Schede riassuntive di
Elementi di Probabilità e Statistica

A cura di Gabriel Antonio Videtta[1]
`g.videtta1@studenti.unipi.it`

Testo basato sul contenuto del corso del prof. Maurelli
e del prof. Trevisan tenutosi presso l'Università di Pisa.

A.A. 2023-2024

Ultimo aggiornamento: 14 luglio 2024

`https://eps.hearot.it`

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

Indice

Notazioni impiegate

Algebra lineare

- q_φ – dato uno spazio vettoriale V equipaggiato con un prodotto scalare φ, q_φ è la forma quadratica associatogli, ovverosia $q_\varphi(v) = \varphi(v, v)$.
- $\|v\|_\varphi$ – dato uno spazio vettoriale reale V equipaggiato con un prodotto scalare (semi)definito positivo φ, $\|\cdot\|_\varphi$ è la (semi)norma indotta da φ, ovverosia $\|v\|_\varphi = \sqrt{q_\varphi(v)} = \sqrt{\varphi(v, v)}$.
- 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 $\varphi(v, w) = 0$.
- $V_\varphi^\perp$ – Radicale del prodotto scalare (o hermitiano) φ sullo spazio V, ovverosia sottospazio dei vettori ortogonali ai vettori di tutto lo spazio.
- $\mathrm{CI}(\varphi)$ – 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 $x \in \mathbb{R}^n$, ovverosia:
$$\|x\|_p = \left(\sum_{i\in[n]} |x_i|^p\right)^{\frac{1}{p}}.$$
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, ovverosia $f^+(a)$ è uguale a $f(a)$ se $f(a) \geq 0$ e 0 altrimenti.
- f^- – parte negativa di una mappa f a valori reali, ovverosia $f^-(a)$ 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.
- C^n, $C^n(\mathbb{R})$ – classe delle funzioni derivabili n volte con n-esima derivata continua. Per $n = 0$, classe di funzioni continue.
- C^∞ – classe delle funzioni derivabili un numero illimitato di volte.
- C_b, $C_b(\mathbb{R})$ – classe delle funzioni reali, continue e limitate.
- $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ – funzione gamma. È tale per cui $\Gamma(n+1) = n!$ per ogni $n \in \mathbb{N}$.
- $f * g$ – convoluzione di funzioni; tale che $(f * g)(z)$ sia pari a $\int_{\mathbb{R}} f(x)g(z-x)\,dx$.

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, ovverosia 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

- $\mathscr{P}(\Omega)$ – insieme delle parti di Ω, ossia insieme dei sottoinsiemi di Ω.
- $f|_A$ – restrizione della funzione al dominio A.
- $A \uplus B$ – 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).
- $A \Delta B = A \setminus B \uplus B \setminus A$ – differenza simmetrica tra A e B.
- $[n]$ – l'insieme $\{1, \ldots, n\}$.
- $\prod_{i \in I} S_i$ con S_i insieme e I ordinato – prodotto cartesiano degli S_i, ordinato secondo I.
- $[[n]]$ – l'insieme $\{0, \ldots, n\} = \{0\} \cup [n]$.
- $\#A$, $|A|$ – numero di elementi di A, o semplicemente la cardinalità di A.
- insieme finito – insieme in bigezione con $[n]$ per qualche $n \in \mathbb{N}$.
- insieme numerabile – insieme in bigezione 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 $\Omega = A \uplus A^c$.
- $X^{-1}(A)$ – controimmagine dell'insieme $A \subseteq C$ in riferimento alla funzione $X : D \to C$, ovverosia $X^{-1}(A) = \{\omega \in D \mid X(\omega) \in A\}$.
- S_X, $\operatorname{im} X$ – immagine della funzione X.
- $\operatorname{supp} X$ – supporto di X, ovverosia sottoinsieme del dominio degli elementi che non annullano X.
- 1_A, I_A – funzione indicatrice di A, ovverosia la funzione $1_A : B \to [[1]] \subseteq \mathbb{R}$ riferita ad $A \subseteq B$ tale per cui:

$$1_A(b) = \begin{cases} 1 & \text{se } b \in A, \\ 0 & \text{altrimenti.} \end{cases}$$

- $1_{\texttt{exp}}$ – 1 se `exp` è vera, 0 altrimenti.
- $\rightrightarrows$ – simbolo utilizzato al posto $\to$ quando si elencano più funzioni che condividono o lo stesso dominio o lo stesso codominio (e.g. $f, g : A, B \rightrightarrows C$ elenca una funzione $f : A \to C$ e una $g : B \to C$; $f, g : A \rightrightarrows B, C$ elenca una funzione $f : A \to B$ e una $g : A \to C$).

Topologia generale

- $\tau(X)$ – dato X spazio metrico, insieme degli aperti di X, ossia topologia di X.
- spazio separabile – spazio topologico contenente un denso, ossia un insieme la cui chiusura è tutto lo spazio (e.g. $\mathbb{Q}$ per $\mathbb{R}$).
- spazio II-numerabile – spazio topologico che ammette una base numerabile.

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 $\tau \subseteq \mathscr{P}(\Omega)$.
- $\sigma\{A_1, \ldots, A_n\}$ – σ-algebra generata dalla famiglia $\tau = \{A_1, \ldots, A_n\} \subseteq \mathscr{P}(\Omega)$.
- $\mathscr{B}(X)$ – σ-algebra dei boreliani, ossia σ-algebra generata dagli aperti di X spazio metrico separabile.
- $\mathscr{F}$ – σ-algebra relativa a Ω, ossia l'insieme dei possibili eventi.
- $(\Omega, \mathscr{F})$ – spazio misurabile.
- π-sistema – insieme $I \subseteq \mathscr{F}$, $I \neq \emptyset$ con $(\Omega, \mathscr{F})$ spazio misurabile, $\sigma(I) = \mathscr{F}$ e I chiuso per intersezioni.
- μ – misura su uno spazio misurabile.
- m – misura di Lebesgue sullo spazio misurabile $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. È tale per cui $m([a,b]) = b - a$ per $b > a$.
- m – misura di Lebesgue sullo spazio misurabile $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ con $d \geq 1$. È tale per cui $m([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdots (b_d - a_d)$ con a_i, $b_i \in \mathbb{R}$ e $b_i > a_i$ per $1 \leq i \leq d$. Non si distingue generalmente la notazione dal caso unidimensionale.
- P – misura di probabilità su uno spazio misurabile.
- $(\Omega, \mathscr{F}, P)$ – spazio di probabilità.

- q.c.– quasi certo/quasi certamente.
- q.o.– quasi ovunque.
- p – per Ω discreto, funzione di densità discreta; per una probabilità discreta P, la densità discreta della probabilità ristretta all'insieme Ω_0 su cui è concentrata P o, con abuso di notazione, la mappa $x \mapsto P(\{x\})$ (che coincide sui termini di Ω_0 con p e che è 0 negli altri punti).
- δ_a – delta di Dirac; dato uno spazio misurabile $(\Omega, \mathscr{F})$ e $a \in \Omega$, probabilità tale per cui $\delta_a(A) = 1$ se $a \in A$ e 0 altrimenti (tale probabilità è concentrata in $\{a\}$ ed è dunque discreta).
- f.d.r. – funzione di ripartizione, rispetto a una probabilità reale.
- F, F_P – per una probabilità reale, funzione di ripartizione.
- f – densità (in senso reale) della probabilità.
- AC – assolutamente continua, riferito a una probabilità.
- 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. $X : \Omega \to S$, $X \in A$ è l'insieme $X^{-1}(A)$. Si estende naturalmente al caso $\notin$.
- $X = a$ – per una v.a. $X : \Omega \to S$, $X = a$ è l'insieme $X^{-1}(a)$. Si estende naturalmente al caso $\neq$.
- $X = Y$ – per due v.a. $X, Y : \Omega \rightrightarrows S$ 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 $X : \Omega \to \mathbb{R}$, $X > a$ è l'insieme $X^{-1}((a, \infty))$; per una v.a. discreta $X : \Omega \to \mathbb{R}$ 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 $\varphi \circ X$.
- $\overset{(d)}{=}$, $\sim$ – per due v.a. $X, Y : \Omega_1, \Omega_2 \rightrightarrows S$ indica l'uguaglianza di legge, ovverosia $P^X_{\Omega_1} = P^Y_{\Omega_2}$.
- 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.
- $(X_1, \ldots, X_n)$ – per una famiglia $(X_i : \Omega \to S_i)_{i \in [n]}$ di v.a. indica la v.a. congiunta (multivariata) $(X_1, \ldots, X_n) : \Omega \to \prod_{i \in [n]} S_i$, $\omega \mapsto (X_1(\omega), \ldots, X_n(\omega))$. Se la famiglia è composta da due variabili, si dice anche *coppia bivariata*.
- $P(A, B) \overset{\text{def}}{=} P(A \cap B)$ – 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.
- $L(A, B) \overset{\text{def}}{=} \frac{P(A|B)}{P(A)}$ – rapporto di influenza tra A e B.
- $\bigotimes_{i \in [n]} P_i = P_1 \otimes \cdots \otimes P_n$ – Date P_i probabilità su S_i discreto, $P_1 \otimes \cdots \otimes P_n \overset{\text{def}}{=} P$ 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.
- $\mathbb{E}[X \mid A] = \overset{\text{def}}{=} \frac{\mathbb{E}[X \cdot 1_A]}{P(A)}$ – valore atteso di X condizionato a A.
- $\mathrm{Cov}(X, Y) \overset{\text{def}}{=} \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ – covarianza di X e Y.
- $\mathrm{Var}(X) \overset{\text{def}}{=} \mathrm{Cov}(X, X)$ – varianza di X.
- $\sigma(X) \overset{\text{def}}{=} \sqrt{\mathrm{Var}(X)}$ – deviazione standard di X.
- $\rho(X, Y)$ – coefficiente di correlazione di Pearson, ovverosia $\cos_{\mathrm{Cov}}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X) \cdot \sigma(Y)}$.
- 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:

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

  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 e teoria della misura

- **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 $a_i \to \sup\{a_i \mid i \in \mathbb{N}\}$ per $i \to \infty$ (e dunque converge se la successione è limitata dall'alto); se $(a_i)_{i\in\mathbb{N}}$ è decrescente, allora $a_i \to \inf\{a_i \mid i \in \mathbb{N}\}$ per $i \to \infty$ (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 $f : \mathbb{R} \supseteq S \to \mathbb{R}$ una funzione convessa a valori reali. Allora vale che:

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

  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:

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

  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:

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

  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:

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

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

- **Una funzione crescente ammette un insieme discreto di discontinuità** – È possibile costruire facilmente una funzione iniettiva da tale insieme a $\mathbb{Q}$ sfruttando i limiti sinistri e destri nelle discontinuità.

- **Lemma di Dynkin, versione probabilistica** – Se due misure di probabilità P e Q su $(\Omega, \mathscr{F})$ coincidono su un π-sistema di $\mathscr{F}$ contenente Ω, allora $P \equiv Q$.

- **Esistenza e unicità della misura di Lebesgue** – Esiste ed è unica la misura m su $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ tale per cui $m([a,b]) = b - a$. Segue dalla versione più generale del lemma di Dynkin.

- **Esistenza e unicità della misura di Lebesgue d-dimensionale** – Esiste ed è unica la misura m su $\left(\mathbb{R}^d, \mathscr{B}\left(\mathbb{R}^d\right)\right)$ tale per cui $m([a_1,b_1] \times \cdots \times [a_d,b_d]) = (b_1 - a_1)\cdots(b_d - a_d)$ con a_i, $b_i \in \mathbb{R}$ e $b_i > a_i$ per $1 \leq i \leq d$.

- **Lemma di Fatou** – Sia $(X, \mathscr{F})$ uno spazio misurabile e sia $(f_i : X \to \mathbb{R})_{i\in\mathbb{N}}$ una successione di funzioni misurabili rispetto a $(\mathbb{R}, m)$ con $f_i \geq 0$ e con $f_i \nearrow f$ puntualmente. Allora $\int_X \liminf_{i\to\infty} f_i \, dm \leq \liminf_{i\to\infty} \int_X f_i \, dm$.

- **Teorema di convergenza monotona, o di Beppo Levi** – Sia $(X, \mathscr{F})$ uno spazio misurabile e sia $(f_i : X \to \mathbb{R})_{i\in\mathbb{N}}$ una successione di funzioni misurabili rispetto a $(\mathbb{R}, m)$ con $f_i \geq 0$ e con $f_i \nearrow f$ puntualmente. Allora f è misurabile e $\int_X f \, dm = \lim_{i\to\infty} \int_X f_i \, dm$.

- **Teorema di convergenza dominata** – Sia $(X, \mathscr{F})$ uno spazio misurabile e sia $(f_i : X \to \mathbb{R})_{i\in\mathbb{N}}$ una successione di funzioni misurabili rispetto a $(\mathbb{R}, m)$. Sia $f : X \to \mathbb{R}$ tale per cui $f_i \to f$ puntualmente. Se esiste una $g : X \to \mathbb{R}$ Lebesgue-integrabile con $g \geq 0$ con $|f_i| \leq g$ per ogni $i \in \mathbb{N}$, allora le f_i e f sono Lebesgue-integrabili e $\lim_{i\to\infty} \int_X f_i \, \mathrm{dm} = \int_X \lim_{i\to\infty} f_i \, \mathrm{dm} = \int_X f \, \mathrm{dm}$.

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:

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

  Inoltre vale che $\left|\bigcup_{i\in[n]} A_i\right| = \sum_{i\in[n]} |A_i|$ 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 $f : [n+1] \to [n]$ 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 $f : [kn+1] \to [n]$ 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 $x_1 + \ldots + x_k = n$ con $x_i \in \mathbb{N}$, esistono esattamente $\binom{n+k-1}{k-1}$ soluzioni. Alternativamente, data la disequazione $x_1 + \ldots + x_k \leq n$ con $x_i \in \mathbb{N}$, esistono esattamente $\binom{n+k}{k}$ soluzioni (dacché ha le stesse soluzioni di $x_1 + \ldots + x_k + y = n$, 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**

  - $\#\{(a_1, \ldots, a_n) \in [k]^n \mid a_1 < a_2 < \ldots < a_n\} = \binom{n}{k}$ se $k \leq n$ – Infatti data una classe di disposizione, esiste un unica lista ordinata in tale classe.
  - $\#\{(a_1, \ldots, a_n) \in [k]^n \mid a_1 \leq a_2 \leq \ldots \leq a_n\} = \binom{n+k-1}{k-1}$. – È sufficiente osservare che si sta contando esattamente le combinazioni con ripetizione in perfetta analogia con la precedente cardinalità.

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 $X : D \to C$ è una funzione e $\mathscr{F} = (A_i)_{i\in I}$ è una famiglia di sottoinsiemi di C, allora vale che $X^{-1}(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} X^{-1}(A_i)$, $X^{-1}(\bigcap_{i\in I} A_i) = \bigcap_{i\in I} X^{-1}(A_i)$, $X^{-1}(A_i^c) = X^{-1}(A_i)^c$, ovverosia 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 $Y : C \to C'$ vale che $(Y \circ X)^{-1}(A) = X^{-1}(Y^{-1}(A))$, per $A \subseteq C'$.

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.
- $\sum_{i=0}^{n} \binom{n}{i} \mathrm{p}^i (1-\mathrm{p})^{n-i} = 1$ per $\mathrm{p} \in [0,1]$ – 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.
- $\sum_{i=0}^{n} i a^i = \frac{a}{(1-a)^2}\left[1-(n+1)a^n + na^{n+1}\right]$ – Segue derivando la somma delle potenze.
- $\sum_{i=0}^{n} i^2 a^i = \frac{a}{(1-a)^3}\left[(1+a)-(n+1)^2a^n + (2n^2+2n-1)a^{n+1} - n^2a^{n+2}\right]$ – 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 $n \to \infty$ 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, spazi e funzioni misurabili

Definizione 1.3 (σ-algebra)**.**
Una σ-algebra $\mathscr{F}$ di Ω è un sottoinsieme $\mathscr{F} \subseteq \mathscr{P}(\Omega)$ tale per cui:

(i.) $\Omega \in \mathscr{F}$,

(ii.) $A \in \mathscr{F} \implies A^c \in \mathscr{F}$,

(iii.) per $(A_i)_{i\in\mathbb{N}}$ famiglia numerabile di insiemi in $\mathscr{F}$, $\bigcup_{i\in\mathbb{N}} A_i \in \mathscr{F}$ ($\mathscr{F}$ è chiuso per unioni numerabili).

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

Definizione 1.4 (Spazio misurabile)**.**
Si definisce **spazio misurabile** una coppia $(\Omega, \mathscr{F})$, dove $\mathscr{F}$ è una σ-algebra di Ω. Gli elementi di $\mathscr{F}$ sono detti **insiemi misurabili** (e nel caso della probabilità, **eventi**).

Definizione 1.5 (Funzione misurabile)**.**
Data una funzione f dallo spazio misurabile $(X, \mathscr{F})$ allo spazio $(Y, \mathscr{S})$ si dice **misurabile** se $f^{-1}(A) \in \mathscr{F}$ per ogni $A \in \mathscr{S}$, ovverosia se la controimmagine di un insieme misurabile è misurabile.

Osservazione 1.6.
Se $\mathscr{G}$ genera $\mathscr{S}$, allora è sufficiente verifica che $f^{-1}(A) \in \mathscr{F}$ per ogni $A \in \mathscr{G}$ affinché f sia misurabile.

Osservazione 1.7.
Un insieme A è misurabile in $(\Omega, \mathscr{F})$ se e solo se 1_A è misurabile rispetto a $\{0, 1\}$ e le sue parti (infatti $1_A^{-1}(1) = A$ e $1_A^{-1}(0) = A^c$).

1.1.3 Insiemi discreti e σ-algebra naturale

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

Definizione 1.8 (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 $\mathscr{P}(\Omega)$.

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

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

Osservazione 1.10.
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 1.11 (Proprietà di $\mathscr{F}$)**.**
Se $\mathscr{F}$ è una σ-algebra di Ω, allora:

(i.) $\emptyset \in \mathscr{F}$,

(ii.) per $(A_i)_{i\in\mathbb{N}}$ *famiglia numerabile di insiemi in* $\mathscr{F}$, $\bigcap_{i\in\mathbb{N}} A_i \in \mathscr{F}$ *($\mathscr{F}$ è chiuso per intersezioni numerabili),*

(iii.) $A \setminus B = A \cap B^c \in \mathscr{F} \impliedby A, B \in \mathscr{F}$.

1.2 Corrispondenze logiche e relazionali tra eventi

Osservazione 1.12 (Corrispondenze affermazioni ed eventi).
Ad alcune affermazioni logiche su A e B eventi di $\mathscr{F}$ 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 $A \setminus B \cupdot B \setminus A = A \Delta B$ (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 1.13 (Probabilità P su $(\Omega, \mathscr{F})$ secondo Kolmogorov)**.**
Dato $(\Omega, \mathscr{F})$ spazio misurabile, una **misura di probabilità** P, detta semplicemente *probabilità*, è una funzione $P : \mathscr{F} \to \mathbb{R}$ tale per cui:

(i.) $P(\Omega) = 1$,

(ii.) $0 \leq P(A) \leq 1$ per ogni $A \in \mathscr{F}$ (ossia P può restringersi su $[0,1]$ al codominio),

(iii.) $P(\bigcupdot_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} P(A_i)$ (σ-additività).

In particolare P è una misura per cui $P(\Omega) = 1$.

Definizione 1.14 (Spazio di probabilità)**.**
Si dice **spazio di probabilità** una tripla $(\Omega, \mathscr{F}, P)$ dove $(\Omega, \mathscr{F})$ è uno spazio misurabile e P è una probabilità su $(\Omega, \mathscr{F})$.

1.3.2 Proprietà della probabilità P

Proposizione 1.15 (Proprietà di P)**.**
Se P è una probabilità su $(\Omega, \mathscr{F})$, allora:

(i.) $P(\emptyset) = 0$,

(ii.) $P(\bigcupdot_{i \in [n]} A_i) = \sum_{i \in [n]} P(A_i)$ *(σ-additività finita),*

(iii.) $P(A) + P(A^c) = 1$,

(iv.) $A \subseteq B \implies P(A) \leq P(B)$ *e* $P(B \setminus A) = P(B) - P(A)$ *(segue da (iii.)),*

(v.) $P(B \setminus A) = P(B) - P(A \cap B)$ *(segue da (iv) considerando che* $B \setminus A = B \setminus (A \cap B)$*),*

(vi.) $P(A \cup B) = P(A \Delta B \cupdot A \cap B) = P(A) + P(B) - P(A \cap B)$ *(segue da (v.)),*

(vii.) $P(\bigcup_{i \in [n]} A_i) = \sum_{j \in [n]} (-1)^{j+1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} P(\bigcap_{k \in [j]} A_{i_k})$ *(segue da (vi.) per induzione, Principio di inclusione-esclusione "probabilistico"),*

(viii.) $P(\bigcup_{i \in \mathbb{N}} A_i) \leq \sum_{i \in \mathbb{N}} P(A_i)$ *(σ-subadditività).*

Osservazione 1.16.
Per Ω finito, la σ-additività finita implica la σ-additività per il Principio della piccionaia.

Proposizione 1.17 (Comportamento di P al limite)**.**
Sia $(A_i)_{i \in \mathbb{N}}$ una famiglia numerabile di eventi in $\mathscr{F}$ sullo spazio di probabilità $(\Omega, \mathscr{F}, P)$. Allora:

(i.) $A_i \nearrow A \implies P(A_i) \nearrow P(A)$,

(ii.) $A_i \searrow A \implies P(A_i) \searrow P(A)$.

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

Definizione 1.18 (Eventi trascurabili e quasi certi)**.**
Sia $A \in \mathscr{F}$. Allora A si dice **trascurabile** se $P(A) = 0$; si dice **quasi certo** se $P(A) = 1$.

Definizione 1.19 (Eventi incompatibili)**.**
Due eventi $A, B \in \mathscr{F}$ si dicono **incompatibili** se $A \cap B = \emptyset$.

Definizione 1.20 (q accade q.c.)**.**
Si dice che una proprietà q **accade quasi certamente (q.c.)** se esiste $A \in \mathscr{F}$ quasi certo che soddisfa q.

Osservazione 1.21.
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 $\mathbb{R}$, ovverosia q accade q.c. se esiste $A \in \mathscr{F}$ trascurabile tale per cui A^c soddisfi q.

1.4 Probabilità condizionata

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

Definizione 1.22 (Probabilità condizionata su B)**.**
Dato $B \in \mathscr{F}$ evento non trascurabile (i.e. $P(B) \neq 0$), la **probabilità condizionata** su B è la misura di probabilità $P(\cdot \mid B)$ sullo stesso spazio misurabile tale per cui:

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

Proposizione 1.23.
$P(\cdot \mid B)$ è una misura di probabilità su $(\Omega, \mathscr{F})$.

Osservazione 1.24.
La probabilità condizionata su Ω coincide con P.

Osservazione 1.25.
In generale $P(A \mid \cdot)$ non è una probabilità, dacché per Ω si ricava che $P(A \mid \Omega) = P(A)$, che potrebbe non essere 1.

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

Lemma 1.26 (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* $j \le n$. *Inoltre vale che:*

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

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

Osservazione 1.27.
Per esempio, la regola della catena per A, B e C si riduce a:

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

Definizione 1.28 (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 $\Omega = \dot\bigcup_{i\in I} B_i$ 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 1.29 (Formula delle probabilità totali, o formula della partizione)**.**
Sia $(B_i)_{i\in I}$ *un sistema di alternative per* Ω. *Allora vale che:*

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

Nella maggior parte dei casi è possibile "invertire" una probabilità condizionata, ovverosia ricavare una probabilità tra $P(A \mid B)$, $P(B \mid A)$, $P(A)$ e $P(B)$ conoscendone tre, a patto che A e B non siano trascurabili, come mostra il:

Teorema 1.30 (di Bayes)**.**
Siano A *e* B *due eventi non trascurabili. Allora vale che:*

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

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

Osservazione 1.31.
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:

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

Osservazione 1.32.
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 cononoscendone 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 1.33 (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:

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

ed è tale per cui:

$$P(A \mid B) = L(A,B)P(A).$$

Proposizione 1.34.
$L(\cdot,\cdot)$ *è simmetrica, ovverosia* $L(A,B) = L(B,A)$ *per ogni evento* A *e* B. *Segue dal Teorema di Bayes.*

Definizione 1.35 (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 $P(A \mid B) \ge P(A)$ (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 $P(A \mid B) \le P(A)$ (ovverosia se $L(A,B) < 1$).

Osservazione 1.36.
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 1.37.
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 $P(A \mid B) \le P(A)$, anche $P(B \mid A) \le P(B)$, 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 1.38 (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 1.39.
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 $\mathscr{F}$ è indipendente se e solo se il "verificarsi" di alcuni eventi della famiglia non influenza il "verificarsi" degli altri.

Osservazione 1.40.
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 1.41.
Se $P(B) > 0$, allora A e B sono indipendenti se e solo se $P(A \mid B) = P(A)$. 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 $P(\bigcap_{j\in J} A_j) \neq 0$ e $P(A \mid \bigcap_{j\in J} A_j) = P(A)$.

Proposizione 1.42.
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 1.43.
Sia $(A_i)_{i\in I}$ una famiglia di eventi indipendenti. Allora, se I è partizionato dagli I_j, ovverosia $I = \dot\bigcup_{j\in J} I_j$, allora $(\bigcap_{i\in I_j} A_i)_{j\in J})$ è ancora una famiglia di eventi indipendenti (ossia intersecando alcuni elementi della famiglia e lasciandone invariati altri, la famiglia ottenuta è ancora indipendente).

Teorema 1.44.
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 1.45.
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 1.46.
Un'evento A è indipendente da ogni evento $B \in \mathscr{F}$, 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 1.47.
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 $\mathscr{P}(\Omega)$.

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 brevemenete di densità) la funzione $p : \Omega \to \mathbb{R}$ tale per cui:

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

Proposizione 2.2 (P è univocamente determinata da p)**.**
Sia $p : \Omega \to \mathbb{R}$ una funzione tale per cui:

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

(ii.) $p(\omega) \geq 0$ *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, ovverosia:

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

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 $P|_{R_P} : \mathscr{P}(R_P)$ 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 $(\Omega, \mathscr{F}, P)$ spazio di probabilità con $\{\omega\} \in \mathscr{F}$ per ogni $\omega \in \Omega$, la probabilità P si dice **discreta** su Ω se esiste $\Omega_0 \in \mathscr{F}$ 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 $(\Omega_0, \mathscr{P}(\Omega_0))$ 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:

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

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 $P(A \cup \Omega_0) = 1$ 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.

Osservazione 2.12.
Se P è una probabilità discreta e, per $a \in \Omega$, δ_a è il **delta di Dirac**, ovverosia la probabilità per cui $\delta_a(A) = 1$ se $a \in A$ e $\delta_a(A) = 0$ se $a \notin A$, allora vale la seguente identità:

$$P = \sum_{\omega \in R_P} p(\omega)\, \delta_\omega,$$

dove si osserva che R_P è numerabile (dacché P è discreta).

2.2 Variabili aleatorie discrete

2.2.1 Definizione di v.a. discreta e composizione

Definizione 2.13 (Variabile aleatoria discreta)**.**
Dato $S \neq \emptyset$, si definisce **variabile aleatoria** (discreta) su Ω discreto, abbreviata v.a. , una funzione $X : \Omega \to S$. 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 $\mathrm{VA}(\Omega, S)$ come l'insieme delle v.a. discrete di Ω che hanno S per codominio.

Osservazione 2.14.
Si può dotare $\mathrm{VA}(\Omega,\mathbb{R})$ 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 $\mathrm{VA}(\Omega,\mathbb{R})$ è naturalmente isomorfo a $\mathbb{R}^{\#\Omega}$ come spazio vettoriale, mentre nel caso di Ω numerabile $\mathrm{VA}(\Omega,\mathbb{R})$ ammette una base non numerabile.

Definizione 2.15 (Composizione di v.a. discrete)**.**
Data $X \in \mathrm{VA}(\Omega, S)$ e una funzione $F : S \to S'$, si definisce la **composizione di** X **tramite** F come $F(X) = F \circ X \in \mathrm{VA}(\Omega, S')$.

2.2.2 Legge di una v.a. X e costruzione canonica

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

Proposizione 2.16.
Sia $X : \Omega \to S$ una v.a. discreta di Ω. Sia P' la misura di probabilità sullo spazio misurabile $(S, \mathscr{P}(S))$ 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 2.17 (Legge di X)**.**
Data una v.a. $X : \Omega \to S$, si definisce **legge di** X (o *distribuzione di* X) la discretizzazione $P^X = P'|_{S_X}$ che agisce sullo spazio misurabile $(S_X, \mathscr{P}(S_X))$, dove P' è tale per cui $P'(A) = P(X \in A) = P(X^{-1}(A))$. Equivalentemente vale che:

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

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 2.18.
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 $X \in A = \bigcup_{x \in A} (X = x)$.

Osservazione 2.19.
Il range di P^X è:

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

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

Osservazione 2.20.
Dato uno spazio di probabilità $(S, \mathscr{P}(S), Q)$ con Ω discreto è sempre possibile trovare uno spazio di probabilità $(\Omega, \mathscr{P}(\Omega), P)$ e una v.a. $X : \Omega \to S$ tale per cui $P^X = Q$.

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

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

Proposizione 2.21.
Data una v.a. $X : \Omega \to S$ e una funzione $f : S \to E$, 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 2.22 (Uguaglianza quasi certa tra v.a.)**.**
Date X, $Y \in \mathrm{VA}(\Omega, S)$, 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 2.23 (Comportamento delle uguaglianze q.c. sulla composizione)**.**
Sia $F : S \to S'$. Siano X, $Y \in \mathrm{VA}(\Omega, S)$. 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 2.24 (Uguaglianza di leggi tra v.a.)**.**
Data $X \in \mathrm{VA}(\Omega_1, S)$ e $Y \in \mathrm{VA}(\Omega_2, S)$, si dice che X **e** Y **hanno la stessa legge**, e si scrive che $X \overset{(d)}{=} Y$ o che $X \sim Y$, se $P^X_{\Omega_1} \equiv P^Y_{\Omega_2}$.

Definizione 2.25 (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 2.26.
Se $X = Y$ q.c., allora $X \overset{(d)}{=} Y$. 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.

[1] Nella 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 2.27.
Per $X, Y \in \mathrm{VA}(\Omega, S)$ v.a. non è generalmente vero che $X \overset{(d)}{=} Y$ implica $X = Y$ q.c.

Proposizione 2.28 (Comportamento delle uguaglianze di legge sulla composizione)**.**
Sia $F : S \to S'$. Siano $X, Y : \Omega_1, \Omega_2 \rightrightarrows S$ v.a. Allora se $X \overset{(d)}{=} Y$, $F(X) \overset{(d)}{=} F(Y)$.

2.2.4 Variabile aleatoria multivariata, leggi congiunte e marginali

Definizione 2.29 (Variabile aleatoria multivariata, o congiunta)**.**
Data una famiglia $(X_i : \Omega \to S_i)_{i\in I}$ 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 $(X_1, \ldots, X_n)$ al posto di $(X_i)_{i\in I}$. Sottintenderemo sempre che I è ordinato quando si nomina una famiglia di v.a. discrete.

Definizione 2.30 (Legge e densità congiunta)**.**
Data una famiglia $(X_i : \Omega \to S_i)_{i\in I}$ 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 2.31 (Leggi e densità marginali)**.**
Data una famiglia $(X_i : \Omega \to S_i)_{i\in I}$ 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 2.32.
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 2.33.
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:

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

Proposizione 2.34.
Ogni densità marginale è univocamente determinata dalla densità congiunta. In particolare nel caso finito vale che:

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

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

Definizione 2.35 (Indipendenza tra v.a. discrete)**.**
Sia $(X_i : \Omega \to S_i)_{i\in I}$ 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:

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

Equivalentemente tale famiglia è indipendente se:

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

Definizione 2.36 (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 2.37.
La definizione è equivalente a richiedere che per ogni scelta di $A_{i_j} \subseteq S_{i_j}$, $X_{i_1} \in A_{i_1}$, ..., $X_{i_n} \in A_{i_n}$ 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 2.38.
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 2.39.
Si osserva che vale la seguente identità:

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

Proposizione 2.40.
Sia $(X_i : \Omega \to S_i)_{i\in I}$ 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:

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

Equivalentemente, sono indipendenti se e solo se:

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

Segue dalla precedente osservazione.

Proposizione 2.41.
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 2.42.
Sia $(X_i : \Omega \to S_i)_{i\in I}$ una famiglia di v.a. discrete e sia $(f_i : S_i \to S'_i)_{i\in I}$ 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 $(f_i(X_i) \in A_i) = (X_i \in f^{-1}(A_i))$.

Proposizione 2.43.
Sia $(X_i : \Omega \to S_i)_{i\in I}$ una famiglia di v.a. discrete e sia I partizionato dagli I_j, ovverosia $I = \bigsqcup_{j\in J} I_j$. 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 2.44.
Le ultime due proposizioni permettono di ricavare molto velocemente l'indipendenza di una certa famiglia di v.a. discrete. Per esempio, se $X_1, X_2, X_3, X_4, X_5 \in \mathrm{VA}(\Omega, \mathbb{R})$ 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:

$$\begin{tikzcd}
X_1 \arrow[r, squiggly] & X_1 \arrow[r, "\mathrm{id}"] & X_1 \\
X_2 \arrow[r, squiggly] & (X_2, X_3) \arrow[r, "+"] & X_2 + X_3 \\
X_3 \arrow[ur, squiggly] & (X_4, X_5) \arrow[r, "\max"] & \max(X_4, X_5) \\
X_4 \arrow[ur, squiggly] & & \\
X_5 \arrow[uur, squiggly] & &
\end{tikzcd}$$

Infatti la prima operazione restituisce una famiglia indipendente per la *Proposizione 2.43*, e la seconda fa lo stesso per la *Proposizione 2.42*.

Osservazione 2.45.
Data una famiglia di probabilità $(P_i)_{i\in[n]}$ su spazi misurabili discreti $(S_i, \mathscr{P}(S_i))$ è sempre possibile costruire uno spazio discreto di probabilità $(\Omega, \mathscr{P}(\Omega), P)$ equipaggiato di una famiglia di v.a. $(X_i : \Omega \to S_i)_{i\in[n]}$ 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:

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

Infatti in tal caso varrebbe che:

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

Tale costruzione si indica come $P \overset{\text{def}}{=} \bigotimes_{i\in[n]} P_i = P_1 \otimes \cdots \otimes P_n$.

2.3 Valore atteso e momenti

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

Definizione 2.46 (Variabile aleatoria integrabile)**.**
Sia X v.a. reale. Si dice che X è **integrabile** (in senso discreto) se:

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

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 2.47 (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:

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

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 $\mathbb{E}[X] \overset{\text{def}}{=} -\mathbb{E}[X^-]$. In questo modo ammettiamo eventualmente i valori di ∞ o $-\infty$.

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

Osservazione 2.48.
Il valore atteso è da associarsi a un "baricentro" della distribuzione di X, ovverosia, su una popolazione Ω, misura quanto vale in media la caratteristica data da X.

Osservazione 2.49.
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 2.50.
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 2.51 (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 2.52 (Valore atteso della composizione tramite la legge)**.**
Sia $X : \Omega \to S$ v.a. discreta e sia $\varphi : S \to \mathbb{R}$. 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 2.53 (Uguaglianza di valori attesi per leggi uguali).
Dal momento che $\mathbb{E}[\varphi(X)]$ dipende soltanto dalla legge di p_X, $X \overset{(d)}{=} Y \implies \mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$.

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

Proposizione 2.54.
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 $E[X] \leq E[Y]$,*

(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 2.55.
Siano $X, Y : \Omega \rightrightarrows S, S'$, due v.a. indipendenti. Se $g, h : S, S' \rightrightarrows \mathbb{R}$ sono funzioni e $g(X)$ e $h(Y)$ ammettono valore atteso[3]*, 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 2.56.
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 2.57.
Dalla *Proposizione 2.54* si deduce che $\mathbb{E}[\cdot]$ è un funzionale di $\mathrm{VA}(\Omega,\mathbb{R})$ (ovverosia $\mathbb{E}[\cdot] \in \mathrm{VA}(\Omega,\mathbb{R})^*$).

Proposizione 2.58.
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 2.59 (Valore atteso condizionale)**.**
Sia X una v.a. reale. Dato allora un evento $A \in \mathscr{P}(\Omega)$, si definisce il **valore atteso condizionale** $\mathbb{E}[X \mid A]$ in modo tale che:

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

Alternativamente vale che:

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

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 2.60 (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:

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

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

2.3.4 Momenti (assoluti) n-esimi

Definizione 2.61 (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 2.62 (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 2.63.
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 $\mathbb{E}[|X|^p \cdot 1_{|X|>1} + |X|^p \cdot 1_{|X|\leq 1}]$; 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 2.64.
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 $\mathbb{E}[|X|^n] \leq M^m$).

Osservazione 2.65.
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 2.66.
Se $\mathbb{E}[|X|^p]$, $\mathbb{E}[|X|^p] < \infty$, 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] Si assume la convenzione per cui $0 \cdot \infty = 0$, $a \cdot \infty = \mathrm{sgn}(a)\infty$ per $a \neq 0$.
[3] Si ammette in questo caso la convenzione per cui $\infty \cdot \infty = \infty$ e che $-\infty \cdot \infty = -\infty$.

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

Proposizione 2.67 (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 2.68.
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 2.69.
Sia X v.a. reale. Allora, se $f : \mathbb{R} \to [0, \infty)$ è crescente, $\forall a \in \operatorname{supp} f$ (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 2.70 (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 2.71 (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 2.72 (Disuguaglianza di Jensen)**.**
Sia X una v.a. reale che ammette valore atteso. Allora, se $g : \mathbb{R} \to \mathbb{R}$ 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 2.73 (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 2.74 (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 2.75.
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 2.76 (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:

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

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

Definizione 2.77 (Varianza)**.**
Data una v.a. reale X con momento secondo finito, si definisce **varianza di** X il termine:

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

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

Proposizione 2.78.
$\mathbb{E}[X]$ è il termine che sostituito a m minimizza il valore $\mathbb{E}[(X - m)^2]$.

Definizione 2.79 (Deviazione standard)**.**
Data una v.a. reale X che ammette varianza, si definisce **deviazione standard di** X il termine:

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

Osservazione 2.80.
La deviazione standard misura quanto X si discosta mediamente da $\mathbb{E}[X]$, se esiste.

Osservazione 2.81.
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 2.82.
Sia X una v.a. reale che ammette varianza. Allora $\mathrm{Var}(X) = 0$ *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., ovverosia se e solo se X è una costante.*

2.5.2 Identità sulla (co)varianza e disuguaglianza di Chebyshev

Proposizione 2.83.
$\mathrm{Cov}(\cdot,\cdot)$ *è una funzione simmetrica e lineare in ogni suo argomento. In particolare per X e Y con momento secondo finito vale che:*

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

Pertanto due v.a. indipendenti hanno covarianza nulla (i.e. sono scorrelate) per l'Osservazione 2.56. In particolare, la covarianza tra una qualsiasi costante q.c. e un'altra v.a. reale è nulla.

Osservazione 2.84.
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 2.85.
A partire dalla precedente proposizione si ricava che per X v.a. reale con momento secondo finito vale che:

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

Osservazione 2.86.
Viste le proprietà discusse nella precedente proposizione si può concludere che la covarianza sul sottospazio di $\mathrm{VA}(\Omega, \mathbb{R})$ delle v.a. con momento secondo finito corrisponde a una forma bilineare simmetrica semidefinita positivo, ovverosia a un prodotto scalare.

Due v.a. indipendenti sono ortogonali tramite Cov per la *Proposizione 2.83*.

Al cono isotropo e al radicale di questo prodotto appartengono solo le costanti per la *Proposizione 2.82*.

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

Lemma 2.87.
Siano X_1, ..., X_n v.a. reali con momento secondo finito. Allora vale che:

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

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

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

Lemma 2.88.
Sia $aX + b$ una v.a. reale con X che ammette momento secondo finito. Allora vale che:

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

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

Proposizione 2.89 (Disuguaglianza di Chebyshev)**.**
Sia X v.a. reale con momento secondo finito. Allora $\forall\, a > 0$ vale che:

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

Segue dall'immediata applicazione della disuguaglianza di Markov.

2.5.3 Coeff. di correlazione e retta di regressione lineare

Definizione 2.90 (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, ovverosia:

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

Lemma 2.91.
Date X, Y v.a. reali non costanti q.c. e con momento secondo finito vale che:

(i.) $\left|\rho(X, Y)\right| \le 1$ *(per la disuguaglianza di Cauchy-Schwarz),*

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

Teorema 2.92.
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:

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

Inoltre il valore di tale minimo è:

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

Definizione 2.93 (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 2.94.
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 $\operatorname{sgn}(a^*) = \operatorname{sgn}(\rho(X,Y))$.

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

2.6.1 Definizioni ed enunciato

Definizione 2.95 (Media campionaria n-esima)**.**
Data una famiglia di v.a. reali $(X_i)_{i\in\mathbb{N}}$ i.i.d. dotate di momento secondo finito[5] si definisce **media campionaria n-esima** il termine:

$$\overline{X_n} \overset{\text{def}}{=} \frac{1}{n}\sum_{i\in[n]} X_i,$$

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

Definizione 2.96 (Limite probabilistico)**.**
Data una successione di v.a. reali $(Y_i : \Omega \to \mathbb{R})_{i\in\mathbb{N}}$ e data una v.a. reale $Y : \Omega \to \mathbb{R}$ si dice che Y_n tende (probabilisticamente) a Y ($Y_n \xrightarrow{\mathbb{P}} Y$) per $n \to \infty$ se:

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

Osservazione 2.97.
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 2.98 (Legge (debole) dei grandi numeri, LGN)**.**
Sia $(X_i)_{i\in\mathbb{N}}$ una famiglia di v.a. reali scorrelate e i.d. (e.g. i.i.d.) dotate di momento secondo finito, ovverosia con $\mathbb{E}[X_1^2] < \infty$. Allora vale che:

$$\overline{X_n} \xrightarrow{\mathbb{P}} \mathbb{E}[X_1], \quad \textit{per } n \to \infty.$$

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

$$P\left(\left|\overline{X_n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \leq \frac{\operatorname{Var}(\overline{X_n})}{\varepsilon^2} = \frac{\operatorname{Var}(X_1)}{\varepsilon^2 n}.$$

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

Osservazione 2.99.
In alcune occasioni, ovverosia quando $\operatorname{Var}(\overline{X_n}) \to 0$ per $n \to \infty$, è ancora possibile applicare la LGN seguendo la stessa dimostrazione.

Osservazione 2.100.
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 $n \to \infty$ vale che:

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

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 $P(\overline{X_n} - m > \varepsilon)$ con $\varepsilon > 0$, $(X_i)_{i\in\mathbb{N}}$ famiglia di v.a. i.i.d. e $m = \mathbb{E}[X_1]$ finito.

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

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

Applicando la disuguaglianza di Markov si ottiene che:

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

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

Definizione 2.101 (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:

$$P(\overline{X_n} - m > \varepsilon) \leq e^{-n\cdot I(\varepsilon)}.$$

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

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

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 $m = \mathbb{E}[X_1]$, $n \to \infty$ 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 $\mathrm{Var}(\overline{X_n})$.

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

$$\mathrm{Var}(n^\alpha(\overline{X_n} - m)) = n^{2\alpha}\mathrm{Var}(\overline{X_n}) = n^{2\alpha-1}\mathrm{Var}(X_1).$$

Pertanto, riapplicando la disuguaglianza di Chebyshev:

$$P\left(n^\alpha\left|\overline{X_n} - m\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}n^{2\alpha-1}\mathrm{Var}(X_1).$$

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 2.102 (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. ($\mathrm{Var}(X_1) > 0$). Sia $\sigma = \sigma(X_1)$ e sia $m = \mathbb{E}[X_1]$. Allora per ogni scelta di a, b tali per cui $-\infty \le a \le b \le \infty$[6] vale che per $n \to \infty$:

$$P\left(a \le \frac{\sqrt{n}}{\sigma}\left(\overline{X_n} - m\right) \le b\right) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

Equivalentemente vale che:

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

⚠ **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 4.1 allegata nelle ultime pagine di queste schede riassuntive.*

Corollario 2.103 (Teorema di De Moivre-Laplace)**.**
Sia $Y_n \sim B(n,\mathrm{p})$. Allora per ogni scelta di a, b tali per cui $-\infty \le a \le b \le \infty$ vale che per $n \to \infty$:

$$P\left(n\mathrm{p} + \sqrt{n\mathrm{p}(1-\mathrm{p})}a \le Y_n \le n\mathrm{p} + \sqrt{n\mathrm{p}(1-\mathrm{p})}b\right) \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

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

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

2.8 Modelli probabilistici classici

2.8.1 Probabilità uniforme

Definizione 2.104 (Probabilità uniforme)**.**
Dato Ω finito, si definisce **probabilità uniforme** l'unica probabilità $P : \mathscr{F} \to \mathbb{R}$ la cui funzione di densità è costante (*equiprobabile*). Equivalentemente è la probabilità P tale per cui:

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

Osservazione 2.105.
Non è possibile dotare Ω numerabile di una probabilità uniforme. Infatti, se l'unica immagine della funzione $p : \Omega \to \mathbb{R}$ 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:

$$\Omega = \left\{(\omega_1, \omega_2, \ldots) \,\middle|\, \omega_1 \in \Omega_1, \omega_2 \in \Omega_2^{(\omega_1)}, \omega_3 \in \Omega_3^{(\omega_1,\omega_2)}, \ldots\right\},$$

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 ω_1, …, ω_{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 $p_i : [[1]] \to \mathbb{R}$ la funzione di densità associata alla misura di probabilità dell'esperimento Ω_i. Associamo allora ad Ω la σ-algebra $\mathscr{F} = \sigma(A_i)_{i\in I}$ generata dagli A_i (che è al più numerabile). Se I è finito, $\mathscr{F} = \mathscr{P}(\Omega)$.

Definizione 2.106 (Modello della sequenza di prove)**.**
Si definisce **probabilità del modello della sequenza di prove** l'unica probabilità P sullo spazio misurabile $(\Omega, \mathscr{F})$ 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 2.107.
Tale probabilità è univocamente determinata dal momento che gli A_i generano $\mathscr{F}$ e che sono indipendenti.

Definizione 2.108 (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 $\mathrm{p} \overset{\text{def}}{=} p_1(1)$ è il **parametro di Bernoulli**.

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

Parte 3

Probabilità sulla retta reale

Discutiamo in questa sezione la teoria della probabilità sulla retta reale, uscendo dunque dal caso discreto.

Per restringere la σ-algebra su cui lavoreremo (ossia l'insieme degli eventi interessanti), siamo costretti a limitarci a una σ-algebra molto più piccola di $\mathscr{P}(\mathbb{R})$, la σ-algebra dei boreliani, che ci permette di escludere "casi meno interessanti".

⚠ **Attenzione.** *Eccetto che nella prima sezione, assumeremo se non detto altrimenti di star lavorando sullo spazio misurabile* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *dotato eventualmente della misura di Lebesgue m. $\mathscr{B}(\mathbb{R})$ ed m sono definiti nella sezione seguente.*

3.1 Cenni di teoria della misura

3.1.1 La σ-algebra di Borel e funzioni boreliane

Definizione 3.1 (σ-algebra dei boreliani)**.**
Dato uno spazio metrico separabile[1] $X \neq \emptyset$ si definisce la **σ-algebra $\mathscr{B}(X)$ dei boreliani di** X (o σ-algebra di Borel) come la σ-algebra generata dai suoi aperti, ovverosia:

$$\mathscr{B}(X) \stackrel{\text{def}}{=} \sigma\{A \subseteq X \mid A \text{ aperto}\}.$$

Gli elementi della σ-algebra di Borel sono detti *boreliani*.

Proposizione 3.2 (Proprietà di $\mathscr{B}(X)$)**.**
Sia $X \neq \emptyset$ uno spazio metrico separabile. Allora valgono le seguenti affermazioni:

(i.) *$\mathscr{B}(X)$ contiene tutti gli aperti e i chiusi di X (infatti metrico e separabile implica II-numerabile), pertanto se $\tau(X)$ è la topologia di X vale che $\tau(X) \subseteq \mathscr{B}(X)$,*

(ii.) *$\mathscr{B}(X) = \sigma\{A \subseteq X \mid A \text{ chiuso}\}$, ossia $\mathscr{B}(X)$ è generata anche dai chiusi di X (infatti $\mathscr{B}(X)$ è chiuso per complementare),*

(iii.) *se $Y \subseteq X$, $Y \neq \emptyset$ ha metrica indotta da X, allora $\mathscr{B}(Y) = \sigma\{Y \cap B \mid B \in \mathscr{B}(X)\} \subseteq \mathscr{B}(X)$ (segue dal fatto che gli aperti di Y sono tutti e solo gli aperti di X intersecati a Y).*

Proposizione 3.3 (Proprietà di $\mathscr{B}(\mathbb{R}^d)$)**.**
Valgono le seguenti affermazioni:

(i.) *$\mathscr{B}(\mathbb{R})$ contiene tutti gli intervalli e tutte le semirette (infatti si ammettono anche intersezioni infinite di aperti),*

(ii.) *$\mathscr{B}(\mathbb{R})$ è generato dagli intervalli semiaperti, ovverosia $\mathscr{B}(\mathbb{R}) = \sigma\{(a, b] \mid a, b \in \mathbb{R}, a < b\}$,*

(iii.) *$\mathscr{B}(\mathbb{R})$ è generato dalle semirette, ovverosia $\mathscr{B}(\mathbb{R}) = \sigma\{(-\infty, a) \mid a \in \mathbb{R}\}$,*

(iv.) *$\mathscr{B}(\mathbb{R}^d) = \sigma\{(-\infty, a_1) \times \ldots \times (-\infty, a_n) \mid a_1, \ldots, a_n \in \mathbb{R}\}$ (segue da (iii.)),*

(v.) *$\mathscr{B}(\mathbb{R}^d) \neq \mathscr{P}(\mathbb{R}^d)$ (segue dal controesempio di Vitali, oltre che da considerazioni sulle cardinalità).*

Definizione 3.4 (Funzione boreliana)**.**
Data una funzione $f : X \to Y$ con X e Y spazi metrici separabili, si dice che f è una **funzione boreliana** se $f^{-1}(A)$ è boreliano per ogni A boreliano di Y. Equivalentemente f è boreliana se la controimmagine di ogni boreliano è un boreliano.

In particolare una funzione è boreliana se e solo se è misurabile rispetto alle σ-algebre di Borel.

Proposizione 3.5.
Sia $f : X \to Y$ con X e Y spazi metrici separabili una funzione continua. Allora f è boreliana.

Segue dal fatto che $\mathscr{B}(Y)$ è generato dagli aperti di Y, le cui controimmagini sono aperte, e dunque boreliane.

3.1.2 Definizione e proprietà di misura, π-sistemi per σ-algebre e lemma di Dynkin

Definizione 3.6 (Misura)**.**
Dato $(\Omega, \mathscr{F})$ spazio misurabile, una misura μ su $(\Omega, \mathscr{F})$ è una funzione $\mu : \mathscr{F} \to [0, \infty]$ con $\mu(\emptyset) = 0$ e per cui valga la σ-additività, ovverosia:

$$\mu\left(\dot{\bigcup_{i \in \mathbb{N}}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i), \quad A_i \in \mathscr{F}.$$

Osservazione 3.7 (Proprietà basilari di una misura)**.**
Dal momento che si richiede per una misura valga $\mu(\emptyset) = 0$, si verifica facilmente che vale la σ-additività finita.

Inoltre, se $A \subseteq B$, allora $\mu(B) = \mu(B \setminus A \cup A) = \mu(B \setminus A) + \mu(A)$, e dunque vale sempre che $\mu(A) \leq \mu(B)$. Vale inoltre ancora la σ-subadditività, con la stessa dimostrazione data per la

[1] Si può generalizzare in modo naturale tale definizione a un qualsiasi spazio topologico. Dal momento che considereremo solo spazi metrici separabili (in particolare $X \subseteq \mathbb{R}^d$), concentreremo le proprietà e le definizioni su questa classe di spazi topologici.

probabilità, e dunque:

$$\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i\in\mathbb{N}} \mu(A_i).$$

Osservazione 3.8 (Comportamento di μ al limite).
Se $(A_i)_{i\in\mathbb{N}}$ è una famiglia numerabile di insiemi in $\mathscr{F}$, allora, seguendo la stessa dimostrazione data per le misure di probabilità, che:

(i.) $A_i \nearrow A \implies \mu(A_i) \nearrow \mu(A)$,

(ii.) $A_i \searrow A \implies \mu(A_i) \searrow \mu(A)$.

Definizione 3.9.
Una misura μ su $(\Omega, \mathscr{F})$ si dice **misura finita** se $\mu(\Omega)$ è finito.

Proposizione 3.10 (Proprietà di una misura finita μ)**.**
Sia μ una misura finita su $(\Omega, \mathscr{F})$. Allora valgono le seguenti affermazioni:

(i.) $P(A) = \frac{\mu(A)}{\mu(\Omega)}$ *è una misura di probabilità,*

(ii.) $\mu(A)$ *è sempre finito e* $\mu(\Omega) = \mu(A) + \mu(A^c)$,

(iii.) $A \subseteq B \implies \mu(B) = \mu(B \setminus A) + \mu(A)$,

(iv.) $\mu(B \setminus A) = \mu(B) - \mu(A \cap B)$,

(v.) $\mu(A \cup B) = \mu(A \Delta B \uplus A \cap B) = \mu(A) + \mu(B) - \mu(A \cap B)$,

(vi.) $\mu\left(\bigcup_{i\in[n]} A_i\right) = \sum_{j\in[n]} (-1)^{j+1} \sum_{1\leq i_1<\cdots<i_j\leq n} \mu\left(\bigcap_{k\in[j]} A_{i_k}\right)$ *(Principio di inclusione-esclusione per le misure finite).*

Tutte le affermazioni seguono immediatamente dalla prima.

Definizione 3.11 (Insiemi μ-trascurabili e proprietà che accadono μ-quasi ovunque)**.**
Un insieme $A \in \mathscr{F}$ si dice **μ-trascurabile** se $\mu(A) = 0$. Una proprietà M si dice che accade μ-quasi ovunque (μ-q.o.) se esiste $A \in \mathscr{F}$ μ-trascurabile per cui M accade per A^c.

Definizione 3.12 (π-sistema di una σ-algebra)**.**
Sia $(\Omega, \mathscr{F})$ uno spazio misurabile. Allora un sottoinsieme $\mathscr{C} \subseteq \mathscr{F}$ si dice **π-sistema di $\mathscr{F}$** se:

(i.) $A, B \in \mathscr{C} \implies A \cap B \in \mathscr{C}$ (chiusura per intersezioni),

(ii.) $\sigma(C) = \mathscr{F}$ (genera $\mathscr{F}$).

Osservazione 3.13.
Un π-sistema contenente di una σ-algebra contenente Ω svolge lo "stesso ruolo" che una base svolge per una topologia.

Lemma 3.14 (di Dynkin, versione probabilistica)**.**
Sia $(\Omega, \mathscr{F})$ uno spazio misurabile e sia $\mathscr{C}$ un suo π-sistema contenente Ω. Siano P e Q due probabilità sullo spazio misurabile di Ω. Se P e Q coincidono su $\mathscr{C}$, allora $P \equiv Q$.

Esempio 3.15.
Alcuni esempi di π-sistemi contenenti $\mathbb{R}$ per $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ sono:

- gli aperti, ovverosia $\mathscr{C} = \{A \in \mathscr{F} \mid A \text{ aperto}\}$ (oppure i chiusi),
- le semirette (a sinistra), ovverosia $\mathscr{C} = \{(-\infty, a] \mid a \in \mathbb{R} \cup \{\infty\}\}$ (oppure le semirette a destra),
- gli intervalli semiaperti (a sinistra) con aggiunti $\emptyset$ e $\mathbb{R}$, ovverosia $\mathscr{C} = \{(a, b] \mid a, b \in \mathbb{R}, b > a\} \cup \{\emptyset, \mathbb{R}\}$ (oppure semiaperti a destra).

3.1.3 La misura di Lebesgue

Teorema 3.16 (Esistenza e unicità della misura di Lebesgue)**.**
*Esiste ed è unica la misura m su $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ tale per cui $m([a, b]) = b - a$ per ogni a, $b \in \mathbb{R}$ con $b > a$. Tale misura è detta **misura di Lebesgue** e corrisponde al concetto "primitivo" di lunghezza.*

L'unicità segue dall'enunciato generale del lemma di Dynkin.

Osservazione 3.17.
Dal momento che $m([0, 1]) = 1$, la misura $m|_{[0,1]}$ è una misura di probabilità su $([0, 1], \mathscr{B}([0, 1]))$, detta *probabilità uniforme su* $[0, 1]$. Analogamente per a, $b \in \mathbb{R}$ con $b > a$, $m([a, b]) = b - a$ e dunque $P = \frac{1}{b-a} m|_{[a,b]}$ è una misura di probabilità (detta *probabilità uniforme su* $[a, b]$).

Assumendo l'assioma della scelta si può dimostrare che non si può estendere in modo coerente $m|_{[0,1]}$ a $([0, 1], \mathscr{P}([0, 1]))$.

Osservazione 3.18.
La misura di Lebesgue è nulla su un punto di $\mathbb{R}$. Infatti, se $a \in \mathbb{R}$, $m(\{a\}) = m([a-1, a] \cap [a, a+1]) = m([a-1, a]) + m([a, a+1]) - m([a-1, a] \cup [a, a+1]) = 1 + 1 - m([a-1, a+1]) = 1 + 1 - 2 = 0$. Dunque, m è in particolare nulla su insiemi numerabili (dacché si partizionano in modo numerabili sui punti).

3.2 Probabilità reale, funzione di ripartizione (f.d.r.) e proprietà

3.2.1 Definizioni e proprietà della f.d.r.

Definizione 3.19.
Si dice **probabilità reale** una qualsiasi probabilità P su $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Definizione 3.20 (Funzione di ripartizione di P)**.**
Data una probabilità reale P si definisce allora la sua **funzione di ripartizione (f.d.r.)** come la funzione $F : \mathbb{R} \to [0, 1]$ tale per cui:

$$F(x) = P((-\infty, x]), \quad \forall x \in \mathbb{R}.$$

Si definisce inoltre $F(\pm\infty) \stackrel{\text{def}}{=} \lim_{x\to\pm\infty} F(x)$. Indicheremo F come F_P, e quando P sarà nota dal contesto ci limiteremo a scrivere F.

Proposizione 3.21 (Proprietà della f.d.r.)**.**
Sia P una probabilità reale. Allora, se F è la sua f.d.r. vale che:

(i.) *F è crescente, ovvero* $F(x) \geq F(y) \impliedby x \geq y$ *(infatti* $(-\infty, x] \supseteq (-\infty, y]$*),*

(ii.) *F è continua a destra, ovverosia per ogni $\tilde{x} \in \mathbb{R}$ vale che* $\lim_{x\to\tilde{x}^+} F(x) = F(\tilde{x})$,

(iii.) $F(-\infty) = 0 \impliedby ((-\infty, -i])_{i\in\mathbb{N}} \searrow \emptyset$,

(iv.) $F(\infty) = 1 \impliedby ((-\infty, i])_{i\in\mathbb{N}} \nearrow \mathbb{R}$.

L'affermazione (ii.) segue dal fatto che per ogni successione decrecente da destra $(x_i)_{i\in\mathbb{N}} \searrow \tilde{x}$, che esclude $\tilde{x}$, è tale per cui $((-\infty, x_i])_{i\in\mathbb{N}} \searrow (-\infty, \tilde{x}]$, e dunque $(F(x_i))_{i\in\mathbb{N}} = (P((-\infty, x_i]))_{i\in\mathbb{N}} \searrow P((-\infty, \tilde{x}]) = F(\tilde{x})$.

3.2.2 Corrispondenza tra f.d.r. e probabilità, calcolo di P tramite F e probabilità continue

Proposizione 3.22 (P è univocamente determinata da F)**.**
Sia $F : \mathbb{R} \to \mathbb{R}$ una funzione tale per cui:

(i.) *F è crescente,*

(ii.) *F è continua a destra,*

(iii.) $\lim_{x\to\infty} F(x) = 1$,

(iv.) $\lim_{x\to-\infty} F(x) = 0$.

Allora $0 \le F \le 1$ ed esiste un'unica probabilità reale P avente F come funzione di ripartizione.

L'unicità segue dal lemma di Dynkin. Per l'esistenza è utile considerare che l'insieme delle discontinuità di F è discreto dacché F è crescente.

Osservazione 3.23.
La continuità a sinistra della f.d.r. non è invece garantita dacché per ogni successione da sinistra crescente $(x_i)_{i\in\mathbb{N}} \nearrow \tilde{x}$, che esclude $\tilde{x}$, vale che $((-\infty, x_i])_{i\in\mathbb{N}} \nearrow (-\infty, \tilde{x})$, e non $(-\infty, \tilde{x}]$. Dunque $\lim_{x\to\tilde{x}^-} F(x)$ esiste ed è $P((-\infty, x))$, indicato comunemente come $F(x^-)$, che può non coincidere con $F(x)$.

Dal momento che:

$$P(\{x\}) = P((-\infty, x] \setminus (-\infty, x)) = F(x) - F(x^-),$$

si deduce che F è continua se e solo se $P(\{x\}) = 0$ (ossia se e solo se $F(x) = F(x^-)$).

Osservazione 3.24.
Si deducono immediatamente dalla precedente osservazione le seguenti identità:

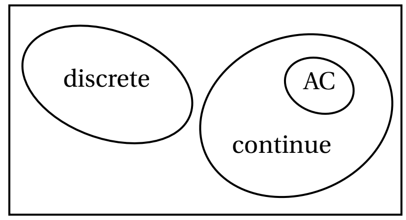
- $P([a, b]) = F(b) - F(a^-)$,
- $P((a, b)) = F(b^-) - F(a)$,
- $P([a, b)) = F(b^-) - F(a^-)$,
- $P((a, b]) = F(b) - F(a)$.

Definizione 3.25 (P continua)**.**
Si dice che una probabilità reale P è **continua** se la sua f.d.r. F lo è, ossia se $P(\{a\}) = 0$ per ogni $a \in \mathbb{R}$ (quest'ultima equivalenza deriva dalla penultima osservazione).

Osservazione 3.26.
Per una probabilità P continua la misura di un intervallo con estremi a e b è semplificata a $F(b) - F(a)$ in tutti i casi (infatti $F(a^-) = F(a)$ e $F(b^-) = F(b)$).

3.3 Classi principali di probabilità reale

Esistono due classi importanti, ma non esaustive, di probabilità reale: le probabilità discrete e quelle assolutamente continue, contenute tra quelle continue. Le classi di probabilità reali si dividono dunque secondo il seguente schema:



Esempio 3.27 (Esempio di probabilità né discreta né continua)**.**
Sia $F : \mathbb{R} \to \mathbb{R}$ tale per cui:

$$F(x) = \begin{cases} 0 & \text{se } x < 0, \\ x + \frac{1}{2} & \text{se } 0 \le x \le \frac{1}{2}, \\ 1 & \text{altrimenti.} \end{cases}$$

Allora F è crescente, continua a destra e tale per cui $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. Pertanto esiste un'unica probabilità P avente F come f.d.r. per la *Proposizione 3.22*.

Dal momento che F non è continua a sinistra in 0, F non è continua, e dunque P non è continua. Inoltre F non induce una probabilità discreta dacché non è costante a tratti in $[0, 1/2]$. Pertanto P non è né continua né discreta.

3.4 Probabilità discreta e rappresentazione della f.d.r.

Come già discusso nella sezione della *Discretizzazione*, una probabilità reale P si dice *discreta* se esiste $\Omega_0 \subseteq \mathbb{R}$ discreto per cui P è concentrata su Ω_0. In tal caso, come già visto, $P(A) = P(A \cap \Omega_0)$ per ogni $A \in \mathscr{B}(\mathbb{R})$, e dunque P è univocamente determinata dalla densità discreta di $P|_{\mathscr{P}(\Omega_0)}$, che chiameremo p.

In questo caso il range R_P è dunque numerabile e, se F è la f.d.r. di P, vale che:

$$F(x) = P((-\infty, x]) = \sum_{\substack{y\in R_P \\ y\le x}} p(y).$$

Osservazione 3.28.
Se P è discreta, come già osservato nella sezione della *Discretizzazione*, allora vale che:

$$P = \sum_{x\in R_P} p(x)\,\delta_x.$$

Osservazione 3.29.
Se R_P non ha punti di accumulazione, allora F è costante a tratti con salti negli $y \in R_P$ di ampiezza $p(y)$.

Al contrario, presa una successione $(p_r)_{r\in\mathbb{Q}}$ con $\sum_{r\in\mathbb{Q}} p_r = 1$, la probabilità $P = \sum_{r\in\mathbb{Q}} p_r\,\delta_r$ è una probabilità discreta con f.d.r. non costante a tratti (infatti tutti i punti di $\mathbb{Q}$ sono punti di accumulazione).

Pertanto, se una probabilità reale è discreta, ci si può effettivamente restringere a tutti i risultati della *Parte 2*.

3.5 Probabilità assolutamente continue (AC)

⚠ **Attenzione.** *Si ricorda che con il simbolo $\int$ si intende l'integrale secondo Lebesgue e che si assume di star lavorando sempre con la misura di Lebesgue m.*

3.5.1 Probabilità AC e funzione di densità

Definizione 3.30 (Probabilità assolutamente continua (AC) e densità)**.**
Una probabilità P si dice **assolutamente continua (AC)** se esiste una funzione boreliana $f : \mathbb{R} \to \mathbb{R}$ tale per cui:

$$P(A) = \int_A f(x)\,dx,$$

dove si impiega l'integrale secondo Lebesgue. Tale funzione f è detta **densità** di P.
Si assume implicitamente che $\int_{\mathbb{R}} |f(x)|\,dx$ sia finito.

Osservazione 3.31.
Se P è AC, allora la sua f.d.r. F è in particolare assolutamente continua, e dunque anche continua.

Osservazione 3.32.
Nella pratica l'integrale $\int_A f(x)\,dx$ si riduce in molti casi al più semplice integrale di Riemann, eventualmente improprio.

3.5.2 Proprietà e caratterizzazione della densità

Proposizione 3.33 (Unicità della densità a meno di m-trascurabilità)**.**
Se P è una probabilità AC con densità f e g, allora $f = g$ q.o. (e dunque $m(f \neq g) = 0$, ossia l'insieme $f \neq g$ è m-trascurabile).

Osservazione 3.34.
Si osserva che se P è una probabilità AC con densità f, allora $f \geq 0$ q.o. per continuità (altrimenti P potrebbe assumere valori negativi) e $\int_{\mathbb{R}} f(x)\,dx = P(\mathbb{R}) = 1$.

Proposizione 3.35 (La densità determina univocamente la probabilità)**.**
Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione boreliana tale per cui:

(i.) $f \geq 0$,

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

Allora esiste un'unica probabilità reale P avente f come densità.

Proposizione 3.36 (Relazioni tra la densità e la f.d.r.)**.**
Sia P una probabilità reale con f.d.r. F. Allora valgono le seguenti affermazioni:

(i.) Se P è AC con densità f, allora $F(x) = \int_{-\infty}^{x} f(y)\,dy$. Viceversa se esiste f per cui $F(x) = \int_{-\infty}^{x} f(y)\,dy$, allora P è AC con densità f.

(ii.) Se F è continua e C^1 a tratti (ovverosia si restringe a una funzione C^1 eccetto che per un insieme di punti isolati), allora P è AC con densità f t.c. $f = F'$ dove è definibile F' e $f = 0$ altrimenti (segue dal Teorema fondamentale del calcolo integrale).

Osservazione 3.37.
Se P è AC con densità f, allora $P(f = 0) = \int_{f=0} f(x)\,dx = 0$ e dunque l'insieme $f = 0$ è trascurabile rispetto a P. Dunque, ristringendo il range si ottiene che:

$$P(A) = P(A \cap (f > 0)).$$

3.6 Variabili aleatorie in generale

In questa sezione cerchiamo di generalizzare il concetto di variabile aleatoria al caso più generale usando il linguaggio delle σ-algebre e delle funzioni misurabili in modo tale da estendere coerentemente le v.a. discrete.

3.6.1 Definizione e legge di una v.a.

Definizione 3.38 (Variabile aleatoria)**.**
Sia $(\Omega, \mathscr{F}, P)$ uno spazio di probabilità. Allora, se $(S, \mathscr{S})$ è uno spazio misurabile e X è una funzione misurabile da Ω a S, allora si dice che X è una **variabile aleatoria (v.a.)**. Se $S = \mathbb{R}$ e $\mathscr{S} = \mathscr{B}(\mathbb{R})$, si dice che X è una *v.a. reale*, se $S = \mathbb{R}^d$ e $\mathscr{S} = \mathscr{B}(\mathbb{R}^d)$, si dice che X è una *v.a. vettoriale* o *vettore aleatorio*.

Osservazione 3.39.
Se Ω è discreto, $\mathscr{F} = \mathscr{P}(\Omega)$, e dunque ogni funzione $X : \Omega \to S$ è una variabile aleatoria. Pertanto la definizione espressa è una perfetta estensione del concetto di v.a. discreta.

Definizione 3.40 (Legge di X)**.**
Sia $(\Omega, \mathscr{F}, P)$ uno spazio di probabilità e sia $X : (\Omega, \mathscr{F}) \to (S, \mathscr{S})$ una v.a. Allora si dice **legge di** X (o *distribuzione*) la probabilità su P^X su $(S, \mathscr{S})$ tale per cui:

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

3.6.2 F.d.r. di una v.a. reale, v.a. discrete, continue e AC

Definizione 3.41 (F.d.r. di una v.a. reale)**.**
Si definisce la **funzione di ripartizione (f.d.r.) di una v.a.** X come la f.d.r. F^X associata alla probabilità reale P^X, ovverosia:

$$F^X(x) = P(X^{-1}((-\infty, x])) = P(X \leq x).$$

Definizione 3.42 (V.a. discreta)**.**
Una v.a. $X : \Omega \to S$ si dice **discreta** se la probabilità P^X è discreta con densità $p_X : S \ni x \mapsto P(X = x)$.
Tale definizione coincide con l'analoga definizione di v.a. discreta data precedentemente se ci restringiamo a Ω discreto. Non è detto in generale che X sia una v.a. discreta se e solo se Ω è discreto.

Esempio 3.43.
Sia P una probabilità reale. Allora $X : \mathbb{R} \to [1]$ che associa a tutti i reali il numero 1 è una funzione misurabile. Inoltre X è discreta dacché $[1]$ è discreto, ma $\mathbb{R}$ non lo è.

Definizione 3.44 (V.a. continue e AC)**.**
Una v.a. reale X si dice **continua** se P^X è continua. Analogamente X si dice **assolutamente continua (AC)** se P^X è AC.

3.6.3 Composizione di v.a.

Definizione 3.45.
Sia $X : (\Omega, \mathscr{F}) \to (S, \mathscr{S})$ una v.a. Allora, se $\varphi : (S, \mathscr{S}) \to (S', \mathscr{S}')$ è una funzione tale per cui $\varphi \circ X$ sia misurabile, si definisce la v.a. composta $\varphi(X) = \varphi \circ X$.

Osservazione 3.46.
Se φ è anch'essa misurabile, allora $\varphi \circ X$ è sicuramente misurabile, e dunque $\varphi(X)$ è una v.a.

Osservazione 3.47.
Se X è discreta, anche $\varphi(X)$ lo è, con range $\varphi(R_X)$. Non è detto che se X è continua (o AC), $\varphi(X)$ sia continua (o AC).

3.6.4 Costruzione canonica, uguaglianza q.c. e in legge

I concetti espressi nel titolo di questa sottosezione si estendono in modo del tutto naturale dal caso discreto, e pertanto si rimanda alla *Parte 2*.

3.7 Valore atteso come integrale secondo la misura P

Cerchiamo in questa sottosezione di dare una definizione di valore atteso che estende la particolare nozione di valore atteso discreto in modo del tutto coerente. Successivamente tutte le disuguaglianze e tutti i risultati espressi nella sezione riguardante il caso discreto saranno tutti validi seguendo le stesse dimostrazioni o idee di dimostrazione.

3.7.1 Costruzione dell'integrale secondo la misura P

Questa sezione tornerà familiare per i lettori che avranno già costruito l'integrale secondo Lebesgue (ovverosia l'integrale secondo la misura m). Infatti le stesse definizioni e le stesse proposizioni si estendono all'integrale secondo una misura generica μ (a patto che μ assuma valori finiti per le funzioni semplici).

Definizione 3.48 (Funzione semplice)**.**
Data una v.a. reale X dello spazio misurabile $(\Omega, \mathscr{F})$, si dice che X è una **funzione semplice** (o *v.a. semplice*) se X assume un numero finito di valori, ovverosia se esistono A_1, ..., $A_n \in \mathscr{F}$ e a_1, ..., $a_n \in \mathbb{R}$ tali per cui:

$$X = \sum_{i \in [n]} a_i 1_{A_i}.$$

Osservazione 3.49.
Si verifica alquanto agevolmente che si può ridefinire la semplicità di X richiedendo che gli A_i siano disgiunti (per esempio, se i b_i rappresentano i valori finiti e distinti assunti da X, gli insiemi $X = b_i$ sono dei possibili candidati).

Proposizione 3.50.
Data X una v.a. semplice sullo spazio di probabilità $(\Omega, \mathscr{F}, P)$, allora, per ogni scrittura $\sum_{i \in [n]} a_i 1_{A_i}$ di X, il valore $\sum_{i \in [n]} a_i P(A_i)$ è lo stesso (ossia non dipende dagli a_i e dagli A_i).

Segue dalle proprietà della σ-algebre e delle misure.

Definizione 3.51 (Integrale secondo la misura P di X v.a. semplice)**.**
Data X una v.a. semplice sullo spazio di probabilità $(\Omega, \mathscr{F}, P)$, allora si definisce l'**integrale di X su Ω secondo la misura P** $\int_\Omega X\, dP$ come il valore $\sum_{i \in [n]} a_i P(A_i)$, dove $\sum_{i \in [n]} a_i 1_{A_i}$ è una scrittura di X.

Lemma 3.52.
Data X v.a. reale con $X \geq 0$, allora esiste una successione $(X_i)_{i \in \mathbb{N}}$ di v.a. semplici con $X_i \geq 0$ tale per cui $X_i \nearrow X$ puntualmente (ovverosia $X_i(\omega) \nearrow X(\omega)$ per ogni $\omega \in \Omega$).

Proposizione 3.53.
Data X una v.a. reale con $X \geq 0$ e data una successione $(X_i)_{i \in \mathbb{N}}$ di v.a. semplici con $X_i \geq 0$ tale per cui $X_i \nearrow X$ puntualmente (ovverosia $X_i(\omega) \nearrow X(\omega)$ per ogni $\omega \in \Omega$), allora il valore $\lim_{i \to \infty} \int_\Omega X_i\, dP$ esiste, è finito non negativo o infinito e non dipende dalla successione $(X_i)_{i \in \mathbb{N}}$.

Definizione 3.54 (Integrale su Ω secondo la misura P di $X \geq 0$)**.**
Data X v.a. reale con $X \geq 0$, si definisce l'**integrale di X su Ω secondo la misura P**, $\int_\Omega X\, dP$, il valore $\lim_{i \to \infty} \int_\Omega X_i\, dP$ come ottenuto dal lemma e la proposizione precedente.

Definizione 3.55 (V.a. integrabili e integrale in generale)**.**
Una v.a. X si dice **integrabile (secondo P)** se $\int_\Omega |X|\, dP$ è finito. In tal caso si definisce l'**integrale di X su Ω secondo la misura P** come:

$$\int_\Omega X\, dP = \int_\Omega X^+\, dP - \int_\Omega X^-\, dP,$$

dove X^+ e X^- sono la parte positiva e negativa di X ed entrambi gli addendi del secondo membro sono finiti dacché $\int_\Omega |X|\, dP$ lo è (infatti $|X| = X^+ + X^-$).

Definizione 3.56 (Integrale su un sottoinsieme $A \subseteq \Omega$)**.**
Data X v.a. reale, si definisce l'integrale $\int_A X\, dP$ come il valore $\int_\Omega 1_A \cdot X\, dP$, qualora definito.

Osservazione 3.57.
Si osserva che $\int_A 1\, dP = \int_\Omega 1_A\, dP = P(A)$, ossia $\int_\Omega$ misura in questo caso l'insieme A secondo P.

3.7.2 Definizione di valore atteso e teoremi correlati

Definizione 3.58 (Valore atteso come integrale secondo la misura P)**.**
Sia X una v.a. integrabile o tale per cui $X \geq 0$. Allora si definisce il **valore atteso di X** $\mathbb{E}[X]$ come il valore $\int_\Omega X\, dP$.

Osservazione 3.59.
In questo modo X è integrabile se e solo se $\mathbb{E}[|X|]$ è finito.

Proposizione 3.60.
I risultati della Proposizione 2.54 passano al caso reale.

Lemma 3.61 (di Fatou)**.**
Sia $(X_i)_{i\in\mathbb{N}}$ una successione di v.a. reali con $X_i \geq 0$. Allora vale che:

$$\mathbb{E}\left[\liminf_{i\to\infty} X_i\right] \leq \liminf_{i\to\infty} \mathbb{E}[X_i].$$

Segue la stessa idea di dimostrazione per il lemma nella sua forma per l'integrale di Lebesgue.

Teorema 3.62 (di convergenza monotona, o di Beppo Levi)**.**
Sia $(X_i)_{i\in\mathbb{N}}$ una successione di v.a. reali non negative q.c. con $X_i \nearrow X$ q.c. (cioè la successione è crescente e $X_i(\omega) \to X(\omega)$ P-quasi ovunque). Allora $\mathbb{E}[X_i] \nearrow \mathbb{E}[X]$.

Segue la stessa idea di dimostrazione per il teorema nella sua forma per l'integrale di Lebesgue.

Teorema 3.63 (di convergenza dominata)**.**
Sia $(X_i)_{i\in\mathbb{N}}$ una successione di v.a. reali e sia X una v.a. reale tale per cui $X_i \to X$ q.c. (cioè $X_i(\omega) \to X(\omega)$ P-quasi ovunque). Se esiste una v.a. integrabile $Y \geq 0$ con $|X_i| \leq Y$ q.c. per ogni $i \in \mathbb{N}$. Allora X_n e X sono integrabili e $\mathbb{E}[X_i] \to \mathbb{E}[X]$.

Segue la stessa idea di dimostrazione per il teorema nella sua forma per l'integrale di Lebesgue.

3.7.3 Calcolo del valore atteso

Proposizione 3.64.
Sia $X : \Omega \to \mathbb{R}$ una v.a. assolutamente continua con densità f e sia $\varphi : \mathbb{R} \to \mathbb{R}$ una funzione boreliana. Allora valgono le seguenti affermazioni:

(i.) *$\varphi(X)$ è integrabile se e solo se $\int_{\mathbb{R}} \left|\varphi(x)\right| f(x)\, dx$ è finito.*

(ii.) *se $\varphi(X)$ ammette valore atteso, allora $\mathbb{E}[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f(x)\, dx$.*

Il risultato segue considerando in ordine a) le funzioni indicatrici, b) le funzioni semplici, c) le funzioni non negative e d) le funzioni integrabili, applicando il Teorema di convergenza monotona.

Osservazione 3.65.
In particolare, per X v.a. assolutamente continua vale che:

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\, dx.$$

Osservazione 3.66.
La formula presentata si estende in modo naturale al caso di più variabili:

$$\mathbb{E}[\varphi(X_1,\ldots,X_n)] = \int_{\mathbb{R}^n} \varphi(x_1,\ldots,x_n) f(x_1,\ldots,x_n)\, dx_1 \cdots dx_n.$$

3.8 Momenti e disuguaglianze, varianza, covarianza, dev. standard, mediana e moda

Tutti le disuguaglianze sul valore atteso (e.g. Markov) e tutti i risultati riguardanti i momenti (assoluti e non), la covarianza, la varianza, la deviazione standard, la mediana e la moda passano direttamente al caso reale a partire dalle proprietà del funzionale lineare $\mathbb{E}[\cdot]$. Si rimanda dunque alla *Parte 2*.

Proposizione 3.67.
Sia X v.a. reale e siano A e B tali per cui:

$$A = \left\{t \in \mathbb{R} \mid F(t) < \frac{1}{2}\right\}, \quad B = \left\{t \in \mathbb{R} \mid F(t) > \frac{1}{2}\right\}.$$

Allora m è una mediana se e solo se $m \in [\underline{m}, \overline{m}]$, dove $\underline{m} = \sup A$ e $\overline{m} = \inf B$.

Infatti ogni mediana m è maggiorante di A e minorante di B per la monotonia di F (e dunque $m \in [\underline{m}, \overline{m}]$). Si verifica poi che ogni elemento di tale intervallo è mediana.

3.9 Trasformazioni di variabili aleatorie

Proposizione 3.68 (Formula del cambio di variabili)**.**
Sia $X : \Omega \to \mathbb{R}^d$ una v.a. assolutamente continua con densità f_X tale per cui $f_X \equiv 0$ fuori da un aperto $O \subseteq \mathbb{R}^d$. Sia $\varphi : O \to O'$ un diffeomorfismo con $O' \subseteq \mathbb{R}^d$ aperto (i.e. φ è C^1, invertibile e φ^{-1} è C^1). Allora la v.a. $Y = \varphi(X)$ è assolutamente continua con densità:

$$f_Y(y) = f_X(\varphi^{-1}(y)) \left|\det D\varphi^{-1}(y)\right| 1_O(y), \quad y \in \mathbb{R}^d.$$

Segue dall'usuale formula del cambio di variabili per l'integrale di Lebesgue.

Proposizione 3.69 (Densità della somma)**.**
Siano X, $Y : \Omega \rightrightarrows \mathbb{R}$ v.a. reali con (X,Y) AC di densità $f_{(X,Y)}$. Allora $X+Y$ è assolutamente continua con densità:

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_{(X,Y)}(x, z-x)\, dx = \int_{\mathbb{R}} f_{(X,Y)}(z-y, y)\, dy.$$

Corollario 3.70.
Siano X, $Y : \Omega \rightrightarrows \mathbb{R}$ v.a. reali, AC con densità f_X e f_Y, e indipendenti. Allora $X+Y$ è assolutamente continua con densità:

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\, dx = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\, dy.$$

Osservazione 3.71.
Per il caso discreto vale una formula analoga. In particolare:

$$p_{X+Y}(k) = \sum_{j\in\mathbb{Z}} p_{(X,Y)}(j, k-j) = \sum_{j\in\mathbb{Z}} p_{(X,Y)}(k-j, j).$$

Dunque, se X e Y sono indipendenti:

$$p_{X+Y}(k) = \sum_{j\in\mathbb{Z}} p_X(j) p_Y(k-j) = \sum_{j\in\mathbb{Z}} p_X(k-j) p_Y(j).$$

3.9.1 Standardizzazione e riproducibilità di v.a. gaussiane

Proposizione 3.72.
Sia $X \sim N(m, \sigma^2)$ e siano a, $b \in \mathbb{R}$ con $a \neq 0$. Allora $aX + b \sim N(am+b, a^2\sigma^2)$.

A partire da questa proposizione si enuncia il corollario riguardante la *standardizzazione*, ovverosia il processo con cui si riconduce una qualsiasi gaussiana alla gaussiana standard:

Corollario 3.73 (Standardizzazione)**.**
$X \sim N(m,\sigma^2) \iff \frac{X-m}{\sigma} \sim N(0,1)$. Equivalentemente $Z \sim N(0,1) \iff \sigma Z + m \sim N(m,\sigma^2)$.

Osservazione 3.74.
Pertanto, tramite il processo di standardizzazione, si ricava facilmente che per $X \sim N(m,\sigma^2)$ vale che:

$$P(a \le X \le b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

Proposizione 3.75 (Riproducibilità delle gaussiane)**.**
Siano $X \sim N(m_1,\sigma_1^2)$ e $Y \sim N(m_2,\sigma_2^2)$ indipendenti. Allora $X+Y \sim N(m_1+m_2,\sigma_1^2+\sigma_2^2)$.

3.10 Legge dei grandi numeri (LGN)

La legge (debole) dei grandi numeri segue lo stesso enunciato (e la stessa dimostrazione) del caso discreto, e si rimanda dunque alla *Parte 2*.

3.10.1 Metodo di Monte-Carlo per il calcolo di integrali

Tramite l'uso delle variabili aleatorie è possibile approssimare integrali sfruttando modelli di probabilità uniforme. Supponiamo di voler calcolare $\int_a^b \varphi(x)\,dx$ con $\varphi : (a,b) \to \mathbb{R}$ boreliana e integrabile. Si assuma anche che $\int_a^b |\varphi(x)|^2\,dx < \infty$.

Si osserva che $\frac{1}{b-a}\int_a^b \varphi(x)\,dx = \mathbb{E}[\varphi(U)]$ con $U \sim U(a,b)$. Infatti vale che:

$$\mathbb{E}[\varphi(U)] = \int_a^b \varphi(x)\underbrace{\frac{1}{b-a}}_{=f_U(x)}dx.$$

Per la LGN[2], prese U_i i.i.d. come $U(a,b)$, allora le $\varphi(U_i)$ sono a loro volta i.i.d. e vale che $\frac{1}{n}\sum_{i=1}^n \varphi(U_i) \xrightarrow{\mathbb{P}} \mathbb{E}[\varphi(U)]$, ovverosia:

$$\frac{1}{n}\sum_{i=1}^n \varphi(U_i) \xrightarrow{\mathbb{P}} \frac{1}{b-a}\int_a^b \varphi(x)\,dx.$$

Pertanto è sufficiente calcolare un "grande numero" di $\varphi(U_i)$ uniformi per approssimare l'integrale scelto.

Lo stesso procedimento si può applicare per calcolare integrali della forma $\int_{(a,b)^d} \varphi(x)\,dx$ con $\varphi : (a,b)^d \to \mathbb{R}$ boreliana, integrabile e tale per cui $\int_{(a,b)^d} |\varphi(x)|^2\,dx$, approssimandolo con v.a. i.i.d. come $U((a,b)^d)$.

3.11 Teorema centrale del limite (TCL, o TLC)

Definizione 3.76 (Convergenza in legge)**.**
Si dice che una successione $(X_i)_{i\in I}$ di v.a. reali con $I \subseteq \mathbb{N}$ converge in legge a X se, detta F_n la f.d.r. di X_n e F la f.d.r. di X, allora:

$$\lim_{n\to\infty} F_n(x) = F(x), \quad \forall x \in \mathbb{R},$$

ossia se F_n converge puntualmente a F.

Teorema 3.77 (Teorema centrale del limite)**.**
Sia $(X_i)_{i\in\mathbb{N}^+}$ una successione di v.a. reali i.i.d dotate di momento secondo finito e q.c. non costanti. Siano $m = \mathbb{E}[X_1]$, $\sigma^2 = \mathrm{Var}(X_1)$. Si definisca la successione $(Z_i)_{i\in\mathbb{N}^+}$ in modo tale che:

$$Z_n = \frac{\left(\sum_{i=1}^n X_i\right) - nm}{\sqrt{n}\sigma} = \frac{\sqrt{n}}{\sigma}(\overline{X_n} - m).$$

Allora la successione $(Z_i)_{i\in\mathbb{N}^+}$ converge in legge a $N(0,1)$, ovverosia:

$$\lim_{n\to\infty} F_n(x) = \Phi(x), \quad \forall x \in \mathbb{R},$$

dove F_n è la f.d.r. di Z_n.

Osservazione 3.78.
Si può dimostrare che vale anche:

$$\mathbb{E}\left[\varphi\left(\frac{\sqrt{n}}{\sigma}(\overline{X_n} - m)\right)\right] \to \mathbb{E}[\varphi(Z)],$$

per ogni $\varphi \in C_b(\mathbb{R})$, ovverosia per ogni φ continua limitata.

Osservazione 3.79.
Sfruttando il procedimento di standardizzazione si può dimostrare che $\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n X_i - nm\right) = \sqrt{n}(\overline{X_n} - m)$ converge in legge a $Z_\sigma \sim N(0,\sigma^2)$.

[2] In questo momento si sfrutta l'ipotesi per cui $\varphi(U_i)$ ha momento secondo finito, ossia che $\int_a^b |\varphi(x)|^2\,dx < \infty$. Questa ipotesi è tuttavia rimovibile.

Parte 4

Statistica inferenziale

Lo scopo della statistica inferenziale è quello di ottenere informazioni riguardanti la distribuzione di probabilità di un esperimento a partire dagli esiti di n ripetizioni di quest'ultimo.

Nel caso di questo corso, studieremo situazioni di statistica inferenziale *parametrica*, ovverosia situazioni in cui è conosciuto il modello di probabilità del singolo esperimento a meno di un singolo parametro (e.g. l'esperimento X è in legge uguale a $B(p)$, ma p non è noto).

4.1 Definizioni preliminari

Si considerino dei dati statistici $x_1, \ldots, x_n \in \mathbb{R}$. Si consideri come spazio di probabilità lo spazio discreto relativo a $[n]$ con distribuzione uniforme.

Si definisca su tale spazio la v.a. $X : [n] \to \mathbb{R}$ tale per cui $i \mapsto x_i$. Si osserva facilmente che X ha range $r_x = \{x_1, \ldots, x_n\}$, e dunque il calcolo di tutti i suoi indici può essere ristretto a r_x.

Analogamente definiamo per dei dati $y_1, \ldots, y_n \in \mathbb{R}$ la v.a. Y.

4.1.1 Indici di centralità e di dispersione sui singoli dati

Definizione 4.1 (Media campionaria)**.**
Si definisce **media campionaria** il seguente indice di centralità:

$$\overline{x} \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Tale media coincide con il valore atteso di X.

Definizione 4.2 (Mediana campionaria)**.**
Si definisce **mediana campionaria** il seguente indice di centralità:

$$m_x \overset{\text{def}}{=} \begin{cases} x_{(n+1)/2} & \text{se } n \text{ dispari,} \\ (x_{n/2}+x_{(n+2)/2})/2 & \text{se } n \text{ pari.} \end{cases}$$

Tale indice è una mediana per X.

Definizione 4.3 (Varianza campionaria *corretta*)**.**
Si definisce **varianza campionaria (corretta)** il seguente indice di dispersione:

$$s^2 = s_x^2 = \sigma_x^2 \overset{\text{def}}{=} \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

⚠ **Attenzione.** *A differenza della media e della mediana, la varianza campionaria appena descritta non coincide con la varianza che si calcolerebbe sulla v.a. X. Infatti vale che:*

$$\text{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2,$$

e dunque:

$$s^2 = \frac{n}{n-1} \text{Var}(X).$$

4.1.2 Indici su coppie di dati

Definizione 4.4 (Coeff. di correlazione campionario)**.**
Date delle coppie di dati $(x_i, y_i)_{i \in [n]}$, si definisce il **coefficiente di correlazione campionario** come:

$$r \overset{\text{def}}{=} \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 \cdot \sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}}.$$

Tale valore coincide con l'usuale coefficiente di correlazione lineare di Bearson su X e Y, ovverosia:

$$r = \cos_{\text{Cov}}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}},$$

che, per la disuguaglianza di Cauchy-Schwarz, appartiene all'intervallo $[-1, 1]$.

4.2 Modello statistico

Come già osservato, la statistica inferenziale parametrica studia situazioni in cui è necessario ricavare o stimare un singolo parametro su un dato modello di probabilità al fine di ricavare la distribuzione di probabilità dei dati $x_1, \ldots, x_n$.

Notazione 4.5 (Parametri θ e probabilità Q_θ)**.**
Denotiamo con Θ l'insieme dei possibili parametri θ per la distribuzione di probabilità sui dati $x_1, \ldots, x_n$.

Denotiamo con Q_θ la probabilità che si otterrebbe utilizzando il parametro σ nel modello di probabilità noto a meno di parametro.

Definizione 4.6.
Si definisce **modello statistico parametrico** una terna $(S, \mathscr{S}, (Q_\theta)_{\theta \in \Theta})$, dove $(S, \mathscr{S})$ è uno spazio misurabile e $(Q_\theta)_{\theta \in \Theta}$ è una famiglia di misure di probabilità.

Esempio 4.7.
Supponiamo di star cercando di ricavare la probabilità p con cui esce testa per una data moneta. Allora un modello statistico che possiamo associare a questo problema è dato da $S = [1]$, $\mathscr{S} = \mathscr{P}([1])$ e $Q_\theta \sim B(\theta)$, con $\Theta = [0,1]$, dove 1 identifica la testa e 0 la croce.

4.3 Teoria degli stimatori su campioni di taglia n

4.3.1 Campione, statistica e stimatore

D'ora in avanti, sottintenderemo di star lavorando sul modello statistico $(S, \mathscr{S}, (Q_\theta)_{\theta\in\Theta})$.

Definizione 4.8 (Campione i.i.d. di taglia n)**.**
Dato un modello statistico, si dice che una famiglia di v.a. $(X_i : \Omega \to S)_{i\in[n]}$ i.i.d. è un **campione i.i.d. di taglia** n se per ogni $\sigma \in \Sigma$ esiste uno spazio di probabilità $(\Omega, \mathscr{F}, P_\sigma)$ tale per cui $(P_\sigma)^{X_i}$ è uguale in legge a Q_θ.

Dato un campione di taglia n, useremo P_σ per riferirci alla misura di probabilità su $(\Omega, \mathscr{F})$ appena descritta. Scriveremo come apice σ per indicare di star lavorando nello spazio di probabilità $(\Omega, \mathscr{F}, P_\theta)$ (e.g. $\mathbb{E}^\sigma$ è riferito a P_θ).

Definizione 4.9 (Statistica e stimatore)**.**
Dato un campione i.i.d. $(X_i)_{i\in[n]}$, si dice **statistica** una v.a. dipendente dalle v.a. X_i ed eventualmente dal parametro σ. Si dice **stimatore** una statistica non dipendente direttamente da σ.

4.3.2 Correttezza di uno stimatore

Definizione 4.10 (Stimatore corretto)**.**
Si dice che uno stimatore U è **corretto** (o *non distorto*) rispetto a $h : \Sigma \to \mathbb{R}$ se per ogni $\sigma \in \Sigma$ vale che:

(i.) U è P_σ-integrabile (i.e. ammette valore atteso),

(ii.) $\mathbb{E}^\sigma[U] = h(\sigma)$.

Osservazione 4.11.
La media campionaria è uno stimatore corretto del valore atteso ($h : \sigma \mapsto \mathbb{E}^\sigma[X_1]$). Infatti:

$$\mathbb{E}^\sigma\left[\overline{X}\right] = \mathbb{E}^\sigma[X_1].$$

Osservazione 4.12.
La varianza campionaria è uno stimatore corretto della varianza ($h : \sigma \mapsto \mathrm{Var}^\sigma(X_1)$). Infatti:

$$\mathbb{E}^\sigma[S^2] = \frac{1}{n-1}\left(n\mathbb{E}^\sigma[X_1^2] - \mathbb{E}^\sigma[X_1^2] - (n-1)\mathbb{E}^\sigma[X_1]^2\right) = \mathrm{Var}^\sigma(X_1).$$

Si verifica analogamente che il coeff. di correlazione campionario è uno stimatore corretto del coeff. di correlazione tra X_i e X_j.

4.3.3 Consistenza e non distorsione di una successione di stimatori

Definizione 4.13 (Successione non distorta di stimatori)**.**
Una successione di stimatori $(U_k)_{k\in\mathbb{N}^+}$ di $h(\sigma)$ si dice **asintoticamente non distorta** se U_k è P_σ-integrabile (i.e. ammette valore atteso) e:

$$\lim_{k\to\infty} \mathbb{E}^\sigma[U_k] = h(\sigma).$$

Definizione 4.14 (Successione consistente di stimatori)**.**
Una successione di stimatori $(U_k)_{k\in\mathbb{N}^+}$ di $h(\sigma)$ si dice **consistente** se:

$$\lim_{k\to\infty} P_\sigma(|U_k - h(\sigma)| > \varepsilon) = 0, \quad \forall \varepsilon > 0,$$

ovverosia se U_k converge in P_σ-probabilità a $h(\sigma)$.

Osservazione 4.15.
La successione di stimatori $(\overline{X_n})_{n\in\mathbb{N}^+}$, corretti per il valore atteso, è sia consistente che asintoticamente non distorta, per la LGN.

Osservazione 4.16.
La successione di stimatori $(S_n^2)_{n\in\mathbb{N}^+}$, corretti per la varianza, consistente, sempre per la LGN.

4.3.4 Rischio quadratico e preferibilità

Definizione 4.17 (Rischio quadratico di uno stimatore)**.**
Dato uno stimatore U di $h : \Sigma \to \mathbb{R}$, si definisce **rischio quadratico** di U per σ il seguente valore:

$$R_\sigma(U) = \mathbb{E}[(U - h(\sigma))^2].$$

Osservazione 4.18.
Se U è uno stimatore corretto di h, allora $R_\sigma(U) = \mathrm{Var}^\sigma(U)$.

Definizione 4.19 (Preferibilità)**.**
Dati due stimatori U, V di $h : \Sigma \to \mathbb{R}$, si dice che U è **preferibile** rispetto a V se $R_\sigma(U) \le R_\sigma(V)$ per ogni $\sigma \in \Sigma$.

4.3.5 Stimatore di massima verosomiglianza

D'ora in avanti sottintenderemo di star lavorando sullo spazio misurabile $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Notazione 4.20.
Data la famiglia di probabilità $(Q_\sigma)_{\sigma\in\Sigma}$, usiamo scrivere m_σ per riferirci alla densità discreta q_σ (o p_σ) di Q_σ, qualora sia discreta, oppure alla sua funzione di densità f_σ, qualora Q_σ sia assolutamente continua.

Definizione 4.21 (Funzione di verosomiglianza)**.**
Dato un campione $(X_i)_{i\in[n]}$ i.i.d. , si definisce **funzione di verosomiglianza** la funzione $L : \Sigma \times \mathbb{R}^n$ tale per cui:

$$(\sigma, (x_i)_{i\in[n]}) \overset{L}{\mapsto} L_\sigma(x_1, \ldots, x_n) \overset{\text{def}}{=} m_\sigma(x_1) \cdots m_\sigma(x_n).$$

Equivalentemente, $L_\sigma(x_1, \ldots, x_n)$ rappresenta la densità congiunta su Q_σ di x_1, ..., x_n.

Notazione 4.22.
Scriveremo $L_U(X_1, \ldots, X_n)$ con U v.a. e $(X_i)_{i\in[n]}$ famiglia di v.a. reali sottintendendo l'insieme $L_{U(\omega)}(X_1(\omega), \ldots, X_n(\omega))$, assumendo $U(\omega) \in \Sigma$.

Definizione 4.23 (Stimatore di massima verosomiglianza di σ)**.**
Si dice che uno stimatore U è di **massima verosomiglianza di** σ su un campione i.i.d. $(X_i)_{i\in[n]}$ se:

$$L_U(X_1,\ldots,X_n) = \sup_{\theta\in\Theta} L_\theta(X_1,\ldots,X_n), \quad \forall\omega\in S.$$

In altre parole, uno stimatore U è di massima verosomiglianza su un campione se per dei dati x_1, ..., x_n restituisce il parametro θ che massimizza $L_\theta(x_1,\ldots,x_n)$, ovverosia la densità consiunta dei dati x_1, ..., x_n (i.e. la probabilità che si ottenga x_1, ..., x_n).

Esempio 4.24 (Prova di Bernoulli)**.**
Sia $Q_\theta \sim B(\theta)$. Dati gli esiti x_1, ..., x_n di n prove, ricaviamo che:

$$L_\theta(x_1,\ldots,x_n) = \theta^{\sum_i x_i}(1-\theta)^{n-\theta^{\sum_i x_i}},$$

da cui:

$$\log L_\theta(x_1,\ldots,x_n) = n\overline{x}\log(\theta) + n(1-\overline{x})\log(1-\theta).$$

Tale funzione ha massimo per $\theta = \overline{x}$, e dunque $\overline{X}$ è uno stimatore di massima verosomiglianza di θ.

In altre parole, la migliore stima di σ data una sequenza di n prove di Bernoulli è la frequenza relativa di successi.

Esempio 4.25.
Sia $Q_\theta \sim U([0,\theta])$ con $\theta > 0$. Dati gli esiti x_1, ..., x_n ricaviamo che:

$$L_\theta(x_1,\ldots,x_n) = \frac{1}{\theta^n}\prod_i 1_{[0,\theta]}(x_i) = \frac{1}{\theta^n} 1_{0\le\min_i x_i\le\max_i x_i\le\theta},$$

che ha massimo per $\theta = \max_i x_i$. Pertanto $\max\{X_1,\ldots,X_n\}$ è uno stimatore di massima somiglianza di θ.

In altre parole, dati degli esiti x_1, ..., x_n, una delle migliori stime che possiamo fare su θ è $\max_i x_i$.

Tabella e proprietà delle distribuzioni discrete

Nome distribuzione	Caso di utilizzo	Parametri	Densità discreta	Valore atteso e momenti	Varianza
Distr. di Bernoulli $X \sim B(p)$	Esperimento con esito di successo (1) o insuccesso (0).	p – probabilità di successo.	$P(X=1)=p$, $P(X=0)=1-p$	$\mathbb{E}[X]=\mathbb{E}[X^2]=p$	$\mathrm{Var}(X)=p(1-p)$
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 $0 \le k \le n$ e 0 altrimenti.	$\mathbb{E}[X]=np$ (è somma di n Bernoulliane) $\mathbb{E}[X^2]=np+n(n-1)p^2$	$\mathrm{Var}(X)=np(1-p)$ (è somma di n Bernoulliane indipendenti)
Distr. binomiale negativa $X \sim \mathrm{BinNeg}(h,p)$	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) $\mathbb{E}[X^2]=\frac{h(1+h-p)}{p^2}$	$\mathrm{Var}(X)=\frac{h(1-p)}{p^2}$ (è somma di h Geometriche indipendenti)
Distr. geometrica $X \sim \mathrm{Geom}(p)$	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 $\mathrm{BinNeg}(1,p)$	$p \in (0,1)$ – probabilità di successo all'i-esimo esperimento.	$P(X=k)=p(1-p)^{k-1}$ per $k \ge 1$ e 0 per $k=0$.	$\mathbb{E}[X]=\frac{1}{p}$ $\mathbb{E}[X^2]=\frac{2-p}{p^2}$	$\mathrm{Var}(X)=\frac{1-p}{p^2}$
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) $X \sim \mathrm{Poisson}(\lambda)$	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.	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	$\mathbb{E}[X]=\lambda$ $\mathbb{E}[X^2]=\lambda(\lambda+1)$	$\mathrm{Var}(X)=\lambda$

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

[1] Ovverosia la probabilità che non vi siano mai successi è nulla.

Tabella e proprietà delle distribuzioni assolutamente continue

Nome distribuzione	Caso di utilizzo	Parametri	Densità	Funzione di ripartizione	Valore atteso e momenti	Varianza
Distr. uniforme $X \sim U(B)$	Estrazione di un punto reale a caso su B senza preferenze.	$B \in \mathscr{B}(\mathbb{R})$ – insieme da cui estrarre.	$f(x) = \frac{1}{m(B)} 1_B(x)$	$F(x) = \frac{m((-\infty, x] \cap B)}{m(B)}$	$\mathbb{E}[X] = \frac{a+b}{2}$ se $B = (a, b)$	$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$ se $B = (a, b)$
Distr. esponenz. $X \sim \mathrm{Exp}(\lambda)$	Processo di Poisson in senso continuo.	$\lambda > 0$ – parametro di Poisson.	$f(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x)$	$F(x) = 1 - e^{-\lambda x}$ per $x \geq 0$, 0 altrimenti.	$\mathbb{E}[X] = \frac{1}{\lambda}$	$\mathrm{Var}(X) = \frac{1}{\lambda^2}$
Distr. gamma $X \sim \Gamma(r, \lambda)$	Estensione della distr. binomiale in senso continuo.	$r > 0$, $\lambda > 0$.	$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} 1_{(0,\infty)}(x)$		$\mathbb{E}[X] = \frac{r}{\lambda}$	$\mathrm{Var}(X) = \frac{r}{\lambda^2}$
Distr. normale $X \sim N(m, \sigma^2)$		m – media. $\sigma^2 > 0$ – varianza.	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2}$	$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ per $N(0,1)$ e si standardizza le altre distr. con il cambio di var. $z = (x-m)/\sigma$.	$\mathbb{E}[X] = m$. Inoltre $\mathbb{E}[(X-m)^p] = 0$ se p è dispari e $\mathbb{E}[(X-m)^p] = \sigma^p (p-1)!!$ se p è pari, dove !! indica il *doppio fattoriale*.	$\mathrm{Var}(X) = \sigma^2$

Si ricorda che la funzione Γ è definita in modo tale che $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$; si tratta di un'estensione della nozione di fattoriale ai valori reali (infatti, $\Gamma(n+1) = n!$ per $n \in \mathbb{N}$).

Valgono inoltre le seguenti altre proprietà:

- La distribuzione $\mathrm{Exp}(\lambda)$ coincide con la distribuzione $\Gamma(1, \lambda)$.

Tabella e proprietà della f.d.r. $\Phi(x)$ della distr. normale standard $N(0,1)$

Per $a \in \mathbb{R}$ si definisce la funzione $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx$. L'integrale $\Phi(\infty) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$ 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 $\frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-x^2/2}\, dx$, che risulta essere $\Phi(b) - \Phi(a)$. Se $a > 0$, allora $\frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{-x^2/2}\, dx = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$.

Tabella 4.1: 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	+0,00	+0,01	+0,02	+0,03	+0,04	+0,05	+0,06	+0,07	+0,08	+0,09
0,0	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	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	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	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,999