \sqrt{\text{Var}(X)}$ – deviazione standard di X .
- $\rho(X, Y)$ – coefficiente di correlazione di Pearson, ovvero sia $\cos_{\text{Cov}}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}$.
- a^*, b^* – date due v.a. X, Y , a^* e b^* sono i parametri della retta di regressione $y = a^* x + b^*$.
- $I(t)$ – trasformata di Cramer.
- LGN - Legge dei Grandi Numeri.
- TCL, TLC - Teorema Centrale del Limite.
- m, σ – spesso nel contesto della LGN e del TCL si usa m per indicare $\mathbb{E}[X_1]$ e σ per indicare $\sigma(X_1)$.

Prerequisiti matematici

Algebra lineare

- **Disuguaglianza di Cauchy-Schwarz** – Se $\varphi(\cdot, \cdot)$ è un prodotto scalare (o hermitiano) definito positivo su uno spazio vettoriale V , allora vale la seguente disuguaglianza:

$$\varphi(v, v)\varphi(w, w) \geq |\varphi(v, w)|^2, \quad \forall v, w \in V.$$

Inoltre vale l'uguaglianza se e solo se v è multiplo di w , o viceversa. Per prodotti semidefiniti positivi la disuguaglianza vale ugualmente, ma in tal caso v si scrive come somma di un vettore del cono isotropo e del prodotto di w per uno scalare.

- **Proprietà di $\cos(v, w)$** – Vale che $\cos(v, w) \in [-1, 1]$ per ogni $v, w \in V$ in spazi vettoriali reali dove \cos è ben definito. Segue dalla disuguaglianza di Cauchy-Schwarz.

Analisi matematica

- **Limite delle successioni monotone** – Se una successione $(a_i)_{i \in \mathbb{N}}$ è monotona, allora ammette limite. Se $(a_i)_{i \in \mathbb{N}}$ è crescente, allora $a_i \rightarrow \sup\{a_i \mid i \in \mathbb{N}\}$ per $i \rightarrow \infty$ (e dunque converge se la successione è limitata dall'alto); se $(a_i)_{i \in \mathbb{N}}$ è decrescente, allora $a_i \rightarrow \inf\{a_i \mid i \in \mathbb{N}\}$ per $i \rightarrow \infty$ (e dunque converge se la successione è limitata dal basso).
- **Convergenza delle serie a termini positivi** – Se una serie è a termini positivi, allora la successione delle somme parziali è crescente, e dunque la serie ammette come valore un valore reale o ∞ .
- **Convergenza assoluta** – Se una serie $\sum_{i \in \mathbb{N}} |a_i|$ converge (l'unica altra opzione è che diverga, per la proprietà sopracitata), allora $\sum_{i \in \mathbb{N}} a_i$ converge. Non è vero il viceversa in generale.
- **Disuguaglianza di Jensen** – Sia $f : \mathbb{R} \ni S \rightarrow \mathbb{R}$ una funzione convessa a valori reali. Allora vale che:

$$f\left(\sum_{i \in [n]} a_i x_i\right) \leq \sum_{i \in [n]} a_i f(x_i), \quad \sum_{i \in [n]} a_i = 1, x_i.$$

Se invece f è concava, vale la disuguaglianza con \geq al posto di \leq .

- **Disuguaglianza di Young** – Sia $p \geq 1$ e sia p' il suo esponente coniugato. Allora vale che:

$$ab \leq \frac{a^p}{p} + \frac{b^p}{p}, \quad \forall a, b > 0.$$

Segue dalla disuguaglianza di Jensen applicata a e^x , che è convessa.

- **Disuguaglianza di Hölder** – Sia $p > 1$ e sia p' il suo esponente coniugato. Allora vale che:

$$\sum_{i \in [n]} |x_i y_i| \leq \|x\|_p \|y\|_{p'}, \quad \forall x, y \in \mathbb{R}^n, \forall n \in \mathbb{N}.$$

Per $p = 2$, è equivalente alla disuguaglianza di Cauchy-Schwarz sul prodotto scalare canonico di \mathbb{R}^n . Segue dalla disuguaglianza di Young.

- **Disuguaglianza sulle potenze** – Siano $x, y \in \mathbb{R}$ e sia $p \geq 1$. Allora vale che:

$$|x + y|^p \leq 2^{p-1}(|x|^p + |y|^p).$$

Segue dalla disuguaglianza di Jensen applicata a $f(t) = t^p$ per $|x|$ e $|y|$ (t^p è convessa per $t \geq 0$).

Combinatoria

- **Principio di *double counting*** – Principio di dimostrazione per il quale se vi sono due modi diversi, ma equivalenti, di contare lo stesso numero di scelte di un qualsiasi sistema, allora le formule ricavate dai due modi devono essere identicamente uguali.
- **Principio di inclusione-esclusione** – Teorema da cui discende che per $(A_i)_{i \in [n]}$ vale che:

$$\left| \bigcup_{i \in [n]} A_i \right| = \sum_{j \in [n]} (-1)^{j+1} \sum_{1 \leq i_1 < \dots < i_j \leq n} \left| \bigcap_{k \in [j]} A_{i_k} \right|.$$

Inoltre vale che $|\bigcup_{i \in [n]} A_i| = \sum_{i \in [n]} |A_i|$ se e solo se gli A_i sono a due a due disgiunti. Per $n = 2$, $|A \cup B| = |A| + |B| - |A \cap B|$.

- **Principio della piccionaia (*Pigeonhole principle*)** – Teorema che asserisce che per ogni funzione $f : [n+1] \rightarrow [n]$ esistono $i, j \in [n+1]$ tali per cui $f(i) = f(j)$. Più informalmente, se si hanno $n+1$ oggetti da posizionare in n buchi, esiste per forza un buco con due oggetti.
- **Principio della piccionaia generalizzato** – Teorema che asserisce che per ogni funzione $f : [kn+1] \rightarrow [n]$ esistono $k+1$ elementi di $[kn+1]$ che condividono la stessa immagine. Più informalmente, se si hanno $kn+1$ oggetti da posizionare in n buchi, esiste per forza un buco con $k+1$ oggetti. Segue per induzione dal Principio della piccionaia.

- **Principio moltiplicativo** – Se una scelta può essere fatta in N passi e all' i -esimo passo corrispondono n_i scelte, allora la scelta globale può essere fatta in $\prod_{i \in [N]} n_i$ modi.
- **Permutazioni di n oggetti** – Dati n oggetti, esistono $n!$ modi di permutarli. Segue dal Principio moltiplicativo.
- **Disposizioni semplici di n oggetti in k posti** – Dati n oggetti e k posti, allora esistono $D_{n,k}$ modi di disporre gli n oggetti nei k posti se $k \leq n$. Se $k = n$, ci si riduce a contare le permutazioni.
- **Disposizioni con ripetizione di n oggetti in k posti** – Dati n oggetti e k posti, allora esistono n^k modi di disporre con ripetizione gli n oggetti nei k posti. Segue dal Principio moltiplicativo.
- **Combinazioni di n oggetti in k posti** – Dati n oggetti e k posti, allora esistono $C_{n,k} = \binom{n}{k} = \frac{n!}{(n-k)!k!}$ modi di disporre gli n oggetti nei k posti non facendo contare l'ordine, se $k \leq n$. Segue dal Principio moltiplicativo.
- **Combinazioni con ripetizione di n oggetti in k buchi** – Data l'equazione $x_1 + \dots + x_k = n$ con $x_i \in \mathbb{N}$, esistono esattamente $\binom{n+k-1}{k-1}$ soluzioni. Alternativamente, data la disequazione $x_1 + \dots + x_k \leq n$ con $x_i \in \mathbb{N}$, esistono esattamente $\binom{n+k}{k}$ soluzioni (dacché ha le stesse soluzioni di $x_1 + \dots + x_k + y = n$, dove $y \in \mathbb{N}$). È un'applicazione di una tecnica combinatorica standard denominata *stars and bars*.
- **Numero di scelte possibili per un'estrazione di n palline rosse e nere da un insieme di N_1 palline rosse unito a un insieme di $N - N_1$ palline nere** – Se k è il numero di palline rosse estratte, le scelte possibili sono $\binom{N_1}{k} \binom{N-N_1}{n-k}$. Si può generalizzare il problema a un insieme di N palline divise in m gruppi da N_i palline ciascuno (e dunque $\sum_{i \in [m]} N_i = N$) dove se ne estrae n e k_i è il numero di palline estratte dall' i -esimo gruppo (dunque $\sum_{i \in [m]} k_i = n$; in tal caso le scelte possibili sono $\prod_{i \in [m]} \binom{N_i}{k_i}$. Segue dal Principio moltiplicativo.
- **Identità sulle cardinalità**
 - $\#\{(a_1, \dots, a_n) \in [k]^n \mid a_1 < a_2 < \dots < a_n\} = \binom{n}{k}$ se $k \leq n$ – Infatti data una classe di disposizione, esiste un'unica lista ordinata in tale classe.
 - $\#\{(a_1, \dots, a_n) \in [k]^n \mid a_1 \leq a_2 \leq \dots \leq a_n\} = \binom{n+k-1}{k-1}$. – È sufficiente osservare che si sta contando esattamente le combinazioni con ripetizione in perfetta analogia con la precedente cardinalità.

$X^{-1}(\bigcap_{i \in I} A_i) = \bigcap_{i \in I} X^{-1}(A_i)$, $X^{-1}(A_i^c) = X^{-1}(A_i)^c$, ovvero X^{-1} commuta con unioni (\cup), intersezioni (\cap) e complementare (c). $X^{-1}(\emptyset) = \emptyset$, e dunque $A_i \cap A_j = \emptyset \implies X^{-1}(A_i) \cap X^{-1}(A_j) = \emptyset$. Inoltre per $Y : C \rightarrow C'$ vale che $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A))$, per $A \subseteq C'$.

Teoria degli insiemi

- **Leggi di De Morgan** – Se A e B sono insiemi, allora $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$.
- **Operazioni con X^{-1} controimmagine** – Se $X : D \rightarrow C$ è una funzione e $\mathcal{F} = (A_i)_{i \in I}$ è una famiglia di sottoinsiemi di C , allora vale che $X^{-1}(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} X^{-1}(A_i)$,

Lista delle identità sulle sommatorie

Identità sulle sommatorie

- $\binom{n}{k} = \binom{n}{n-k}$ – ogni scelta di k oggetti corrisponde a non sceglierne $n - k$, e dunque vi è un principio di “dualità”.
- $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ – le combinazioni di n oggetti in k posizioni si ottengono facendo la somma delle combinazioni ottenute fissando un oggetto e combinando gli altri $n - 1$ oggetti sui $k - 1$ posti rimanenti, e delle combinazioni ottenute ignorando lo stesso oggetto, ossia combinando gli altri $n - 1$ oggetti su tutti e k i posti.
- $(1 + x)^n = \sum_{i=0}^n \binom{n}{i} x^i$ – Teorema del binomio di Newton.
- $2^n = \sum_{i=0}^n \binom{n}{i}$ – Segue immediatamente dal Teorema del binomio di Newton; è coerente col fatto che si stanno contando le parti di $[n]$.
- $\sum_{i=0}^n (-1)^i \binom{n}{i} = 0$ – Segue immediatamente dal Teorema del binomio di Newton (infatti $(1 - 1)^n = 0$).
- $\sum_{i=0}^n i \binom{n}{i} = n2^{n-1}$ – Segue derivando rispetto a x l'identità del Teorema del binomio di Newton.
- $\sum_{i=0}^n \binom{n}{i} p^i (1 - p)^{n-i} = 1$ per $p \in [0, 1]$ – Segue dal Teorema del binomio di Newton.
- $\sum_{i=0}^n \binom{n}{i}^2 = \sum_{i=0}^n \binom{n}{i} \binom{n}{n-i} = \binom{2n}{n}$ – Dato un gruppo di n maschi e di n femmine, si vuole contare quanti team di n persone si possono costruire prendendo persone da entrambi i gruppi. Chiaramente la risposta è $\binom{2n}{n}$, ma si può contare lo stesso numero di scelte fissando a ogni passo l'indice i , che conta il numero di maschi nel team, a cui corrispondono $\binom{n}{i} \binom{n}{n-i}$ scelte. L'identità segue dunque dal Principio del *double counting*.
- $\sum_{i=r}^n \binom{i}{r} = \binom{n+1}{r+1}$ – Dato un gruppo di r persone distinguibili e di n bastoni indistinguibili, per contare le possibili distribuzioni con cui si possono affidare gli n bastoni è sufficiente applicare la combinazione con ripetizione, ottenendo $\binom{n+r-1}{r-1}$; un altro modo di far ciò è fissare i bastoni da affidare a una persona fissata in precedenza e distribuire gli $n - i$ bastoni rimanenti tra gli altri, che a ogni i si può fare in $\binom{n-i+k-2}{k-2}$ modi. L'identità segue dunque dal Principio del *double counting* riparametrizzando la somma ottenuta.
- $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ – Somma dei numeri da 1 a n .
- $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$ – Somma dei quadrati da 1 a n .
- $\sum_{i=1}^n i^3 = \left[\sum_{i=1}^n i \right]^2 = \frac{n^2(n+1)^2}{4}$ – Somma dei cubi da 1 a n .
- $\sum_{i=0}^n a^i = \frac{a^{n+1}-1}{a-1}$ per $a \neq 1$, n altrimenti – Somma delle potenze di a con esponente da 0 a n .
- $\sum_{i=0}^n i a^i = \frac{a}{(1-a)^2} [1 - (n+1)a^n + na^{n+1}]$ – Segue derivando la somma delle potenze.
- $\sum_{i=0}^n i^2 a^i = \frac{a}{(1-a)^3} [(1+a) - (n+1)^2 a^n + (2n^2 + 2n - 1)a^{n+1} - n^2 a^{n+2}]$ – Segue derivando due volte la somma delle potenze.
- $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$ per $|x| < 1$ – Serie geometrica. Deriva prendendo il limite per $n \rightarrow \infty$ della somma di potenze.
- $\sum_{i=0}^{\infty} i x^i = \frac{x}{(1-x)^2}$ per $|x| < 1$ – Segue derivando la serie geometrica.
- $\sum_{i=0}^{\infty} i^2 x^i = \frac{x(x+1)}{(1-x)^3}$ per $|x| < 1$ – Segue derivando due volte la serie geometrica.

Parte 1

Spazi di probabilità in generale

1.1 Definizioni preliminari

1.1.1 Esperimento aleatorio, spazi campionari

Definizione 1.1 (Esperimento aleatorio).

Si dice **esperimento aleatorio** un fenomeno il cui esito non è determinabile a priori.

Definizione 1.2 (Spazio campionario).

Si definisce **spazio campionario**, spesso indicato con Ω , un insieme non vuoto che contiene gli esiti di un esperimento aleatorio.

1.1.2 σ -algebre e spazi misurabili

Definizione 1.3 (σ -algebra).

Una σ -algebra \mathcal{F} di Ω è un sottoinsieme $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ tale per cui:

- (i.) $\Omega \in \mathcal{F}$,
- (ii.) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
- (iii.) per $(A_i)_{i \in \mathbb{N}}$ famiglia numerabile di insiemi in \mathcal{F} ,
 $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$ (\mathcal{F} è chiuso per unioni numerabili).

Una σ -algebra \mathcal{F} di uno spazio campionario Ω rappresenta l'insieme degli **eventi accettabili**. In particolare:

Definizione 1.4 (Spazio misurabile).

Si definisce **spazio misurabile** una coppia (Ω, \mathcal{F}) , dove \mathcal{F} è una σ -algebra di Ω .

1.1.3 Insiemi discreti e σ -algebra naturale

In alcuni casi la scelta della σ -algebra \mathcal{F} è naturale, come nel caso in cui si considera uno spazio campionario discreto:

Definizione 1.5 (Insieme discreto).

Diciamo che un insieme Ω è discreto se è finito o numerabile. Se non viene esplicitato altrimenti, per Ω si considererà sempre la σ -algebra naturale $\mathcal{P}(\Omega)$.

1.1.4 Proprietà di una σ -algebra e σ -algebra generata

In casi non discreti, è invece più naturale considerare σ -algebre molto meno grandi dell'insieme delle parti; in particolare,

come vedremo nella *Parte 3*, sarà naturale chiedersi qual è la σ -algebra più piccola che contiene una certa famiglia di insiemi:

Definizione 1.6 (σ -algebra generata da una famiglia di insiemi).

Sia τ una famiglia di sottoinsiemi di $\mathcal{P}(\Omega)$. Allora si definisce la σ -algebra generata da τ , detta $\sigma(\tau)$, come la più piccola σ -algebra contenente τ . Equivalentemente:

$$\sigma(\tau) = \bigcap_{\substack{\mathcal{F} \subseteq \mathcal{P}(\Omega) \\ \tau \subseteq \mathcal{F} \\ \mathcal{F} \text{ } \sigma\text{-alg.}}} \mathcal{F}.$$

Osservazione 1.7.

La definizione data è una buona definizione dal momento che si verifica facilmente che l'intersezione di σ -algebre è ancora una σ -algebra.

Proposizione 1.8 (Proprietà di \mathcal{F}).

Se \mathcal{F} è una σ -algebra di Ω , allora:

- (i.) $\emptyset \in \mathcal{F}$,
- (ii.) per $(A_i)_{i \in \mathbb{N}}$ famiglia numerabile di insiemi in \mathcal{F} ,
 $\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{F}$ (\mathcal{F} è chiuso per intersezioni numerabili),
- (iii.) $A \setminus B = A \cap B^c \in \mathcal{F} \iff A, B \in \mathcal{F}$.

1.2 Corrispondenze logiche e relazionali tra eventi

Osservazione 1.9 (Corrispondenze affermazioni ed eventi).

Ad alcune affermazioni logiche su A e B eventi di \mathcal{F} corrispondono degli eventi ben precisi o delle relazioni:

- “Si verificano A e B ” corrisponde a $A \cap B$,
- “Si verifica A o B ” corrisponde a $A \cup B$,
- “Si verifica esattamente uno tra A e B ” corrisponde a $A \setminus B \cup B \setminus A = A \Delta B$ (differenza simmetrica),
- “Non si verifica A ” corrisponde a A^c ,
- “Si verifica qualcosa” corrisponde a Ω ,
- “Non si verifica niente” corrisponde a \emptyset ,

- “Se succede A , allora succede B ” corrisponde a $A \subseteq B$,
- “Non succedono A e B contemporaneamente” corrisponde a $A \cap B = \emptyset$.

1.3 Misure di probabilità

1.3.1 La probabilità P su Ω e spazi di probabilità

Definizione 1.10 (Probabilità P su (Ω, \mathcal{F}) secondo Kolmogorov).

Dato (Ω, \mathcal{F}) spazio misurabile, una **misura di probabilità** P , detta semplicemente *probabilità*, è una funzione $P: \mathcal{F} \rightarrow \mathbb{R}$ tale per cui:

- $P(\Omega) = 1$,
- $0 \leq P(A) \leq 1$ per ogni $A \in \mathcal{F}$ (ossia P può restringersi su $[0, 1]$ al codominio),
- $P(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i)$ (σ -additività).

In particolare P è una misura.

Definizione 1.11 (Spazio di probabilità).

Si dice **spazio di probabilità** una tripla (Ω, \mathcal{F}, P) dove (Ω, \mathcal{F}) è uno spazio misurabile e P è una probabilità su (Ω, \mathcal{F}) .

1.3.2 Proprietà della probabilità P

Proposizione 1.12 (Proprietà di P).

Se P è una probabilità su (Ω, \mathcal{F}) , allora:

- $P(\emptyset) = 0$,
- $P(\bigcup_{i \in [n]} A_i) = \sum_{i \in [n]} P(A_i)$ (σ -additività finita),
- $P(A) + P(A^c) = 1$,
- $A \subseteq B \implies P(A) \leq P(B)$ e $P(B \setminus A) = P(B) - P(A)$ (segue da (iii)),
- $P(B \setminus A) = P(B) - P(A \cap B)$ (segue da (iv) considerando che $B \setminus A = B \setminus (A \cap B)$),
- $P(A \cup B) = P(A \Delta B \cup A \cap B) = P(A) + P(B) - P(A \cap B)$ (segue da (v)),
- $P(\bigcup_{i \in [n]} A_i) = \sum_{j \in [n]} (-1)^{j+1} \sum_{1 \leq i_1 < \dots < i_j \leq n} P(\bigcap_{k \in [j]} A_{i_k})$ (segue da (vi) per induzione, Principio di inclusione-esclusione “probabilistico”),
- $P(\bigcup_{i \in \mathbb{N}} A_i) \leq \sum_{i \in \mathbb{N}} P(A_i)$ (σ -subadditività).

Osservazione 1.13.

Per Ω finito, la σ -additività finita implica la σ -additività per il Principio della piccionaia.

Proposizione 1.14 (Comportamento di P al limite).

Sia $(A_i)_{i \in \mathbb{N}}$ una famiglia numerabile di eventi in \mathcal{F} sullo spazio di probabilità (Ω, \mathcal{F}, P) . Allora:

- $A_i \nearrow A \implies P(A_i) \nearrow P(A)$,
- $A_i \searrow A \implies P(A_i) \searrow P(A)$.

1.3.3 Eventi incompatibili, quasi certi e trascurabili, proprietà che accadono q.c.

Definizione 1.15 (Eventi trascurabili e quasi certi).

Sia $A \in \mathcal{F}$. Allora A si dice **trascurabile** se $P(A) = 0$; si dice **quasi certo** se $P(A) = 1$.

Definizione 1.16 (Eventi incompatibili).

Due eventi $A, B \in \mathcal{F}$ si dicono **incompatibili** se $A \cap B = \emptyset$.

Definizione 1.17 (q accade q.c.).

Si dice che una proprietà q **accade quasi certamente (q.c.)** se esiste $A \in \mathcal{F}$ quasi certo che soddisfa q .

Osservazione 1.18.

Si osserva che la nozione di proprietà che accade q.c. è perfettamente coerente con la nozione di proprietà che accade q.c. riferita a P come misura (e non specificatamente come misura di probabilità) su \mathbb{R} , ovvero q accade q.c. se esiste $A \in \mathcal{F}$ trascurabile tale per cui A^c soddisfi q .

1.4 Probabilità condizionata

1.4.1 Definizione di $P(\cdot | B)$

Definizione 1.19 (Probabilità condizionata su B).

Dato $B \in \mathcal{F}$ evento non trascurabile (i.e. $P(B) \neq 0$), la **probabilità condizionata** su B è la misura di probabilità $P(\cdot | B)$ sullo stesso spazio misurabile tale per cui:

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

Proposizione 1.20.

$P(\cdot | B)$ è una misura di probabilità su (Ω, \mathcal{F}) .

Osservazione 1.21.

La probabilità condizionata su Ω coincide con P .

Osservazione 1.22.

In generale $P(A | \cdot)$ non è una probabilità, dacché per Ω si ricava che $P(A | \Omega) = P(A)$, che potrebbe non essere 1.

1.4.2 Regola della catena, formula delle probabilità totali e Teorema di Bayes

Lemma 1.23 (Regola della catena, o della torre).

Dati $(A_i)_{i \in [n]}$ con $P(\bigcap_{i \in [n]} A_i) > 0$, allora vale che $P(\bigcap_{i \in [j]} A_i) > 0$ per ogni $j \leq n$. Inoltre vale che:

$$P\left(\bigcap_{i \in [n]} A_i\right) = \left(\prod_{j \in [n-1]} P\left(A_j \mid \bigcap_{i=j+1}^n A_i\right)\right) P(A_n),$$

che segue per induzione applicando $P(A \cap B) = P(A | B)P(B)$.

Osservazione 1.24.

Per esempio, la regola della catena per A, B e C si riduce a:

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

Definizione 1.25 (Sistema di alternative).

Una famiglia $(B_i)_{i \in I}$ con $I = \mathbb{N}$ o $I = [n]$ si dice **sistema di alternative** per Ω se $\Omega = \bigcup_{i \in I} B_i$ e $P(B_i) > 0$ per ogni $i \in I$ (ovverosia B_i non è mai trascurabile).

Un sistema di alternative permette di calcolare più agevolmente la probabilità di un evento riducendosi alle probabilità condizionate, come mostra il:

Lemma 1.26 (Formula delle probabilità totali, o formula della partizione).

Sia $(B_i)_{i \in I}$ un sistema di alternative per Ω . Allora vale che:

$$P(A) = \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I} P(A | B_i)P(B_i).$$

Nella maggior parte dei casi è possibile “invertire” una probabilità condizionata, ovvero si ricava una probabilità tra $P(A | B)$, $P(B | A)$, $P(A)$ e $P(B)$ conoscendone tre, a patto che A e B non siano trascurabili, come mostra il:

Teorema 1.27 (di Bayes).

Siano A e B due eventi non trascurabili. Allora vale che:

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

Segue considerando le due scritte possibili di $P(A \cap B)$.

Osservazione 1.28.

Applicando il Teorema di Bayes e la formula delle probabilità totali, si ricava che per un sistema di alternative $(B_i)_{i \in I}$ e A non trascurabile vale che:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j \in I} P(A | B_j)P(B_j)}, \quad \forall i \in I.$$

Osservazione 1.29.

Applicando la regola della catena, la formula delle probabilità totali e il Teorema di Bayes è possibile calcolare agevolmente la probabilità di un’intersezione di eventi conoscendone l’albero di sviluppo probabilistico. In particolare, per calcolare la probabilità di un nodo è sufficiente moltiplicare le probabilità dei rami facenti parte del percorso dal nodo alla radice.

1.4.3 Rapporto di influenza, correlazione positiva e negativa

Definizione 1.30 (Rapporto di influenza).

Siano A e B due eventi non trascurabili. Allora il **rapporto di influenza** di A e B (o più brevemente, la loro *influenza*) è il parametro:

$$L(A, B) \stackrel{\text{def}}{=} \frac{P(A | B)}{P(A)},$$

ed è tale per cui:

$$P(A | B) = L(A, B)P(A).$$

Proposizione 1.31.

$L(\cdot, \cdot)$ è simmetrica, ovvero sia $L(A, B) = L(B, A)$ per ogni evento A e B . Segue dal Teorema di Bayes.

Definizione 1.32 (Correlazione positiva e negativa tra A e B).

Se A e B sono due eventi non trascurabili, si dice che A è **positivamente correlato** a B (o che si *dilata probabilisticamente* rispetto a B) se $P(A | B) \geq P(A)$ (ovverosia se $L(A, B) > 1$).

Analogamente si dice che A è **negativamente correlato** a B (o che si *contrae probabilisticamente* rispetto a B) se $P(A | B) \leq P(A)$ (ovverosia se $L(A, B) < 1$).

Osservazione 1.33.

Il caso in cui $L(A, B) = 1$ è discusso nella sezione *Indipendenza stocastica tra eventi* e corrisponde all’indipendenza tra A e B .

Osservazione 1.34.

Si può parlare più generalmente di correlazione tra A e B senza scegliere un evento “rispetto” a cui analizzarla, dacché $L(\cdot, \cdot)$ è simmetrica per il Teorema di Bayes. Infatti, se $P(A | B) \leq P(A)$, anche $P(B | A) \leq P(B)$, cioè A è correlato positivamente a B se e solo se B è correlato positivamente ad A .

Una correlazione positiva tra A e B indica che, accadendo B , si amplifica la probabilità che accada A ; viceversa, una correlazione negativa inficia ridimensionando in contrazione la probabilità che accada A se accade B .

1.5 Indipendenza stocastica tra eventi

Definizione 1.35 (Famiglia di eventi indipendenti).

Una famiglia $(A_i)_{i \in I}$ di eventi si dice **stocasticamente indipendente**, o più semplicemente indipendente, se per ogni $J \subseteq I$ finito vale che:

$$P(\cap_{j \in J} A_j) = \prod_{j \in J} P(A_j).$$

Nel caso di due eventi questo si riduce a verificare che $P(A \cap B) = P(A)P(B)$. Si dice che gli A_i sono **collettivamente indipendenti**.

Osservazione 1.36.

Generalmente non è sufficiente verificare che ogni coppia di eventi distinti è indipendente per verificare che la famiglia è globalmente indipendente. Infatti, il significato dell’indipendenza in termini puramente probabilistici è che una famiglia \mathcal{F} è indipendente se e solo se il “verificarsi” di alcuni eventi della famiglia non influenza il “verificarsi” degli altri.

Osservazione 1.37.

Se $(A_i)_{i \in I}$ è una famiglia di eventi indipendenti, allora per $J \subseteq I$, $(A_j)_{j \in J}$ è ancora una famiglia di eventi indipendenti (l’indipendenza si tramanda per restrizione).

Proposizione 1.38.

Se $P(B) > 0$, allora A e B sono indipendenti se e solo se $P(A | B) = P(A)$. Inoltre, se $(A_j)_{j \in J} \cup \{A\}$ è una famiglia finita di eventi non trascurabili (eccetto eventualmente per A) indipendenti tra loro, allora $P(\cap_{j \in J} A_j) \neq 0$ e $P(A | \cap_{j \in J} A_j) = P(A)$.

Proposizione 1.39.

Se A e B sono indipendenti, allora anche A^c e B sono indipendenti. Analogamente lo sono A e B^c , così come A^c e B^c .

Da ciò segue che se $(A_i)_{i \in I}$ è una famiglia di eventi indipendenti, allora $(A_i^{\alpha_i})_{i \in I}$ è una famiglia di eventi indipendenti per qualsiasi scelta di α_i in $\{1, c\}$.

Proposizione 1.40.

Sia $(A_i)_{i \in I}$ una famiglia di eventi indipendenti. Allora, se I è partizionato dagli I_j , ovvero sia $I = \cup_{j \in J} I_j$, allora $(\cap_{i \in I_j} A_i)_{j \in J}$ è ancora una famiglia di eventi indipendenti (ossia intersecando alcuni elementi della famiglia e lasciandone invariati altri, la famiglia ottenuta è ancora indipendente).

Teorema 1.41.

Sia $(A_i)_{i \in I}$ una famiglia di eventi indipendenti. Allora, ogni operazione di unione, intersecazione o complementare di alcuni elementi della famiglia restituisce una famiglia ancora indipendente.

Segue dalle due proposizioni precedenti (infatti $A \cup B = (A^c \cap B^c)^c$).

Esempio 1.42.

Per esempio, se A , B e C sono indipendenti, anche $A \cup B$, C^c è indipendente. Se A , B , C e D sono indipendenti, anche $(A \cap B) \cup C^c$ e D^c lo sono.

Osservazione 1.43.

Un evento A è indipendente da ogni evento $B \in \mathcal{F}$, incluso sé stesso, se e solo se $P(A) \in \{0, 1\}$, ovvero se e solo se A è trascurabile o quasi certo (infatti si avrebbe che $P(A) = P(A \cap A) = P(A)^2$).

Osservazione 1.44.

Due eventi incompatibili A e B sono indipendenti se e solo se uno dei due è trascurabile.

Parte 2

Probabilità discreta

Consideriamo in questa sezione soltanto i casi in cui Ω è un insieme discreto, cioè finito o numerabile. Gli associamo in modo naturale la σ -algebra $\mathcal{P}(\Omega)$.

2.1 Funzione di densità discreta

2.1.1 Definizione per il caso discreto

Definizione 2.1 (Funzione di densità discreta).

Per una probabilità P su Ω si definisce **funzione di densità discreta** (o di massa, o più brevemente di densità) la funzione $p : \Omega \rightarrow \mathbb{R}$ tale per cui:

$$p(\omega) = P(\{\omega\}), \quad \forall \omega \in \Omega.$$

Proposizione 2.2 (P è univocamente determinata da p).

Sia $p : \Omega \rightarrow \mathbb{R}$ una funzione tale per cui:

$$(i.) \sum_{\omega \in \Omega} p(\omega) = 1,$$

$$(ii.) p(\omega) \geq 0 \text{ per ogni } \omega \in \Omega.$$

Allora esiste un'unica probabilità P la cui funzione di densità è p , e vale che:

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

2.1.2 Range di una probabilità discreta e restrizione

Definizione 2.3 (Range di P).

Sia P una probabilità su Ω discreto e sia p la sua funzione di densità. Si definisce allora **range** R_P di P il supporto di p , ovvero:

$$R_P \stackrel{\text{def}}{=} \text{supp } p = \{\omega \in \Omega \mid p(\omega) > 0\} \subseteq \Omega.$$

Definizione 2.4 (Restrizione di P sul range).

Data P probabilità su Ω discreto, si definisce **probabilità ristretta sul range** R_P la misura di probabilità $P|_{R_P} : \mathcal{P}(R_P)$ tale per cui:

$$P|_{R_P}(A) = P(A).$$

Osservazione 2.5.

La definizione data è una buona definizione dal momento che $P(R_P) = 1$.

Proposizione 2.6 (Proprietà della restrizione di P sul range).

Sia P una probabilità su Ω discreto e sia p la sua funzione di densità. Allora vale che $P(A) = P|_{R_P}(A \cap R_P)$.

2.1.3 Misure di probabilità discrete su spazi campionari non discreti e discretizzazione

Definizione 2.7 (Probabilità discreta su spazio campionario non discreto).

Dato (Ω, \mathcal{F}, P) spazio di probabilità con $\{\omega\} \in \mathcal{F}$ per ogni $\omega \in \Omega$, la probabilità P si dice **discreta** su Ω se esiste $\Omega_0 \in \mathcal{F}$ discreto e quasi certo ($P(\Omega_0) = 1$). In tal caso si dice che P si *concentra* su Ω_0 .

Definizione 2.8 (Discretizzazione di P discreta su Ω).

Se P è una probabilità discreta su Ω concentrata su Ω_0 , si definisce **discretizzazione di P** la misura di probabilità P_0 su $(\Omega_0, \mathcal{P}(\Omega_0))$ la cui funzione di densità discreta è la mappa p per la quale $\Omega_0 \ni \omega_0 \mapsto P(\{\omega_0\})$. Equivalentemente vale che:

$$P_0(A) = \sum_{a \in A} p(a) = P(A), \quad \forall A \in \mathcal{P}(\Omega_0).$$

Proposizione 2.9 (Proprietà della discretizzazione di P).

Se P è una probabilità discreta su Ω concentrata su Ω_0 , allora vale che:

$$P(A) = P(A \cap \Omega_0) = P_0(A \cap \Omega_0) = \sum_{a \in A \cap \Omega_0} p(a),$$

dove p è la funzione di densità di P_0 . Segue dall'identità $P(A \cup \Omega_0^c) = 1$ e dalla definizione di discretizzazione.

Osservazione 2.10.

In perfetta analogia al caso totalmente discreto, la discretizzazione di P discreta su Ω e concentrata su Ω_0 è univocamente determinata da p .

Osservazione 2.11.

Se Ω è discreto, allora si può sempre discretizzare P al suo range R_P .

2.2 Variabili aleatorie discrete

2.2.1 Definizione di v.a. discreta e composizione

Definizione 2.12 (Variabile aleatoria discreta).

Dato $S \neq \emptyset$, si definisce **variabile aleatoria** (discreta) su Ω discreto, abbreviata v.a., una funzione $X : \Omega \rightarrow S$. X si dice **variabile aleatoria reale** (v.a. reale) se $S \subseteq \mathbb{R}$ o **variabile aleatoria vettoriale** (v.a. vettoriale, o *vettore aleatorio*) se $S \subseteq \mathbb{R}^n$ per qualche $n \in \mathbb{N}$.

Dato $S \neq \emptyset$, definiamo $VA(\Omega, S)$ come l'insieme delle v.a. discrete di Ω che hanno S per codominio.

Osservazione 2.13.

Si può dotare $VA(\Omega, \mathbb{R})$ di una struttura di algebra, oltre che di spazio vettoriale, dove le operazioni di somma vettoriale, di prodotto esterno e di prodotto tra vettori sono completamente naturali.

Se Ω è finito, allora $VA(\Omega, \mathbb{R})$ è naturalmente isomorfo a $\mathbb{R}^{\#\Omega}$ come spazio vettoriale, mentre nel caso di Ω numerabile $VA(\Omega, \mathbb{R})$ ammette una base non numerabile.

Definizione 2.14 (Composizione di v.a. discrete).

Data $X \in VA(\Omega, S)$ e una funzione $F : S \rightarrow S'$, si definisce la **composizione di X tramite F** come $F(X) = F \circ X \in VA(\Omega, S')$.

2.2.2 Legge di una v.a. X

Nel caso di Ω discreto, S_X , ossia l'immagine della v.a. X , è ancora un insieme discreto. Questo ci porta alla:

Proposizione 2.15.

Sia $X : \Omega \rightarrow S$ una v.a. discreta di Ω . Sia P' la misura di probabilità sullo spazio misurabile $(S, \mathcal{P}(S))$ tale per cui:

$$P'(A) = P(X \in A) = P(X^{-1}(A)).$$

Allora P' si concentra su S_X e dunque vale che:

$$P'(A) = P'(A \cap S_X).$$

Definizione 2.16 (Legge di X).

Data una v.a. $X : \Omega \rightarrow S$, si definisce **legge di X** (o *distribuzione di X*) la discretizzazione $P^X = P'|_{S_X}$ che agisce sullo spazio misurabile $(S_X, \mathcal{P}(S_X))$, dove P' è tale per cui $P'(A) = P(X \in A) = P(X^{-1}(A))$. Equivalentemente vale che:

$$P^X : \mathcal{P}(S_X) \ni A \mapsto P(X \in A) = P(X^{-1}(A)).$$

Si indica con p_X la funzione di densità discreta di P^X . Per $P^X(A)$ con $A \subseteq S$ si intenderà $P^X(A \cap S_X)$, e analogamente $p_X(x)$ si estende in modo tale che valga 0 per $x \notin S_X$.

Osservazione 2.17.

Dalla definizione della legge di X si ricava immediatamente che:

$$P(X \in A) = P^X(A) = \sum_{x \in A} p_X(x) = \sum_{x \in A} P(X = x),$$

dove si osserva che $X \in A = \bigcup_{x \in A} (X = x)$.

¹Nella definizione compare due volte la scrittura $X = Y$: la prima volta si intende dire che la v.a. X è uguale a quella Y q.c., mentre dove compare la seconda volta si intende l'insieme $(X = Y) \subseteq \Omega$.

Osservazione 2.18.

Il range di P^X è:

$$R_X \stackrel{\text{def}}{=} R_{P^X} = \{x \in S \mid p_X(x) = P(X = x) > 0\},$$

ovverosia R_{P^X} è composto dagli elementi di S le cui controimmagini non siano trascurabili rispetto a P .

Osservazione 2.19.

Dato uno spazio di probabilità $(S, \mathcal{P}(S), Q)$ con Ω discreto è sempre possibile trovare uno spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$ e una v.a. $X : \Omega \rightarrow S$ tale per cui $P^X = Q$.

È sufficiente porre $\Omega = S$, $P = Q$ e $X = \text{id}_S$ (**costruzione canonica**). Infatti vale che:

$$P^X(A) = P(X \in A) = P(A) = Q(A).$$

Proposizione 2.20.

Data una v.a. $X : \Omega \rightarrow S$ e una funzione $f : S \rightarrow E$, vale la seguente identità:

$$p_{f(X)}(e) = \sum_{x \in f^{-1}(e)} p_X(x).$$

Equivalentemente vale che:

$$P(f(X) = e) = \sum_{x \in f^{-1}(e)} P(X = x).$$

Segue dal fatto che $(f(X) = e) = (X \in f^{-1}(e))$.

2.2.3 Uguaglianza q.c., medesima legge e stabilità per composizione

Definizione 2.21 (Uguaglianza quasi certa tra v.a.).

Date $X, Y \in VA(\Omega, S)$, si dice che X è **uguale a Y quasi certamente** ($X = Y$ q.c.¹) rispetto alla probabilità P se l'insieme $(X = Y) = \{\omega \in \Omega \mid X(\omega) = Y(\omega)\}$ è quasi certo rispetto a P .

Proposizione 2.22 (Comportamento delle uguaglianze q.c. sulla composizione).

Sia $F : S \rightarrow S'$. Siano $X, Y \in VA(\Omega, S)$. Allora se $X = Y$ q.c., $F(X) = F(Y)$ q.c.

Segue considerando la seguente relazioni di insiemi: $(X = Y) \subseteq (F(X) = F(Y))$.

Definizione 2.23 (Uguaglianza di leggi tra v.a.).

Data $X \in VA(\Omega_1, S)$ e $Y \in VA(\Omega_2, S)$, si dice che X e Y **hanno la stessa legge**, e si scrive che $X \stackrel{(d)}{=} Y$ o che $X \sim Y$, se $P_{\Omega_1}^X \equiv P_{\Omega_2}^Y$.

Definizione 2.24 (Variabili aleatorie i.d.).

Si dice che una famiglia di v.a. sono **identicamente distribuite** (**i.d.**) se condividono la stessa legge.

Spesso sottintenderemo che tali v.a. sono costruite sullo stesso Ω .

Proposizione 2.25.

Se $X = Y$ q.c., allora $X \stackrel{(d)}{=} Y$. Segue considerando che P è concentrata sull'insieme $X = Y$, e quindi ci si può sempre restringere su questo insieme, interscambiando eventualmente le v.a.

Osservazione 2.26.

Per $X, Y \in \text{VA}(\Omega, S)$ v.a. non è generalmente vero che $X \stackrel{(d)}{=} Y$ implica $X = Y$ q.c.

Proposizione 2.27 (Comportamento delle uguaglianze di legge sulla composizione).

Sia $F: S \rightarrow S'$. Siano $X, Y: \Omega_1, \Omega_2 \Rightarrow S$ v.a. Allora se $X \stackrel{(d)}{=} Y$, $F(X) \stackrel{(d)}{=} F(Y)$.

2.2.4 Variabile aleatoria multivariata, leggi congiunte e marginali

Definizione 2.28 (Variabile aleatoria multivariata, o congiunta).

Data una famiglia $(X_i: \Omega \rightarrow S_i)_{i \in I}$ di v.a. discrete di Ω con I ordinato, si definisce la **v.a. congiunta** (o *blocco multivariato*) la variabile discreta $(X_i)_{i \in I}$ tale per cui:

$$(X_i)_{i \in I}: \Omega \ni \omega \mapsto (X_i(\omega))_{i \in I} \in \prod_{i \in I} S_i.$$

Se $I = [n]$, scriviamo (X_1, \dots, X_n) al posto di $(X_i)_{i \in I}$. Sottintenderemo sempre che I è ordinato quando si nomina una famiglia di v.a. discrete.

Definizione 2.29 (Legge e densità congiunta).

Data una famiglia $(X_i: \Omega \rightarrow S_i)_{i \in I}$ di v.a. discrete di Ω e P probabilità su Ω discreto, si dice **legge congiunta** delle X_i la legge relativa alla loro v.a. congiunta, ovverosia $P^{(X_i)_{i \in I}}$. Analogamente, con il termine **densità congiunta** ci si riferirà alla densità discreta della legge congiunta.

Definizione 2.30 (Leggi e densità marginali).

Data una famiglia $(X_i: \Omega \rightarrow S_i)_{i \in I}$ di v.a. discrete di Ω e P probabilità su Ω discreto, ci si riferisce con il termine di **legge marginale** a una qualsiasi legge P^{X_i} e con il termine di **densità marginale** alla relativa funzione di densità discreta.

Osservazione 2.31.

La legge congiunta restituisce *sempre* più informazioni rispetto all'insieme delle leggi marginali. Infatti, si può sempre ricostruire una legge marginale data la legge congiunta, ma non è sempre vero il viceversa.

Osservazione 2.32.

Si osserva che vale la seguente identità:

$$P^{(X_i)_{i \in I}} \left(\prod_{i \in I} A_i \right) = P \left(\bigcap_{i \in I} (X_i \in A_i) \right), \quad \forall A_i \subseteq S_i.$$

Pertanto, nel caso finito vale che:

$$P^{(X_1, \dots, X_n)} \left(\prod_{i \in I} A_i \right) = P(X_1 \in A_1, \dots, X_n \in A_n), \quad \forall A_i \subseteq S_i.$$

Proposizione 2.33.

Ogni densità marginale è univocamente determinata dalla densità congiunta. In particolare nel caso finito vale che:

$$p_{X_i}(x_i) = \sum_{\substack{x_j \in S_j \\ j \neq i}} p_{(X_1, \dots, X_n)}(x_1, \dots, x_n).$$

2.2.5 Indipendenza di variabili aleatorie discrete e stabilità per congiunzione e composizione

Definizione 2.34 (Indipendenza tra v.a. discrete).

Sia $(X_i: \Omega \rightarrow S_i)_{i \in I}$ una famiglia di v.a. discrete. Si dice che tale famiglia di v.a. è **indipendente** se per ogni n e ogni famiglia finita di indici distinti $(i_j)_{j \in [n]} \subseteq I$ vale che:

$$P(X_{i_1} \in A_{i_1}, \dots, X_{i_n} \in A_{i_n}) = \prod_{j \in [n]} P(X_{i_j} \in A_{i_j}), \quad \forall A_{i_j} \subseteq S_{i_j}.$$

Equivalentemente tale famiglia è indipendente se:

$$P^{(X_{i_1}, \dots, X_{i_n})}(A_{i_1} \times \dots \times A_{i_n}) = \prod_{j \in [n]} P^{X_{i_j}}(A_{i_j}), \quad \forall A_{i_j} \subseteq S_{i_j}.$$

Definizione 2.35 (Variabili aleatorie i.i.d.).

Data una famiglia di variabili aleatorie, si dice che queste sono **indipendenti e identicamente distribuite (i.i.d.)** se formano una famiglia di v.a. indipendenti e se condividono la stessa legge.

Spesso sottintenderemo che tali v.a. sono costruite sullo stesso Ω .

Osservazione 2.36.

La definizione è equivalente a richiedere che per ogni scelta di $A_{i_j} \subseteq S_{i_j}$, $X_{i_1} \in A_{i_1}, \dots, X_{i_n} \in A_{i_n}$ formino una famiglia di eventi collettivamente indipendenti. Pertanto è possibile sfruttare tutte le proposizioni viste nella sottosezione *Indipendenza stocastica tra eventi*.

Inoltre, se la famiglia $(X_i)_{i \in I}$ è indipendente, lo è chiaramente anche $(X_{\sigma(i)})_{i \in I}$ per ogni $\sigma \in S(I)$ (in riferimento in particolare alla seconda identità presente nella definizione di indipendenza tra v.a.).

Osservazione 2.37.

Una v.a. costante è sempre indipendente con altre v.a., dal momento che le sue uniche controimmagini sono Ω e \emptyset , che sono indipendenti da ogni evento.

Osservazione 2.38.

Si osserva che vale la seguente identità:

$$P(X_1 \in A_1, \dots, X_n \in A_n) = \sum_{x_i \in A_i} P(X_1 = x_1, \dots, X_n = x_n).$$

Proposizione 2.39.

Sia $(X_i: \Omega \rightarrow S_i)_{i \in I}$ una famiglia di v.a. discrete. Allora tale famiglia è indipendente se per ogni n e ogni famiglia finita di indici distinti $(i_j)_{j \in [n]} \subseteq I$ vale che:

$$P(X_{i_1} = x_{i_1}, \dots, X_{i_n} = x_{i_n}) = \prod_{j \in [n]} P(X_{i_j} = x_{i_j}), \quad \forall x_{i_j} \in S_{i_j}.$$

Equivalentemente, sono indipendenti se e solo se:

$$p_{(X_{i_1}, \dots, X_{i_n})}(x_{i_1}, \dots, x_{i_n}) = \prod_{j \in [n]} p_{X_{i_j}}(x_{i_j}), \quad \forall x_{i_j} \in S_{i_j}.$$

Segue dalla precedente osservazione.

Proposizione 2.40.

Sia $(A_i)_{i \in I}$ una famiglia di eventi. Allora tale famiglia è indipendente se e solo se la famiglia di v.a. $(1_{A_i})_{i \in I}$ è indipendente.

Segue dalla precedente proposizione; infatti $(1_{A_i} = 1) = A_i$ e $(1_{A_i} = 0) = A_i^c$.

Proposizione 2.41.

Sia $(X_i : \Omega \rightarrow S_i)_{i \in I}$ una famiglia di v.a. discrete e sia $(f_i : S_i \rightarrow S'_i)_{i \in I}$ una famiglia di funzioni. Allora se $(X_i)_{i \in I}$ è una famiglia di v.a. indipendenti, anche $(f_i(X_i))_{i \in I}$ lo è.

Segue dal fatto che $(f_i(X_i) \in A_i) = (X_i \in f_i^{-1}(A_i))$.

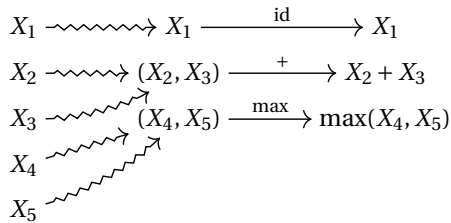
Proposizione 2.42.

Sia $(X_i : \Omega \rightarrow S_i)_{i \in I}$ una famiglia di v.a. discrete e sia I partizionato dagli I_j , ovvero $I = \cup_{j \in J} I_j$. Allora se $(X_i)_{i \in I}$ è una famiglia di v.a. indipendenti, anche $((X_i)_{i \in I_j})_{j \in J}$ è una famiglia di v.a. indipendenti.

Segue applicando la definizione.

Osservazione 2.43.

Le ultime due proposizioni permettono di ricavare molto velocemente l'indipendenza di una certa famiglia di v.a. discrete. Per esempio, se $X_1, X_2, X_3, X_4, X_5 \in VA(\Omega, \mathbb{R})$ sono indipendenti, si ricava immediatamente che $X_1, X_2 + X_3$ e $\max(X_4, X_5)$ sono indipendenti a partire dal seguente albero, dove ogni colonna è una famiglia di v.a. indipendenti:



Infatti la prima operazione restituisce una famiglia indipendente per la *Proposizione 2.42*, e la seconda fa lo stesso per la *Proposizione 2.41*.

Osservazione 2.44.

Data una famiglia di probabilità $(P_i)_{i \in [n]}$ su spazi misurabili discreti $(S_i, \mathcal{P}(S_i))$ è sempre possibile costruire uno spazio discreto di probabilità $(\Omega, \mathcal{P}(\Omega), P)$ equipaggiato di una famiglia di v.a. $(X_i : \Omega \rightarrow S_i)_{i \in [n]}$ tale per cui

1. la famiglia $(X_i)_{i \in [n]}$ è una famiglia di v.a. indipendenti,
2. $P^{X_i} \equiv P_i$.

È infatti sufficiente porre $\Omega = \prod_{i \in [n]} S_i$ (il prodotto finito di discreti è discreto), $X_i = \pi_i$ (la proiezione dal prodotto cartesiano all'insieme S_i) con P probabilità univocamente determinata dalla relazione:

$$p(x_1, \dots, x_n) = \prod_{i \in [n]} p_i(x_i).$$

Infatti in tal caso varrebbe che:

$$P(X_1 = x_1, \dots, X_n = x_n) = p(x_1, \dots, x_n) = \prod_{i \in [n]} P(X_i = x_i).$$

Tale costruzione si indica come $P \stackrel{\text{def}}{=} \otimes_{i \in [n]} P_i = P_1 \otimes \dots \otimes P_n$.

2.3 Valore atteso e momenti

2.3.1 Valore atteso su v.a. integrabili e/o non negative

Definizione 2.45 (Variabile aleatoria integrabile).

Sia X v.a. reale. Si dice che X è **integrabile** (in senso discreto)

se:

$$\mathbb{E}[|X|] \stackrel{\text{def}}{=} \sum_{\omega \in \Omega} |X(\omega)| p(\omega) < \infty,$$

ovverosia se $\mathbb{E}[|X|]$, detto il **momento primo assoluto**, converge (l'unica altra possibilità è che diverga, dacché è una serie a termini positivi).

Definizione 2.46 (Valore atteso di una v.a.).

Sia X v.a. reale. Se X è integrabile si definisce il **valore atteso** di X (o *momento primo*) come:

$$\mathbb{E}[X] \stackrel{\text{def}}{=} \sum_{\omega \in \Omega} X(\omega) p(\omega) \in \mathbb{R},$$

dove l'ultima appartenenza è data proprio dal fatto che $\mathbb{E}[|X|] < \infty$ (e dunque vi è convergenza assoluta, dacché $p(\omega) \geq 0$).

Se $X \geq 0$ q.c., si definisce allora stesso modo $\mathbb{E}[X]$, che però può assumere come valore anche ∞ ; e così per $X \leq 0$ q.c. si pone $\mathbb{E}[X] \stackrel{\text{def}}{=} -\mathbb{E}[X^-]$. In questo modo ammettiamo eventualmente i valori di ∞ o $-\infty$.

Diciamo che X **ha valore atteso**, se esiste un $\mathbb{E}[X]$ associatogli.

Osservazione 2.47.

Il valore atteso è da associarsi a un "baricentro" della distribuzione di X , ovvero, su una popolazione Ω , misura quanto vale in media la caratteristica data da X .

Osservazione 2.48.

Per la v.a. 1_A con $A \subseteq \Omega$ vale che $\mathbb{E}[1_A] = 1 \cdot P(1_A = 1) + 0 \cdot P(1_A = 0) = P(A)$.

Osservazione 2.49.

Per X tale per cui $\mathbb{E}[X^+], \mathbb{E}[X^-] < \infty$ vale che:

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Come vedremo, questo è un caso particolare della linearità di $\mathbb{E}[\cdot]$ (infatti $X = X^+ - X^-$).

Lemma 2.50 (Valore atteso tramite la legge).

Per X con valore atteso vale la seguente identità:

$$\mathbb{E}[X] = \sum_{x \in R_X} x \cdot p_X(x) = \sum_{x \in R_X} x \cdot P(X = x).$$

Segue dal fatto che $\mathbb{E}[X] = \sum_{x \in R_X} \sum_{s \in X^{-1}(x)} x \cdot p(s)$.

Questa proposizione può estendersi facilmente alla:

Proposizione 2.51 (Valore atteso della composizione tramite la legge).

Sia $X : \Omega \rightarrow S$ v.a. discreta e sia $\varphi : S \rightarrow \mathbb{R}$. Allora vale che:

- (i.) $\varphi(X)$ è integrabile se e solo se $\sum_{x \in R_X} |\varphi(x)| P(X = x) < \infty$,
- (ii.) se $\varphi(X)$ ha valore atteso, allora:

$$\mathbb{E}[\varphi(X)] = \sum_{x \in R_X} \varphi(x) \cdot p_X(x) = \sum_{x \in R_X} \varphi(x) \cdot P(X = x).$$

Segue dal fatto che $\mathbb{E}[\varphi(X)] = \sum_{x \in R_X} \sum_{s \in X^{-1}(x)} \varphi(x) \cdot p(s)$.

Osservazione 2.52 (Uguaglianza di valori attesi per leggi uguali).

Dal momento che $\mathbb{E}[\varphi(X)]$ dipende soltanto dalla legge di p_X , $X \stackrel{(d)}{=} Y \implies \mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$.

2.3.2 Proprietà del valore atteso e moltiplicatività per v.a. indipendenti

Proposizione 2.53.

Siano X e Y due v.a. reali con valore atteso. Allora vale che:

- (i.) Se $X = c$ q.c., allora $\mathbb{E}[X] = c$,
- (ii.) Se $X \geq 0$ q.c./integrabile, allora per $a \in \mathbb{R}^+$, $aX \geq 0$ q.c./integrabile,
- (iii.) Se X ha valore atteso, allora per $a \in \mathbb{R}$ pure aX lo ha e $\mathbb{E}[aX] = a\mathbb{E}[X]^2$
- (iv.) Se $X \geq 0$ q.c. o $X \leq 0$ q.c. e $\mathbb{E}[X] = 0$, allora $X = 0$ q.c.,
- (v.) Se $X \leq Y$ q.c., allora $\mathbb{E}[X] \leq \mathbb{E}[Y]$,
- (vi.) Se X e Y hanno valore atteso e non sono uno ∞ e l'altro $-\infty$, allora $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Proposizione 2.54.

Siano $X, Y : \Omega \Rightarrow S, S'$, due v.a. indipendenti. Se $g, h : S, S' \Rightarrow \mathbb{R}$ sono funzioni e $g(X)$ e $h(Y)$ ammettono valore atteso³, allora vale che:

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \cdot \mathbb{E}[h(Y)].$$

Usando che $\mathbb{E}[g(X)h(Y)] = \sum_{(x,y) \in R_{(X,Y)}} g(x)h(y)P(X=x, Y=y)$, segue, per l'indipendenza di X e Y , dal fatto che $R_{(X,Y)} = R_X \times R_Y$ e che $P(X=x, Y=y) = P(X=x)P(Y=y)$.

Osservazione 2.55.

In particolare, per v.a. reali X, Y indipendenti che ammettono valore atteso vale che:

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

Osservazione 2.56.

Dalla Proposizione 2.53 si deduce che $\mathbb{E}[\cdot]$ è un funzionale di $VA(\Omega, \mathbb{R})$ (ovverosia $\mathbb{E}[\cdot] \in VA(\Omega, \mathbb{R}^*)$).

Proposizione 2.57.

Sia X una v.a. reale che assume valori naturali quasi certamente. Allora vale che:

$$\mathbb{E}[X] = \sum_{n \in \mathbb{N}} P(X > n).$$

In generale se X è una v.a. reale che assume valori positivi il cui range ordinato è $(x_i)_{i \in I}$ (con $I = \mathbb{N}^+$ o $I = [k]$), allora, posto $x_0 = 0$, vale che:

$$\mathbb{E}[X] = \sum_{n \in \mathbb{N}} (x_{n+1} - x_n)P(X > x_n).$$

2.3.3 Valore atteso condizionale

Definizione 2.58 (Valore atteso condizionale).

Sia X una v.a. reale. Dato allora un evento $A \in \mathcal{P}(\Omega)$, si definisce il **valore atteso condizionale** $\mathbb{E}[X | A]$ in modo tale che:

$$\mathbb{E}[X | A] \stackrel{\text{def}}{=} \frac{\mathbb{E}[X \cdot \mathbf{1}_A]}{P(A)} = \sum_{\omega \in A} X(\omega) \cdot P(\{\omega\} | A).$$

Alternativamente vale che:

$$\mathbb{E}[X | A] = \sum_{x \in R_X} x \cdot \frac{P((X=x) \cap A)}{P(A)} = \sum_{x \in R_X} x \cdot P(X=x | A).$$

² Si assume la convenzione per cui $0 \cdot \infty = 0$, $a \cdot \infty = \text{sgn}(a)\infty$ per $a \neq 0$.

³ Si ammette in questo caso la convenzione per cui $\infty \cdot \infty = \infty$ e che $-\infty \cdot \infty = -\infty$.

Il valore atteso condizionale rimodula il valore atteso in modo tale da considerare solamente le immagini di X possibili sotto l'ipotesi che sia accaduto l'evento A . Pertanto è naturale aspettarsi il seguente:

Lemma 2.59 (Formula dei valori attesi totali, o formula della partizione dei valori attesi).

Sia X una v.a. reale e sia $(A_i)_{i \in [n]}$ un sistema di alternative finito per Ω . Allora vale che:

$$\mathbb{E}[X] = \sum_{i \in [n]} \mathbb{E}[X | A_i]P(A_i).$$

Segue considerando che $X = X \cdot (\sum_{i \in [n]} \mathbf{1}_{A_i})$.

2.3.4 Momenti (assoluti) n -esimi

Definizione 2.60 (Momento n -esimo assoluto).

Data X v.a. reale e $n \in \mathbb{R}^+$, definiamo il **momento assoluto di ordine n** (momento n -esimo assoluto, se esiste, $\mathbb{E}[|X|^n]$).

Generalmente si pone più attenzione ai momenti n -esimi assoluti con n intero positivo.

Definizione 2.61 (Momento n -esimo).

Data X v.a. reale e $n \in \mathbb{R}^+$, se X ammette momento n -esimo assoluto, allora X^n ammette $\mathbb{E}[X^n]$, che viene detto **momento n -esimo di X^n** .

Lemma 2.62.

Data X v.a. reale e $1 \leq p \leq q$ in \mathbb{R} , se $\mathbb{E}[|X|^q] < \infty$ allora $\mathbb{E}[|X|^p] < \infty$.

Segue dal fatto che $\mathbb{E}[|X|^p]$ è uguale a $\mathbb{E}[|X|^p \cdot \mathbf{1}_{|X|>1} + |X|^p \cdot \mathbf{1}_{|X|\leq 1}]$; applicando la linearità di $\mathbb{E}[\cdot]$ e che $x^p \leq x^q$ per $x \geq 1$, si ricava così che $\mathbb{E}[|X|^p] \leq \mathbb{E}[|X|^q] + 1$.

Osservazione 2.63.

Se X è limitata quasi certamente ($|X| \leq M$ q.c. con $M > 0$), allora X ammette momento n -esimo assoluto per ogni $n \in \mathbb{R}^+$ (segue dal fatto che $\mathbb{E}[|X|^n] \leq M^n$).

Osservazione 2.64.

La disuguaglianza impiegata nello scorso lemma ha una generalizzazione più ampia, che non dimostriamo, ma che segue dalla *Disuguaglianza di Hölder*:

$$\mathbb{E}[|X|^p]^{\frac{1}{p}} \leq \mathbb{E}[|X|^q]^{\frac{1}{q}}, \quad 1 < p < q.$$

Lemma 2.65.

Se $\mathbb{E}[|X|^p], \mathbb{E}[|Y|^p] < \infty$, allora $\mathbb{E}[|aX + Y|^p] < \infty$ per ogni $a, b \in \mathbb{R}$.

Segue dal fatto che $|aX + Y|^p \leq 2^{p-1}(|a|^p|X|^p + |Y|^p)$.

2.3.5 Disuguaglianza di Markov, di Hölder, di Cauchy-Schwarz e di Jensen

Proposizione 2.66 (Disuguaglianza di Markov).

Sia $X \geq 0$ v.a. reale. Allora $\forall a > 0$ vale che:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Segue considerando che $X \geq a \cdot 1_{X \geq a}$, e dunque $\mathbb{E}[X] \geq a \cdot \mathbb{E}[1_{X \geq a}] = a \cdot P(X \geq a)$.

Corollario 2.67.

Sia X v.a. reale. Allora $\forall a \neq 0, \forall p > 0$ vale che:

$$P(|X| \geq |a|) \leq \frac{\mathbb{E}[|X|^p]}{|a|^p}.$$

Segue dalla disuguaglianza di Markov.

In generale la disuguaglianza di Markov si può esprimere per composizione con funzioni crescenti:

Corollario 2.68.

Sia X v.a. reale. Allora, se $f: \mathbb{R} \rightarrow [0, \infty)$ è crescente, $\forall a \in \text{supp } f$ (i.e. $f(a) \neq 0$) vale che:

$$P(X \geq a) \leq \frac{\mathbb{E}[f(X)]}{f(a)}.$$

Segue dalla disuguaglianza di Markov. Si osserva in particolare che non si è richiesto che X fosse t.c. $X \geq 0$.

Proposizione 2.69 (Disuguaglianza di Hölder).

Siano X, Y v.a. reali. Siano $p, q > 1$ coniugati (ossia t.c. $\frac{1}{p} + \frac{1}{q} = 1$). Allora, se X ammette momento p -esimo assoluto e Y ammette momento q -esimo assoluto, entrambi finiti, vale che:

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}} \cdot \mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$

Segue dalla usuale disuguaglianza di Hölder in analisi.

Proposizione 2.70 (Disuguaglianza di Cauchy-Schwarz).

Siano X, Y v.a. reali. Allora, se X e Y ammettono momento secondo assoluto finito, vale che:

$$\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{\frac{1}{2}} \cdot \mathbb{E}[Y^2]^{\frac{1}{2}}.$$

Segue dalla usuale disuguaglianza di Cauchy-Schwarz in analisi o dalla disuguaglianza di Hölder per $p = q = \frac{1}{2}$.

Proposizione 2.71 (Disuguaglianza di Jensen).

Sia X una v.a. reale che ammette valore atteso. Allora, se $g: \mathbb{R} \rightarrow \mathbb{R}$ è una funzione convessa che ammette valore atteso vale che:

$$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

Equivalentemente, se g è concava vale la disuguaglianza con \geq al posto di \leq . Segue dall'usuale disuguaglianza di Jensen.

2.4 Altri indici di centralità: moda e mediana

Il valore atteso $\mathbb{E}[X]$ è considerato un **indice di centralità** dacché fornisce un'idea del baricentro della distribuzione di X . Di seguito sono definiti altri due indici di centralità celebri.

Definizione 2.72 (Moda).

Data una v.a. reale X , si dice che $x \in S_X$ è una **moda** se x è un massimo per P_X . Una distribuzione in generale può avere più mode.

Definizione 2.73 (Mediana).

Data una v.a. reale X , si dice che $x \in S_X$ è una **mediana** se $P(X \leq x) \geq \frac{1}{2}$ e $P(X \geq x) \geq \frac{1}{2}$.

Proposizione 2.74.

Esistono sempre almeno una moda e almeno una mediana per X v.a. reale.

2.5 Indici di dispersione: covarianza, varianza, dev. standard e coeff. di correlazione

2.5.1 Definizioni e covarianza come forma bilineare simmetrica

Definizione 2.75 (Covarianza e v.a. scorrelate).

Date due v.a. reali X, Y con momento secondo finito, si definisce **covarianza di X e Y** il termine:

$$\text{Cov}(X, Y) \stackrel{\text{def}}{=} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Si dice che X e Y sono **scorrelate** se $\text{Cov}(X, Y) = 0$.

Definizione 2.76 (Varianza).

Data una v.a. reale X con momento secondo finito, si definisce **varianza di X** il termine:

$$\text{Var}(X) \stackrel{\text{def}}{=} \text{Cov}(X, X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \geq 0,$$

dove la non negatività segue dal fatto che $(X - \mathbb{E}[X])^2 \geq 0$.

Proposizione 2.77.

$\mathbb{E}[X]$ è il termine che sostituito a m minimizza il valore $\mathbb{E}[(X - m)^2]$.

Definizione 2.78 (Deviazione standard).

Data una v.a. reale X che ammette varianza, si definisce **deviazione standard di X** il termine:

$$\sigma(X) \stackrel{\text{def}}{=} \sqrt{\text{Var}(X)}.$$

Osservazione 2.79.

La deviazione standard misura quanto X si discosta mediamente da $\mathbb{E}[X]$, se esiste.

Osservazione 2.80.

La varianza e la deviazione standard sono detti **indici di dispersione** della distribuzione di X , dacché misurano quanto le immagini di X distano mediamente dal valore atteso $\mathbb{E}[X]$.

Proposizione 2.81.

Sia X una v.a. reale che ammette varianza. Allora $\text{Var}(X) = 0$ se e solo se X è costante q.c.

Segue dal fatto che $\mathbb{E}[(X - \mathbb{E}[X])^2] = 0$ se e solo se $\mathbb{E}[X] = X$ q.c., ovvero se e solo se X è una costante.

2.5.2 Identità sulla (co)varianza e disuguaglianza di Chebyshev

Proposizione 2.82.

$\text{Cov}(\cdot, \cdot)$ è una funzione simmetrica e lineare in ogni suo argomento. In particolare per X e Y con momento secondo finito vale che:

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Pertanto due v.a. indipendenti hanno covarianza nulla (i.e. sono scorrelate) per l'Osservazione 2.55. In particolare, la covarianza tra una qualsiasi costante q.c. e un'altra v.a. reale è nulla.

Osservazione 2.83.

La precedente proposizione mette ancora in luce come sia determinante la legge congiunta $p_{(X,Y)}$, usata per calcolare $\mathbb{E}[XY]$, che in generale le leggi p_X e p_Y , che pure si usano per calcolare $\mathbb{E}[X]$ e $\mathbb{E}[Y]$, non riescono a ricostruire.

Osservazione 2.84.

A partire dalla precedente proposizione si ricava che per X v.a. reale con momento secondo finito vale che:

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

Osservazione 2.85.

Viste le proprietà discusse nella precedente proposizione si può concludere che la covarianza sul sottospazio di $\text{VA}(\Omega, \mathbb{R})$ delle v.a. con momento secondo finito corrisponde a una forma bilineare simmetrica semidefinita positivo, ovvero a un prodotto scalare.

Due v.a. indipendenti sono ortogonali tramite Cov per la Proposizione 2.82.

Al cono isotropo e al radicale di questo prodotto appartengono solo le costanti per la Proposizione 2.81.

Se $\varphi \stackrel{\text{def}}{=} \text{Cov}$, vale che $q_\varphi \equiv \text{Var}$ e $\|\cdot\|_\varphi \equiv \sigma$, ovvero la varianza Var è la forma quadratica associata alla covarianza Cov , mentre σ ne è la norma.

Lemma 2.86.

Siano X_1, \dots, X_n v.a. reali con momento secondo finito. Allora vale che:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i \in [n]} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

In particolare, se $(X_i)_{i \in [n]}$ è una famiglia di v.a. scorrelate a due a due (e.g. indipendenti) vale che:

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i \in [n]} \text{Var}(X_i).$$

Lemma 2.87.

Sia $aX + b$ una v.a. reale con X che ammette momento secondo finito. Allora vale che:

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

Segue dal fatto che aX e b sono indipendenti, che $\text{Var}(b) = 0$ e che Var è la forma quadratica di Cov.

Proposizione 2.88 (Disuguaglianza di Chebyshev).

Sia X v.a. reale con momento secondo finito. Allora $\forall a > 0$ vale che:

$$P(|X - \mathbb{E}[X]| > a) \leq \frac{\text{Var}(X)}{a^2}.$$

Segue dall'immediata applicazione della disuguaglianza di Markov.

2.5.3 Coeff. di correlazione e retta di regressione lineare

Definizione 2.89 (Coefficiente di correlazione di Pearson, PCC).

Date X, Y v.a. reali non costanti q.c.⁴ e con momento secondo finito si definisce il **coefficiente di correlazione di Pearson** (PCC) $\rho(X, Y)$, o più brevemente *coefficiente di correlazione*, come il coseno di X e Y rispetto a Cov, ovvero:

$$\rho(X, Y) \stackrel{\text{def}}{=} \cos_{\text{Cov}}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)}.$$

Lemma 2.90.

Date X, Y v.a. reali non costanti q.c. e con momento secondo finito vale che:

(i.) $|\rho(X, Y)| \leq 1$ (per la disuguaglianza di Cauchy-Schwarz),

(ii.) $\rho(aX + b, cX + d) = \rho(X, Y)$ (per verifica diretta).

Teorema 2.91.

Siano X, Y v.a. reali con momento secondo finito e non costanti q.c. Allora la funzione:

$$\mathbb{R}^2 \ni (a, b) \mapsto \mathbb{E}[(Y - (aX + b))^2] \in \mathbb{R}$$

è ben definita e ammette un unico punto di minimo (a^*, b^*) , dove:

$$a^* = C_{\text{Cov}}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad b^* = \mathbb{E}[Y] - a^* \mathbb{E}[X].$$

Inoltre il valore di tale minimo è:

$$\mathbb{E}[(Y - (a^* X + b^*))^2] = \text{Var}(Y) \cdot (1 - \rho(X, Y)^2).$$

Definizione 2.92 (Retta di regressione (lineare)).

Date X, Y v.a. reali con momento secondo finito e non costanti q.c. si definisce **retta di regressione** (lineare) la retta $y = a^* x + b^*$.

Osservazione 2.93.

Dal precedente teorema si può ottenere una caratterizzazione della correlazione lineare tra due v.a. reali X e Y non costanti q.c. e con momento secondo finito. Infatti vale che:

⁴ Infatti il coseno è definito solo per coppie di vettori anisotropi ed il cono isotropo di Cov è costituito dalle sole costanti q.c.

- la retta di regressione di X e Y rappresenta la migliore approssimazione lineare di Y tramite X ,
- $\rho(X, Y) \approx 0$ (X, Y quasi scorrelate) \implies poca correlazione lineare ($\mathbb{E}[(Y - (a^*X + b^*))^2]$ assume approssimativamente il valore massimo possibile e dunque Y dista mediamente tanto da ogni retta di X),
- $\rho(X, Y) \approx 1 \implies$ forte correlazione lineare (infatti se $\rho = 1$, $\mathbb{E}[(Y - (a^*X + b^*))^2] = 0$, e dunque $Y = a^*X + b^*$ q.c.).

Si osserva inoltre che $\text{sgn}(a^*) = \text{sgn}(\rho(X, Y))$.

2.6 Legge dei grandi numeri (LGN), media campionaria e limite in senso probabilistico

2.6.1 Definizioni ed enunciato

Definizione 2.94 (Media campionaria n -esima).

Data una famiglia di v.a. reali $(X_i)_{i \in \mathbb{N}}$ i.i.d. dotate di momento secondo finito⁵ si definisce **media campionaria n -esima** il termine:

$$\bar{X}_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i \in [n]} X_i,$$

ovverosia la media aritmetica delle prime n v.a. della famiglia.

Definizione 2.95 (Limite probabilistico).

Data una successione di v.a. reali $(Y_i : \Omega \rightarrow \mathbb{R})_{i \in \mathbb{N}}$ e data una v.a. reale $Y : \Omega \rightarrow \mathbb{R}$ si dice che Y_n tende (probabilisticamente) a Y ($Y_n \xrightarrow{\mathbb{P}} Y$) per $n \rightarrow \infty$ se:

$$\lim_{n \rightarrow \infty} P(|Y_n - Y| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Osservazione 2.96.

Una successione di v.a. reali $(Y_i)_{i \in \mathbb{N}}$ tende a Y se si può sempre scegliere un n arbitrariamente grande tale per cui la probabilità che Y_i sia pari a Y (eccetto per un errore assoluto ε fissato) è certa entro un errore arbitrario.

Teorema 2.97 (Legge (debole) dei grandi numeri, LGN).

Sia $(X_i)_{i \in \mathbb{N}}$ una famiglia di v.a. reali scorrelate e i.i.d. (e.g. i.i.d.) dotate di momento secondo finito, ovverosia con $\mathbb{E}[X_1^2] < \infty$. Allora vale che:

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X_1], \quad \text{per } n \rightarrow \infty.$$

Dimostrazione.

Si osserva che $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X_1]$ e che $\text{Var}(\bar{X}_n) = \frac{1}{n} \text{Var}(X_1)$. Allora, se $\varepsilon > 0$, per la disuguaglianza di Chebyshev vale che:

$$P\left(\left|\bar{X}_n - \mathbb{E}[X_1]\right| > \varepsilon\right) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\text{Var}(X_1)}{\varepsilon^2 n}.$$

Dal momento che $\frac{\text{Var}(X_1)}{\varepsilon^2 n} \rightarrow 0$ per $n \rightarrow \infty$, si ottiene la tesi. \square

Osservazione 2.98.

In alcune occasioni, ovverosia quando $\text{Var}(\bar{X}_n) \rightarrow 0$ per $n \rightarrow \infty$, è ancora possibile applicare la LGN seguendo la stessa dimostrazione.

⁵Dal momento che le X_i sono i.i.d. è sufficiente che X_1 sia dotata di momento secondo finito.

Osservazione 2.99.

La legge dei grandi numeri ci permette di ricondurre la definizione assiomatica di Kolmogorov di probabilità a quella frequentista. Se infatti fissiamo una probabilità P e costruiamo un modello di prove ripetute (come definito successivamente) il cui successo è dipeso da se accade l'evento A , considerando come famiglia di v.a. i.i.d. la famiglia $(1_{A_i})_{i \in \mathbb{N}}$, dove A_i è l'evento di successo di A nella prova i -esima, per la legge dei grandi numeri si ottiene che per $n \rightarrow \infty$ vale che:

$$\frac{\overline{1_{A_n}}}{\text{numero di prove}} = \frac{\text{numero di volte che accade } A}{\text{numero di prove}} \xrightarrow{\mathbb{P}} \mathbb{E}[1_{A_1}] = P(A).$$

2.6.2 Trasformata di Cramer per l'ottimizzazione della stima

Cerchiamo in questa sezione di ottenere, utilizzando la funzione esponenziale, una stima ottimale per $P(\bar{X}_n - m > \varepsilon)$ con $\varepsilon > 0$, $(X_i)_{i \in \mathbb{N}}$ famiglia di v.a. i.i.d. e $m = \mathbb{E}[X_1]$ finito.

Dacché $\exp : \mathbb{R} \rightarrow (0, \infty)$ è crescente, vale che, per $\lambda > 0$:

$$\begin{aligned} P(\bar{X}_n - m > \varepsilon) &= P\left(\lambda \sum_{i \in [n]} (X_i - m) > \lambda n \varepsilon\right) = \\ &= P\left(\exp\left(\lambda \sum_{i \in [n]} (X_i - m)\right) > \exp(\lambda n \varepsilon)\right). \end{aligned}$$

Applicando la disuguaglianza di Markov si ottiene che:

$$\begin{aligned} P(\bar{X}_n - m > \varepsilon) &\leq \frac{1}{e^{\lambda n \varepsilon}} \mathbb{E}\left[\exp\left(\lambda \sum_{i \in [n]} (X_i - m)\right)\right] = \\ &= \frac{1}{e^{\lambda n \varepsilon}} \mathbb{E}[\exp(\lambda(X_1 - m))]^n = \\ &= \exp\left(-n\left(\lambda \varepsilon - \log \mathbb{E}\left[e^{\lambda(X_1 - m)}\right]\right)\right), \end{aligned}$$

dove si è utilizzato che le v.a. sono indipendenti e identicamente distribuite.

Definizione 2.100 (Trasformata di Cramer).

Dato $\varepsilon > 0$, $(X_i)_{i \in \mathbb{N}}$ famiglia di v.a. i.i.d. e $m = \mathbb{E}[X_1]$ finito, si definisce **trasformata di Cramer** il valore:

$$I(t) = \sup_{\lambda > 0} \left(\lambda t - \log \mathbb{E}\left[e^{\lambda(X_1 - m)}\right]\right).$$

Ottimizzando dunque in λ , la precedente disuguaglianza di scrive come:

$$P(\bar{X}_n - m > \varepsilon) \leq e^{-n \cdot I(\varepsilon)}.$$

Se dunque esiste $\lambda > 0$ per cui $\mathbb{E}\left[e^{\lambda(X_1 - m)}\right]$ è finito, allora $I(\varepsilon) > 0$, e dunque $P(\bar{X}_n - m > \varepsilon)$ tende esponenzialmente a 0 per $n \rightarrow \infty$.

2.7 Teorema centrale del limite (TCL, o TLC)

2.7.1 Intuizione del TCL: zoom-in e scaling

Per la legge dei grandi numeri sappiamo già che $\overline{X}_n - m \xrightarrow{\mathbb{P}} 0$ per $m = \mathbb{E}[X_1]$, $n \rightarrow \infty$ e $(X_i)_{i \in [n]}$ famiglia di v.a. i.i.d. Ciò è dipeso, come illustrato dalla dimostrazione, dal fatto che è presente un fattore $\frac{1}{n}$ in $\text{Var}(\overline{X}_n)$.

Se $\alpha > 0$ e consideriamo lo *scaling* (o *zoom-in*) $n^\alpha(\overline{X}_n - m)$ vale che:

$$\text{Var}(n^\alpha(\overline{X}_n - m)) = n^{2\alpha} \text{Var}(\overline{X}_n) = n^{2\alpha-1} \text{Var}(X_1).$$

Pertanto, riapplicando la disuguaglianza di Chebyshev:

$$P\left(n^\alpha \left| \overline{X}_n - m \right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} n^{2\alpha-1} \text{Var}(X_1).$$

Per $\alpha < \frac{1}{2}$ si riottiene una tesi analoga a quella della LGN. È lecito dunque aspettarsi che per $\alpha = \frac{1}{2}$ possa accadere qualcosa di diverso, da cui l'intuizione del TCL.

2.7.2 Enunciato del TCL e Teorema di De Moivre-Laplace per la distr. binomiale

Teorema 2.101 (Teorema centrale del limite, TCL; oppure Teorema del limite centrale, TLC).

Sia $(X_i)_{i \in \mathbb{N}}$ una famiglia di v.a. i.i.d. dotate di momento secondo finito ($\mathbb{E}[X_1^2] < \infty$) e non costanti q.c. ($\text{Var}(X_1) > 0$). Sia $\sigma = \sigma(X_1)$ e sia $m = \mathbb{E}[X_1]$. Allora per ogni scelta di a, b tali per cui $-\infty \leq a \leq b \leq \infty$ ⁶ vale che per $n \rightarrow \infty$:

$$P\left(a \leq \frac{\sqrt{n}}{\sigma} (\overline{X}_n - m) \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Equivalentemente vale che:

$$P\left(a \leq \frac{1}{\sqrt{ns}} \left[\left(\sum_{i \in [n]} X_i \right) - nm \right] \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

⚠ Attenzione. Per il calcolo di $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$ mediante la funzione $\Phi(x)$ si rimanda alla Tabella 2.1 allegata nelle ultime pagine di queste schede riassuntive.

Corollario 2.102 (Teorema di De Moivre-Laplace).

Sia $Y_n \sim B(n, p)$. Allora per ogni scelta di a, b tali per cui $-\infty \leq a \leq b \leq \infty$ vale che per $n \rightarrow \infty$:

$$\begin{aligned} P\left(np + \sqrt{np(1-p)}a \leq Y_n \leq np + \sqrt{np(1-p)}b\right) \\ \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Dimostrazione.

Segue dal TCL dal momento che Y_n è somma di n v.a. X_i i.i.d. con $X_i \sim B(p)$. In particolare $m = \mathbb{E}[X_1] = p$ e $\sigma = \sigma(X_1) = \sqrt{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2} = \sqrt{p(1-p)}$. □

⁶Si ammettono dunque anche i casi $\pm\infty$.

2.8 Modelli probabilistici classici

2.8.1 Probabilità uniforme

Definizione 2.103 (Probabilità uniforme).

Dato Ω finito, si definisce **probabilità uniforme** l'unica probabilità $P : \mathcal{F} \rightarrow \mathbb{R}$ la cui funzione di densità è costante (*equiprobabile*). Equivalentemente è la probabilità P tale per cui:

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

Osservazione 2.104.

Non è possibile dotare Ω numerabile di una probabilità uniforme. Infatti, se l'unica immagine della funzione $p : \Omega \rightarrow \mathbb{R}$ è c , $\sum_{\omega \in \Omega} p(\omega) = c \sum_{\omega \in \Omega} 1$, che può valere solo 0 o ∞ , e dunque non 1 (e pertanto non può indurre una probabilità).

2.8.2 Sequenze di esperimenti e modello delle prove ripetute di Bernoulli

Cerchiamo di modellare una sequenza ordinata (e potenzialmente infinita, ma al più numerabile) di esperimenti. Data una famiglia $(\Omega_i)_{i \in I}$, con $I = \mathbb{N}$ o $I = [n]$, dove ciascuno Ω_i indica l' i -esimo esperimento, definiamo in tal caso:

$$\Omega = \left\{ (\omega_1, \omega_2, \dots) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2^{(\omega_1)}, \omega_3 \in \Omega_3^{(\omega_1, \omega_2)}, \dots \right\},$$

dove la notazione $\Omega_i^{(\omega_j)_{j \in [i-1]}}$ indica il sottoinsieme di Ω_i degli esiti dell'esperimento possibili una volta che nei precedenti esperimenti sono successi $\omega_1, \dots, \omega_{i-1}$. Se i precedenti esperimenti non condizionano gli esiti dei successivi, allora $\Omega = \prod_{i \in I} \Omega_i$.

Riduciamoci al caso di una sequenza (finita o infinita) di esperimenti tra di loro non condizionati, ciascuno con esito successo (1) o insuccesso (0). Un tale esperimento è detto **prova di Bernoulli**. In tal caso $\Omega = \prod_{i \in I} \{[1]\}$.

Sia A_i l'evento "successo all' i -esima prova", ossia:

$$A_i = \{\omega \in \Omega \mid \omega_i = 1\}.$$

Sia $p_i : \{[1]\} \rightarrow \mathbb{R}$ la funzione di densità associata alla misura di probabilità dell'esperimento Ω_i . Associamo allora ad Ω la σ -algebra $\mathcal{F} = \sigma(A_i)_{i \in I}$ generata dagli A_i (che è al più numerabile). Se I è finito, $\mathcal{F} = \mathcal{P}(\Omega)$.

Definizione 2.105 (Modello della sequenza di prove).

Si definisce **probabilità del modello della sequenza di prove** l'unica probabilità P sullo spazio misurabile (Ω, \mathcal{F}) tale per cui $(A_i)_{i \in I}$ è una famiglia di eventi indipendenti e per la quale $P(A_i) = p_i(1)$.

Osservazione 2.106.

Tale probabilità è univocamente determinata dal momento che gli A_i generano \mathcal{F} e che sono indipendenti.

Definizione 2.107 (Modello delle prove ripetute).

Se P è una probabilità del modello della sequenza di prove e $p_i(1) = p_j(1)$ per ogni coppia i, j , allora il modello prende il nome di **modello delle prove ripetute** e si dice che $p \stackrel{\text{def}}{=} p_1(1)$ è il **parametro di Bernoulli**.

A partire dal modello delle prove ripetute si possono formalizzare numerose distribuzioni, come quelle della sezione successiva.

Tabella e proprietà delle distribuzioni discrete

Nome distribuzione	Caso di utilizzo	Parametri	Densità discreta	Valore atteso	Momento secondo	Varianza
Distr. di Bernoulli $X \sim B(p)$	Esperimento con esito di successo (1) o insuccesso (0).	p – probabilità di successo.	$P(X=1) = p, P(X=0) = 1-p$	$\mathbb{E}[X] = p$	$\mathbb{E}[X^2] = p$	$\text{Var}(X) = p(1-p)$
Distr. binomiale $X \sim B(n, p)$	In una serie di n esperimenti col modello delle prove ripetute, X conta il numero di successi. X è in particolare somma di n v.a. i.i.d. distribuite come $B(p)$.	n – numero di esperimenti p – probabilità di successo dell' i -esimo esperimento	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ per $0 \leq k \leq n$ e 0 altrimenti.	$\mathbb{E}[X] = np$ (è somma di n Bernoulliane)	$\mathbb{E}[X^2] = np + n(n-1)p^2$	$\text{Var}(X) = np(1-p)$ (è somma di n Bernoulliane indipendenti)
Distr. binomiale negativa $X \sim \text{BinNeg}(h, p)$	In una serie di infiniti esperimenti col modello delle prove ripetute, X conta l'esperimento in cui si ha l' h -esimo successo.	h – numero dei successi da misurare $p \in (0, 1)$ – probabilità di successo dell' i -esimo esperimento	$P(X=k) = \binom{k-1}{h-1} p^h (1-p)^{k-h}$ laddove definibile e 0 altrimenti.	$\mathbb{E}[X] = \frac{h}{p}$ (è somma di h Geometriche)	$\mathbb{E}[X^2] = \frac{h(1+h-p)}{p^2}$	$\text{Var}(X) = \frac{h(1-p)}{p^2}$ (è somma di h Geometriche indipendenti)
Distr. geometrica $X \sim \text{Geom}(p)$	In una serie di infiniti esperimenti col modello delle prove ripetute, X conta l'esperimento in cui si ha il primo successo. È pari a $\text{BinNeg}(1, p)$	$p \in (0, 1)$ – probabilità di successo all' i -esimo esperimento.	$P(X=k) = p(1-p)^{k-1}$ per $k \geq 1$ e 0 per $k=0$.	$\mathbb{E}[X] = \frac{1}{p}$	$\mathbb{E}[X^2] = \frac{2-p}{p^2}$	$\text{Var}(X) = \frac{1-p}{p^2}$
Distr. ipergeometrica $X \sim H(N, N_1, n)$	In un'estrazione di n palline in un'urna di N palline, di cui N_1 sono rosse, X conta il numero di palline rosse estratte.	N – numero di palline nell'urna N_1 – numero di palline rosse nell'urna n – numero di palline estratte	$P(X=k) = \frac{\binom{N_1}{k} \binom{N-N_1}{n-k}}{\binom{N}{n}}$ laddove definibile e 0 altrimenti.			
Distri. di Poisson (o degli eventi rari) $X \sim \text{Poisson}(\lambda)$	In una sequenza di $n \gg 1$ esperimenti di parametro $p \ll 1$ con $np \approx \lambda$, X misura il numero di successi. Si può studiare come distribuzione limite della distribuzione binomiale.	λ – tasso di successo.	$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$	$\mathbb{E}[X] = \lambda$	$\mathbb{E}[X^2] = \lambda(\lambda+1)$	$\text{Var}(X) = \lambda$

Valgono inoltre le seguenti altre proprietà:

- Una somma di n v.a. i.i.d. distribuite come $B(p)$ si distribuisce come $B(n, p)$.
- Se $X \sim B(n, p)$ e $Y \sim B(m, p)$ sono indipendenti, $X + Y$ si distribuisce come $B(n + m, p)$.
- Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ sono indipendenti, $X + Y$ si distribuisce come $\text{Poisson}(\lambda + \mu)$.
- Se $X \sim \text{Geom}(p)$, allora $P(X = \infty) = 0^7$. Da ciò si deduce che $P(X > k) = (1-p)^k$.
- Una $X \sim \text{BinNeg}(h, p)$ è somma di h v.a. i.i.d. distribuite come $\text{Geom}(p)$.
- Una v.a. X sui numeri naturali si dice che ha la *proprietà di perdita di memoria* se $P(X > n+k | X > k) = P(X > n)$. Una v.a. ha la proprietà di perdita della memoria se e solo se è distribuita come la distribuzione geometrica.

⁷ Ovverosia la probabilità che non vi siano mai successi è nulla.

Tabella e proprietà della f.d.r. $\Phi(x)$ di una normale standard

Per $a \in \mathbb{R}$ si definisce la funzione $\Phi(a) = \int_{-\infty}^a e^{-x^2/2} dx$. L'integrale $\Phi(\infty) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} e^{-x^2/2} dx$ vale esattamente 1, mentre $\Phi(-\infty) \stackrel{\text{def}}{=} 0$. Per la parità di $e^{-x^2/2}$ vale che $\Phi(a) = 1 - \Phi(-a)$ (**simmetria**). A partire da questa funzione si può calcolare $\int_a^b e^{-x^2/2} dx$, che risulta essere $\Phi(b) - \Phi(a)$. Se $a > 0$, allora $\int_{-a}^a e^{-x^2/2} dx = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$.

Tabella 2.1: Tabella z di alcuni valori di $\Phi(x)$ per x non negativo. Per x negativo utilizzare **simmetria**. Si prendono le cifre fino al decimo e si legge la riga corrispondente, in base al centesimo si individua poi l'approssimazione da usare.

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,5279	0,53188	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,70884	0,71226	0,71566	0,71904	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,77337	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,87493	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,9608	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,97441	0,975	0,97558	0,97615	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,9803	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,98422	0,98461	0,985	0,98537	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,99461	0,99477	0,99492	0,99506	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,9972	0,99728	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999