Probabilità

Marco Ambrogio Bergamo

Anno 2023-2024

Parte I
Teoria della misura

Algebra Sia E un insieme e $\mathcal{A} \subset \mathcal{P}(E)$ una famiglia di sottoinsiemi. $\mathcal{A}$ è un'algebra se è chiusa rispetto al complementare, unioni **finite** e intersezioni **finite**.
Per dimostrare che è un'algebra verificare le 3 prop.:

1. $\emptyset \in \mathcal{A}$ oppure $E \in \mathcal{A}$
2. $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$
3. $A \in \mathcal{A} \implies A^C \in \mathcal{A}$

σ-algebra Sia E un insieme e $\mathcal{S} \subset \mathcal{P}(E)$ (famiglia delle parti). $\mathcal{S}$ è una σ-algebra se è chiuso rispetto al complementare, unioni **infinite numerabili** e intersezioni **infinite numerabili**.

Spazio misurabile È la coppia $(E, \mathcal{S})$ dove $\mathcal{S}$ è una σ-algebra su E.

σ-algebra generata da $\mathcal{C}$ Si indica con $\sigma(\mathcal{C})$ ed è la più piccola σ-algebra contenente $\mathcal{C}$, ovvero l'intersezione di tutte le σ-algebra contenenti $\mathcal{C}$.

σ-algebra di Borel Si indica con $\mathcal{B}(\mathbb{R}^n)$ ed è la σ-algebra generata dalla famiglia degli aperti di $\mathbb{R}^n$, ovvero dalla topologia. È quindi $\sigma(\mathcal{T})$.

Classe monotona È una famiglia $\mathcal{M} \subset \mathcal{P}(E)$ tale che:

(M1) $E \in \mathcal{M}$
(M2) $A, B \in \mathcal{M}$ e $A \subset B \implies B \setminus A \in \mathcal{M}$
(M3) è stabile per limite crescente, ovvero $(A_n)_n$ successione di insiemi crescenti contenuti in $\mathcal{M} \implies A = \bigcup A_n \subset \mathcal{M}$

Teorema delle classi monotone Sia $\mathcal{A}$ una famiglia di insiemi stabile per **intersezioni finite** e sia $\mathcal{M}$ una classe monotona contenente $\mathcal{C}$. Allora $\sigma(\mathcal{C}) \subset \mathcal{M}$.

Funzione misurabile Un'applicazione $f : (E_1, \mathcal{S}_1) \to (E_2, \mathcal{S}_2)$ tra spazi misurabili si dice misurabile se $\forall A \in \mathcal{S}_2 \implies f^{-1}(A) \in \mathcal{S}_1$.
In realtà basta verificare questa proprietà $\forall A \subset \mathcal{C} \mid \sigma(\mathcal{C}) = \mathcal{S}$, quindi nel caso $\mathcal{S}_2$ sia su uno sp. topologico, basta verificare la proprietà sugli aperti di E_2.

Lemma Perché una funzione $f : (E, \mathcal{S}) \to \mathbb{R}$ sia misurabile, basta che $f^{-1}((-\infty, a]) \in \mathcal{S}$ per ogni $a \in \mathbb{R}$.
MOLTO IMPORTANTE per def. di variabile aleatoria.

Misura Una misura su $(E, \mathcal{S})$, sp. misurabile, è un'applicazione $\mu : \mathcal{S} \to \overline{\mathbb{R}^+}$ tale che:

1. $\mu(\emptyset) = 0$
2. per ogni successione $(A_n)_n$ di insiemi a due a due disgiunti vale: $\mu(\cup_{n\geq 1} A_n) = \sum \mu(A_n)$

finita se $\mu(E) < +\infty$

σ-finita

di probabilità se $\mu(E) = 1$

Spazio di misura È la tripla $(E, \mathcal{S}, \mu)$

Parte II
Eventi e loro probabilità

Definizioni

Spazio campionario Ω = Insieme dei casi elementari di un esperimento

Casi elementari $\omega \in \Omega$

Eventi $A \subseteq \Omega$ (appartengono alla σ-algebra)

Famiglia di eventi Si indica con $\mathcal{F}$ ed è una σ-algebra su Ω, ovvero:

1. $\emptyset \in \mathcal{F}$
2. $A_1, A_2 ... \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
3. $A \in \mathcal{F} \implies A^C \in \mathcal{F}$

È quindi la famiglia di sottoinsiemi di Ω, che sono i possibili "eventi composti".

Probabilità Dato un insieme Ω e un'algebra di parti $\mathcal{A}$ (NON una σ-algebra), la probabilità è una funzione $\mathbb{P} : \mathcal{A} \to [0, 1]$ tale che

(P1) $\mathbb{P}(\Omega) = 1$

(P2) $\forall E_1, E_2 \in \mathcal{A} \mid E_1 \cap E_2 = \emptyset \implies \mathbb{P}(E_2 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$

Misura di probabilità Come probabilità **ma con σ-additività**. $\mathbb{P}$ su $(\Omega, \mathcal{F})$ con $\mathcal{F}$ una σ-algebra è una funzione $\mathbb{P} : \mathcal{F} \to [0, 1]$ tale che:

(P1) $\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(\Omega) = 1$

(P2) $A_1, A_2 ... \in \mathcal{F}$ **disgiunti**, ovvero $A_i \cap A_j = 0 \quad \forall i, j \implies \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

TEOREMA (DI CONTINUITÀ DELLA MISURA DI PROB.) Vale:

- Se $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ e $A \coloneqq \cup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i \implies \mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$
- Se $B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$ e $B \coloneqq \cap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i \implies \mathbb{P}(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$

Probabilità condizionata Se $\mathbb{P}(B) \neq 0$ allora la probabilità che avvenga A sapendo che è avvenuto B è

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

Quindi più la probabilità che avvenga B è grande, più quella che poi avvenga A è piccola, a parità di intersezione. Arriviamo a questo risultato da due necessità: logicamente la probabilità di A dato B deve essere proporzionale alla loro intersezione, in secondo luogo $\mathbb{P}(\Omega \mid B)$ deve fare 1, quindi scaliamo l'intersezione per $1/\mathbb{P}(B)$ essendo $\Omega \cap B = B$.

TEOREMA (DELLE PROB. TOTALI / DELLA PARTIZIONE PER LE PROB.) $\forall A, B \mid 0 < \mathbb{P}(B) < 1:$

$$\mathbb{P}(A) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^C)\mathbb{P}(B^C)$$

In generale, se $B_1, \ldots, B_n$ una partizione di $\Omega \mid \mathbb{P}(B_i) > 0$ allora:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

TEOREMA (DI BAYES) $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}, \quad$ con $\mathbb{P}(B), \mathbb{P}(A) > 0$

Eventi indipendenti Per una coppia di eventi vale: A e B sono indipendenti se $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, ovvero se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Per una famiglia di eventi ci sono due tipi di indipendenza:

- **indipendenza a coppie:** Una famiglia $\{A_i : i \in I\}$ lo è se $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ per ogni $i \neq j$.
  Ciò non implica che la famiglia è indipendente.
- **(mutua) indipendenza:** In generale, una famiglia $\{A_i : i \in I\}$ è indipendente se

$$\mathbb{P}(\bigcap_{i\in J} A_i) = \prod_{i\in J} \mathbb{P}(A_i)$$

  per ogni sottoinsieme finito J di I, ovvero per tutte le i-uple di indici, dove $2 \leq i \leq \#I$

Indipendenza condizionata A e B sono indipendenti dato C (con $\mathbb{P}(C) > 0$) se $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C) \cdot \mathbb{P}(B \mid C)$

Spazio di prob. completo $(\Omega, \mathcal{F}, \mathbb{P})$ lo è se ogni sottoinsieme di ogni evento nullo ($A \in \mathcal{F} \mid \mathbb{P}(A) = 0$) è a sua volta un evento ($\in \mathcal{F}$).
Ogni spazio non completo può essere completato: se $\mathcal{N}$ è la famiglia di tutti i sottoinsiemi degli eventi ($\in \mathcal{F}$) nulli, allora $\mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{N})$ è la più piccola σ-algebra che contiene le due famiglie. $(\Omega, \mathcal{G}, \mathbb{P})$ è il **completamento** di $(\Omega, \mathcal{F}, \mathbb{P})$.

Spazio prodotto Abbiamo due (o più) esperimenti con rispettivi spazi di prob. $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ e $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Consideriamo il nuovo spazio di probabilità sul nuovo esperimento che consiste nel fare la coppia (disgiunta) di esperimenti. Esso è $(\Omega_1 \times \Omega_2, \mathcal{G}, \mathbb{P}_{12})$ dove:

- Spazio degli eventi: è il prodotto cartesiano $\Omega_1 \times \Omega_2$, ovvero le 2-uple di eventi elementari.
- σ-algebra: $\mathcal{G} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$ (in generale il prodotto da solo non è una σ-algebra). Quindi contiene sicuramente gli insiemi del tipo $\{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$
- Misura di probabilità: $\mathbb{P}_{12} : \mathcal{F}_1 \times \mathcal{F}_2 \to [0, 1] \quad | \quad \mathbb{P}_{12}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$

Teoremi

0.1 Probabilità del complementare

$$\mathbb{P}(E^C) = 1 - \mathbb{P}(E)$$

Dimostrazione. $1 = \mathbb{P}(\Omega) = \mathbb{P}(E \cup E^C) \stackrel{\star}{=} \mathbb{P}(E) + \mathbb{P}(E^C) \implies$ tesi
$\star$ = essendo $E \cap E^C = 0$, per secondo assioma della def. di prob.

0.2 Decomposizione della probabilità di un evento

$$\mathbb{P}(E_1) = \mathbb{P}(E_1 \setminus E_2) + \mathbb{P}(E_1 \cap E_2)$$

Dimostrazione.

$$\begin{cases} (E_1 \setminus E_2) \cap (E_1 \cap E_2) = \emptyset \\ (E_1 \setminus E_2) \cup (E_1 \cap E_2) = E_1 \end{cases} \stackrel{def.prob}{\Longrightarrow} \mathbb{P}(E_1) = \mathbb{P}((E_1 \setminus E_2) \cup (E_1 \cap E_2)) = \mathbb{P}(E_1 \setminus E_2) + \mathbb{P}(E_1 \cap E_2)$$

$\implies$ tesi

0.3 Monotonia della prob.

$$E_2 \subseteq E_1 \implies \mathbb{P}(E_1) \leq \mathbb{P}(E_2)$$

Dimostrazione.

- applicare decompos. delle prob. a questi due eventi: $\mathbb{P}(E_1) = \mathbb{P}(E_1 \setminus E_2) + \mathbb{P}(\overbrace{E_1 \cap E_2}^{=E_2})$
- La prob. è sempre positiva

$\implies$ tesi

0.4 Estensione del secondo assioma

Se $A_i \cap A_j = 0 \quad \forall i, j$, allora

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$$

Dimostrazione.

- Per induzione: prima $n = 2$, poi $n \implies n + 1$
- Definire A l'unione da 1 a n e B l'evento $n + 1$-esimo e usare prima il secondo assioma per A e B, poi ip. induttiva.

0.5 Probabilità dell'unione qualsiasi

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Dimostrazione.

- $A \setminus B, B \setminus A, A \cap B$ sono a due a due disgiunti, quindi la prob di $A \cup B$ è la somma delle tre prob.
- sostituire la prob. della differenza con decomposizione della prob. e si ottiene il risultato

0.6 Disuguaglianza di Boole

$$\mathbb{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$$

Dimostrazione. Per induzione, procedere come in estensione del secondo assioma

0.7 Principio di inclusione-esclusione

$$\mathbb{P}(\bigcup_{i=1}^n E_i) = \sum_i \mathbb{P}(E_i) - \sum_{i<j} \mathbb{P}(E_i \cap E_j) + \sum_{i<j<k} \mathbb{P}(E_i \cap E_j \cap E_k) - \cdots + (-1)^{n+1}\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n)$$

Dimostrazione.

- Per induzione: prima $n = 2$, poi $n \implies n + 1$
- Definire A l'unione da 1 a n e B l'evento $n + 1$-esimo
- $\mathbb{P}(\cup_{i=1}^{n+1}) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B)}_{(1)} - \mathbb{P}(A \cap B)$
- Applicare passo induttivo a $\mathbb{P}(A)$:

$$\mathbb{P}(A) = \underbrace{\sum_{i=1}^n \mathbb{P}(E_i)}_{(1)} \underbrace{- \sum_{1\leq i<j\leq n} \mathbb{P}(E_i \cap E_j)}_{(2)} \underbrace{+ \sum_{1\leq i<j<k\leq n} \mathbb{P}(E_i \cap E_j \cap E_k)}_{(3)} - \cdots + (-1)^{n+1}\mathbb{P}(\cap_{i=1}^n E_i)$$

- Osservare che $A \cap B = \cup_{i=1}^n \overbrace{(E_i \cap E_{n+1})}^{:=F_i}$ per prop. distributiva
- Applicare di nuovo passo induttivo a $A \cap B = \cup_{i=1}^n (F_i)$:

$$\mathbb{P}(A \cap B) = \underbrace{\sum_{i=1}^n \mathbb{P}(F_i)}_{(2)} \underbrace{- \sum_{1\leq i<j\leq n} \mathbb{P}(F_i \cap F_j)}_{(3)} + \sum_{1\leq i<j<k\leq n} \mathbb{P}(F_i \cap F_j \cap F_k) - \cdots + \underbrace{(-1)^{n+1}\mathbb{P}(\cap_{i=1}^n F_i)}_{\cdot(-1)=(-1)^{n+2}\mathbb{P}(\cap_{i=1}^{n+1} E_i)}$$

- Mettere tutto in quella blu, sommando (1) con (1), sottraendo (2) con (2) ... l'addendo j-esimo della prima meno l'addendo $(j-1)$-esimo della seconda (diventano uguali alla tesi ma con $n+1$) e ricordando come cambia segno l'ultimo addendo della seconda, ottengo la tesi con $n + 1$

Formula delle probabilità totali (prima versione)

$\{B_1, B_2, \dots, B_n\}$ partizione $\implies P(E) = \sum_{i=1}^{n} P(E \cap B_i)$

Dimostrazione. Notare che $E = \cup_{i=1}^{n}(E \cap B_i)$ e che tutti questi eventi sono a due a due disgiunti. Tesi segue dall'estensione del secondo assioma.

Formula di Bayes (versione semplice)

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

Dimostrazione. Basta applicare la def. di prob condizionale e poi al numeratore scriverla al contrario per $B \cap A$

Formula delle probabilità totali (seconda versione)

$\{B_1, B_2, \dots, B_n\}$ partizione $\implies P(E) = \sum_{i=1}^{n} P(E|B_i) \cdot P(B_i)$

Dimostrazione. Basta applicare la prima versione e poi la def. girata di prob. condizionata

Formula di Bayes (versione estesa)

$\{B_1, B_2, \dots, B_n\}$ partizione $\implies P(B_i|E) = \frac{P(E|B_i) \cdot P(B_i)}{\sum_{k=1}^{n} P(E|B_k) \cdot P(B_k)}$

Dimostrazione. Applicare Bayes a $P(B_i \mid E)$ e formula prob. totali a $P(E)$

Caratterizzazione dell'indipendenza mediante condizionamento

$P(A|B) = P(A)$

Dimostrazione. Applicare def. di prob. condizionata e di eventi indip.

Conseguenza dell'indipendenza tra due eventi

E_1, E_2 indipendenti $\implies E_1^C, E_2$ indipendenti (e tutte le altre combinazioni di complementare o non).

Dimostrazione. $\mathbb{P}(E_1^C \cap E_2) = \mathbb{P}(E_2 \setminus E_1) = \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2) \overset{indip}{=} \mathbb{P}(E_2)(1 - \mathbb{P}(E_!)) = \mathbb{P}(E_2)\mathbb{P}(E_1^C) \implies$ tesi

Conseguenza della mutua indipendenza tra una famiglia di eventi

Se $E_1, \dots, E_n \in \mathcal{A}$ eventi mutualmente indipendenti $\implies$ lo sono anche

1. $E_1^{\alpha_1}, \dots, E_n^{\alpha_n} \quad \forall \alpha_i \in \{0, 1\}$ che rappresentano niente (1) o complementare (0)
2. $F_1, \dots, F_k$ con $F_j = \cap_{i \in I_j} E_i$ dove $I_1 \dots I_k$ è una partizione di $\{1, \dots, n\}$
3. $G_1, \dots, G_k$ con $F_j = \cup_{i \in I_j} E_i$ dove $I_1 \dots I_k$ è una partizione di $\{1, \dots, n\}$

Dimostrazione.

1. Per induzione sul numero dei complementari.
   $k = 1$: pongo $\alpha_n = 0$ e tutti gli altri $=1$. Fisso $I \subseteq \{1 \dots n\} = \{i_1 \dots i_k, n\}$ (se non contenesse n nulla da dim), faccio
   $$\mathbb{P}(\underbrace{E_{i_1} \cap \dots \cap E_{i_k}}_{:=B} \cap E_n^C) = \mathbb{P}(B \cap E^C)$$
   ed è analogo al caso di due eventi visto sopra. Ciò vale per ogni scelta di I, quindi tesi.
   $k \implies k{+}1$: vogliamo dimostrare $E_1^C \dots E_k^C, E_{k+1} \dots E_n$ mutual. indip. $\implies E_1^C \dots E_k^C, E_{k+1}^C \dots E_n$ mutual. indip. per ogni scelta di $I \subseteq \{1 \dots n\}$. Non è restrittivo metterci nel caso $I = \{1 \dots n\}$, quindi:
   $$\mathbb{P}(E_1^C \cap \dots \cap E_{k+1}^C \cap E_{k+2} \cap \dots \cap E_n) = \mathbb{P}(B \cap E_{k+1}^C)$$
   avendo rinominato tutto con B. Seguiamo i passaggi di prima (=sottrazione degli eventi ecc) e applichiamo ipotesi induttiva.

2. facile, l'intersez. degli F_j è l'intersezione di tutti gli E_i ma intersecati a gruppetti.
3. Dimostrazione in un caso particolare: $k = 2$ e $G_1 = \cup_{i=1}^{n-1} E_i, \quad G_2 = E_n$. Devo dim. che $P(G_1 \cap G_2)$ è il prodotto delle prob. Scrivere primo membro mettendo tutto nell'unione, poi applicare principio di inclusione-esclusione ...

Parte III
Modelli probabilistici

1 Probabilità su un insieme finito

Spazio degli eventi $\#\Omega = N < +\infty$

Algebra $\mathcal{A} = \mathscr{P}(\Omega)$

Simplesso N-1 dimensionale Δ_{N-1} vettore di N dimensioni che ha per componenti le probabilità degli N casi elementari ω_i. Ha un grado di libertà in meno ($N-1$ gradi di libertà) poiché c'è il vincolo che la somma delle componenti deve essere =1, quindi l'ultimo termine è determinato dai primi $N-1$.

TEOREMA Eiste corrispondenza biunivoca tra le probabilità su Ω e i punti di Δ_{N-1}

1.1 Modello uniforme

Definizione

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

Entropia

1.2 Modello ipergeometrico

Urna con N palline, di cui B bianche e R rosse ($B + R = N$)

Approccio statico

- estrazione in blocco di $n \leq N$ palline (senza reimmissione)
- **numero** tutte le palline da 1 a N (le rendo tutte distinte) e dopo l'estrazione non mi interessa dell'ordine, quindi assumo di ordinarle in ordine crescente, ottenendo un vettore dell'estrazione.

Spazio degli eventi $\Omega = \{(i_1 \dots i_n) \in \{1 \dots N\}^n \mid i_1 < i_2 < \dots < i_n\}$. $\#\Omega = \binom{N}{n}$

Algebra $\mathcal{A} = \mathscr{P}(\Omega)$

Probabilità scegliamo una prob. uniforme su $\Omega \implies P(\omega) = \frac{1}{\#\Omega} = \frac{1}{\binom{N}{n}}$

Evento di interesse E_b = "nella n-upla di palline estratte ce ne sono esattamente b bianche e $r = n - b$ rosse"

Troviamo che:

$$P(E_b) = \frac{\#E_b}{\#\Omega} = \frac{\binom{B}{b}\binom{R}{r}}{\binom{N}{n}}$$

Dimostrazione. Devo dimostrare che i due numeratori hanno stessa cardinalità. In due passaggi:

1. $E_b \overset{biuniv.}{\longleftrightarrow} S(B, b) \times S(R, r)$ dove $S(N, n)$ sono tutti i sottoinsiemi di $\{1 \dots N\}$ contenenti n oggetti. Lo faccio costruendo una funzione biunivoca tra i due insiemi.
2. uso principio fondamentale dell calcolo combinatorio: $\#(A \times B) = \#A \cdot \#B$

Approccio dinamico

- palline **non numerate**
- ne estraggo una ad una senza reimmissione, ogni volta scrivendo 1 se bianca, 0 se rossa.

Spazio degli eventi $\Omega = \{0,1\}^n = \{(x_1, \dots, x_n) \mid x_i \in \{0,1\}\}$. $\#\Omega = 2^n$

Algebra $\mathcal{A} = \mathscr{P}(\Omega)$

Evento considerati A_i = "esce palla bianca all'i-esima estrazione" $= \{\omega = (x_1, \dots, x_n) \in \Omega \mid x_i = 1\}$

Probabilità deve soddisfare:

1. $P(A_1) = \frac{B}{N}$
2. $P(A_{k+1} \mid A_1^{\alpha_1} \cap \cdots \cap A_k^{\alpha_k}) = \frac{B-s_k}{N-k}$ dove $s_k = \sum_{i=1}^{k} \alpha_i$ = numero di bianche estratte delle k estrazioni precedenti.

Evento di interesse E_b = "nella n-upla di palline estratte ce ne sono esattamente b bianche e $r = n - b$ rosse"

$$= \bigcup_{(\alpha_1 \dots \alpha_n) \mid \sum \alpha_i = b} A_1^{\alpha_1} \cap \cdots \cap A_k^{\alpha_k}$$

Il numero di vettori di α_i che soddisfano tale somma è $\binom{n}{b}$ in quanto contengono solo 0 e 1.

Troviamo che la prob. è la stessa

Dimostrazione.

1. Ricostruire E_b in termini degli A_i. Per induzione sul numero n di estrazioni.
2. notare che l'unione è di eventi disgiunti.

2 Modello di Bernoulli

Vederlo come l'ipergeometrico dinamico, ma con mutua idipendenza degli A_i (senza reimmissione).

Spazio degli eventi $\Omega = \{0,1\}^n = \{(x_1, \dots, x_n) \mid x_i \in \{0,1\}\}$. $\#\Omega = 2^n$

Algebra $\mathcal{A} = \mathscr{P}(\Omega)$

Evento considerati A_i = "esce palla bianca all'i-esima estrazione" $= \{\omega = (x_1, \dots, x_n) \in \Omega \mid x_i = 1\}$

Probabilità pongo $p \coloneqq \frac{B}{N} \in (0,1)$ e la prob. è caratterizzata da una delle prossime:

- $p = P(A_1) = P(A_{k+1} \mid A_1^{\alpha_1} \cap \cdots \cap A_k^{\alpha_k}) \quad \forall k, \alpha_i$
- $P(A_i) = p \quad \forall i$ e A_i sono mutualmente indip.

  Dimostrazione. (dell'equivalenza)

Evento di interesse E_b = "nella n-upla di palline estratte ce ne sono esattamente b bianche e $r = n - b$ rosse"

$$= \bigcup_{(\alpha_1 \dots \alpha_n) \mid \sum \alpha_i = b} A_1^{\alpha_1} \cap \cdots \cap A_k^{\alpha_k}$$

Si trova che:

$$\mathbb{P}(E_b) = \binom{n}{b} p^b (1-p)^{n-b}$$

Dimostrazione. Scrivere E_b in termini di A_i, notare che è unione disgiunta, usare indipendenza e ricordare che il numero di vettori di α_i che soddisfano tale somma (=b) è $\binom{n}{b}$ in quanto contengono solo 0 e 1.

3 Probabilità su insieme discreto infinito (costruzione Bernoulli infinito)

Consideriamo una successione infinita di esperimenti di Bernoulli, ovvero solo due esiti: **successo (=1) e insuccesso (=0)**.

Spazio di probabilità $\Omega = \{0,1\}^\infty = \{\omega = (x_1, x_2, \dots) \mid x_i \in \{0,1\} \, \forall i \in \mathbb{N}\}$
Insieme di tutte le successioni di 0 e 1. Ω ha cardinalità **infinita non numerabile**.

Eventi A_k = "successo alla k-esima prova" $= \{\omega \in \Omega \mid x_k = 1\}$
Insieme delle successioni il cui k-esimo elemento è 1

Vogliamo costruire su Ω:

- un'**algebra** $\mathcal{A}$ che contenga tutti gli A_k
- una **probabilità** $\mathbb{P} : \mathcal{A} \to [0,1]$ che sia come nello schema di Bernoulli finito, ovvero che ci sia distribuzione uniforme sugli A_k e che tali eventi siano tutti mutualmente indipendenti.

$\implies$ algebra dei cilindri

Algebra dei cilindri

- n-cilindro di base $B^{(n)}$: $C_n(B^{(n)}) \coloneqq \{\omega = (x_1, x_2, \dots) \in \Omega \,|\, (x_1, \dots, x_n) \in B^{(n)}\}$ con $B^{(n)} \in \mathscr{P}(\{0,1\}^n)$, $n \in \mathbb{N}$
Se $B^{(n)} \in \{0,1\}^n$, ovvero se la base è un solo vettore, il cilindro sono le successioni di zeri e uno con le prime n coordinate fisse ed uguali, ovvero tolgo n gradi di libertà ("cilindro base"). Se invece la base appartiene alle parti di $\{0,1\}^n$ abbiamo che la base è un insieme di vettori, quindi il cilindro è un insieme di "cilindri base".
Pensare un cilindro come a un array/tabella di successioni, con infinite righe (ogni colonna ha i primi n termini uguali) e al più 2^n colonne, che è la cardinalità massima di $\{0,1\}^n$, ovvero se la base è tutto $\{0,1\}^n$ (ovvero tutti i vettori possibili di n dimensioni di 0 e 1).

- Algebra degli n-cilindri: $\mathscr{C}_n \coloneqq \{C_n(B^{(n)}) \,|\, B^{(n)} \in \mathscr{P}(\{0,1\}^n)\}$, da cui $\mathscr{C}_1 \subset \mathscr{C}_2 \subset \dots$

  Dimostrazione. ($\mathscr{C}_n$ è un'algebra) Per dire che è un'algebra devo verificare le 3 proprietà:
  - $\Omega \in \mathscr{C}$ sì, infatti $\Omega = C_n(\{0,1\}^n)$, ovvero l'insieme di tutte le successioni con i primi n termini fissati, ma per tutti i possibili vettori di n termini (è un cilindro poiché $\{0,1\}^n \in \mathscr{P}(\{0,1\}^n)$
  - complementare: $(C_n(B^{(n)}))^C = C_n((B^{(n)})^C)$ ma anche questo è un cilindro poiché $(B^{(n)})^C \in \mathscr{P}(\{0,1\}^n)$
  - unione: come sopra

  Per vedere che $\mathscr{C}_n \subset \mathscr{C}_{n+1}$ osservare che $C_n(B^{(n)}) = C_{n+1}(B^{(n)} \times \{0,1\})$

- Algebra dei cilindri: $\mathscr{C} \coloneqq \bigcup_{n=1}^{\infty} \mathscr{C}_n$

  Dimostrazione. ($\mathscr{C}$ è un'algebra) Classico ragionamento per dimostrare algebre di unioni infinite di algebre, ovvero che se un elemento sta in $\mathscr{C} \implies$ sta in un qualche $\mathscr{C}_n$ per def di $\mathscr{C}$.

  Dimostrazione. ($\mathscr{C}$ non è una σ-algebra) Basta mostrare che, se $A_k = \{\omega = (x_1, x_2, \dots) \in \Omega \mid x_k = 1\}$, allora $\cap_{k=1}^{\infty} A_k = (1, 1, \dots) \notin \mathscr{C}$ in quanto ogni cilindro è un numero infinito di successioni (vedi rappresentazione in array/tabella dei cilindri, che ha sempre infinite righe) e questa è una sola successione.

Probabilità (Daniell - Kolmogorov)

Volgiamo assegnare valori numerici alle prob. $\mathbb{P}(C_n(B^{(n)}))$ per ogni $n \in \mathbb{N}$ e per ogni $B^{(n)} \in \mathscr{P}(\{0,1\}^n)$: ricorriamo allo schema di Bernoulli finito (binomiale) con una **successione di spazi di prob.** $(\{0,1\}^n, \mathscr{P}(\{0,1\}^n), \pi_n)$. Dobbiamo "incollare" questa successione.

Probabilità Poniamo

$$P(C_n(B^{(n)})) \coloneqq \pi_n(B^{(n)})$$

ovvero la probabilità del cilindro è la probabilità di Bernoulli finito (binomiale) della base. È ben definita (ovvero ogni cilintro ha una sola probabilità), con $(\{0,1\}^n, \mathscr{P}(\{0,1\}^n), \pi_n)$ spazio di probabilità di Bernoulli finito.
Dobbiamo dimostrarlo perché ci sono cilindri con scrittura diversa che però sono uguli, per opportune basi: ad esempio se la base del primo cilindro è un solo vettore di dimensione n, posso scriverlo anche con una base di più dimensioni (quindi meno gradi di libertà), ma di più vettori (perché la base si prende nelle parti di $\{0,1\}^m$. Per esempio, se $m = n+1$ ciò accade con:

- $B_1^{(n)} \in \{0,1\}^n$ fissata
- $B_2^{(n+1)} = B_1^{(n)} \times \{0,1\} \in \mathscr{P}(\{0,1\}^{n+1})$

  Dimostrazione. (Buona definizione) Vogliamo dimostrare che

  $$C_n(B_1^{(n)}) = C_m(B_2^{(m)}) \implies \pi_n(B_1^{(n)}) = \pi_m(B_2^{(m)})$$

  La dimostriamo nel caso in cui
  - $B^{(n)} = \{(x_1, \dots, x_m)\} \in \mathscr{P}(\{0,1\}^n) \mid \sum_{i=1}^{n} = z$ fissato
  - $B^{(m)} = B^{(n)} \times \{0,1\}^{m-n} \in \mathscr{P}(\{0,1\}^m)$

Procedo nel seguente modo:

1. scrivo $B^{(m)}$ come unione di successioni $\{(x_1, \ldots, x_n, y_1, \ldots, y_{m-n})\}$ al variare di $(y_1, \ldots, y_{m-n} \in \{0,1\}^{m-n}$, che sono tutte disgiunte
2. procedo con trovare la prob. $\pi_m(B^{(m)})$ che, essendo disgiunti, è la sommatoria delle prob. al variare della stessa roba.
3. scompongo la sommatoria in due sommatorie, una in cui il vettore delle y ha somma r e l'altra per r che va da 0 a $m-n$.
4. riscrivo $\pi_m(\{(x_1, \ldots, x_n, y_1, \ldots, y_{m-n})\})$ come prob. di avere quella esatta stringa
5. faccio in modo di tirar fuori dalla sommatoria $p^k(1-p)^{n-k}$ (aggiungo e sottraggo n a un esponente)
6. noto che ciò che rimane nella sommatoria fa 1 (somma di tutte le prob. di una binomiale) e quindi mi rimane solo $p^k(1-p)^{n-k}$ che è $\pi_n(B^{(n)})$

Dimostrazione. (Che è una probabilità) Per $\mathbb{P}(\Omega) = \mathbb{P}(C_n(\{0,1\}^n)) = \pi_n(\{0,1\}^n) = 1$.
Per l'unione di eventi disgiunti notare che l'unione di due cilindri disgiunti è il cilindro che ha per base l'unione delle basi disgiunte, quindi tutto segue.

Distribuzioni speciali (geometrica e binomiale negativa)

Volgiamo trovare le probabilità (ovvero le distribuzioni) dei seguenti eventi particolari:

- E_k= "il **primo** successo si verifica alla prova k-esima"
- $E_{n,k}$ = "l'n**-esimo** successo si verifica alla prova k-esima"

Vedi costruzioni di relativi modelli.

4 Misura di probabilità su un insieme discreto infinito (estensione di Bernoulli infinito)

Abbiamo visto che $\mathscr{C}$ non è una σ-algebra. Allora, ponendo $\mathscr{F} = \sigma(\mathscr{C})$ la σ-algebra generata da $\mathscr{C}$, ci chiediamo se si possa estendere la P di prima a una $\overline{P}$ che gode della σ-additività.

Estensione di probabilità $\overline{P} : \mathscr{F} \to [0,1]$ è un'estensione di P se $\overline{P}(E) = P(E)$ per ogni $E \in \mathscr{C}$

Teorema (di Caratheodory). *$P : \mathcal{A} \to [0,1]$ probabilità, esiste unica $\overline{P} : \mathcal{F} = \sigma(\mathcal{A}) \to [0,1]$ estensione di P $\iff$ P continua nell'insieme vuoto, ovvero $\forall E_1 \supseteq E_2 \supseteq \ldots$ con $\bigcap_{i=1}^{\infty} E_i = \emptyset$ si ha $\lim_{n\to\infty} P(E_n) = 0$.*
Inoltre, tale estensione è unica.

Quindi tutto il punto sta nel vedere se la prob. di Daniel-Kolmogorov è continua nell'insieme vuoto. Il teorema segue immediatamente dal seguente lemma:

Lemma. *$\{E_n\}_{n\geq 1}$ è una succ. di eventi in $\mathscr{C}$ (algebra dei cilindri) tali che $E_1 \supseteq E_2 \supseteq \ldots$ e $\cap_{i=1}^{\infty} = \emptyset \implies E_n = \emptyset\ \forall n \geq \overline{n}$, per qualche $\overline{n} \in \mathbb{N}$.*

Dimostrazione. Tosta

5 Probabilità su un insieme continuo (costruzione)

Modello continuo ci serve quando abbiamo un esperimento che coinvolge una grandezza continua, come lunghezza/peso/tempo...

- **Spazio degli eventi**: Ω è un segmento (a,b), oppure una semiretta $[0,+\infty)$ o tutto $\mathbb{R}$

- **Algebra:** non possiamo usare le parti di Ω perché contiene i punti, e non vogliamo definire una probabilità sui singoli punti. Usiamo **algebra dei segmenti**:

$$\mathscr{L} := \{\cup_{i=1}^n S_i \mid n \in \mathbb{N}_0,\ S_i \text{ sono segmenti inclusi in } \Omega\}$$

dove se $n = 0$ abbiamo l'insieme vuoto e i segmenti possono essere aperti/chiusi/semiaperti/semichiusi e in cui ogni estremo può essere anche $\pm\infty$

Dimostrazione. (Che $\mathscr{L}$ è un'algebra di parti) Per $\Omega \in \mathscr{L}$ basta prendere $n = 1$ e $S_1 = \Omega$. Poi verificare **l'intersezione** di due eventi (e sottointendere che per induzione possiamo estendere a più eventi). Scrivere $E_1 = \cup_{i=1}^n S_i^{(1)}$ ed $E_2 = \cup_{j=1}^m S_j^{(2)}$, allora l'intersezione è la doppia unione delle intersezioni di segmenti, che sono a loro volta segmenti. Per E^C è il complementare di un'unione, quindi usare legge de Morgan (ricordando che il complementare di un intervallo è al più un'unione di 2 segmenti=semirette).

Dimostrazione. ($\mathscr{L}$ NON è una σ-algebra) Basta mostrare un controesempio: per esempio, l'unione infinita di **sementi disgiunti essenzialmente** (intersezione delle due chiusure è vuota) non può essere scritta come unione finita di altri segmenti, quindi questa unione infinita non sta in $\mathscr{L}$.

- **Probabilità:** fissiamo una densità di probabilità (vedi def. dopo) $f : \Omega \to \mathbb{R}$. Allora definiamo la probabilità di un evento

$$P(E) := \sum_{i=1}^n \int_{S_i} f(x)dx \quad \text{se} \quad \begin{cases} E = \cup_{i=1}^n S_i \\ S_i \cap S_j = \emptyset \quad \forall i \neq j \end{cases}$$

Nonostante l'unione di segmenti si può scrivere in più modi, tale definizione è **ben posta** (la prob. di due scritture diverse della stessa unione è uguale) ed **una probabilità**.

Dimostrazione. (P è ben posta)
Dobbiamo dimostrare che $E = \cup_{i=1}^n S_i^{(1)} = \cup_{j=1}^m S_j^{(2)}$ dove $S_i^{(1)}$ sono tutti disgiunti come pure gli $S_j^{(2)} \implies$ gli ultimi due membri hanno la stessa P. Impostare nel seguente modo:

$$S_i^{(1)} = S_i^{(1)} \cap E = S_i^{(1)} \cap (\sqcup_{j=1}^m S_j^{(2)}) = \bigsqcup_{j=1}^m (S_i^{(1)} \cap S_j^{(2)})$$

Ora la $P(E)$ è la sommatoria degli integrali sugli $S_i^{(1)}$, che posso quindi scrivere come sopra e quindi uso additività degli integrali per ottenere doppia sommatoria su i e j. Scambio le sommatorie per avere all'interno quella in i, riuso additività dell'int. togliendo la sommatoria in i e integrando su $\cup_i(S_i^{(1)} \cap S_j^{(2)}) = S_j^{(2)}$, che è la tesi.

Dimostrazione. (P è una probabilità)
Per $P(\Omega) = \int_\Omega f(x)dx = 1$ per def. di densità.
Per il secondo assioma devo verificare che l'unione disgiunta di due eventi è la somma delle prob. Ma i due eventi sono a loro volta unione disgiunta di segmenti, quindi l'unione degli eventi è l'unione disgiunta di tutti questi segmenti: $E_1 \sqcup E_2 = \bigsqcup_{k=1}^{m+n} T_k$ rinominando i vari segmenti. Ora trovo $P(E_1 \sqcup E_2) =$ sommatoria degli int. sui T_k, uso prop. degli integrali per arrivare a $P(E_1) + P(E_2)$.

6 Misura di prob. su un insieme continuo (estendibilità del modello continuo)

Analogo a ciò fatto per Bernoulli infinito. Notare che la σ-algebra generata da $\mathscr{L}$ coincide con quella di Borel generata dagli aperti.

Parte IV

Variabili aleatorie

Fissiamo uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ dove $\mathcal{F}$ è una σ-algebra e $\mathbb{P} : \mathcal{F} \to [0,1]$ è una misura di probabilità.

Variabile aleatoria È una funzione $X : \Omega \to \mathbb{R}$ misurabile. Abbiamo due condizioni equivalenti di misurabilità:

- $X^{-1}((-\infty, x]) \in \mathcal{F}$ per ogni $x \in \mathbb{R}$.
- $X^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R})$, che è la σ-algebra di Borel (quella generata dagli aperti di $\mathbb{R}$).

NB: ha come dominio Ω, non $\mathcal{F}$. Quindi ha come input i casi elementari ω, non gli eventi A.
NB: è una **funzione**, quindi ogni ω ha una e una sola immagine (valore in $\mathbb{R}$). Per questo le preimmagini di due numeri diversi sono sempre due insiemi disgiunti (se non lo fossero, tali insiemi avrebbero intersezione non nulla, quindi ci sarebbe un ω che vi appartiene, ovvero che avrebbe come immagine entrambi i numeri, andando contro l'ipotesi di funzione)

Funzione di ripartizione (cumulative distribution f.) di una variabile aleatoria X è la funzione

$$F : \mathbb{R} \to [0,1] \quad | \quad F(x) = \mathbb{P}(X^{-1}(-\infty, x]) = \mathbb{P}(X \leq x)$$

Tale probabilità è sempre definita grazie alla condizione di misurabilità di X.

Funzione di ripartizione discreta $F : \mathbb{R} \to [0,1]$ è discreta se, per un **insieme finito o numerabile** I esistono:

- $D \subset \mathbb{R} = \{x_i\}_{i \in I}$ collezione di punti
- $\{p_i\}_{i \in I}$ collezione di numeri in $(0,1] \mid \sum_{i \in I} p_i = 1$

ed F è della forma:

$$F(z) = \sum_{i \in I | x_i \leq z} p_i \qquad \forall z \in \mathbb{R}$$

Funzione di ripartizione assolutamente continua $F : \mathbb{R} \to [0,1]$ si dice assolutamente continua

- su $[a,b]$ se $\forall \varepsilon > 0 \quad \exists \delta > 0 \mid \forall E \subset [a,b]$ rappresentabile come unione disgiunta finita di segmenti, $E = \sqcup_{i=1}^{n} (\alpha_i, \beta_i)$ si ha
$$\sum_{i=1}^{n} (\beta_i - \alpha_i) \leq \delta \implies \sum_{i=1}^{n} [F(\beta_i) - F(\alpha_i)] \leq \varepsilon$$
- su $\mathbb{R}$ se lo è su $[a,b] \quad \forall a < b \in \mathbb{R}$

Funzione di ripartizione continua singolare $F : \mathbb{R} \to [0,1]$ si dice continua singolare se valgono tutte:

- $F \in C^0(\mathbb{R})$ (continua in ogni punto)
- derivabile quasi ovunque (l'insieme $D^C \subset \mathbb{R}$ dei punti di non der. ha misura nulla)
- $F'(z) = 0 \quad \forall z \in D$

Come fa tale funzione ad essere monotona non decrescente (per essere f. di rip) essendo costante (derivata nulla) su quasi tutti i punti? Vuol dire che cresce su un **insieme frattale**, invisibile alla misura lineare. Esempio: **scala di Cantor / del diavolo**, dove si vede proprio il concetto di frattale.

Teoremi

Equivalenza tra condizioni di misurabilità

$X^{-1}((-\infty, x]) \in \mathcal{F} \quad \forall x \in \mathbb{R} \iff X^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R})$

Dimostrazione.

- $\Leftarrow$) ovvio, poiché $(-\infty, x] \in \mathcal{B}(\mathbb{R})$, poiché è un chiuso e tale algebra ha i chiusi.

- $\Rightarrow$) Osservo inanzitutto che $\mathcal{B}(\mathbb{R}) = \sigma\{(-\infty, z] \mid z \in \mathbb{R}\}$. Prendo poi $\mathcal{C} \coloneqq \left\{C \subseteq \mathbb{R} \mid X^{-1}(C) \in \mathcal{F}\right\}$, noto che $(-\infty, z] \in \mathcal{C}$ (per ipotesi) e che $\mathcal{C}$ è σ-algebra poiché:

  $X^{-1}(\mathbb{R}) \in \mathcal{F} \implies \Omega \in \mathcal{F}$
  $X^{-1}(C) \in \mathcal{F} \implies X^{-1}\left(C^C\right) = X^{-1}(C)^C \in \mathcal{F}$ preimm. commuta con compl.
  $X^{-1}(C_i) \in \mathcal{F} \implies X^{-1}\left(\bigcup_{i=1}^{+\infty} C_i\right) = \bigcup_{i=1}^{+\infty} X^{-1}(C_i) \in \mathcal{F}$ preimm. commuta con unione.

  quindi, essendo che
  - $\mathcal{B}(\mathbb{R})$ è la più piccola σ-algebra che contiene $(-\infty, x]$
  - $(-\infty, x] \subset \mathcal{C}$
  - $\mathcal{C}$ è una σ-algebra

  $\implies \mathcal{B}(\mathbb{R}) \subseteq \mathcal{C}$, dove $\mathcal{C}$ è l'insieme di tutti gli insiemi la cui preimmagine sta in $\mathcal{F} \implies$ tesi

Lemma continuità dal basso/alto

Dal basso (successione crescente di eventi):

$$E_1 \subseteq E_2 \subseteq \ldots \implies \lim_{n\to+\infty} P(E_n) = P\left(\bigcup_{i=1}^{+\infty} E_i\right)$$

Dall'alto (successione descrescente di eventi):

$$E_1 \supseteq E_2 \supseteq \ldots \implies \lim_{n\to+\infty} P(E_n) = P\left(\bigcap_{i=1}^{+\infty} E_i\right)$$

Dimostrazione.

- dal basso: Prendo $F_1 \coloneqq E_1$, $F_n \coloneqq E_n \setminus E_{n-1}$, osservo che $F_i \cap F_j = \emptyset$ per $i < j$. Per induzione dimostro che $E_n = \bigsqcup_{i=1}^{n} F_i$ (prima con $n = 1$, poi con $n \implies n+1$), quindi applicando la σ-additività
  $$P\left(\bigcup_{i=1}^{+\infty} E_i\right) = P\left(\bigsqcup_{i=1}^{+\infty} F_i\right) = \sum_{i=1}^{+\infty} P(F_i) = \lim_{n\to+\infty} \sum_{i=1}^{n} P(F_i) = \lim_{n\to+\infty} P\left(\bigsqcup_{i=1}^{n} F_i\right) = \lim_{n\to+\infty} P(E_n)$$
- dall'alto: considero $G_i = E_i^c$, e diventa dal basso (ovvero dall'insieme 1 tolgo buchi sempre più piccoli, quindi l'insieme n ingrandisce).

Prorpietà f. di ripartizione

F è una f. di ripartizione, allora valgono:

(a) $x \leq y \implies F(x) \leq F(y)$ (è monotona non decrescente)

(b) $\lim_{h\to x^+} F(x+h) = F(x)$ (F è continua da destra)

(c) $\lim_{x\to-\infty} F(x) = 0, \quad \lim_{x\to\infty} F(x) = 1$

Dimostrazione.

(a) $x \leq y \implies E_1 = \{X \leq x\} \subseteq E_2 = \{X \leq y\}$ (ragionamento facile) e quindi, per **monotonia della prob.** $F(x) = \mathbb{P}(E_1) \leq \mathbb{P}(E_2) = F(y)$, ovvero la tesi.

(b) Essendo F monotona, la tesi è equivalente a

$$\lim_{n\to+\infty} F(z + \frac{1}{n}) = F(z)$$

ovvero ridurre h che (varia nel continuo) ad una successione (che varia nel discreto).
Notare che la tesi originale implica sempre quetsa, ma il viceversa necessita della monotonia (vedi un seno che in z ha frequenza che aumenta come $1/n$, quindi possiamo costruire una successione che sta sempre sulle creste e tende ad un limite diverso da quello se variassimo nel continuo).
Definiamo $E_n \coloneqq \{X \leq z + 1/n\}$, per (a) vediamo che $E_1 \supseteq E_2 \supseteq \ldots$ quindi applichiamo **continuità dall'alto**:

$$\lim_{n\to+\infty} F(z + \frac{1}{n}) = \lim_{n\to+\infty} \mathbb{P}(E_n) = \mathbb{P}(\cap_{i=1}^{+\infty} E_i)$$

Ora dobbiamo dimostrare che $\{X \leq z\} = (\cap_{i=1}^{+\infty} E_i)$:

$\subseteq$) Tale relazione vale per ogni E_i, e quindi anche per l'intersezione.

$\supseteq$) Dobbiamo vedere che se preso un ω nell'intersezione., allora $X(\omega) \leq z$. Infatti $X(\omega) \leq z + 1/i \quad \forall i \in \mathbb{N}$, allora basta fare il limite ad ambo i membri e arrivare alla tesi.

(c) (come (b)) Sempre per la monotonia le due tei sono equivalenti a:

$$\lim_{n\to+\infty} F(-n) = 0 \qquad \text{e} \qquad \lim_{n\to+\infty} F(n) = 1$$

1. Definiamo $E_n \coloneqq \{X \leq -n\}$, per (a) vediamo che $E_1 \supseteq E_2 \supseteq \dots$ quindi applichiamo **continuità dall'alto**:

$$\lim_{n\to+\infty} F(-n) = \lim_{n\to+\infty} \mathbb{P}(E_n) = \mathbb{P}(\cap_{i=1}^{+\infty} E_i) = \mathbb{P}(\emptyset) = 0$$

2. Definiamo $E_n \coloneqq \{X \leq n\}$, per (a) vediamo che $E_1 \subseteq E_2 \subseteq \dots$ quindi applichiamo **continuità dal basso**:

$$\lim_{n\to+\infty} F(n) = \lim_{n\to+\infty} \mathbb{P}(E_n) = \mathbb{P}(\cup_{i=1}^{+\infty} E_i) = \mathbb{P}(\Omega) = 1$$

Formule probabilità di una v.a. dalla CDF

1. $\mathbb{P}(X \in (a,b]) = F(b) - F(a)$
2. $\mathbb{P}(X \in (a,b)) = \lim_{z\to b^-} F(z) - F(a)$
3. $\mathbb{P}(X \in [a,b]) = F(b) - \lim_{z\to a^-} F(z)$
4. $\mathbb{P}(X \in [a,b)) = \lim_{z\to b^-} F(z) - \lim_{w\to a^-} F(w)$
5. $\mathbb{P}(X = a) = F(a) - \lim_{z\to a^-} F(z)$

Dimostrazione.

1. $(a,b] = \overbrace{(-\infty,b]}^{=E} \setminus \overbrace{(-\infty,a]}^{=F}$, con $F \subset E \implies \mathbb{P}(X \in (a,b]) = \mathbb{P}(E \setminus F) = \mathbb{P}(E) - \mathbb{P}(F) = F(b) - F(a)$
2. scrivere $(a,b) = \cup_{n=1}^{+\infty}(a, b - 1/n]$, poi usare continuità dal basso, usare 1) e poi monotonia di F per togliere $1/n$
3. Scrivere $[a,b] = \cap_{n=1}^{+\infty}(a - 1/n, b]$, procedere come prima usando continuità dall'alto.
4. Scrivere $[a,b) = \{a\} \sqcup (a,b)$, procedere come sopra usando 2) e 5)
5. Porre $b = a$ in 3)

Teorema di Rappresentazione di Skorokhod

Monotonia non decrescente, continuità da destra e $\lim_{z\to-\infty} F(z) = 0$, $\lim_{z\to+\infty} F(z) = 1$ caratterizzano le funzioni di ripartizione.
In particolare, se una funzione F ha tali caratteristiche, esiste sempre una v.a. di cui F è la funzione ri ripartizione.

Dimostrazione. Prendo $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, densità di probabilità uniforme $f(x) = 1$, $\forall x \in [0,1]$ ed estensione della probabilità sull'algebra dei segmenti in $[0,1]$ a una misura su $\mathcal{F}$ (si può fare ed è unica per teorema). Definisco $X(\omega) \coloneqq \inf\{x \in \mathbb{R} \mid F(x) \geq \omega\}$, e dimostro che $\{X(\omega) \leq z\} = [0, F(z)]$, ovvero

$$X(\omega) \leq z \iff \omega \leq F(z)$$

(visualizzare tutto ciò nel plot della f. di ripartizione con ω preso tra 0 e 1 sull'asse delle y e $X(\omega)$ sull'asse delle x)

$$\dots$$

Una volta dimostrato $\implies F_X(z) = P(\{X \leq z\}) = P([0, F(z)]) = F(z)$, ovvero la tesi.

Proprietà aggiuntive delle funzioni di ripartizione

- Discontinuità solo di tipo salto
- Finitezza o numerabilità dei punti di discontinuità
- Derivabilità in quasi ogni punto

Lipschitzianità $\Rightarrow$ Assoluta continuità

Condizione sufficiente di assoluta continuità (Lebesgue)

Sia $f : \mathbb{R} \to \mathbb{R}$ densità di probabilità, se

$$\forall z \in \mathbb{R} : \quad F(z) = \int_{-\infty}^{z} f(x)\, dx = \lim_{M \to +\infty} \int_{-M}^{z} f(x)\, dx$$

allora F assolutamente continua. Quindi $C^1(\mathbb{R}) \Rightarrow$ assolutamente continua per il teorema fondamentale del calcolo.

Derivabilità quasi ovunque delle funzioni assolutamente continue

F funzione di ripartizione assolutamente continua, allora $\exists f : \mathbb{R} \to \mathbb{R}$ densità di probabilità tale che $F(x) = \int_{-\infty}^{x} f(x)\, dx$, e $F'(x) = f(x)$ q.o. $\forall x \in \mathbb{R}$

Teorema di decomposizione di Lebesgue

Ogni $F : \mathbb{R} \to [0,1]$ funzione di ripartizione è scrivibile come combinazione convessa di $F_{\mathrm{d}}, F_{\mathrm{ac}}, F_{\mathrm{cs}}$ funzioni di ripartizione rispettivamente discreta, assolutamente continua e continua singolare

Parte V
Speranza matematica (integrale di Riemann-Stieltjes)

Definziamo la speranza matematica / valore atteso di X a seconda del codominio di X

Codominio finito $Im(X) = \{x_1, \dots, x_n\}$, con $x_1, \dots, x_n$ numeri distinti, si pone:

$$\mathbb{E}[X] := \sum_{i=1}^{n} x_i p_i$$

dove $p_i := \mathbb{P}[X^{-1}(x_i)]$. (dim. che vale anche per numeri uguali)

Somma di Riemann-Stieltjes relativa alla f. di rip. F e alla suddivisione data dai punti $0 = x_0 < x_1 < x_2 < \dots < x_n$:

$$S_n(F; \Delta_n) := \sum_{i=1}^{n-1} x_i(F(x_{i+1}) - F(x_i)) + x_n(1 - F(x_n))$$

dove $\Delta_n = \{x_i\}_{i=0,1,\dots,n}$

Codominio incluso in $[0, +\infty)$ allora poniamo

$$\mathbb{E}[X] := \lim_{n\to\infty} S_n(F; \Delta_n) \overset{TEO}{=} \int_0^{+\infty} [1 - F(x)]dx$$

Ciò estende la prima definizione (la prima è un caso particolare di quetsa)

Codominio qualunque Definiamo:

- $X_+(\omega) := \max\{X(\omega), 0$
- $X_-(\omega) := \max\{-X(\omega), 0$

Se:

- $\mathbb{E}[X_+] < +\infty$ (ben definita) $\overset{Teo}{\iff} \int_0^{+\infty}[1 - F_+(x)]dx = \int_0^{+\infty}[1 - F(x)]dx < +\infty$
- $\mathbb{E}[X_-] < +\infty$ (ben definita) $\overset{Disp.}{\iff} \int_0^{+\infty}[F(-x)]dx < +\infty$

allora poniamo

$$\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-] \overset{disp.}{=} \int_0^{+\infty} [1 - F(x)]dx - \int_{-\infty}^{0} F(x)dx$$

Prop. se F_{ac} è assolutamente continua con densità $f : \mathbb{R} \to \mathbb{R}$ allora:

$$\mathbb{E}[X_{ac}] = \int_{-\infty}^{+\infty} xf(x)dx$$

Prop. Se X ha codominio numerabile e vale $\sum_{i=1}^{+\infty} |x_i|p_i < +\infty$ allora:

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i p_i$$

Prop. (valore atteso di funzioni di v.a.) Se $g : \mathbb{R} \to \mathbb{R}$

- **codominio finito:** $\mathbb{E}[g(X)] = \sum_{i=1}^{n} g(x_i)p_i$

  Dimostrazione. Scrivere $g(X) = \sum_{i=1}^{n} g(x_i)\mathbb{1}_{\{X=x_i\}}$ e o usare linearità oppure per def. di valore atteso discreto (non cambiano le prob. degli eventi).
- **codominio numerabile:** $\mathbb{E}[g(X)] = \sum_{i=1}^{+\infty} g(x_i)p_i$ se assolutamente convergente
- **v.a. assolutamente continua:** $\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx$ se assolutamente convergente.

  Dimostrazione. Dimostrare nel caso in cui f ha supporto su $[a, b]$ e g è un diffeomorfismo C^1 con inversa C^1) su tale intervallo.

Teoremi

Proprietà della speranza matematica

Linearità	$\mathbb{E}[aX+bY]=a\mathbb{E}[X]+b\mathbb{E}[Y]$ $\left(X,Y\in L^1(\Omega)\right)$	Normalizzazione	$\mathbb{E}[1]=1$	
Positività	$X\geq 0 \implies \mathbb{E}[X]\geq 0$	Jensen	$\lvert\mathbb{E}[X]\rvert\leq\mathbb{E}[\lvert X\rvert]$	$\left(\begin{array}{c}\lvert\cdot\rvert\to\varphi\\ \text{per }\varphi\text{ convessa}\end{array}\right)$
Monotonia	$X\leq Y \implies \mathbb{E}[X]\leq\mathbb{E}[Y]$	Beppo Levi	$\underset{(X_n\geq 0)}{X_n}\overset{n\to\infty}{\longrightarrow}X \implies \mathbb{E}[X_n]\overset{n\to\infty}{\longrightarrow}\mathbb{E}[X]$	

Costruzione della speranza matematica per X con codominio $\subseteq[0,+\infty)$

$\mathbb{E}[X]=\int_0^{+\infty}[1-F(x)]\,dx$ oppure $\mathbb{E}[X]=\sup_{0\leq Z\leq X}\mathbb{E}[Z]$ con Z v.a. discreta

Dimostrazione. Prendo $[0,+\infty)=\{0\}\cup(0,x_1]\cup\cdots\cup(x_{n-1},x_n]\cup(x_n,+\infty)$, e prendo variabile discreta $X^{(d)}$che approssima per difetto X, da cui definisco la somma di

$$S_n(F;\Delta_n):=\mathbb{E}\left[X^{(d)}\right]=\sum_{i=1}^{n-1}x_i\left(F(x_{i+1})-F(x_i)\right)+x_n\left(1-F(x_n)\right)$$

Considerando ora $\{\Delta_n\}_{n\geq 1}$ con $\Delta_n=\{x_i^{(n)}\}_{i=0,1,\ldots,k_n}$tale che $\Delta_n\subset\Delta_{n+1}$, $\|\Delta_n\|\overset{n\to\infty}{\longrightarrow}0$, $x_{k_n}^{(n)}\overset{n\to\infty}{\longrightarrow}+\infty$, allora se converge l'integrale $\int_0^{+\infty}$ si ha che

$$\mathbb{E}[X]:=\lim_{n\to+\infty}S_n(F;\Delta_n)=\int_0^{+\infty}[1-F(x)]\,dx$$

Questo perchè la successione $S_n(F;\Delta_n)$ è monotona non decrescente e per la formula di Abel si ha

$$S_n(F;\Delta_n)=\sum_{i=1}^{n}(x_i-x_{i-1})(1-F(x_i))\leq\int_0^{x_n}[1-F(x)]\,dx\leq\int_0^{+\infty}[1-F(x)]\,dx$$

Calcolo della speranza matematica per una v.a. assolutamente continua

$\int_0^{+\infty}[1-F_{\text{ac}}(x)]\,dx-\int_{-\infty}^0F_{\text{ac}}(x)\,dx=\int_{-\infty}^{+\infty}xf(x)\,dx$

Dimostrazione

$$\int_0^M[1-F_{\text{ac}}(x)]\,dx=\int_0^M\left[\int_x^{+\infty}f(y)\,dy\right]dx=\int_0^M\left[\lim_{T\to+\infty}\int_x^{M+T}f(y)\,dy\right]dx=\lim_{T\to+\infty}\int_0^M\left[\int_x^{M+T}f(y)\,dy\right]dx$$

$$\int_0^M\left[\int_x^{M+T}f(y)\,dy\right]dx=\int_0^Mf(y)\left[\int_0^ydx\right]dy+\int_M^{M+T}f(y)\left[\int_0^Mdx\right]dy=\int_0^Myf(y)\,dy+M\int_M^{M+T}f(y)\,dy$$

$$\int_0^M[1-F_{\text{ac}}(x)]\,dx=\int_0^Myf(y)\,dy+M[1-F_{\text{ac}}(M)]\,dy\overset{M\to+\infty}{\longrightarrow}\int_0^{+\infty}yf(y)\,dy$$

Valore atteso della funzione di una v.a.

Disuguaglianza di Cauchy - Schwartz

Dati $X,Y\in L^2$, allora $XY\in L^1$ e $\lvert\mathbb{E}[XY]\rvert\leq\left(\mathbb{E}\left[X^2\right]\right)^{1/2}\left(\mathbb{E}\left[Y^2\right]\right)^{1/2}$, e in particolare $L^2\subset L^1$ e $(\mathbb{E}[X])\leq\mathbb{E}\left[X^2\right]$

Dimostrazione $XY\in L^1$ in quanto $\lvert XY\rvert\leq\frac{X^2}{2}+\frac{Y^2}{2}$, e con $Y=1$ ho $\lvert X\rvert\leq\frac{X^2}{2}+\frac{1}{2}\Rightarrow L^2\subset L^1$. Per dimostrare la disuguaglianza scrivo

$$\mathbb{E}[xX+Y]^2=x^2\mathbb{E}\left[X^2\right]+2x\mathbb{E}[XY]+\mathbb{E}\left[Y^2\right]\geq\mathbf{0}\qquad\Delta/4=\mathbb{E}[XY]^2-\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]\leq\mathbf{0}$$

Disuguaglianza di Markov

Siano $X:\Omega\longrightarrow\mathbb{R}$ v.a. e $a\geq 0$, allora $\mathbb{P}(\lvert X\rvert>a)\leq\frac{\mathbb{E}[\lvert X\rvert]}{a}$

Dimostrazione Preso $A := \{\omega \in \Omega | \, |X(\omega)| > a\}$, ho $|X| \geq a \cdot \mathbf{1}_A$, dunque

$$\mathbb{E}[|X|] \geq \mathbb{E}[a \cdot \mathbf{1}_A] = a\mathbb{E}[\mathbf{1}_A] = a\mathbb{P}(A) = a\mathbb{P}(|X| > a)$$

Disuguaglianza di Chebychev

Siano $X : \Omega \longrightarrow \mathbb{R}$ v.a. e $a \geq 0$, allora $\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}[X^2]}{a^2}$ (equivalentemente $\mathbb{P}(|X - \mathbb{E}[X]| > a) \leq \frac{\mathrm{Var}(X)}{a^2}$)

Dimostrazione Ho che $X^2 \geq a^2 \cdot \mathbf{1}_{\{|X|>A\}}$, dunque $\mathbb{E}\left[|X|^2\right] \geq \mathbb{E}\left[a^2 \cdot \mathbf{1}_A\right] = a^2\mathbb{P}(|X| > a)$

Parte VI
Vettori aleatori

Vettore aleatorio Dato uno spazio di probabilità $(\Omega, \mathscr{F}, \mathbb{P})$ un vettore aleatorio è una funzione $(X, Y) : \Omega \to \mathbb{R}^2$ misurabile. Come condizione di misurabilità sono equivalenti:

- X, Y sono v.a. (misurabili)
- (X, Y) soddisfa misurabilità congiunta, ovvero $\forall B \in \mathcal{B}(\mathbb{R}^2)$ (ovvero la più piccola σ-algebra generata dai rettangoli aperti di $\mathbb{R}^2$) : $(X, Y)^{-1}(B) \in \mathscr{F}$

Funzione di ripartizione congiunta di X e Y è $F_{X,Y}(x, y) = \mathbb{P}[X \leq x, Y \leq y]$

Prop. Per la funzione di rip. congiunta valgono:

1. Quasi monotona
2. Continua da destra
3. Limite per le due variabili all'inf. è 1, limite per una fissa e l'altra a meno inf. è 0.

Funzioni di ripartizione marginali Sono:

- $F_X(x) = \mathbb{P}[X \leq x] = \lim_{y \to +\infty} F_{X,Y}(x, y)$
- $F_Y(y) = \mathbb{P}[Y \leq y] = \lim_{x \to +\infty} F_{X,Y}(x, y)$

Fdr dicreta

Fdr assolutamente continua

Prop. X, Y indipendenti $\implies \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Dimostrazione.

- **codominio finito:** scrivere variabili come sommatorie di funz. indicatrici, su valori distinti di X o Y. Fare valore atteso del prodotto, usare linearità e il prodotto di due funz. indicatrici è la funz. indicatrie dell'intersezione.
- **codominio positivo:** prendere successioni X_n, Y_n con cod. finito che tendono a X, Y. Per continuità del valore atteso, esso è il limite dei valori attesi, che possiamo calcolare poiché in codominio finito.
- **in generale:** scrivere le v.a. come $X = X_+ - X_-$ ecc. Fare il val atteso proprio del prodotto delle due differenze, usare linearità e poi il fatto che sono tutte a codominio positivo.

Parte VII
Variabili aleatorie discrete

Variabile aleatoria discreta Se ha come immagine un sottoinsieme numerabile $\{x_1, x_2, \dots\} \subset \mathbb{R}$

Funzione di probabilità di massa (probability mass f.) di una variabile aleatoria discreta X è la funzione

$$f : \mathbb{R} \to [0,1] \quad | \quad f(x) = \mathbb{P}(X = x)$$

Lemma (caratterizzazione f. di prob. di massa) $f : \mathbb{R} \to [0,1]$ è una funzione di prob. di una v.a. discreta se e solo se:

(a) L'insieme $\{x \in \mathbb{R} \mid f(x) \neq 0\}$ è numerabile

(b) $\sum_i f(x_i) = 1$ dove $x \in \{x \in \mathbb{R} \mid f(x) \neq 0\}$

Oss. Funzione di ripartizione e funzione di probabilità sono legate dalle seguenti:

CDF data PMF $F(x) = \sum_{i|x_i \leq x} f(x_i)$

PMF data CDF $f(x) = F(x) - \lim_{y \uparrow x} F(y)$

Lemma La funzione di probabilità soddisfa:

(a) L'insieme $\{x \mid f(x) \neq 0\}$ è numerabile

(b) $\sum_i f(x_i) = 1$ dove $x_1, x_2 \dots$ sono le $x \mid f(x) \neq 0$

V.a. indipendenti X e Y sono indipendenti se gli eventi $\{X = x\} = X^{-1}(x) \in \mathcal{F}$ e $\{Y = y\} = Y^{-1}(y) \in \mathcal{F}$ sono indipendenti per ogni x, y.

Teorema X e Y indipendenti, $g, h : \mathbb{R} \to \mathbb{R} \implies g(X)$ e $h(Y)$ sono indipendenti.

Famiglia di v.a.d. indipendenti $\{X_i \mid i \in I\}$ sono indipendenti se gli eventi $\{X_i = x_i\}$ sono indipendenti per tutte le possibili scelte degli insiemi $\{x_i \mid i \in I\}$. Ovvero $\{X_i \mid i \in I\}$ è una famiglia indipendente se e solo se:

$$\mathbb{P}(X_i = x_i \forall i \in J) = \prod_{i \in J} \mathbb{P}(X_i = x_i)$$

Valore atteso/media (expectet value/expectation/mean) di una variabile aleatoria con PMF f è:

$$\mathbb{E}(X) \coloneqq \sum_{x|f(X)>0} x f(x) \quad \text{se è assolutamente convergente}$$

È una media pesata rispetto alle probabilità dei vari valori di X. Se f fosse una densità di massa allora $\mathbb{E}(X)$ sarebbe il centro di massa.

Lemma (formula cambio di variabile) Se $g : \mathbb{R} \to \mathbb{R}$ allora: $\mathbb{E}(g(X)) = \sum_x g(x) f(x)$

Momento $k \in \mathbb{N}_0$, il k-momento di X è: $m_k = \mathbb{E}(X^k)$

$m_0 = 1$ (Massa, normalizzata)

$m_1 = \mathbb{E}(X)$ **valore atteso** (centro di massa). Misura il valore medio che assume X. Indicata con μ

m_2 (Momento d'inerzia rispetto all'asse $x = 0$, infatti è la media pesata rispetto alle prob. delle distanze al quadrato da $x = 0$)

Momento centrale $k \in \mathbb{N}_0$, il k-momento centrale di X è: $\sigma_k = \mathbb{E}((X - m_1)^k)$.
I concetti sono identici a prima solo che ci siamo spostati nel sistema di riferimento del centro di massa (ovvero della media, attraverso la traslazione $X - \mathbb{E}X$)

$\sigma_0 = 1$

$\sigma_1 = 0$ (coordinata del centro di massa nel sist. di rif. del c.m.)

$\sigma_2 = \mathbb{E}((X - \mathbb{E}X)^2)$ **varianza**. Misura quanta tendenza ha X ad assumere valori diversi da quello medio. Indicata con $\mathrm{var}(X)$. È il momento d'inerzia rispetto all'asse $x = \mu$

Deviazione standard $\sigma = \sqrt{var(X)} = \sqrt{\sigma_2}$. È una normalizzazione della varianza.

LEMMA i momenti centrali $\{\sigma_i\}$ possono essere scritti in termini di momenti $\{m_i\}$. Ad esempio

$$\sigma_2 = m_2 - m_1^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2$$

Per dimostrarlo basta applicare la def. di σ_2 e usare la formula di cambio variabile.

TEOREMA L'operatore $\mathbb{E}$ ha le seguenti proprietà:

(a) $X \geq 0 \implies \mathbb{E}(X) \geq 0$

(b) $a, b \in \mathbb{R}$ allora $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ (è un **operatore lineare**)

(c) La variabile aleatoria 1 (ha sempre 1 come immagine) ha $\mathbb{E}(1) = 1$

NB: tutta la teoria della probabilità può essere riformulata partendo da qua, ovvero ponendo queste proprietà come assiomi e quindi definendo la misura di probabilità in funzione della media ovvero: $\mathbb{P}(A) := \mathbb{E}(I_A)$.

LEMMA X e Y indipendenti $\implies \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Il viceversa non è vero in generale.

V.a.d. non correlate (uncorrelated) Se $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
Quindi variabili indipendenti $\implies$ non correlate, ma non viceversa

TEOREMA X e Y due variabili aleatorie. Allora:

(a) $var(aX) = a^2 var(X)$ (**non linearità** dell'operatore var)

(b) se X e Y non correlate $\implies var(X + Y) = var(X) + var(Y)$

Dim: usare var in funzione di $\mathbb{E}$ e poi la linearità di $\mathbb{E}$.

Vettori aleatori e condizionamenti (caso discreto)

Vettori aleatori Siano X e Y due variabili aleatorie sullo stesso spazio di probabilità. Allora il vettore aleatorio è: $\mathbf{X} = [X, Y] : \Omega \to \mathbb{R}^2$

Funzione di ripartizione congiunta (Joint CDF) di X e Y è

$$F : \mathbb{R}^2 \to [0, 1] \quad | \quad F(x, y) = \mathbb{P}(X \leq x \wedge Y \leq y)$$

Denominata con $F_{X,Y}$

Funzione di probabilità congiunta (Joint PMF) di X e Y è

$$f : \mathbb{R}^2 \to [0, 1] \quad | \quad \begin{aligned} f(x, y) &= \mathbb{P}(X = x \wedge Y = y) \\ &= \mathbb{P}(\mathbf{X} = (x, y)) \end{aligned}$$

Denominata con $f_{X,Y}$
NB: è una **funzione**, ovvero le preimmagini di due punti del piano distinti sono distinte. Infatti se così non fosse ci sarebbe un $\omega \in \overbrace{(X_1^{-1} \cap Y_1^{-1})}^{\text{preim. primo punto}} \cap \overbrace{(X_2^{-1} \cap Y_2^{-1})}^{\text{preim. secondo punto}}$, ma ciò implica che appartiene anche all'intersezione $X_1^{-1} \cap X_2^{-1}$, andando contro l'ipotesi che X (e Y) siano funzioni.

LEMMA X e Y sono indipendenti $\iff f_{X,Y}(x, y) = f_X(x)f_Y(y)$ per ogni $x, y \in \mathbb{R}$
Più in generale sono indipendenti $\iff f_{X,Y}(x, y)$ può essere fattorizzato come prodotto $g(x)h(y)$ di due funzioni, una solo in x e l'altra solo in y

LEMMA $\mathbb{E}(g(X, Y)) = \sum_{x,y} g(x, y) f_{X,Y}(x, y)$

Covarianza di X e Y è:

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

Il concetto di covarianza generalizza quello di varianza, in quanto $cov(X, X) = var(X)$.
Espandendo la def e usando la linearità di $\mathbb{E}$ otteniamo:

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

Coefficiente di correlazione di X e Y è:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

È una normalizzazione della covarianza, in modo che ρ assuma valori in $[-1,1]$

OSS. X e Y sono non correlate $\iff cov(X,Y) = 0$

LEMMA $|\rho(X,Y)| \leq 1$, con uguale $\iff \exists a,b,c \in \mathbb{R} \mid \mathbb{P}(aX + bY = c) = 1$
Si dimostra con la dis. di Cauchy-Schwarz.

TEOREMA (DISUGUALIANZA DI CAUCHY-SHWARZ) $\{\mathbb{E}(XY)\}^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$ con uguale $\iff \exists a,b \in \mathbb{R}$ con almeno uno $\neq 0 \mid \mathbb{P}(aX = bY) = 1$

OSS. $\rho = +1(-1) \iff Y$ cresce (decresce) **linearmente** quando X cresce.

Funzione di ripartizione condizionata di Y dato X è $F_{Y|X}(y \mid x) := \mathbb{P}(Y \leq y \mid X = x)$ per ogni $x \mid \mathbb{P}(X = x) > 0$

Funzione di probabilità condizionata di Y dato X è $f_{Y|X}(y \mid x) := \mathbb{P}(Y = y \mid X = x)$ per ogni $x \mid \mathbb{P}(X = x) > 0$.
Usando la def. di prob. condizionata si può anche scrivere come $f_{Y|X} = f_{X,Y}/f_X$

OSS. X e Y sono indipendenti $\iff f_{Y|X} = f_Y$

Valore atteso condizionato di Y dato X è $\psi(x) := \mathbb{E}(Y \mid X = x) = \sum_y y f_{Y|X}(y \mid x)$ al variare di x.
Non è un numero, ma funzione di X, quindi una variabile aleatoria a sua volta.

TEOREMA $\mathbb{E}(\psi(X)) = \mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(Y)$
Da ciò segue una nuova espressione per il valore atteso (analogo alla formula della prob. totale):

$$\mathbb{E}(Y) = \sum_x \mathbb{E}(Y \mid X = x)\mathbb{P}(X = x)$$

Più in generale: $\mathbb{E}[\psi(X)g(X)] = \mathbb{E}[Yg(X)]$

TEOREMA (CONVOLUZIONE) $\mathbb{P}(X + Y = z) = \sum_x f_{X,Y}(x, z - x) = \sum_x \mathbb{P}(X = x \wedge Y = z - x)$
Se X e Y sono indipendenti, allora

$$\mathbb{P}(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x)f_Y(z - x) = \sum_y f_X(z - y)f_Y(y)$$

Gli ultimi due membri sono detti **convoluzione discreta** di f_X e f_Y

Parte VIII

Variabili aleatorie continue

Variabile aleatoria continua Se la sua funzione di ripartizione può essere espressa come integrale, ovvero nella forma:

$$F(x) = \int_{-\infty}^{x} f(u)du \quad x \in \mathbb{R}$$

dove $f(u)$ è una funzione integrabile, $f : \mathbb{R} \to [0, \infty)$, chiamata funzione di densità di probabilità (prob. density f.)

Funzione di densità di probabilità (pr. density f.) $f : \Omega = \mathbb{R} \to \mathbb{R}$ è una densità di probabilità se soddifa:

i) $f(x) \geq 0$ per ogni $x \in \Omega$

ii) f è integrabile secondo Riemann su ogni segmento limitato $S \subseteq \Omega$, in senso proprio (f limitata) o in senso improprio (f illimitata).

iii) $\int_\Omega f(x)dx = 1$

Quando f è illimitata, per esempio su $[a, b)$ limitato, si dice che è integrabile se è finito il limite $\lim_{k\to+\infty} \int_S f_k(x)dx$ dove $f_k(x) \coloneqq min\{f(x), b - \frac{1}{k}\}$. Se il segmento è illimitato, è integrabile se è finito il limite per un estremo o entrambi che tendono a $\pm\infty$

Funzione di ripartizione

Oss. Probabilità su un insieme $B \subset \mathbb{R}$: $\int_B f(x)dx$

Valore atteso $\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx$

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

Lemma Se X ha p.d.f. $f(x) = 0$ se $x = 0$, allora $\mathbb{E}(X) = \int_0^{+\infty} [1 - F(x)]dx$

Vettori aleatori continui

Densità marginale

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)dx$$

Valore atteso Se $g : \mathbb{R}^2 \to \mathbb{R}$ allora:

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y)dxdy$$

(vedi caso in cui siano indip)

Cambio variabile di densità congiunta densità(funz-cambiovariabile) per det(jacobiano) del cambio

Somma di var. Vedi caso discreto e usa integrale da -inf a +inf

Variabili aleatorie miste (discrete+continue)

Sono v.a. che hanno una parte continua (ovvero per quei valori la CDF è continua ed c'è il concetto di densità di prob.) e parte discreta (ovvero esiste un singolo valore la cui rpob. è $\neq 0$, e in quel punto si ha concetto di prob. di massa e la CDF ha un salto).

Esempio: `https://www.probabilitycourse.com/chapter4/4_3_1_mixed.php`

Se derivo le parti continue della CDF (quindi saltando i punti di discontinuità) esse non saranno una PDF perché l'integrale non vale 1, ma vale uno se ci sommiamo la sommatoria delle probabilità dei punti in cui è discreta. La CDF di ogni v.a. mista può essere scritta come **somma di una parte continua e una discontinua**.

Parte IX

Funzioni generatrici e caratteristiche

Funzione generatrice delle probabilità

Ordinary generating function (OGF) di una successione a_n è:

$$G(a_n, x) = \sum_{n=0}^{\infty} a_n x^n$$

È la serie di potenze con coefficienti i termini della successione.

Funzione generatrice delle probabilità Definita per una v.a. X discreta a valori in $\mathbb{N}_0$ (**discreta positiva**) è

$$\begin{aligned} G(z) &= \mathbb{E}[z^X] \qquad \text{per } z \in \mathbb{R} \text{ in cui converge} \\ &= \sum_{n=0}^{\infty} z^n f(n) \end{aligned}$$

È la OGF della successione $\{f(n)\}_n$ delle probabilità di massa. È una **serie di potenze**, che quindi converge sicuramente per $z \in (-1, 1)$, ma inoltre per $z = 1$ converge in quanto la sommatoria (delle probabilità) vale 1, quindi converge anche per $-1 \implies$ **definita almeno in $-1 \leq z \leq 1$**. Il raggio di convergenza può essere anche maggiore, per esempio nel caso in cui X assume un numero finito di valori.

Oss. Le funzioni generatrici caratterizzano la PMF, ovvero

$$X, Y \text{ hanno la stessa PMF} \iff \text{hanno la stessa funzione caratteristica}$$

Dimostrazione. $\Rightarrow$) ovvio
$\Leftarrow$) vedi teoria serie di potenze. Con la seguente formula posso ricavare il coefficiente n-esimo della serie:

$$f(n) = \frac{1}{n!} \frac{d^n G_X}{dx^n}(0)$$

Calcolare la derivata n-esima serve per rimuovere tutti i termini a sinistra della serie (minori di n), calcolarla in zero serve per rimuovere tutti i termini a destra (maggiori di n).

Prop. Se X, Y sono indipendenti, allora $G_{X+Y}(z) = G_X(z)G_Y(z)$

Dimostrazione.

$$G_{X+Y}(z) \overset{\text{def}}{=} \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] \overset{\star}{=} \mathbb{E}[z^X]\mathbb{E}[z^Y] \overset{\text{def}}{=} G_X(z)G_Y(z)$$

$\star$ = per teo. funzioni di v.a. indip sono indip, e per lemma indip $\implies$ scorrelate

Prop. (momenti dalla funz generatrice) Se il raggio di convegenza è > 1 (devo poter fare la derivata in 1):

(a) $\mathbb{E}[X] = G'_X(1)$
(b) $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = G''_X(1) + G'_X(1) - G'_X(1)^2$

Dimostrazione. Basta calcolare le derivate prima e seconda di $G_X(z)$ e poi in 1, derivata prima in 1 è esattamente il valore atteso, combinando la prima e seconda trovo la varianza.

Funzione generatrice dei momenti

Exponential generating function (EGF) di una successione a_n è:

$$EG(a_n, x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n$$

Parte X

Convergenza di Successioni di Variabili Aleatorie e principali Teoremi Limite del Calcolo delle Probabilità

Definizioni

Convergenze di variabili aleatorie

Quasi certa	$X_n \xrightarrow{\text{q.c.}} X$	$P\left(\lim\limits_{n\to+\infty} X_n(\omega) = X(\omega)\right) = 1$	
In media di ordine $p \geq 1$	$X_n \xrightarrow{L^p} X$	$\lim\limits_{n\to+\infty} \mathbb{E}\left[\lvert X_n - X\rvert^p\right] = 0$	$\mathbb{E}\lvert X\rvert^p, \mathbb{E}\lvert X_n\rvert^p < +\infty \quad \forall n$
In probabilità	$X_n \xrightarrow{P} X$	$\lim\limits_{n\to+\infty} P(\lvert X_n - X\rvert > \varepsilon) = 0$	$\forall \varepsilon > 0$
In distribuzione / In legge	$X_n \xrightarrow{\text{d}} X$	$F_{X_n}(x) \overset{n\to+\infty}{\longrightarrow} F_X(x)$	punt. $\forall x \mid F(x)$ continua

Successioni di variabili aleatorie indipendenti identicamente distribuite (i.i.d.)

Teoremi

Implicazioni tra convergenze

$$\begin{array}{ccccc}
\overbrace{X_n \xrightarrow{L^q} X}^{q>p} & \Longrightarrow & X_n \xrightarrow{L^p} X & & \\
& & \Downarrow & & \\
X_n \xrightarrow{\text{q.c.}} X & \Longrightarrow & X_n \xrightarrow{P} X & \Longrightarrow & X_n \xrightarrow{d} X
\end{array}$$

Dimostrazione

- $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{P} X$: per la disuguaglianza di Markov, $P(\lvert X_n - X\rvert > \varepsilon) \leq \frac{\mathbb{E}\lvert X_n - X\rvert}{\varepsilon} \longrightarrow 0$
- $X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X$: prendo $F_n(x) = P(X_n \leq x)$ e $F(x) = P(X \leq x)$. Se riesco a dimostrare per un certo $\alpha_n^\varepsilon \overset{n\to+\infty}{\longrightarrow} 0$ vale

$$F(x-\varepsilon) - \alpha_n^\varepsilon \leq F_n(x) \leq F(x+\varepsilon) + \alpha_n^\varepsilon$$

ho dimostrato, in quanto

$$\liminf_{n\to+\infty} \left[F(x-\varepsilon) - \alpha_n^\varepsilon\right] \leq \liminf_{n\to+\infty} F_n(x) \leq \limsup_{n\to+\infty} F_n(x) \leq \limsup_{n\to+\infty} \left[F(x+\varepsilon) + \alpha_n^\varepsilon\right]$$

$$F(x-\varepsilon) \leq \liminf_{n\to+\infty} F_n(x) \leq \limsup_{n\to+\infty} F_n(x) \leq F(x+\varepsilon), \text{ che per } x \text{ di continuità è la tesi}$$

Per la stima dall'alto ho

$$F_n(x) = P(X_n \leq x) = \underbrace{P(X_n \leq x, X \leq x+\varepsilon)}_{(1)} + \underbrace{P(X_n \leq x, X > x+\varepsilon)}_{(2)} \leq \underbrace{F(x+\varepsilon)}_{(1)} + \underbrace{\overbrace{P(\lvert X_n - X\rvert > \varepsilon)}^{\alpha_n^\varepsilon}}_{(2)}$$

analogamente per quella dal basso.

- $X_n \xrightarrow{\text{q.c.}} X \implies X_n \xrightarrow{P} X$:

Lemma

1. $X_n \xrightarrow{\text{q.c.}} X \iff \forall \varepsilon > 0 \ \lim_{m\to+\infty} P\left(\bigcup_{n\geq m} \{\lvert X_n - X\rvert > \varepsilon\}\right) = 0$
2. $\sum_n P(\lvert X_n - X\rvert > \varepsilon) < +\infty \implies X_n \xrightarrow{\text{q.c.}} X$

Teorema Centrale del Limite

Sia data una successione $\{Y_n\}_{n\geq 1}$ di v.a. i.i.d definite su $\left(\Omega, \mathscr{F}, \overline{P}\right)$, tali per cui

1. $\mathbb{E}\left[Y_1^2\right] < +\infty$ ($\mathbb{E}\left[Y_n^2\right] < +\infty\ \forall n$)
2. $\sigma^2 \coloneqq \mathrm{Var}\left(Y_1\right) > 0$ (equivalente a $\sigma^2 \coloneqq \mathrm{Var}\left(Y_n\right) > 0 \quad \forall n \in \mathbb{N}$).

Allora, posto

$$X_n \coloneqq \frac{\left(\sum_{i=1}^n Y_i\right) - n\mu}{\sqrt{n\sigma^2}} \qquad \text{con } \mu \coloneqq \mathbb{E}\left[Y_1\right]$$

si ha $X_n \xrightarrow{d} X$, dove X è una qualunque v.a. su $\left(\Omega, \mathscr{F}, \overline{P}\right)$ con distribuzione Gaussiana standard.

Dimostrazione Condizione più restrittiva: richiediamo che $\exists z_* \mid M_{Y_1}(t) \coloneqq \mathbb{E}\left[e^{tY_1}\right] < +\infty\, \forall z \in (0, z_*)$

- Sia $\overline{Y_n} \coloneqq Y_n - \mu$, si deduce che

$$\begin{aligned} M_{X_n}(z) &= \mathbb{E}\left[\exp\left\{\frac{z}{\sqrt{n\sigma^2}}\sum_{i=1}^n \overline{Y_i}\right\}\right] = \mathbb{E}\left[\prod_{i=1}^n \exp\left\{\frac{z}{\sqrt{n\sigma^2}}\overline{Y_i}\right\}\right] = \prod_{i=1}^n \mathbb{E}\left[\exp\left\{\frac{z}{\sqrt{n\sigma^2}}\overline{Y_i}\right\}\right] = \\ &= \prod_{i=i}^n M_{\overline{Y_i}}\left(\frac{z}{\sqrt{n\sigma^2}}\right) = \left(M_{\overline{Y_1}}\left(\frac{z}{\sqrt{n\sigma^2}}\right)\right)^n \end{aligned}$$

$$M_{\overline{Y_i}}\left(\frac{z}{\sqrt{n\sigma^2}}\right) = \mathbb{E}\left[\exp\left\{\frac{z\left(Y_i - \mu\right)}{\sqrt{n\sigma^2}}\right\}\right] = e^{-z\mu} M_{Y_1}\left(\frac{z}{\sqrt{n\sigma^2}}\right)$$

$\forall z > 0\ \exists n_0 \mid \forall n \geq n_0 \quad \frac{z}{\sqrt{n\sigma^2}} \leq z_*$
quindi definitivamente ben definita

- Osservo che

$$M_X(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + xz} dx = e^{z^2/2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-z)^2/2} dx = e^{z^2/2}$$

- Siccome $M_{\overline{Y_i}} < +\infty$ in $(0, z_*)$ allora $M_{\overline{Y_i}} \in C^k\left(-z_*, z_*\right)$, e per Taylor ho

$$\begin{aligned} M_{\overline{Y_1}}(z) &= M_{\overline{Y_1}}(0) + zM'_{\overline{Y_1}}(0) + \frac{z^2}{2} M''_{\overline{Y_1}}(0) + o\left(|z|^3\right) = \\ &= \underbrace{M_{\overline{Y_1}}(0)}_{1} + z\underbrace{\mathbb{E}\left[\overline{Y_1}\right]}_{0} + \frac{z^2}{2}\underbrace{\mathbb{E}\left[\overline{Y_1}^2\right]}_{\sigma^2} + o\left(|z|^3\right) = 1 + \frac{z^2}{2}\sigma^2 + o\left(|z|^3\right) \end{aligned}$$

$$M_{X_n}(z) = \left(1 + \frac{z^2}{2n\cancel{\sigma^2}}\cancel{\sigma^2} + o\left(\frac{|z|^3}{n^{3/2}\sigma^3}\right)\right)^n \overset{n\to+\infty}{\longrightarrow} e^{z^2/2} = M_X(t)$$

- La tesi deriva da $\lim_{n\to+\infty} M_{X_n}(t) = M_X(t)$ e dal teorema di Curtiss

Legge Debole dei Grandi Numeri

Sia data una successione $\{Y_n\}_{n\geq 1}$ di v.a. i.i.d definite su $\left(\Omega, \mathscr{F}, \overline{P}\right)$, tali per cui $\mathbb{E}\left[Y_1^2\right] < +\infty$ ($\mathbb{E}\left[Y_n^2\right] < +\infty\ \forall n$)
Allora, posto $X_n \coloneqq \frac{1}{n}\sum_{i=1}^n Y_i$ si ha $X_n \xrightarrow{L^2} \mu$, con $\mu \coloneqq \mathbb{E}\left[Y_1\right]$, ovvero $\mathbb{E}\left[\left(X_n - \mu\right)^2\right] \overset{n\to+\infty}{\longrightarrow} 0$

Dimostrazione Per linearità si ha

$$\mathbb{E}\left[X_n\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n Y_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[Y_i\right] = \frac{1}{n} n\mu = \mu$$

quindi ho che

$$\mathbb{E}\left[\left(X_n - \mu\right)^2\right] = \mathbb{E}\left[\left(X_n - \mathbb{E}\left[X_n\right]\right)^2\right] = \mathrm{Var}\left(X_n\right)$$

dunque la tesi deriva da

$$\mathrm{Var}\left(X_n\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\left(Y_i\right) = \frac{\cancel{n}\cdot\sigma^2}{n^{\cancel{2}}} \overset{n\to+\infty}{\longrightarrow} 0$$

Legge Forte dei Grandi Numeri

Sia data una successione $\{Y_n\}_{n\geq 1}$ di v.a. i.i.d definite su $\left(\Omega, \mathscr{F}, \overline{P}\right)$, tali per cui $\mathbb{E}\left[Y_1^4\right] < +\infty$ ($\mathbb{E}\left[Y_n^4\right] < +\infty\ \forall n$) [basta in realtà $\mathbb{E}\left[|Y_1|\right] < +\infty$ (Kolmogorov)]

Allora, posto $X_n \coloneqq \frac{1}{n}\sum_{i=1}^{n} Y_i$ si ha $X_n \xrightarrow{\text{q.c.}} \mu$, con $\mu \coloneqq \mathbb{E}\left[Y_1\right]$

Dimostrazione Dimostro usando 2. del lemma, $\sum_{n=1}^{+\infty} P\left(|X_n - X| > \varepsilon\right) < +\infty$, per Markov $P\left(|X_n - X| > \varepsilon\right) \leq \frac{\mathbb{E}\left[(X_n-X)^4\right]}{\varepsilon^4}$, da cui prendendo $\overline{Y_n} \coloneqq Y_n - \mu$

$$\mathbb{E}\left[\left(X_n - X\right)^4\right] = \frac{1}{n^4}\mathbb{E}\left[\left(\sum_{i=1}^{n}\overline{Y_i}\right)^4\right] = \frac{1}{n^4}\sum_{\substack{(r_1,\dots,r_n)\in\mathbb{N}_0^n\\ r_1+\dots+r_n=4}} \frac{4!}{r_1!\dots r_n!}\mathbb{E}\left[\overline{Y_1}^{r_1}\cdot\dots\cdot\overline{Y_n}^{r_n}\right]$$

Osserviamo che $\mathbb{E}\left[\overline{Y_1}^{r_1}\cdot\dots\cdot\overline{Y_n}^{r_n}\right] = \mathbb{E}\left[\overline{Y_1}^{r_1}\right]\cdot\dots\cdot\mathbb{E}\left[\overline{Y_n}^{r_n}\right]$ e che $\mathbb{E}\left[\overline{Y_i}\right] = 0$. Gli r_i si possono partizionare in vari modi, ma se si ha $r_i = 1$, si annulla. Rimangono solo le partizioni (A) $r_i = 4$, $r_{\text{restanti}} = 0$ e (D) $r_i = 2$, $r_{j\neq i} = 2$, $r_{\text{restanti}} = 0$

$$\begin{aligned}\sum_A &= \frac{1}{n^4}\sum_{i=1}^{n}\frac{\cancel{4!}}{\cancel{4!}}\mathbb{E}\left[\overline{Y_i}^4\right] = \frac{1}{n^3}\mathbb{E}\left[\overline{Y_1}^4\right] = O\left(\frac{1}{n^3}\right)\\ \sum_D &= \frac{1}{n^4}\sum_{i\neq j}\frac{4!}{2!\cdot 2!}\mathbb{E}\left[\overline{Y_i}^2\right]\mathbb{E}\left[\overline{Y_j}^2\right] = \frac{1}{n^4}\binom{n}{2}\frac{4!}{2!\cdot 2!}\left(\mathbb{E}\left[\overline{Y_i}^2\right]\right)^2 = O\left(\frac{1}{n^2}\right)\end{aligned}$$

Quindi

$$P\left(|X_n - \mu| > \varepsilon\right) \leq \frac{\mathbb{E}\left[\left(X_n-\mu\right)^4\right]}{\varepsilon^4} \leq \frac{O\left(1/n^2\right)}{\varepsilon^4} < +\infty \quad\Longrightarrow\quad \sum_{n=1}^{+\infty} P\left(|X_n - X| > \varepsilon\right) < +\infty \quad\Longrightarrow\quad X_n \xrightarrow{\text{q.c.}} \mu$$

Parte XI
Controesempi

Paradosso dei prigionieri

Problema di Monty Hall

Parte XII
Esercizi

Figata: `https://stats.libretexts.org/Bookshelves/Probability_Theory/Probability_Mathematical_Statistics_and_Stochastic_Processes_(Siegrist)/04%3A_Expected_Value/4.11%3A_Vector_Spaces_of_Random_Variables`