



Università degli Studi di Pisa - Dipartimento di Matematica

Elementi di probabilità e statistica

Autore

Alessio Del Vigna

`delvigna@mail.dm.unipi.it`

Titolare del corso

Prof. Pietro Majer

Università di Pisa

Note dell'omonimo corso tenuto dal Prof. Pietro Majer nell'anno accademico 2009-2010.

Queste note contengono almeno un errore. Se ne trovaste qualcuno, vi pregherei di segnalarmelo all'indirizzo `delvigna@mail.dm.unipi.it`.

Una precisazione. La frase "Queste note contengono almeno un errore" è un primo esempio di teorema di esistenza. Come ogni teorema che si rispetti, ecco la sua dimostrazione: se ci sono errori, ce ne sono; se non ci sono errori, allora è questa stessa frase a costituire un errore. Così sappiamo che un errore c'è, ma in questo momento non sappiamo dire qual è: questo, del resto, è il prezzo dei teoremi di esistenza! Pertanto, se l'unico errore dovesse essere la frase "Queste note contengono almeno un errore", non segnalatelo: eccezion fatta per questa frase, significherebbe -felicemente- che queste note non contengono altri errori.

Indice

1	Algebre e spazi di misura	1
1.1	Concetti generali	1
1.2	Algebre finite	3
1.3	Algebre generate	5
1.4	Spazi di misura completi	7
2	Spazi di probabilità	11
2.1	Generalità	11
2.2	Probabilità classica e condizionata	13
2.3	Principio di inclusione–esclusione	16
2.4	Impostazione soggettivista in probabilità	17
3	Somme di numeri e funzioni misurabili	19
3.1	Somme di famiglie di numeri reali	19
3.2	Teoremi di passaggio al limite	23
3.3	Funzioni misurabili	25
4	Integrali secondo una misura	29
4.1	Funzioni integrabili rispetto ad una misura	29
4.2	Gli spazi p	34
4.3	Integrazione e misura immagine	36
4.4	Misure prodotto e integrali iterati	37
4.5	Integrali su spazi prodotto	40
5	Probabilità discreta e generale	41
5.1	Richiami di probabilità discreta	41
5.2	Variabili aleatorie	42
5.3	Speranza matematica e varianza	44
5.4	Indipendenza di variabili aleatorie	48
5.5	Variabili aleatorie discrete	51
5.5.1	Legge binomiale e di Bernoulli	52

5.5.2	Legge geometrica	55
5.5.3	Legge di Poisson	57
5.5.4	Legge ipergeometrica	59
5.6	Funzione generatrice di probabilità	61
5.7	Legge dei grandi numeri e disuguaglianza di Chebyshev	62
5.7.1	La formula di Stirling	66
5.7.2	Alcune osservazioni	67
5.8	Il teorema del limite centrale	67
5.9	Leggi di probabilità continue	71
5.10	Convoluzione di funzioni misurabili su \mathbb{R}^n	73
5.11	Misure su \mathbb{R}^n dotate di densità ¹	75
5.12	Variabili aleatorie con densità	76
5.12.1	Caso del blocco di variabili aleatorie	77
5.12.2	Somma di variabili aleatorie	78
5.12.3	Applicazione di diffeomorfismi	79
5.13	Densità continue notevoli	79
5.13.1	Densità uniforme	79
5.13.2	Legge esponenziale	80
5.13.3	Legge gamma	81
5.13.4	Legge normale (o gaussiana)	82
6	Statistica	83
6.1	Introduzione all'inferenza statistica	83
6.2	Intervalli di fiducia	84
6.2.1	Utilizzo della legge dei grandi numeri	85
6.2.2	Utilizzo del teorema di de Moivre–Laplace	86
6.3	Cenni sulla teoria della stima	87
A	Esercizi risolti	89
A.1	Algebre e spazi di misura	89
A.2	Combinatoria	91
A.3	Probabilità condizionata	93
A.4	Variabili aleatorie discrete	98
A.5	Variabili aleatorie reali	104
A.6	Inferenza statistica	109

Capitolo 1

Algebre e spazi di misura

Di fronte ad una situazione che suggerisce l'uso del calcolo delle probabilità incontriamo alcune affermazioni legate tra loro dai connettivi logici *o*, *e*, *non*: è facile convincersi che si può tradurre questo in una famiglia di sottinsiemi (chiamati eventi) di un opportuno insieme Ω , contenente l'insieme vuoto e tutto l'insieme, e stabile per le operazioni di unione (finita), intersezione e complementazione. Una tale famiglia di insiemi si chiama un'algebra di parti. Il grado di fiducia che un sottoinsieme si realizzi (chiamato probabilità), è rappresentato da un numero compreso tra 0 e 1; inoltre è intuitivo supporre che se due eventi sono incompatibili (cioè hanno intersezione vuota) la probabilità che si realizzi uno qualsiasi dei due debba essere la somma delle probabilità dei singoli eventi. Questo equivale a dire che la probabilità è una funzione d'insieme (finitamente) additiva.

1.1 Concetti generali

Definizione 1.1.1. Dato un'insieme Ω , si chiama *algebra di parti* di Ω una famiglia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ tale che

- (1) $\emptyset \in \mathcal{A}$;
- (2) se $A \in \mathcal{A}$ allora $A^C \in \mathcal{A}$;
- (3) se $A, B \in \mathcal{A}$ allora $A \cup B \in \mathcal{A}$.

Osservazione 1.1.1. Dalla proprietà (3) segue immediatamente che \mathcal{A} è chiuso rispetto all'unione di un numero finito di suoi elementi. Da ciò si ha immediatamente che \mathcal{A} è anche chiuso per intersezione finita di suoi elementi in quanto se $\{A_k\}_{k=1}^n \subseteq \mathcal{A}$ allora

$$\bigcap_{k=1}^n A_k = \left(\bigcup_{k=1}^n A_k^C \right)^C,$$

e il secondo membro sta in \mathcal{A} per le proprietà dell'algebra.

Definizione 1.1.2. Sia \mathcal{A} un'algebra di parti di Ω . Una *misura* su (Ω, \mathcal{A}) è un'applicazione $\mu : \mathcal{A} \rightarrow [0, +\infty]$ tale che

- (1) $\mu(\emptyset) = 0$;
- (2) se $A, B \in \mathcal{A}$ e sono disgiunti allora $\mu(A \cup B) = \mu(A) + \mu(B)$.

Le definizioni sopra riportate, oltre ad essere molto intuitive, sono supportate da valide argomentazioni logiche, tuttavia dal punto di vista matematico presentano una difficoltà: l'additività finita (proprietà (2)) non consente di andare al limite, e di conseguenza di calcolare degli integrali. La buona proprietà per poter effettuare queste operazioni è la additività numerabile, detta anche σ -additività. Inoltre la famiglia di parti sulla quale possa essere definita una funzione σ -additiva è opportuno che sia stabile per unione numerabile e non per unione finita. Per questo motivo, seguendo quella che è ormai comunemente chiamata la *definizione assiomatica* di probabilità secondo Kolmogorov, sostituiamo alle precedenti queste definizioni:

Definizione 1.1.3. Dato un'insieme Ω , si chiama σ -algebra di parti di Ω una famiglia $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ tale che

- (1) $\emptyset \in \mathcal{A}$;
- (2) se $A \in \mathcal{A}$ allora $A^C \in \mathcal{A}$;
- (3) se $\{A_n\}_{n=0}^{+\infty} \subseteq \mathcal{A}$ allora $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}$.

Osservazione 1.1.2. Come prima, dalla proprietà (3) della definizione segue che \mathcal{A} è chiusa per intersezione numerabile.

Osservazione 1.1.3. Osserviamo che una σ -algebra è anche un'algebra: infatti se $A = A_0$ e $B = A_1$ sono in \mathcal{A} allora, posto $A_n = \emptyset$ per ogni $n \geq 2$

$$A \cup B = \bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}.$$

Definizione 1.1.4. Sia \mathcal{A} una σ -algebra di parti di Ω . Una *misura* su (Ω, \mathcal{A}) è un'applicazione $\mu : \mathcal{A} \rightarrow [0, +\infty]$ tale che

- (1) $\mu(\emptyset) = 0$;
- (2) se $\{A_n\}_{n=0}^{+\infty} \subseteq \mathcal{A}$ e sono disgiunti allora $\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n)$.

In tal caso la tripletta $(\Omega, \mathcal{A}, \mu)$ prende il nome di *spazio di misura*.

Adesso vediamo i vantaggi che porta con sé l'assunzione della proprietà di σ -additività. Enunciamo e dimostriamo la seguente proposizione:

Proposizione 1.1.1. Sia \mathcal{A} una σ -algebra e μ una misura finita su (Ω, \mathcal{A}) finitamente additiva e sia $\{A_n\}_{n=0}^{+\infty} \subseteq \mathcal{A}$. Allora le proprietà seguenti sono equivalenti:

- (1) μ è σ -additiva;
- (2) se $A_n \subseteq A_{n+1}$ per ogni n allora si ha $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu\left(\bigcup_{n=0}^{+\infty} A_n\right)$;
- (3) se $A_n \supseteq A_{n+1}$ per ogni n allora si ha $\lim_{n \rightarrow +\infty} \mu(A_n) = \mu\left(\bigcap_{n=0}^{+\infty} A_n\right)$.

Dimostrazione. Mostriamo ad esempio l'equivalenza tra (1) e (2). Supponiamo che μ sia σ -additiva, e poniamo

$$\begin{cases} B_1 = A_1 \\ B_n = A_n - A_{n-1} \text{ se } n > 1 \end{cases}$$

La successione $\{B_n\}_{n=1}^{+\infty}$ è costituita da insiemi a due a due disgiunti per costruzione e quindi per questi vale la σ -additività. Osserviamo che $\mu(B_n) = \mu(A_n) - \mu(A_{n-1})$ in quanto

$$A_n = (A_n - A_{n-1}) \cup A_{n-1} = B_n \cup A_{n-1}$$

e l'unione è disgiunta e osserviamo inoltre che $\bigcup_{n=1}^{+\infty} A_n = \bigcup_{n=1}^{+\infty} B_n$. Quindi

$$\mu\left(\bigcup_{n=1}^{+\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{+\infty} B_n\right) = \sum_{n=1}^{+\infty} \mu(B_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mu(B_k) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

Viceversa supponiamo che valga la proprietà (2) e mostriamo che μ è σ -additiva. Sia $\{B_n\}_{n=1}^{+\infty}$ una successione numerabile di insiemi e costruiamo

$$A_n = \bigcup_{k=1}^n B_k$$

per ogni $n \geq 1$ intero. Osserviamo che la successione degli A_n è crescente e quindi vale che

$$\lim_{n \rightarrow +\infty} \mu(A_n) = \mu\left(\bigcup_{n=0}^{+\infty} A_n\right).$$

Adesso ricordando la definizione degli A_n si ha

$$\mu\left(\bigcup_{n=1}^{+\infty} B_n\right) = \mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mu(A_n) = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \mu(B_k) = \sum_{n=1}^{+\infty} \mu(B_n).$$

L'equivalenza tra (2) e (3) segue facilmente per complementazione. \square

1.2 Algebre finite

Vogliamo adesso studiare che struttura ha un'algebra finita. La definizione che diamo ora vale in realtà per qualsiasi algebra:

Definizione 1.2.1. Un *atomo* di un'algebra \mathcal{A} è un elemento non vuoto minimale di \mathcal{A} , ossia è un $A \in \mathcal{A}$ tale che se $B \in \mathcal{A}$ e $B \subseteq A$ allora $B = \emptyset$ o $B = A$. Denotiamo con $\tilde{\Omega}$ l'insieme degli atomi di \mathcal{A} .

Ebbene, la struttura di un'algebra finita è piuttosto semplice, nel senso che tali algebre hanno cardinalità ben precise. Prima però dobbiamo mostrare dei fatti preliminari:

Lemma 1.2.1. *Sia \mathcal{A} un'algebra finita, allora due atomi o coincidono o sono disgiunti; inoltre ogni elemento di Ω è contenuto in un unico atomo.*

Dimostrazione. Siano $A, B \subseteq \mathcal{A}$ due atomi e supponiamo che $A \cap B \neq \emptyset$. Essendo $A \cap B$ contenuto sia in A che in B si ha per definizione di atomo che $A \cap B = A$ e $A \cap B = B$, da cui $A = B$. Sia $\omega \in \Omega$, mostriamo che l'insieme

$$A_\omega = \bigcap_{\substack{A \in \mathcal{A} \\ A \ni \omega}} A$$

è un atomo. Osserviamo che la famiglia degli insiemi intersecati non è vuota in quanto sicuramente Ω vi appartiene. Poi si ha che A_ω è un elemento di \mathcal{A} : l'unica cosa che dobbiamo mostrare è che l'intersezione che lo definisce è al più numerabile. Faremo ciò fra poco in un teorema in cui tale intersezione risulterà al più numerabile in modo semplice (proposizione 1.3.1). Il fatto che sia minimale discende immediatamente dal fatto che abbiamo costruito un'intersezione. \square

Osservazione 1.2.1. Osserviamo che possiamo scrivere ogni $A \in \mathcal{A}$ come unione disgiunta dei suoi atomi, ossia

$$A = \bigcup_{\omega \in A} A_\omega,$$

e che tale scrittura è unica a meno di ripetizioni.

Proposizione 1.2.1. *Sia \mathcal{A} un'algebra finita. Allora la sua cardinalità è 2^r , dove r è il numero degli atomi di \mathcal{A} .*

Dimostrazione. Osserviamo che la corrispondenza

$$\begin{aligned} \mathcal{A} &\longrightarrow \mathcal{P}(\tilde{\Omega}) \\ A &\longmapsto I_A = \{A_\omega \mid \omega \in A\} \end{aligned}$$

è bigettiva. Supponiamo che $I_A = I_B$ per $A, B \in \mathcal{A}$: per l'osservazione che precede questa proposizione si ha che

$$A = \bigcup_{\omega \in A} A_\omega = \bigcup_{\omega \in B} A_\omega = B,$$

e questo mostra l'iniettività. Inoltre prendiamo $I \in \mathcal{P}(\tilde{\Omega})$ insieme di atomi, allora è ovvio per definizione che l'insieme $A = \bigcup_{J \in I} J$ è tale che $I_A = I$: questo mostra la surgettività. Da ciò segue subito la tesi. \square

1.3 Algebre generate

In questo breve paragrafo vogliamo parlare di algebre generate da sottoinsiemi delle parti di Ω . Diamo la seguente definizione:

Definizione 1.3.1. Sia $\mathcal{B} \subseteq \mathcal{P}(\Omega)$. L'algebra minimale che contiene \mathcal{B} si dice *algebra generata* da \mathcal{B} .

Affinché la precedente definizione abbia senso è opportuno osservare che l'algebra generata da un qualsiasi sottoinsieme di $\mathcal{P}(\Omega)$ esiste. In effetti basta considerare

$$\mathcal{A}(\mathcal{B}) = \bigcap_{\substack{\mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{P}(\Omega) \\ \mathcal{A} \text{ algebra}}} \mathcal{A},$$

e osservare che la famiglia di insiemi intersecata non è vuota in quanto sicuramente $\mathcal{P}(\Omega)$ è uno di essi. Quanto detto finora si può ridire anche per le σ -algebre: esiste la σ -algebra generata da qualsiasi $\mathcal{B} \subseteq \mathcal{P}(\Omega)$.

Il tipo di costruzione presentata ora è una costruzione dall'alto ma poco fa capire sulla struttura dell'algebra generata. Proprio a questo scopo esiste un'altra costruzione dell'algebra generata che è effettuata dal basso e consente di capire la struttura di un algebra. Se in effetti consideriamo l'algebra generata riusciamo a darne una descrizione esplicita, mentre se cerchiamo la σ -algebra generata ci serve un procedimento che ha un numero di passi pari a ω_1 (primo ordinale non numerabile). Iniziamo dal primo caso.

Sia $\mathcal{B} \subseteq \mathcal{P}(\Omega)$, consideriamo l'insieme

$$\mathcal{B}' = \mathcal{B} \cup \{B^C \mid B \in \mathcal{B}\},$$

chiaramente chiuso per complementazione. Adesso sia

$$\mathcal{A}' = \left\{ \bigcup_{i \in I} \bigcap_{j \in J} C_{ij} \mid C_{ij} \in \mathcal{B}', I, J \text{ finiti} \right\}.$$

Osserviamo che \mathcal{A}' è contenuto in qualsiasi altra algebra che contiene \mathcal{B} : infatti se un'algebra contiene \mathcal{B} allora deve contenere per definizione tutti gli elementi che siamo andati a considerare in \mathcal{A}' . Adesso non ci resta che verificare la definizione di algebra. Mostriamo la chiusura rispetto all'intersezione:

$$\left(\bigcup_{i \in I} \bigcap_{j \in J} C_{ij} \right) \cap \left(\bigcup_{h \in H} \bigcap_{k \in K} D_{hk} \right) = \bigcup_{i \in I} \left(\bigcap_{j \in J} C_{ij} \cap \bigcup_{h \in H} \bigcap_{k \in K} D_{hk} \right) = \bigcup_{\substack{i \in I \\ h \in H}} \bigcap_{\substack{j \in J \\ k \in K}} (C_{ij} \cap D_{hk}),$$

che, per come l'abbiamo riscritto, è effettivamente un elemento di \mathcal{A}' . In modo analogo si mostrano la chiusura per unione e complementazione.

Il problema che si pone nel considerare le σ -algebre generate è che il processo di costruzione dal basso non ha termine dopo due passi come prima, ma richiede un procedimento che termina con il primo ordinale non numerabile. Di seguito accenneremo soltanto alla questione, in quanto esula dagli scopi del corso. Come prima si definisce l'insieme \mathcal{B}' e per ogni famiglia F si definisce

$$S(F) = \left\{ \bigcup_{i=1}^{+\infty} \bigcap_{j=1}^{+\infty} C_{ij} \mid C_{ij} \in F \right\}.$$

Il procedimento che vogliamo descrivere si dà mediante un'induzione transfinita. Sia infatti $\mathcal{A}_0 = \mathcal{B}'$ e $\mathcal{A}_n = S(\mathcal{A}_{n-1})$ per ogni $n \geq 1$ allora si considera

$$\begin{cases} \mathcal{A}_\omega = \bigcup_{n < \omega} \mathcal{A}_n \\ \mathcal{A}_{\omega+n} = S(\mathcal{A}_{\omega+n-1}) \end{cases};$$

questo procedimento deve proseguire per tutti gli ordinali numerabili, sino ad interrompersi ad $\mathcal{A}_{\omega_1} = \bigcup_{\alpha < \omega_1} \mathcal{A}_\alpha$, in cui finalmente si ottiene la σ -additività.

Osservazione 1.3.1. Ci limitiamo a osservare che esiste un caso in cui anche la descrizione della σ -algebra generata è piuttosto semplice, ossia quando \mathcal{B} è finito o numerabile. Supponiamo $\mathcal{B} = \{B_i\}_{i \in I}$, ove I ha cardinalità finita o numerabile. Allora non è difficile vedere che

$$\sigma\mathcal{A}(\mathcal{B}) = \left\{ B_J = \bigcup_{i \in J} B_i \mid J \subseteq I \right\}.$$

Tra l'altro l'associazione tra B_J e J stesso instaura una bigezione tra la σ -algebra generata e $\mathcal{P}(I)$ come è facile mostrare.

Definizione 1.3.2. Un'algebra si dice *atomica* se è generata dai propri atomi

Questa definizione per il momento non verrà spiegata in maniera esaustiva in quanto la ritroveremo parlando di spazi completi rispetto a una misura. Per ora ci limitiamo a osservare che se ad esempio l'insieme degli atomi è finito o numerabile allora la σ -algebra è essenzialmente $\mathcal{P}(I)$, come nell'osservazione che precede la definizione. Inoltre però vogliamo dare il seguente risultato:

Proposizione 1.3.1. *Sia \mathcal{A} una σ -algebra su Ω finito o numerabile. Allora \mathcal{A} è atomica.*

Dimostrazione. Definiamo su Ω la relazione

$$\omega \sim \omega' \iff \text{per ogni } A \in \mathcal{A} \text{ si ha } \omega \in A \text{ se e solo se } \omega' \in A.$$

Si verifica immediatamente che quella appena definita è una relazione di equivalenza (il lettore lo faccia per esercizio). Ogni elemento dunque appartiene alla sua classe di equivalenza, che è data da

$$[\omega] = A_\omega = \bigcap_{\omega \in A \in \mathcal{A}} A$$

in quanto $\omega' \in [\omega]$ se e solo se $\omega \sim \omega'$ (e questo, tradotto per gli insiemi appena scritti, fornisce l'uguaglianza). Si ha dunque

$$A = \bigcup_{\omega \in A} [\omega]$$

in quanto se $\omega \in A$ allora $[\omega] \subseteq A$, ossia che l'algebra è generata da $[\omega]$. L'unico punto ancora non chiaro è se $[\omega] \in \mathcal{A}$ in quanto nella sua definizione non sappiamo ancora niente sul fatto che l'intersezione sia o meno numerabile. In effetti così è, in quanto si può esprimere $[\omega]$ come intersezione al più numerabile: per ogni $x \in \Omega - [\omega]$ sia $A_x \in \mathcal{A}$ tale che $\omega \in A_x$ e $x \notin A_x$ (esiste per la relazione di equivalenza). Quindi si può scrivere

$$[\omega] = \bigcap_{x \in \Omega - [\omega]} A_x \in \mathcal{A},$$

e abbiamo concluso per l'osservazione precedente, perché abbiamo un'intersezione finita o numerabile. \square

Osservazione 1.3.2. Se due misure coincidono su una base dell'algebra (ossia su un insieme che la genera) allora coincidono su tutta l'algebra? La risposta è no, e lo proveremo con un controesempio. Consideriamo lo spazio $\Omega = \{a, b, c, d\}$ con la base $\mathcal{B} = \{\{a, b\}, \{c, d\}, \{a, c\}, \{b, d\}\}$ sua base; siano poi μ e μ' le misure tali che

$$\mu(a) = \mu(c) = 0, \quad \mu(b) = \mu(d) = \frac{1}{2}, \quad \mu'(a) = \mu'(c) = \frac{1}{2} \quad \text{e} \quad \mu'(b) = \mu'(d) = 0.$$

È ovvio che tali misure coincidono su \mathcal{B} ma in effetti non coincidono sull'algebra.

1.4 Spazi di misura completi

Una nozione importante in teoria della misura è quella di spazio completo rispetto a una misura. La seguente definizione ci dice che significa:

Definizione 1.4.1. Uno spazio di misura $(\Omega, \mathcal{A}, \mu)$ si dice *completo* se ogni sottinsieme di un insieme di misura nulla è misurabile. Ossia se per ogni $A \in \mathcal{A}$ tale che $\mu(A) = 0$ si ha che per ogni $B \subseteq A$ vale $B \in \mathcal{A}$.

Osservazione 1.4.1. Nella definizione precedente è chiaro che ogni sottoinsieme di un insieme di misura nulla è misurabile e ha misura nulla. Infatti $A = B \cup (A - B)$ e l'unione è disgiunta, pertanto $0 = \mu(A) = \mu(B) + \mu(A - B)$, e ricordando che la misura ha valori positivi segue quanto volevamo.

Teorema 1.4.1 (di completamento). *Ogni spazio di misura $(\Omega, \mathcal{A}, \mu)$ si può estendere ad uno spazio $(\Omega, \tilde{\mathcal{A}}, \tilde{\mu})$ completo, con $\tilde{\mathcal{A}} \supseteq \mathcal{A}$ e $\tilde{\mu}|_{\mathcal{A}} = \mu$.*

Dimostrazione. Definiamo

$$\tilde{\mathcal{A}} = \{B \subseteq \Omega \mid \exists A, A' \in \mathcal{A} \text{ tali che } A \subseteq B \subseteq A', \mu(A' - A) = 0\},$$

e notiamo che $\tilde{\mathcal{A}} = \{B \subseteq \Omega \mid A \subseteq B \subseteq A \cup N \text{ con } A, N \in \mathcal{A}, \mu(N) = 0\}$. Intanto dobbiamo mostrare che quella appena definita è una σ -algebra. Sia $\{B_k\}_{k \in \mathbb{N}} \subseteq \tilde{\mathcal{A}}$ una successione numerabile di insiemi, dobbiamo mostrare che la loro unione è ancora contenuta nell'insieme $\tilde{\mathcal{A}}$. Per ogni $n \in \mathbb{N}$ vale $A_n \subseteq B_n \subseteq A_n \cup N_n$ per certi $A_n, N_n \in \mathcal{A}$ e $\mu(N_n) = 0$. Ma allora

$$\bigcup_{k=0}^{+\infty} A_k \subseteq \bigcup_{k=0}^{+\infty} B_k \subseteq \bigcup_{k=0}^{+\infty} A_k \cup \bigcup_{k=0}^{+\infty} N_k,$$

ed essendo $\bigcup_k A_k \in \mathcal{A}$ e $\bigcup_k N_k$ di misura nulla segue che l'unione dei B_k è un elemento di $\tilde{\mathcal{A}}$ per definizione.¹

Infine per la complementazione sia $B \in \tilde{\mathcal{A}}$, allora $A \subseteq B \subseteq A'$ per certi $A, A' \in \mathcal{A}$ come nella definizione. Ma allora passando ai complementari si ha $A'^C \subseteq B^C \subseteq A^C$; osserviamo che

$$\mu(A^C - A'^C) = \mu(A' - A) = 0$$

e dunque per definizione $B \in \tilde{\mathcal{A}}$.

Abbiamo conquistato che $\tilde{\mathcal{A}}$ è un'algebra, adesso dobbiamo definire una misura $\tilde{\mu} : \tilde{\mathcal{A}} \rightarrow [0, +\infty]$ che estenda μ . Il modo naturale di farlo è il seguente: per ogni $B \in \tilde{\mathcal{A}}$ si pone

$$\tilde{\mu}(B) = \mu(A) = \mu(A'),$$

dove $A, A' \in \mathcal{A}$ sono quelli della definizione di $\tilde{\mathcal{A}}$. Osserviamo che quella data è una buona definizione, in quanto se $A \subseteq B \subseteq A'$ e $A_1 \subseteq B \subseteq A'_1$ allora segue che $\mu(A' - A_1) = \mu(A - A'_1) = 0$ considerando le inclusioni incrociate. A questo punto non è difficile verificare che con tale misura lo spazio dato è completo. \square

¹potevamo anche mostrare la chiusura per intersezione: detto $N = \bigcup_{k=1}^{+\infty} N_k$ possiamo scrivere che per ogni n si ha $A_n \subseteq B_n \subseteq A_n \cup N$. Intersecando le famiglie di insiemi si ha

$$\bigcap_{k=0}^{+\infty} A_k \subseteq \bigcap_{k=0}^{+\infty} B_k \subseteq \bigcap_{k=0}^{+\infty} A_k \cup N.$$

Adesso consideriamo nuovamente il caso in cui Ω sia un insieme finito o numerabile: infatti sotto queste ipotesi si riescono a classificare gli spazi completi. Vale infatti la seguente

Proposizione 1.4.1. *Sia $(\Omega, \mathcal{A}, \mu)$ uno spazio di misura con Ω finito o numerabile. Tale spazio è completo se e solo se \mathcal{A} è atomica e se un atomo ha misura nulla allora è un singoletto.*

Dimostrazione. (\Leftarrow) Sia $A \in \mathcal{A}$ di misura nulla. Intanto, poiché \mathcal{A} è atomica, possiamo scrivere

$$A = \bigcup_i N_i$$

dove gli $N_i \subseteq \mathcal{A}$ sono gli atomi di A e l'unione è disgiunta. Dall'additività di μ segue che $\mu(N_i) = 0$ per ogni i e quindi per ipotesi N_i è un singoletto. Prendiamo adesso $B \subseteq A$ allora

$$B = \bigcup_{b \in B} \{b\}$$

e quindi $B \in \mathcal{A}$ in quanto è l'unione (al più numerabile) di atomi, che sono elementi di \mathcal{A} .

(\Rightarrow) Sia N un atomo di \mathcal{A} di misura nulla. Sia $\omega \in N$, allora $N - \{\omega\} \subset N$ ed inoltre $N - \{\omega\} \in \mathcal{A}$ per la completezza dello spazio di misura. Ma allora per definizione di atomo segue che $N - \{\omega\} = \emptyset$, ossia che $N = \{\omega\}$. Questo mostra che ogni atomo a misura nulla è un singoletto. La proposizione 1.3.1 ci garantisce che l'algebra è atomica in quanto definita su Ω finito o numerabile. \square

Osservazione 1.4.2. Esiste anche una dimostrazione a sé stante del fatto che \mathcal{A} è atomica. Affermiamo che in \mathcal{A} esiste un massimo insieme di misura nulla. Definiamo infatti

$$\Omega_0 = \bigcup_{\substack{N \text{ atomo} \\ \mu(N)=0}} N = \bigcup_{\substack{\omega \in \Omega \\ \mu(\{\omega\})=0}} \{\omega\},$$

che è in \mathcal{A} perché l'unione è al più numerabile. Vale certamente $\mu(\Omega_0) = 0$ per additività ed inoltre per ogni B di misura nulla si ha $B \subseteq \Omega_0$: infatti B si scrive come unione disgiunta dei suoi atomi, i quali hanno misura nulla per additività, e quindi tali atomi fanno già parte dell'unione che definisce Ω_0 .

Sia ora $\omega \in \Omega - \Omega_0$ e poniamo

$$\mu = \inf_{\substack{\omega \in A \in \mathcal{A} \\ A \cap \Omega_0 = \emptyset}} \mu(A).$$

Affermiamo che μ in realtà è un minimo. Sia $\{A_k\}$ in \mathcal{A} con $\omega \in A_k \subseteq \Omega - \Omega_0$ una successione di insiemi tale che $\mu(A_k)$ converge decrescendo verso μ . Allora

$\bigcap_{k=0}^{+\infty} A_k \in \mathcal{A}$ è non vuota perché contiene ω , ed inoltre proprio per questo motivo

$$\mu \leq \mu \left(\bigcap_{h=0}^k A_h \right) \leq \mu(A_k),$$

e passando al limite con $k \rightarrow +\infty$ si ottiene che $\bigcap_{k=0}^{+\infty} A_k$ realizza la minima misura. Inoltre affermiamo che l'insieme che realizza il minimo è unico tra gli insiemi disgiunti da Ω_0 e contenenti ω poiché è l'atomo che contiene ω .

Ci stiamo avviando a concludere. Sia $A \in \mathcal{A}$ disgiunto da Ω_0 e supponiamo che $\omega \in A$, allora $\omega \in A_\omega \cap A \subseteq A_\omega$. Allora

$$\mu \leq \mu(A_\omega \cap A) \leq \mu(A_\omega) = \mu,$$

dove l'ultima uguaglianza vale per quanto mostrato prima (l'insieme di misura minima μ è proprio l'atomo che contiene ω). Quindi $\mu(A_\omega \cap A) = \mu(A_\omega)$, da cui $\mu(A_\omega - A) = 0$. Quindi, essendo $A_\omega - A$ disgiunto da Ω_0 , segue che $A_\omega - A = \emptyset$: questo mostra che $A_\omega \subseteq A$. Invece se $\omega \notin A$ si ha $A_\omega \cap A = \emptyset$. Se per $\omega \in \Omega_0$ si pone $A_\omega = \{\omega\}$ si ha in effetti

$$\Omega = \bigcup_{\omega \in \Omega} A_\omega,$$

ossia che Ω è generata dai suoi atomi.

Aldilà di questa seconda ultima dimostrazione, la proposizione ci dice che è possibile estendere la σ -algebra di uno spazio completo $(\Omega, \mathcal{A}, \mu)$ con Ω al più numerabile a tutto $\mathcal{P}(\Omega)$.

Capitolo 2

Spazi di probabilità

2.1 Generalità

Adesso vogliamo cominciare a concretizzare i concetti presentati finora in astratto e andare a dedurre dagli assiomi la definizione di probabilità. Intanto iniziamo con la seguente:

Definizione 2.1.1. Assegnato un insieme Ω ed una algebra \mathcal{A} di parti di Ω , si chiama *probabilità* una funzione $P : \mathcal{A} \rightarrow [0, 1]$ tale che

- (1) $P(\Omega) = 1$;
- (2) se $A, B \in \mathcal{A}$ sono disgiunti allora $P(A \cup B) = P(A) + P(B)$.

Definizione 2.1.2. Assegnato un insieme Ω ed una σ -algebra \mathcal{A} di parti di Ω , si chiama *probabilità* una funzione $P : \mathcal{A} \rightarrow [0, 1]$ tale che

- (1) $P(\Omega) = 1$;
- (2) se $\{A_n\}_{n=0}^{\infty} \subseteq \mathcal{A}$ sono disgiunti allora $P(\bigcup_{n=0}^{\infty} A_n) = \sum_{n=0}^{\infty} P(A_n)$.

Precisiamo che una terna (Ω, \mathcal{A}, P) viene detta *spazio di probabilità*. Vogliamo proporre adesso un esempio piuttosto elementare di tale spazio:

Esempio 2.1.1. Sia $\Omega = \{T, C\}$, allora $\mathcal{P}(\Omega) = \{\emptyset, \Omega, \{T\}, \{C\}\}$. Una possibile misura di probabilità è data dall'applicazione $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$ tale che

$$P(T) = P(C) = \frac{1}{2}.$$

I nomi T e S dati nell'esempio evocano gli eventi “testa” e “croce” nel lancio di una moneta.

Diamo come assodato (ma senza dimostrazione) il seguente fatto: esiste ed è unica una misura μ sui boreliani di \mathbb{R} invariante per traslazioni e tale che $\mu([0, 1]) = 1$. Il prossimo teorema mostra che in effetti su tutti i sottoinsiemi di \mathbb{R} ciò non è possibile:

Teorema 2.1.1 (controesempio di Vitali). *Non esiste una misura di probabilità definita su tutti i sottoinsiemi di \mathbb{R} che sia invariante per traslazioni e tale che $\mu([0, 1]) = 1$.*

Dimostrazione. Consideriamo \mathbb{R} come \mathbb{Q} spazio vettoriale: \mathbb{Q} è sottospazio di \mathbb{R} , e allora esiste V tale che $\mathbb{R} = \mathbb{Q} \oplus V$ ¹. Questo significa che

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (V + q).$$

Supponiamo per assurdo che esista una misura μ su $\mathcal{P}(\mathbb{R})$ con le proprietà dette. Intanto osserviamo che $\mu(V)$ non può essere nulla, sennò in tal caso

$$\mu(V) = 0 \implies \mu(V + q) = 0 \implies \mu(\mathbb{R}) = \sum_{q \in \mathbb{Q}} \mu(V + q) = 0,$$

ed è assurdo. Quindi $\mu(V) > 0$ e dunque esiste un $k \in \mathbb{Z}$ tale che $\mu(V \cap [k, k+1]) = \varepsilon > 0$ in quanto $V = \bigcup_{k \in \mathbb{Z}} (V \cap [k, k+1])$. Adesso

$$(V \cap [k, k+1]) + (\mathbb{Q} \cap [0, 1]) \subseteq [k, k+1] + [0, 1] \subseteq [k, k+2).$$

Poiché per costruzione $(V \cap [k, k+1]) + q$ al variare di $q \in \mathbb{Q}$ sono disgiunti segue che

$$\begin{aligned} \mu((V \cap [k, k+1]) + (\mathbb{Q} \cap [0, 1])) &= \mu \left(\bigcup_{q \in \mathbb{Q} \cap [0, 1]} (V \cap [k, k+1] + q) \right) = \\ &= \sum_{q \in \mathbb{Q} \cap [0, 1]} \mu(V \cap [k, k+1]) = \sum_{q \in \mathbb{Q} \cap [0, 1]} \varepsilon = \infty. \end{aligned}$$

Ma da questo seguirebbe $\mu([k, k+2]) = \infty$ e questo è assurdo in quanto la misura di tale intervallo è 2 (e ciò segue dall'invarianza per traslazioni). \square

La difficoltà appena enunciata (cioè l'impossibilità di estendere la probabilità a tutti i sottoinsiemi di un insieme Ω) non si pone quando l'insieme Ω è finito o numerabile. In tal caso è usuale, anche se non obbligatorio, considerare come σ -algebra l'insieme $\mathcal{P}(\Omega)$ di tutte le parti di Ω . In tal caso la probabilità è univocamente determinata dai suoi valori sui singoletti $\{\omega\}$. Infatti, una volta stabiliti i valori $P(\{\omega\}) \geq 0$ tali che $\sum_i P(\{\omega\}) = 1$ varrà che per ogni $A \in \mathcal{P}(\Omega)$ si ha

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

Come notazione precisiamo che d'ora innanzi scriveremo $P(\omega)$ per indicare la probabilità $P(\{\omega\})$ sul singoletto.

¹l'esistenza di V discende dall'esistenza di basi di \mathbb{R} come spazio vettoriale su \mathbb{Q} , nel senso che ogni insieme di vettori linearmente indipendenti può essere esteso a base dello spazio (segue dal lemma di Zorn). V sarà lo spazio generato dai vettori di completamento.

2.2 Probabilità classica e condizionata

Adesso ricaviamo in due casi particolari di spazio di misura delle misure di probabilità su tali spazi. Partiamo con un caso ancora generale: sia (Ω, \mathcal{A}) uno spazio su cui è definita una misura μ finita. Possiamo sempre definire una misura di probabilità su tale spazio, ossia l'applicazione $P : \mathcal{A} \rightarrow [0, 1]$ tale che

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

Se Ω è finito c'è una misura privilegiata rispetto alle altre, che è quella indotta dalla cardinalità. Tale misura è finita e quindi possiamo compiere la costruzione appena vista. In questo caso la misura di probabilità diventa

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

Teorema 2.2.1. *Su uno spazio (Ω, \mathcal{A}) dove Ω è finito è sempre definita una misura di probabilità uniforme.*

Dimostrazione. Bisogna considerare come misura di probabilità quella indotta dalla cardinalità e appena definita. \square

La misura di probabilità appena introdotta prende il nome di *probabilità classica uniforme*: è la probabilità elementare che sicuramente il lettore già conosceva che si calcola come numero di casi favorevoli al verificarsi dell'evento, diviso il numero di tutti i casi possibili. L'uniformità è dovuta al fatto che ad ogni singolo caso è attribuita la medesima probabilità.

Osservazione 2.2.1. A titolo d'esempio con un insieme infinito: su \mathbb{N} non è possibile definire una misura di probabilità uniforme. Infatti se attribuiamo al singolo caso la misura 0 allora segue che la misura di \mathbb{N} è 0; se invece gli attribuiamo una misura strettamente positiva si ha che \mathbb{N} avrebbe misura infinita.

Adesso vediamo un'altra costruzione. Sia $(\Omega, \mathcal{A}, \mu)$ uno spazio di misura e sia $A \in \mathcal{A}$ di misura non nulla. Si può definire una misura sulla σ -algebra $\mathcal{A}|_A = \{B \in \mathcal{A} \mid B \subseteq A\}$ dello spazio $(\Omega, \mathcal{A}|_A)$ per restrizione. La misura è $\mu' = \mu|_{\mathcal{A}|_A}$. L'applicazione di tale costruzione si ha su uno spazio già probabilizzato (Ω, \mathcal{A}, P) , su cui la nuova probabilità viene ad essere (ricordiamoci di normalizzare):

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

dove con $P(B|A)$ si intende la probabilità dell'evento B condizionata al verificarsi dell'evento A . Possiamo riassumere questa seconda costruzione nel seguente:

Teorema 2.2.2. Dato uno spazio di probabilità (Ω, \mathcal{A}, P) e dato $A \in \mathcal{A}$ con $P(A) > 0$ è ben definita la probabilità condizionata ad A sullo spazio $(\Omega, P(\cdot|A))$ dalla formula $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

Osservazione 2.2.2. Osserviamo che nella costruzione della probabilità condizionata ad A per ogni $B \in \mathcal{A}$ si ha che se B è disgiunto da A vale che $P(B|A) = 0$. Mentre se $B \subseteq A$ si ha che

$$P(B|A) = P(B) \cdot \frac{1}{P(A)},$$

ossia tutto quello che è contenuto in A è in qualche modo normalizzato rispetto ad A .

Adesso vediamo un risultato molto importante, in quanto mette in relazione le probabilità condizionate di due eventi uno rispetto all'altro. Prima però dobbiamo premettere:

Definizione 2.2.1. Sia (Ω, \mathcal{A}) uno spazio misurabile. Un sistema di alternative per Ω è una partizione $\{A_i\}_{i=1}^n \subseteq \mathcal{A}$ di Ω in insiemi misurabili.

Teorema 2.2.3. Sia $\{A_i\}_{i=1}^n$ un sistema di alternative. Allora per ogni $B \in \mathcal{A}$ si ha $P(B) = \sum_{i=1}^n P(A_i \cap B)$ e si può esprimere la probabilità come combinazione convessa delle probabilità condizionate ad A_i .

Dimostrazione. Per ogni $B \in \mathcal{A}$ si ha che $B = \bigcup_{i=1}^n (B \cap A_i)$ dove l'unione è disgiunta, e quindi

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i),$$

e la combinazione è convessa per definizione di probabilità condizionata. \square

Corollario 2.2.1 (formula di Bayes). Sia $\{A_i\}_{i=1}^n$ un sistema di alternative. Vale allora

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)}.$$

Dimostrazione. La dimostrazione è immediata. Infatti si può scrivere

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)},$$

che è la formula di Bayes enunciata. \square

Esempio 2.2.1. Vediamo subito un'applicazione pratica della formula di Bayes. Si vuole provare sperimentalmente l'efficacia di un nuovo esame medico, che serve a capire se un individuo soffre di una malattia M ; questo esame produce solo due risultati: "malato" (o "positivo") che indichiamo con m e "sano" (o "negativo") che indichiamo con s . Definiamo quindi gli eventi (disgiunti) T_m e T_s rispettivamente l'evento di risposta al test. Dalla sperimentazione clinica, effettuata su soggetti comprovatamente sani S e su soggetti comprovatamente malati M , si possono ricavare le probabilità

$$P(T_i|S) \quad \text{e} \quad P(T_i|M)$$

con $i \in \{s, m\}$, ossia le probabilità che il test sia positivo o negativo se provato su persone malate o sane. Supponiamo inoltre di conoscere $P(M)$ probabilità che un individuo sia malato nella regione di interesse (per esempio presa come frequenza di malattia).

Noi siamo interessati alle probabilità che in un certo senso misurano l'*attendibilità* del test clinico, ossia che un individuo sia sano o malato in relazione al risultato del test. Tali probabilità sono proprio $P(S|T_i)$ e $P(M|T_i) = 1 - P(S|T_i)$ e per calcolarle ci è d'aiuto la formula di Bayes. Ad esempio se siamo interessati alla probabilità che un soggetto sia sano dopo che il test ha dato risultato negativo si ha

$$P(S|T_s) = \frac{P(T_s|S)P(S)}{P(T_s|S)P(S) + P(T_s|M)P(M)} = \frac{P(T_s|S)P(S)}{P(T_s|S)P(S) + P(T_s|M)(1 - P(S))},$$

grazie alla formula di Bayes.

Adesso introduciamo il concetto di indipendenza stocastica: vogliamo cioè tradurre in formula l'idea che la conoscenza che si è realizzato l'evento A non modifica la valutazione di probabilità dell'evento B e viceversa. A tale scopo consideriamo due eventi A e B non trascurabili e scriviamo $P(A) = P(A|B)$ e $P(B) = P(B|A)$: un rapidissimo calcolo mostra che queste sono equivalenti tra loro e all'uguaglianza $P(A \cap B) = P(A)P(B)$. A differenza delle precedenti, l'ultima uguaglianza è simmetrica e vale anche nel caso in cui uno degli eventi, o anche entrambi, sono trascurabili; allora assurgiamo questa a definizione di indipendenza:

Definizione 2.2.2. Diciamo due eventi A e B *stocasticamente indipendenti* se vale $P(A \cap B) = P(A)P(B)$.

Definizione 2.2.3. Due eventi A e B si dicono *positivamente* (o *negativamente*) *correlati* se e solo se $P(A|B) > P(A)$ (rispettivamente $P(A|B) < P(A)$).

2.3 Principio di inclusione–esclusione

Presentiamo qui il principio di inclusione–esclusione, ossia una formula che correla la cardinalità dell'unione di più insiemi con quella degli insiemi stessi e delle loro intersezioni. Data la sua importanza daremo due dimostrazioni di questo principio.

Teorema 2.3.1 (principio di inclusione–esclusione). *Sia $\{A_i\}_{i=1}^n$ una famiglia finita di insiemi finiti. Allora vale:*

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| &= \sum_{i=1}^n |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| + \cdots + (-1)^n |A_1 \cap \cdots \cap A_n| \\ &= \sum_{i=1}^n \left((-1)^{i+1} \sum_{1 \leq j_1 < \cdots < j_i \leq n} \left| \bigcap_{k=1}^i A_{j_k} \right| \right). \end{aligned}$$

Dimostrazione. La dimostrazione è molto semplice, si procede per induzione su $n \geq 1$. Se $n = 1$ o $n = 2$ la formula è banale; se $n > 1$ allora basterà scrivere $\bigcup_{i=1}^{n+1} A_i = \bigcup_{i=1}^n A_i \cup A_{n+1}$ e applicare la formula per il caso di due insiemi e l'ipotesi induttiva. \square

La seconda dimostrazione si basa sulla *funzione indicatrice* di un insieme.

Definizione 2.3.1. Sia X un insieme. La funzione $\chi_X : X \rightarrow \{0, 1\}$ definita da

$$\chi_X(x) = \begin{cases} 1 & \text{se } x \in X \\ 0 & \text{se } x \notin X \end{cases}$$

viene detta *funzione indicatrice* (o caratteristica) di X .

Sono di immediata verifica le seguenti proprietà:

- (1) $\chi_{\bigcap_{i=1}^n X_i} = \prod_{i=1}^n \chi_{X_i}$;
- (2) $\chi_{X^c} = 1 - \chi_X$;
- (3) $\chi_{\bigcup_{i=1}^n X_i} = 1 - \prod_{i=1}^n (1 - \chi_{X_i})$;
- (4) $X = Y$ se e solo se $\chi_X = \chi_Y$.

La seconda dimostrazione si basa proprio sulla proprietà (4), in quanto basta calcolare la funzione caratteristica degli insiemi ai due membri e si ha la tesi.

Adesso passiamo ad un caso particolare della formula: presentiamo la formula di inclusione–esclusione nel caso di famiglie con regolarità di intersezione.

Definizione 2.3.2. Una famiglia di insiemi finiti $\{E_i\}_{i=1}^n$ si dice avere *regolarità di intersezione* se l'intersezione di k qualunque di essi abbia cardinalità che dipende solo da k .

Proposizione 2.3.1. Sia $\{E_i\}_{i=1}^n$ una famiglia di insiemi con regolarità di intersezione e per ogni $i_1 < \dots < i_k$ sia $\left| \bigcap_{j=1}^k E_{i_j} \right| = c_k$. Allora

$$\left| \bigcup_{i=1}^n E_i \right| = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} c_k.$$

Dimostrazione. La formula di inclusione–esclusione generale è

$$\left| \bigcup_{i=1}^n E_i \right| = \sum_{i=1}^n |E_i| - \sum_{i_1 < i_2} |E_{i_1} \cap E_{i_2}| + \dots + (-1)^n |E_1 \cap \dots \cap E_n|.$$

Siccome la k -esima somma ha $\binom{n}{k}$ termini (cioè la cardinalità dei k -sottoinsiemi dell'insieme $\{1, \dots, n\}$) e i loro termini sono costanti c_k . \square

Esempio 2.3.1. Qual è il numero di permutazioni $\sigma \in \mathcal{S}_n$ senza punti fissi?

Sia $E_i = \{\sigma \in \mathcal{S}_n \mid \sigma(i) = i\}$: questo ha cardinalità $(n-1)!$ perché una permutazione di E_i è determinata da una permutazione di $I - \{i\}$. Per ogni $J \subseteq I$ definiamo

$$E_J = \bigcap_{i \in J} E_i = \{\sigma(i) = i \text{ per ogni } i \in J\}.$$

Detta $k = |J|$ vale che $|E_J| = (n-k)!$ perché, non considerati quei k elementi che stanno fissi, bisogna permutare i rimanenti $n-k$ elementi. Quindi la famiglia degli E_J ha la proprietà di regolarità di intersezione, e quindi

$$|X^C| = \left| \bigcup_{i=1}^n E_i \right| = \sum_{i=1}^n (-1)^{k+1} \binom{n}{k} (n-k)! = n! \sum_{i=1}^n (-1)^{k+1} \frac{1}{k!}.$$

Allora $|X| = n! - |X^C| = n! \sum_{k=0}^n \frac{(-1)^k}{k!}$. Inoltre per il criterio di Leibniz possiamo fornire la stima

$$\left| \frac{1}{e} - \sum_{k=0}^n \frac{(-1)^k}{k!} \right| \leq \frac{1}{(n+1)!}.$$

2.4 Impostazione soggettivista in probabilità

In un certo momento dell'evoluzione del pensiero matematico riguardo la probabilità, e come ora vedremo sicuramente durante una fase ancora intuitiva e poco formale, si dette all'argomento un'impostazione soggettivista. La definizione seguente è dovuta a Bruno de Finetti²

²precisiamo che De Finetti era un assicuratore, pertanto la sua definizione di probabilità è stata in qualche modo influenzata dalla sua professione.

Definizione 2.4.1. La *probabilità* di un evento E , secondo un individuo, è il prezzo che egli ritiene equo pagare in cambio di un importo unitario esigibile al verificarsi dell'evento.

Definizione 2.4.2. La valutazione degli $E \in \mathcal{A}$, con \mathcal{A} algebra finita, è *coerente* se non è possibile trovare una combinazione di scommesse sugli eventi E che dia un guadagno certo.

Esempio 2.4.1. Se X valuta $\frac{1}{6}$ la probabilità che esca 1, 2 o 3 nel lancio di un dado e valuta $\frac{3}{7}$ la probabilità che esca un numero minore o uguale a 3, questa è coerente? La risposta è no, e il motivo è che esiste una combinazione di scommesse che sicuramente produce un guadagno.

La definizione di coerenza è sufficiente per mostrare le proprietà che vorremmo siano soddisfatte da una probabilità. Infatti dalla coerenza segue che

$$0 \leq P(E) \leq 1$$

e che la probabilità è additiva. Inoltre l'identità $P(B|A)P(A) = P(B \cap A)$ segue anch'essa dalla coerenza. Infatti siano $P(A)$ e $P(B \cap A)$ e $P(B|A)$ le valutazioni sugli eventi A , $A \cap B$ e B nell'ipotesi che si sia già verificato A . Non possono valere le disuguaglianze per la coerenza, e dunque vale l'uguaglianza che cercavamo³.

³anche da un punto di vista classico si ha direttamente dalla definizione che

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

Capitolo 3

Somme di numeri e funzioni misurabili

Il lettore già avrà la nozione di serie: la somma di una serie di numeri positivi è definita come il limite, quando esiste, delle somme parziali. Questa nozione però è limitata a famiglie numerabili di numeri e invece noi vorremmo generalizzare questo concetto: vediamo come ciò sia possibile introducendo le somme su famiglie. In seguito vedremo gli integrali definiti grazie ad una misura, che utilizzeranno proprio questa nozione di somma.

3.1 Somme di famiglie di numeri reali

Ricordiamo alcune proprietà dell'estremo superiore di un insieme¹. Le proprietà che vogliamo ricordare:

- (1) se per ogni $a \in A$ esiste $b \in B$ tale che $b \geq a$ allora $\sup A \leq \sup B$;
- (2) $\sup (\bigcup_{i \in I} A_i) = \sup_i \sup A_i$;
- (3) $\sup(A + B) = \sup A + \sup B$, basta infatti scrivere $A + B = \bigcup_{b \in B} (A + b)$.

Definizione 3.1.1. Dato I insieme di indici indichiamo con $\mathcal{F}(I)$ la *famiglia delle parti finite* di I , ossia l'insieme dei sottoinsiemi finiti di I .

Definizione 3.1.2. Sia $\{a_i\}$ una famiglia di numeri reali non negativi indicizzata su $i \in I$. Definiamo la *somma della famiglia*, e la si indica con $\sum_{i \in I} a_i \in [0, \infty]$, come

$$\sup_{J \in \mathcal{F}(I)} \sum_{i \in J} a_i.$$

Se la somma è finita diremo che la famiglia $\{a_i\}_{i \in I}$ è *sommabile*.

¹ricordiamo che dato $A \subseteq \mathbb{R}$ con $\sup A$ indichiamo il minimo dell'insieme dei maggioranti di A (che esiste per l'assioma di separazione dei numeri reali).

Lemma 3.1.1. *La somma di famiglie è monotona, ossia: se $0 \leq a_i \leq b_i$ per ogni $i \in I$ allora $\sum_{i \in I} a_i \leq \sum_{i \in I} b_i$.*

Dimostrazione. Basta usare la proprietà (1) dell'estremo superiore e la definizione di somma di una famiglia di numeri reali non negativi. \square

Proposizione 3.1.1. *Vale l'associatività sulla somma: se $\{a_i\}_{i \in I}$ è una famiglia sommabile di numeri non negativi e se $\{I_\lambda\}_{\lambda \in \Lambda}$ è una partizione di I si ha che*

$$\sum_{i \in I} a_i = \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i.$$

Dimostrazione. Basta scrivere la seguente catena:

$$\begin{aligned} \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i &= \sup_{K \in \mathcal{F}(\Lambda)} \sum_{\lambda \in K} \sum_{i \in I_\lambda} a_i = \sup_{K \in \mathcal{F}(\Lambda)} \sum_{\lambda \in K} \sup_{J \in \mathcal{F}(I_\lambda)} \sum_{i \in J} a_i = \\ &= \sup_{K \in \mathcal{F}(\Lambda)} \sup_{J \in \mathcal{F}(I_\lambda)} \sum_{\lambda \in K} \sum_{i \in J} a_i = \sup_{K \in \mathcal{F}(\Lambda)} \sup_{J \subseteq \mathcal{F}(\bigcup_{\lambda \in K} I_\lambda)} \sum_{i \in J} a_i = \\ &= \sup_{J \in \mathcal{F}(I)} \sum_{i \in J} a_i, \end{aligned}$$

e la tesi è provata. \square

Osservazione 3.1.1. A parziale motivazione dell'introduzione delle somme di famiglie di numeri non negativi osserviamo il seguente fatto. Se Ω è un insieme e $\{p_\omega\}_{\omega \in \Omega}$ una famiglia di numeri non negativi allora è definita su $(\Omega, \mathcal{P}(\Omega))$ una misura. Per ogni $A \in \mathcal{P}(\Omega)$ si definisce infatti

$$\mu(A) = \sum_{\omega \in A} p_\omega,$$

e si verifica immediatamente che questa è una misura. Ci limitiamo solo a osservare che dall'associatività discende la σ -additività di μ .

Senza perdere di generalità nel caso di misure finite possiamo limitarci a Ω al più numerabile. Infatti vale che se $\{a_i\}_{i \in I}$ è una famiglia sommabile di numeri non negativi allora l'insieme $I' = \{i \mid a_i \neq 0\}$ è al più numerabile².

Adesso vogliamo chiederci come possiamo estendere la nozione di sommabilità a famiglie di numeri che cambiano segno. In generale, se $\{a_i\}_{i \in I}$ è una famiglia di numeri non positivi non è possibile dare un senso alla somma della famiglia. Adesso però aggiungiamo un'ipotesi abbastanza forte e vedremo che la buona definizione è raggiunta. L'ipotesi è la seguente: si suppone che $a_i = p_i - n_i$, con $p_i, n_i \geq 0$ per ogni i , e che $\sum_{i \in I} p_i$ e $\sum_{i \in I} n_i$ siano entrambe finite.

²infatti, se così non fosse, esisterebbe $\varepsilon > 0$ tale che $a_i > \varepsilon > 0$ per un'infinità più che numerabile di indici i . Ma allora $\sum_{i \in I'} a_i > \sum_{i \in I'} \varepsilon = +\infty$.

Definizione 3.1.3. Sia $\{a_i\}_{i \in I}$ una qualsiasi famiglia di numeri reali che verifica le ipotesi appena scritte. Allora si pone per definizione che

$$\sum_{i \in I} a_i = \sum_{i \in I} p_i - \sum_{i \in I} n_i.$$

Per rendere giustizia alla precedente definizione dobbiamo mostrarne la buona definizione. Supponiamo dunque che $a_i = p'_i - n'_i$, con $p'_i, n'_i \geq 0$ per ogni i , e che $\sum_{i \in I} p'_i$ e $\sum_{i \in I} n'_i$ siano entrambe finite. Si ha $a_i = p_i - n_i = p'_i - n'_i$, e allora si può scrivere

$$\sum_{i \in I} p_i + \sum_{i \in I} n'_i = \sum_{i \in I} (p_i + n'_i) = \sum_{i \in I} (p'_i + n_i) = \sum_{i \in I} p'_i + \sum_{i \in I} n_i,$$

da cui $\sum_i p_i - \sum_i n_i = \sum_i p'_i - \sum_i n'_i$.

Osservazione 3.1.2. Supponiamo che valga $a_i = p_i - n_i$ per ogni i con le proprietà già espresse. Allora $a_i \leq p_i$ e $p_i \geq 0$ quindi

$$0 \leq a_i^+ = \max(0, a_i) \leq p_i.$$

Analogamente essendo $-a_i \leq n_i$ e $n_i \geq 0$ si ha $0 \leq a_i^- = \max(0, -a_i) \leq n_i$ e quindi $\sum_{i \in I} a_i^+$ e $\sum_{i \in I} a_i^-$ sono entrambe ancora finite. Dunque la proprietà iniziale equivale a chiedere che queste due somme sono finite.

Teorema 3.1.1 (di Tonelli). *Sia $\{a_{ij}\}_{(i,j) \in I \times J}$ una famiglia sommabile di numeri reali non negativi. Allora*

$$\sum_{(i,j) \in I \times J} a_{ij} = \sum_{j \in J} \sum_{i \in I} a_{ij} = \sum_{i \in I} \sum_{j \in J} a_{ij}.$$

Dimostrazione. Questo è un caso particolare della associatività delle somme infinite. Siccome si può scrivere

$$I \times J = \bigcup_{j \in J} I \times \{j\} = \bigcup_{i \in I} \{i\} \times J,$$

partizionando il prodotto cartesiano, basta applicare l'associatività. \square

Adesso, in analogia con quanto fatto con le serie nel corso di analisi matematica, possiamo introdurre il concetto di assoluta sommabilità:

Definizione 3.1.4. Una famiglia di numeri $\{a_i\}_{i \in I}$ si dice *assolutamente sommabile* se la famiglia $\{|a_i|\}_{i \in I}$ è sommabile, ossia se $\sum_{i \in I} |a_i|$ è finita.

Con un metodo lievemente diverso si può mostrare anche l'associatività per le famiglie assolutamente sommabili. Questo è il contenuto della prossima:

Proposizione 3.1.2. *Vale l'associatività sulla somma: se $\{a_i\}_{i \in I}$ è una famiglia assolutamente sommabile di numeri e se $\{I_\lambda\}_{\lambda \in \Lambda}$ è una partizione di I si ha che*

$$\sum_{i \in I} a_i = \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i.$$

Dimostrazione. Sia $\{a_i\}_{i \in I} \subseteq \mathbb{R}$ con $\sum_{i \in I} |a_i|$ finita e $\{I_\lambda\}_{\lambda \in \Lambda}$ partizione di I . Osserviamo che per ogni $\lambda \in \Lambda$ si ha che $\{a_i\}_{i \in I_\lambda}$ è anch'essa assolutamente sommabile; infatti $\sum_{i \in I_\lambda} |a_i| \leq \sum_{i \in I} |a_i| < \infty$, e quindi vale intanto

$$\sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i = \sum_{\lambda \in \Lambda} \left(\sum_{i \in I_\lambda} a_i^+ - \sum_{i \in I_\lambda} a_i^- \right).$$

Poniamo, per ogni $\lambda \in \Lambda$, $p_\lambda = \sum_{i \in I_\lambda} a_i^+ \geq 0$ e $n_\lambda = \sum_{i \in I_\lambda} a_i^- \geq 0$. Abbiamo che

$$\sum_{\lambda \in \Lambda} p_\lambda = \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i^+ = \sum_{i \in I} a_i^+ \leq \sum_{i \in I} |a_i|$$

e similmente possiamo operare sull'altra. Quindi possiamo andare avanti dalla precedente con il passaggio

$$\sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i = \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i^+ - \sum_{\lambda \in \Lambda} \sum_{i \in I_\lambda} a_i^- = \sum_{i \in I} a_i^+ - \sum_{i \in I} a_i^- = \sum_{i \in I} a_i,$$

e la tesi è provata. \square

In piena analogia con il teorema di Tonelli, che seguiva dall'associatività per famiglie sommabili non negative, esiste anche il corrispondente per famiglie assolutamente sommabili.

Teorema 3.1.2 (di Fubini). *Sia $\{a_{ij}\}_{(i,j) \in I \times J}$ una famiglia assolutamente sommabile di numeri reali. Allora*

$$\sum_{(i,j) \in I \times J} a_{ij} = \sum_{j \in J} \sum_{i \in I} a_{ij} = \sum_{i \in I} \sum_{j \in J} a_{ij}.$$

Dimostrazione. La dimostrazione è identica a quella presentata per il teorema di Tonelli. \square

Osservazione 3.1.3 (somme alla Cauchy). Nel caso particolare in cui $I = J = \mathbb{N}$ si ha

$$\sum_{(i,j) \in \mathbb{N}^2} a_{ij} = \sum_{n=1}^{\infty} \sum_{i=0}^n a_{i,n-i}.$$

Infatti questa regola di associatività corrisponde alla partizione $\mathbb{N} \times \mathbb{N} = \bigcup_{n \in \mathbb{N}} \{(i,j) \mid i+j = n\}$.

Osservazione 3.1.4. Vediamo adesso un caso particolare del teorema di Fubini–Tonelli. Vale che

$$\sum_{i \in I} (a_i + b_i) = \sum_{i \in I} a_i + \sum_{i \in I} b_i.$$

Basta infatti prendere $J = \{1, 2\}$, porre $a_i = a_{i,1}$ ed anche $b_i = a_{i,2}$. In tal modo si osserva che il primo membro è $\sum_{i \in I} \sum_{j=1}^2 a_{ij}$.

Osservazione 3.1.5. Considerata la famiglia $\{a_i\}_{i \in I}$ come un'applicazione $a : I \rightarrow \mathbb{R}$ (quella tale che $a(i) = a_i$) si può scrivere

$$\sum_{i \in I} a(i) = \sum_{t \in \mathbb{R}} tN(t),$$

dove $N(t) = |a^{-1}(t)|$. Basta considerare l'unione disgiunta $I = \bigcup_{t \in \mathbb{R}} a^{-1}(t)$.

Osservazione 3.1.6. Sia $E \subseteq \mathcal{F}(I)$ una sottofamiglia di parti finite di I , cofinale nell'ordinamento di $\mathcal{F}(I)$ (ossia per ogni $J \in \mathcal{F}(I)$ esiste $J' \in E$ tale che $J' \subseteq J$). Se $\{a_i\}_{i \in I}$ è una famiglia di numeri non negativi, allora

$$\sum_{i \in I} a_i = \sup_{J \in E} \sum_{i \in J} a_i.$$

In particolare, se $I = \mathbb{N}$ e $E = \{J_n\}$ con $J_n = \{0, \dots, n\}$ allora

$$\sum_{i \in \mathbb{N}} a_i = \sup_{n \in \mathbb{N}} \sum_{i=0}^n a_i,$$

che è la definizione di serie.

3.2 Teoremi di passaggio al limite

Siano $a_k : I \rightarrow \mathbb{R}$ funzioni convergenti puntualmente ad $a : I \rightarrow \mathbb{R}$. Possiamo concludere che $\sum_{i \in I} a_k(i) \rightarrow \sum_{i \in I} a(i)$? La risposta è no.

Esempio 3.2.1. Consideriamo $I = \mathbb{N}$ e $a_k(i) = \delta_{ki}$ (delta di Kronecker). Per ogni i si ha $a_k(i) \rightarrow 0$ (anzi, è definitivamente nullo). Ma

$$\sum_{i \in \mathbb{N}} a_k(i) = 1$$

ed è diversa dalla somma della funzione limite, che è nulla.

Teorema 3.2.1 (teorema di convergenza dominata). *Siano $a_k : I \rightarrow \mathbb{R}$ convergenti puntualmente ad $a : I \rightarrow \mathbb{R}$. Sia $g : I \rightarrow \mathbb{R}$ tale che per ogni $k \in \mathbb{N}$ vale $|a_k(i)| \leq g(i)$ con $\sum_{i \in I} g(i) < \infty$. Allora $\sum_{i \in I} a_k(i) \rightarrow \sum_{i \in I} a(i)$ quando $k \rightarrow \infty$.*

Dimostrazione. Dimostriamo il teorema nel caso particolare delle serie, ossia con $I = \mathbb{N}$. Siccome per ipotesi $\sum_{i=0}^{\infty} g(i)$ è finita si ha che per ogni $\varepsilon > 0$ esiste $n \in \mathbb{N}$ tale che i resti n -esimi sono infinitesimi, ossia $\sum_{i=n+1}^{\infty} g(i) < \varepsilon$. Si ha, per ogni $k \in \mathbb{N}$

$$\begin{aligned} \left| \sum_{i=1}^{\infty} a_k(i) - \sum_{i=1}^{\infty} a(i) \right| &\leq \left| \sum_{i=0}^n a_k(i) - \sum_{i=0}^n a(i) \right| + \sum_{i=n+1}^{\infty} |a_k(i)| + \sum_{i=n+1}^{\infty} |a(i)| \leq \\ &\leq \left| \sum_{i=0}^n a_k(i) - \sum_{i=0}^n a(i) \right| + 2 \sum_{i=n+1}^{\infty} g(i) \leq \\ &\leq \left| \sum_{i=0}^n a_k(i) - \sum_{i=0}^n a(i) \right| + 2\varepsilon. \end{aligned}$$

Visto che per ogni $i \in \mathbb{N}$ si ha $a_k(i) \rightarrow a(i)$ per $k \rightarrow \infty$, per k sufficientemente grande, si ha che $|a_k(i) - a(i)| < \varepsilon$, e allora segue che

$$\left| \sum_{i=1}^{\infty} a_k(i) - \sum_{i=1}^{\infty} a(i) \right| < 3\varepsilon.$$

La tesi è così provata. \square

Teorema 3.2.2 (di Beppo Levi). *Siano $a_k : I \rightarrow \mathbb{R}$ funzioni positive e che convergono puntualmente crescendo a $a : I \rightarrow \mathbb{R}$. Allora $\sum_{i \in I} a_k(i) \rightarrow \sum_{i \in I} a(i)$ crescendo.*

Dimostrazione. Supponiamo che $\sum_{i \in I} a(i)$ diverga. Allora per ogni $N \geq 0$ esiste $J \in \mathcal{F}(I)$ tale che $\sum_{i \in J} a(i) > N$. Siccome $a_k \rightarrow a$ puntualmente si ha che per qualche k (anzi, definitivamente) le a_k supereranno N e dunque

$$\sum_{i \in I} a_k(i) \geq \sum_{i \in J} a_k(i) > N.$$

E allora $\sum_{i \in I} a_k(i) \rightarrow \infty = \sum_{i \in I} a(i)$.

Se invece la somma è finita allora la convergenza è assicurata dal teorema di convergenza dominata in cui si prende $g(i) = a(i)$. \square

3.3 Funzioni misurabili

Definizione 3.3.1. Siano (Ω, \mathcal{A}) e (Ω', \mathcal{A}') spazi misurabili. Una funzione f si dice *misurabile* se per ogni $A \in \mathcal{A}'$ si ha $f^{-1}(A) \in \mathcal{A}$.

Essendo la definizione di misurabilità analoga a quella di continuità in topologia si ha un risultato simile sulla composizione: la composizione di applicazioni misurabili è misurabile.

Osservazione 3.3.1. Da qui in avanti nel caso di funzioni a valori reali considereremo sempre la σ -algebra dei boreliani. Si faccia attenzione al fatto che non sarebbe lo stesso se prendessimo la σ -algebra degli insiemi misurabili secondo Lebesgue.

Proposizione 3.3.1. Sia $(\Omega, \mathcal{A}, \mu)$ uno spazio di misura e sia $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ una funzione misurabile. Allora esiste una misura μ' su (Ω', \mathcal{A}') definita da

$$\mu'(B) = \mu(f^{-1}(B)).$$

Dimostrazione. Osserviamo che μ' è ben definita in quanto $f^{-1}(B) \in \mathcal{A}$. Inoltre μ' è σ -additiva in quanto se $\{B_i\}_{i \in \mathbb{N}} \subseteq \mathcal{A}'$ allora

$$f^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \bigcup_{i \in \mathbb{N}} f^{-1}(B_i),$$

e questo conclude la dimostrazione. \square

Definizione 3.3.2. La misura μ' definita nel precedente teorema prende il nome di *misura immagine*³ di μ rispetto ad f .

Osservazione 3.3.2. Osserviamo che la proposizione appena enunciata vale anche nel caso particolare degli spazi di probabilità.

Adesso vediamo un criterio di misurabilità che torna utile in svariate situazioni: questo ci dice che se un'algebra è generata allora si può limitare la definizione di misurabilità solo agli insiemi che generano l'algebra:

Proposizione 3.3.2. Se \mathcal{B} è una famiglia di parti di \mathcal{A}' che genera \mathcal{A}' come σ -algebra allora f è misurabile se e solo se per ogni $B \in \mathcal{B}$ si ha che $f^{-1}(B) \in \mathcal{A}$.

Dimostrazione. (\implies) Questa implicazione è ovvia perché se f è misurabile allora $f^{-1}(A) \in \mathcal{A}$ per ogni $A \in \mathcal{A}'$ e dunque a maggior ragione la condizione è verificata per gli insiemi di \mathcal{B} .

³in inglese indicata con *push forward*.

(\Leftarrow) Consideriamo

$$E = \{A \in \mathcal{A}' \mid f^{-1}(A) \in \mathcal{A}\}$$

e mostriamo che $E = \mathcal{A}'$. Per ipotesi $\mathcal{B} \subseteq E \subseteq \mathcal{A}'$, e questo punto basta mostrare che E è una σ -algebra e poi la tesi segue per la minimalità di \mathcal{A}' in quanto algebra generata da \mathcal{B} . Il fatto che E sia una σ -algebra è abbastanza immediato: sia $\{B_i\}_{i \in \mathbb{N}} \subseteq E$, allora anche la loro unione ci sta perché

$$f^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \bigcup_{i \in \mathbb{N}} f^{-1}(B_i) \in E;$$

le altre verifiche sono immediate. \square

Esempio 3.3.1. Consideriamo il caso di $(\Omega', \mathcal{A}') = (\mathbb{R}, B(\mathbb{R}))$ (dove $B(\mathbb{R})$ è la σ -algebra dei boreliani). Questa è generata dagli aperti di \mathbb{R} , ed anche equivalentemente dalla famiglia

$$\mathcal{B} = \{(c, \infty) \mid c \in \mathbb{R}\}^4$$

In tal caso f è misurabile se per ogni $c \in \mathbb{R}$ si ha che $f^{-1}((c, \infty)) = \{f > c\} \in \mathcal{A}$ e si ritrova la definizione che probabilmente il lettore già conosce dal corso di analisi matematica.

Osservazione 3.3.3. Consideriamo il caso di un'applicazione $f : (\Omega, \mathcal{A}) \rightarrow \mathbb{Z}$ (o in \mathbb{N} o qualsiasi insieme numerabile con la topologia discreta). In tal caso f è misurabile se e solo se per ogni $k \in \mathbb{Z}$ si ha $f^{-1}(k) \in \mathcal{A}$. Questo, grazie al criterio, perché l'algebra considerata è generata dai singoletti (essendoci la topologia discreta). Come conseguenza di questo fatto si ha che f può essere scritta come segue:

$$f = \sum_{k \in \mathbb{Z}} k \cdot \chi_{E_k},$$

dove $E_k = \{f = k\}$ (e formano una partizione di Ω). Cioè per ogni $\omega \in \Omega$ si ha che $f(\omega) = \sum_{k \in \mathbb{Z}} k \chi_{E_k}(\omega) = j$ in quanto esiste un unico j tale che $\omega \in E_j$ (è unico proprio perché gli E_k formano una partizione di Ω).

Proposizione 3.3.3. *Siano $f_1, f_2 : \Omega \rightarrow \mathbb{R}$ due applicazioni misurabili. Allora si ha che $(f_1, f_2) : \Omega \rightarrow \mathbb{R}^2$, $f_1 + f_2$ e $f_1 f_2$ sono misurabili.*

Dimostrazione. Siano $I, J \subseteq \mathbb{R}$ intervalli, allora

$$(f_1, f_2)^{-1}(I \times J) = f_1^{-1}(I) \cap f_2^{-1}(J) \in \mathcal{A}.$$

⁴in realtà basterebbe $c \in \mathbb{Q}$. Infatti ogni aperto di \mathbb{R} è unione al più numerabile disgiunta di intervalli aperti (le sue componenti connesse); è numerabile perché per ogni intervallo esiste almeno un punto razionale in esso contenuto.

Ora, $\mathcal{B} = \{I \times J \mid I, J \text{ sono intervalli aperti di } \mathbb{R}\}$ è una base della topologia di \mathbb{R}^2 , ed inoltre può essere ridotta a base numerabile: basta prendere gli intervalli I e J a estremi razionali. Dunque ogni aperto di \mathbb{R}^2 è unione numerabile di elementi di \mathcal{B} , e questo significa che \mathcal{B} genera anche la σ -algebra dei boreliani di \mathbb{R}^2 . Dunque si conclude per il criterio di misurabilità su basi.

Siccome l'applicazione di somma e di prodotto per componenti sono continue si ha che sono anche misurabili. Dunque per composizione con l'applicazione della parte precedente si ha la tesi. \square

Proposizione 3.3.4. *Sia $\{f_n\}_{n \in \mathbb{N}}$ una successione di funzioni $f_n : \Omega \rightarrow \overline{\mathbb{R}}$ per ogni $n \in \mathbb{N}$. Allora $\sup_{n \in \mathbb{N}} f_n$, $\inf_{n \in \mathbb{N}} f_n$ sono misurabili.*

Dimostrazione. Basta osservare che per ogni $c \in \mathbb{R}$ si ha che

$$\left\{ \sup_{n \in \mathbb{N}} f_n \leq c \right\} = \bigcap_{n \in \mathbb{N}} \{f_n \leq c\} \in \mathcal{A}.$$

Infatti, per ogni $\omega \in \Omega$ si ha che $\sup_{n \in \mathbb{N}} f_n(\omega) \leq c \iff (\forall n \in \mathbb{N}, f_n(\omega) \leq c)$, e la tesi è provata in quanto la seconda parte è analoga. \square

Corollario 3.3.1. *Sia $\{f_n\}_{n \in \mathbb{N}}$ una successione di funzioni $f_n : \Omega \rightarrow \overline{\mathbb{R}}$ per ogni $n \in \mathbb{N}$. Allora $\limsup_{n \rightarrow \infty} f_n$, $\liminf_{n \rightarrow \infty} f_n$ sono misurabili.*

Dimostrazione. Basta ricordare che

$$\limsup_{n \rightarrow \infty} f_n = \inf_{k \in \mathbb{N}} \sup_{n \geq k} f_n \quad \text{e} \quad \liminf_{n \rightarrow \infty} f_n = \sup_{k \in \mathbb{N}} \inf_{n \geq k} f_n,$$

e abbiamo concluso grazie alla proposizione precedente. \square

Capitolo 4

Integrali secondo una misura

Dopo aver sviluppato le funzioni misurabili nel precedente capitolo, dobbiamo vedere come misurare tali funzioni. Questo problema porterà alla definizione dell'integrale secondo una misura, che sarà alla base di tutta la teoria che svilupperemo.

4.1 Funzioni integrabili rispetto ad una misura

Adesso vogliamo presentare la nozione di integrale rispetto ad una misura. Ciò presenterà molte analogie con le somme infinite di famiglie di numeri reali: iniziamo infatti da $(\Omega, \mathcal{A}, \mu)$ spazio misurabile e da funzioni misurabili a valori non negativi.

Definizione 4.1.1. Una funzione $\varphi : \Omega \rightarrow \mathbb{R}$ si dice *semplice* se è misurabile e la cardinalità dell'immagine è finita. Denotiamo con S l'insieme delle funzioni semplici.

Lemma 4.1.1. *Una funzione φ è semplice se e solo se è combinazione lineare di funzioni caratteristiche di insiemi misurabili.*

Dimostrazione. (\Leftarrow) Vale che $A \in \mathcal{A}$ allora χ_A è misurabile; quindi anche combinazioni lineari di funzioni caratteristiche risultano misurabili e semplici.

(\Rightarrow) Viceversa, se φ è misurabile a valori nell'insieme $J \subseteq \mathbb{R}$ finito allora si può scrivere

$$\varphi = \sum_{c \in J} c \cdot \chi_{\{\varphi=c\}}$$

e gli insiemi $\{\varphi = c\}$ sono misurabili (controimmagine di un punto, che è misurabile). \square

Osservazione 4.1.1. Osserviamo che la classe S delle funzioni semplici è uno spazio vettoriale. È un sottospazio dello spazio vettoriale di tutte le funzioni misurabili, e precisamente è quello generato dalla funzioni caratteristiche χ_A con $A \in \mathcal{A}$.

Adesso ci apprestiamo a definire l'integrale delle funzioni semplici a valori non negativi. Vogliamo precisare un fatto: se avessimo considerato spazi di misura con misura finita la definizione di integrale per le funzioni semplici poteva essere data sempre. Noi però siamo nel caso generale di misure possibilmente infinite, quindi dobbiamo procedere gradualmente.

Definizione 4.1.2. Sia $\varphi : \Omega \rightarrow [0, \infty)$ una funzione semplice con $\varphi = \sum_{j \in J} a_j \cdot \chi_{A_j}$. Si definisce il suo integrale su Ω come

$$\int_{\Omega} \varphi d\mu = \sum_{j \in J} a_j \mu(A_j).^1$$

Osservazione 4.1.2. Dobbiamo mostrare la buona definizione di integrale, ossia che non dipende dalla scrittura di φ : si prova che se $\varphi = \sum_{j \in J_1} a_j \cdot \chi_{A_j} = \sum_{j \in J_2} a_j \cdot \chi_{A_j}$ con $J_1 \cap J_2 = \emptyset$ insiemi di indici finiti allora vale

$$\sum_{j \in J_1} a_j \mu(A_j) = \sum_{j \in J_2} a_j \mu(A_j).$$

Visto che qualche $\mu(A_j)$ potrebbe essere infinito non si può portare tutto a primo membro, perché magari otterremmo delle differenze indeterminate². Ma una volta

¹la relativa definizione per misure finite è la stessa, solo che φ poteva assumere valori di segno arbitrario.

²nel caso di misure finite invece ciò era possibile: in realtà basta mostrare che se $\varphi = \sum_{j \in J} a_j \cdot \chi_{A_j} = 0$ allora $\int_{\Omega} \varphi d\mu = 0$: infatti se vale la doppia scrittura di φ basta considerare la differenza dei membri. Siano, per ogni $X \in \mathcal{P}(J)$

$$B_X = \bigcap_{i \in X} A_i - \bigcup_{i \notin X} A_i.$$

Tali insiemi sono a due a due disgiunti (infatti per ogni $i \in X$ vale $B_X \subseteq A_i$ mentre se $i \notin X$ allora $B_X \cap A_i = \emptyset$) e inoltre ogni A_i si scrive come $A_i = \bigcup_{i \ni X} B_X$. Posto

$$b_X = \sum_{i \in X} a_i$$

allora $\sum_{j \in J} a_j \cdot \chi_{A_j} = \sum_{X \in \mathcal{P}(J)} b_X \cdot \chi_{B_X}$. Allora

$$\sum_{X \in \mathcal{P}(J)} b_X \mu(B_X) = \sum_{X \in \mathcal{P}(J)} \left(\sum_{j \in X} a_j \right) \mu(B_X) = \sum_{j \in J} a_j \sum_{\substack{X \in \mathcal{P}(J) \\ j \in X}} \mu(B_X) = \sum_{j \in J} a_j \mu(A_j).$$

Dunque se in particolare $\varphi = \sum_{j \in J} a_j \cdot \chi_{A_j}$ è la funzione nulla allora $b_X = 0$ per ogni $X \in \mathcal{P}(J)$ e dunque $\sum_{j \in J} a_j \mu(A_j) = 0$.

definito

$$B_X = \bigcap_{i \in X} A_i - \bigcup_{i \notin X} A_i,$$

con $X \in \mathcal{P}(J_1 \cup J_2)$ si ottiene che entrambi i membri dell'uguaglianza da provare sono uguali a $\sum_{a \in \mathcal{P}(J_1 \cup J_2)} b_X \mu(A_j)$, dove $b_X = \varphi(B_X)$.

Non è difficile vedere che dalla definizione seguono le seguenti proprietà:

- (1) si ha $\int_{\Omega} \varphi d\mu \geq 0$;
- (2) l'integrale è *lineare* per coefficienti positivi, ossia dati $a, b \geq 0$ e $\varphi, \psi \in S$ a valori non negativi si ha

$$\int_{\Omega} (a\varphi + b\psi) d\mu = a \int_{\Omega} \varphi d\mu + b \int_{\Omega} \psi d\mu;$$

- (3) l'integrale è *monotono*, a dire che se $\varphi \geq \psi \geq 0$ sono due funzioni semplici allora

$$\int_{\Omega} \varphi d\mu = \int_{\Omega} \psi d\mu.^3$$

A questo punto diamo la definizione di integrale rispetto a μ per funzioni misurabili a valori non negativi:

Definizione 4.1.3. Sia f una funzione misurabile a valori non negativi. Definiamo il suo valore su Ω come

$$\int_{\Omega} f d\mu = \lim_{k \rightarrow \infty} \int_{\Omega} \varphi_k d\mu,$$

con $\{\varphi_k\}_{k=0}^{\infty} \subseteq S$ successione di funzioni non negative convergente crescendo a f .

Ancora una volta dobbiamo verificare la buona definizione e si devono mostrare due fatti:

- (1) la successione $\{\varphi_k\}_{k=0}^{\infty}$ con le proprietà richieste esiste sempre;
- (2) il limite è indipendente dalla successione scelta.

La verifica di queste proprietà è il contenuto delle prossime due proposizioni.

Proposizione 4.1.1. Sia $f : \Omega \rightarrow [0, +\infty]$ una funzione misurabile. Allora esiste una successione di funzioni semplici $\{\varphi_k\}_{k=0}^{\infty} \subseteq S$ non negative che convergono crescendo a f .

Dimostrazione. Basta considerare per ogni $k \in \mathbb{N}$

$$\varphi_k = \min \left(2^{-k} \lfloor 2^k f \rfloor, k \right).$$

³ basta solo scrivere le definizioni e osservare che se $\varphi = \sum_{j \in J} a_j \cdot \chi_{A_j}$ e $\psi = \sum_{j \in J} b_j \cdot \chi_{A_j}$ segue dall'ipotesi che $a_j \geq b_j$ per ogni $j \in J$. Allora è semplice concludere.

Siccome $\lfloor 2^k f \rfloor \leq 2^k f \leq \lfloor 2^k f \rfloor + 1$ si ha che

$$0 \leq f - 2^{-k} \lfloor 2^k f \rfloor < 2^{-k}.$$

Perciò si ha che $2^{-k} \lfloor 2^k f \rfloor \rightarrow f$ ed anche che $\min(2^{-k} \lfloor 2^k f \rfloor, k) \rightarrow f$. Infine la successione $\{\varphi_k\}$ è crescente: infatti per le proprietà della parte intera si ha

$$2 \lfloor 2^k f \rfloor \leq \lfloor 2^{k+1} f \rfloor$$

e si ottiene la crescita moltiplicando per 2^{-k-1} . \square

Proposizione 4.1.2. *Sia $f : \Omega \rightarrow [0, +\infty]$ una funzione misurabile e $\{\varphi_k\}_{k=0}^{\infty} \subseteq S$ una successione di funzioni non negative convergente crescendo a f . Il limite $\lim_{k \rightarrow \infty} \int_{\Omega} \varphi_k d\mu$ non dipende dalla successione $\{\varphi_k\}_{k=0}^{\infty}$ scelta.*

Dimostrazione. Per fare questo si fa uso di un lemma, che riportiamo subito:

Lemma 4.1.2. *Sia $\{\varphi_k\}_{k=0}^{\infty} \subseteq S$ una successione di funzioni semplici non negative e che convergono crescendo a φ misurabile non negativa. Sia inoltre $\psi \leq \varphi$ semplice. Allora*

$$\lim_{k \rightarrow \infty} \int_{\Omega} \varphi_k d\mu \geq \int_{\Omega} \psi d\mu.$$

Dimostrazione. Iniziamo dal caso di $\psi = \chi_A$ con $A \in \mathcal{A}$. Fissato $\varepsilon > 0$ definiamo per ogni $n \in \mathbb{N}$ l'insieme

$$A_n = \{x \in A \mid \varphi_n(x) \geq 1 - \varepsilon\}.$$

Siccome per ipotesi si ha convergenza crescente si ha $A = \bigcup_{n=0}^{\infty} A_n$: infatti se $x \in A$ allora $\varphi(x) \geq \psi(x) = 1$ e quindi per la convergenza crescente $\varphi_n(x) \geq 1 - \varepsilon$ definitivamente e quindi $x \in A_n$ definitivamente. Essendo l'unione crescente si ha $\mu(A) = \lim_{n \rightarrow \infty} \mu(A_n)$. Siccome $\varphi_n \geq (1 - \varepsilon) \cdot \chi_{A_n}$ si ha per monotonia

$$\int_{\Omega} \varphi_n d\mu \geq (1 - \varepsilon) \int_{\Omega} \chi_{A_n} d\mu = (1 - \varepsilon)\mu(A_n).$$

Passando al limite per $n \rightarrow \infty$ si ha la tesi nel nostro caso. A questo punto se $\psi \in S$ la tesi segue per linearità. \square

Sia ora $\{\psi_k\}_{k=0}^\infty$ un'altra successione che converge crescendo verso f . Essendo $\max(\varphi_j, \psi_j) \leq f$ segue per il lemma la seguente disuguaglianza:

$$\lim_{k \rightarrow \infty} \int_{\Omega} \varphi_k d\mu \geq \int_{\Omega} \max(\varphi_j, \psi_j) d\mu$$

per ogni j . Passando al limite su $j \rightarrow \infty$ e minorando con $\int_{\Omega} \psi_j d\mu$ e per simmetria si ha la tesi. \square

Semplice conseguenza di quest'ultima è che si può definire l'integrale di una funzione misurabile a valori non negativi in modo equivalente come

$$\int_{\Omega} f d\mu = \sup_{\substack{\varphi \in \mathcal{S} \\ 0 \leq \varphi \leq f}} \int_{\Omega} \varphi d\mu.$$

Anche nel caso degli integrali di funzioni misurabili a valori non negativi valgono la linearità rispetto a coefficienti positivi e la monotonia. Entrambe queste due sono conseguenze della definizione o della definizione equivalente che abbiamo appena espresso⁴.

Adesso presentiamo uno dei più importanti teoremi di passaggio al limite dentro il segno di integrale:

Teorema 4.1.1 (di Beppo Levi). *Sia $f \geq 0$ misurabile e siano $f_n \geq 0$ misurabili che convergono crescendo a f . Allora*

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \int_{\Omega} f d\mu.$$

Dimostrazione. Per ogni $n \in \mathbb{N}$ sia $\{g_{nk}\}_{k \in \mathbb{N}}$ una successione di funzioni semplici a valori non negativi che converge crescendo verso f_n per $k \rightarrow \infty$. Consideriamo

$$h_{nk} = \max\{g_{1k}, \dots, g_{nk}\}$$

si ha che h_{nk} è crescente sia in k che in n . Inoltre per ogni $n \in \mathbb{N}$ si ha che h_{nk} converge crescendo verso f_n per $k \rightarrow \infty$ in quanto

$$g_{nk} \leq h_{nk} \leq \max\{f_1, \dots, f_n\} = f_n \leq f$$

per ogni n e per ogni k . Anche h_{kk} converge crescendo verso f poiché

$$f = \sup_{n \in \mathbb{N}} f_n = \sup_{n \in \mathbb{N}} \sup_k h_{nk} = \sup_{(n,k)} h_{nk} = \sup_k h_{kk}.$$
⁵

⁴ad esempio per la monotonia se sono date $f \geq g \geq 0$ misurabili basta prendere due successioni di funzioni semplici che convergono una ad f e l'altra a g e che siano puntualmente la prima maggiore della seconda. A questo punto la definizione fa concludere.

⁵ l'ultima uguaglianza vale perché: (1) la disuguaglianza \geq è ovvia; (2) per l'altra si osserva che per ogni n e per ogni k si ha $h_{nk} \leq h_{\max(n,k), \max(n,k)}$.

Valendo $h_{kk} \leq f_k \leq f$ si ha per monotonia che

$$\int_{\Omega} h_{kk} d\mu \leq \int_{\Omega} f_k d\mu \leq \int_{\Omega} f d\mu,$$

e passando al limite per $k \rightarrow \infty$ abbiamo concluso: l'integrale di sinistra converge a $\int_{\Omega} f d\mu$ perché è la definizione di integrale. \square

Adesso compiremo l'ultimo passo verso la generalizzazione della nozione di integrale, e lo definiremo per funzioni misurabili a valori arbitrari. Ecco la definizione:

Definizione 4.1.4. Sia f una funzione misurabile. Diciamo che f è *integrabile* se e solo se $\int_{\Omega} |f| d\mu$ è finito; equivalentemente si può dire che $\int_{\Omega} f^+ d\mu$ e $\int_{\Omega} f^- d\mu$ sono entrambi finiti. Infine, quando f è integrabile si definisce il suo integrale come

$$\int_{\Omega} f d\mu = \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu.$$

Neanche a dirlo, anche in questo caso più generale valgono linearità e monotonia. Quindi ora l'integrale è una forma lineare sullo spazio vettoriale delle funzioni integrabili ${}^1(\Omega, \mathcal{A}, \mu)$ ⁶. Inoltre possiamo dare la seguente definizione:

Definizione 4.1.5. Per f misurabile definiamo

$$\|f\|_1 = \int_{\Omega} |f| d\mu \in [0, \infty],$$

detta *norma integrale* di f .

Si può mostrare facilmente che la norma integrale è subadditiva e per essa vale l'omogeneità, ma non è vero che è nulla solo sulla funzione nulla: infatti $\|\chi_A\| = 0$, con $A \in \mathcal{A}$ insieme misurabile di misura nulla non vuoto. Su ${}^1(\Omega, \mathcal{A}, \mu)$, dunque, la norma integrale è una *seminorma*.

4.2 Gli spazi p

In modo analogo a quanto fatto per 1 si può definire lo spazio ${}^p(\Omega, \mathcal{A}, \mu)$, dove $p \in [1, \infty]$. Sia $f : \Omega \rightarrow \mathbb{R}$ misurabile, se p è finito si definisce

$$\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}},$$

⁶che questo sia uno spazio vettoriale segue dal fatto seguente. Se $f, g \in {}^1(\Omega, \mathcal{A}, \mu)$ allora $f + \lambda g \in {}^1(\Omega, \mathcal{A}, \mu)$. Infatti questa è misurabile e inoltre, valendo $|f + \lambda g| \leq |f| + |\lambda||g|$, si ha

$$\int_{\Omega} |f + \lambda g| d\mu \leq \int_{\Omega} |f| d\mu + |\lambda| \int_{\Omega} |g| d\mu < +\infty.$$

mentre se $p = \infty$ definiamo

$$\|f\|_\infty = \inf_{\substack{N \in \mathcal{A} \\ \mu(N)=0}} \sup_{x \in \Omega - N} |f(x)|.$$

Grazie a queste applicazioni possiamo definire lo spazio

$${}^p(\Omega, \mathcal{A}, \mu) = \{f \text{ misurabile} \mid \|f\|_p < \infty\}.$$

Affermiamo, ma senza dimostrazione in quanto questa esula dagli scopi del libro, che gli spazi ${}^p(\Omega, \mathcal{A}, \mu)$ sono spazi vettoriali, ed inoltre $\|\cdot\|_p$ è una seminorma su p . Dimostrare la subadditività di $\|\cdot\|_p$ in generale non è banale. Invece nei casi $p \in \{1, 2, \infty\}$ è più semplice, vediamo perché:

Proposizione 4.2.1. *L'applicazione $\|\cdot\|_p$ è subadditiva*

Dimostrazione. Come detto mostreremo la proposizione se $p \in \{1, 2, \infty\}$. Se $p = 1$, valendo

$$|f(x) + g(x)| \leq |f(x)| + |g(x)|$$

per ogni $x \in \Omega$ si può integrare e si ricava la disuguaglianza cercata. Se $p = \infty$ non dobbiamo neanche utilizzare la definizione di integrale e quindi lasciamo la dimostrazione per esercizio al lettore. Supponiamo che $p = 2$, che è il caso per noi più importante. Affermiamo il seguente fatto:

Lemma 4.2.1. *Se $f, g \in {}^2(\Omega, \mathcal{A}, \mu)$ allora $fg \in {}^1(\Omega, \mathcal{A}, \mu)$ e vale*

$$\left| \int_\Omega fg \, d\mu \right| \leq \left(\int_\Omega f^2 \, d\mu \right)^{\frac{1}{2}} \left(\int_\Omega g^2 \, d\mu \right)^{\frac{1}{2}}.$$

Dimostrazione. Intanto fg è misurabile ed inoltre $|fg| \leq |f|^2 + |g|^2$ per ogni $x \in \Omega$. Quindi $fg \in {}^1(\Omega, \mathcal{A}, \mu)$ perché

$$\int_\Omega |fg| \, d\mu \leq \int_\Omega |f|^2 \, d\mu + \int_\Omega |g|^2 \, d\mu.$$

Si consideri ora la funzione di una variabile $t \mapsto \int_\Omega (f(x) - tg(x))^2 \, d\mu$. Visto che $(f(x) - tg(x))^2 = f^2(x) - 2tf(x)g(x) + t^2g^2(x)$ segue che

$$0 \leq \int_\Omega (f(x) - tg(x))^2 \, d\mu = \|f\|_2^2 - 2t \int_\Omega fg \, d\mu + t^2 \|g\|_2^2 \in \mathbb{R}.$$

Visto che questo polinomio di secondo grado in t è positivo ne segue che il suo discriminante deve essere negativo. Da questa condizione segue la tesi. \square

A questo punto dalla disuguaglianza di Cauchy–Schwartz segue la subadditività, in quanto si può scrivere⁷:

$$\|f + g\|_2^2 = \langle f + g, f + g \rangle \leq \|f\|_2^2 + 2\|f\|_2\|g\|_2 + \|g\|_2^2 = (\|f\|_2 + \|g\|_2)^2,$$

e abbiamo concluso. \square

Facciamo un piccolo ragionamento in generale. Se V è uno spazio vettoriale reale e $\|\cdot\|$ è una seminorma su V allora l'insieme

$$N = \{v \in V \mid \|v\| = 0\}$$

è un sottospazio vettoriale. Sul quoziente V/N resta definita una *norma*, che è indotta da $\|\cdot\|$, ponendo $\|[v]\| = \|w\|$, dove w è un qualsiasi rappresentante della classe $[v]$.

Osservazione 4.2.1. Quella appena data è una buona definizione. Sia $[v] = [w]$, cioè $v = w + n$ con $n \in N$; allora

$$\|v\| = \|w + n\| \leq \|w\| + \|n\| = \|w\|$$

per ogni v e w . Essendo chiaramente simmetrica la relazione produce l'uguaglianza cercata.

Ritorniamo al nostro caso e consideriamo $N = \{f \text{ misurabile} \mid f = 0 \text{ quasi ovunque}\}$. Allora sul quoziente

$${}^p(\Omega, \mathcal{A}, \mu)/N = L^p(\Omega, \mathcal{A}, \mu)$$

è ben definita una norma, dove si pone $\|[f]\|_p = \|f\|_p$ per ogni $f \in {}^p$. Gli elementi di $L^p(\Omega, \mathcal{A}, \mu)$ sono classi di funzioni in ${}^p(\Omega, \mathcal{A}, \mu)$ che differiscono per funzioni nulle quasi ovunque.

4.3 Integrazione e misura immagine

Abbiamo visto che se $(\Omega, \mathcal{A}, \mu)$ è uno spazio misurato e $f : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ un'applicazione misurabile allora è ben definita la misura immagine $f(\mu)$ come segue:

$$f(\mu)(A') = \mu(f^{-1}(A')).$$

Inoltre se $u \in \mathcal{M}(\Omega')$ allora $u \circ f \in \mathcal{M}(\Omega)$ ed è ben definita dunque l'applicazione

$$f^* : \begin{array}{ccc} \mathcal{M}(\Omega') & \longrightarrow & \mathcal{M}(\Omega) \\ u & \longmapsto & u \circ f \end{array},$$

e f^* è un operatore lineare positivo (se $u \geq 0$ allora $f^*u \geq 0$). Vale anche un risultato simile per funzioni ¹:

⁷si ricordi che L^2 è uno spazio di Hilbert, nel senso che la norma è indotta da un prodotto scalare, che è $\langle f, g \rangle_2 = \int_{\Omega} f(x)\bar{g}(x) dx$.

Teorema 4.3.1. Se $u \in {}^1(\Omega', \mathcal{A}', f(\mu))$ allora $u \circ f \in {}^1(\Omega, \mathcal{A}, \mu)$ e vale

$$\int_{\Omega} u \circ f \, d\mu = \int_{\Omega'} u \, df(\mu).$$

Dimostrazione. Supponiamo $u = \chi_A$ con $A \in \mathcal{A}'$. Allora

$$\int_{\Omega} \chi_A \circ f \, d\mu = \int_{\Omega} \chi_{f^{-1}(A)} \, d\mu = \mu(f^{-1}(A)) = f(\mu)(A) = \int_{\Omega'} \chi_A \, df(\mu).$$

Poiché i due membri della tesi sono quantità lineari in u , allora la disuguaglianza vale anche almeno per ogni u semplice non negativa.

Se u è misurabile positiva su Ω' allora esiste una successione di funzioni positive $\{\varphi_n\}_{n \in \mathbb{N}} \in S$ che converge crescendo puntualmente ad u . Allora la successione $\{\varphi_n \circ f\} \in S$ converge crescendo puntualmente a $u \circ f$. Quindi

$$\int_{\Omega} (u \circ f) \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} (\varphi_n \circ f) \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega'} \varphi_n \, df(\mu) = \int_{\Omega'} u \, df(\mu).$$

Infine, siccome per $u \in \mathcal{M}(\Omega')$ vale $|u \circ f| = |u| \circ f$, si ha $\|u \circ f\|_{1, \Omega, \mu} = \|u\|_{1, \Omega', f(\mu)}$. A questo punto si conclude utilizzando la parte positiva e quella negativa di f . \square

Adesso presenteremo la costruzione degli spazi prodotto per poi poter avere qualche applicazione di questo importante teorema.

4.4 Misure prodotto e integrali iterati

Adesso consideriamo (S, \mathcal{S}) e (T, \mathcal{T}) spazi misurabili. Allora si può considerare il prodotto $S \times T$ e la σ -algebra generata dalla famiglia di prodotti:

$$\{A \times B \mid A \in \mathcal{S}, B \in \mathcal{T}\}.$$

Tale σ -algebra verrà indicata con $\mathcal{S} \otimes \mathcal{T}$ e si dirà *algebra prodotto*.

Osservazione 4.4.1. Se S e T sono spazi metrici separabili e \mathcal{S} e \mathcal{T} sono le loro σ -algre dei boreliani (ossia $\mathcal{S} = B(S)$ e $\mathcal{T} = B(T)$), allora si può verificare che $\mathcal{S} \otimes \mathcal{T} = B(S \times T)$, ossia la σ -algebra dei boreliani di $S \times T$ come spazio topologico.

Visto che abbiamo a che fare con un prodotto, esisteranno come al solito le due proiezioni canoniche:

$$p: S \times T \longrightarrow S \quad \text{e} \quad q: S \times T \longrightarrow T.$$

C'è una stretta correlazione tra dotare lo spazio prodotto della σ -algebra prodotto $\mathcal{S} \otimes \mathcal{T}$ ed è analoga a quella che c'è tra spazi topologici prodotto e continuità:

nel corso di topologia sicuramente lo studente avrà incontrato che la topologia prodotto è la meno fine che rende continue le proiezioni. Ebbene, analogamente vale:

Lemma 4.4.1. *Le proiezioni canoniche del prodotto $S \times T$ sui fattori sono misurabili. Anzi, $\mathcal{S} \otimes \mathcal{T}$ è la minima σ -algebra che rende misurabili le proiezioni.*

Dimostrazione. Verifichiamo che p e q sono misurabili: dobbiamo mostrare che per ogni $A \in \mathcal{S}$ si ha che $p^{-1}(A) \in \mathcal{S} \otimes \mathcal{T}$, analogo per q . In effetti si ha

$$p^{-1}(A) = A \times T \in \mathcal{S} \otimes \mathcal{T}.$$

Sia \mathcal{A} un'algebra su $S \times T$ per la quale p e q sono misurabili. Allora per ogni $A \in \mathcal{S}$ e per ogni $B \in \mathcal{T}$ si ha

$$A \times B = p^{-1}(A) \cap q^{-1}(B) \in \mathcal{A}.$$

Da questo segue che $\mathcal{S} \otimes \mathcal{T} \subseteq \mathcal{A}$. \square

Lemma 4.4.2. *Sia (Ω, \mathcal{A}) uno spazio misurabile. $f : \Omega \rightarrow (S, \mathcal{S})$ e $g : \Omega \rightarrow (T, \mathcal{T})$ sono applicazioni misurabili se e solo se l'applicazione $(f, g) : \Omega \rightarrow S \times T$ è misurabile*

Dimostrazione. (\implies) Dobbiamo verificare che la controimmagine di un misurabile in $S \times T$ secondo (f, g) è misurabile in Ω . In effetti basta fare questa verifica solo per i generatori di $\mathcal{S} \otimes \mathcal{T}$. Infatti per ogni $A \in \mathcal{S}$ e $B \in \mathcal{T}$ si ha

$$(f, g)^{-1}(A \times B) = f^{-1}(A) \cap g^{-1}(B),$$

e questo sta in \mathcal{A} perché sia f che g sono applicazioni misurabili.

(\impliedby) Questa implicazione la lasciamo come esercizio. \square

Osservazione 4.4.2. La costruzione del prodotto può essere fatta anche più in generale. Se $\{(S_i, \mathcal{S}_i)\}_{i \in I}$ è una famiglia di insiemi misurabili allora sul prodotto

$$\prod_{i \in I} S_i = \{(s_i) \mid s_i \in S_i \forall i\}$$

è definita la σ -algebra $\bigotimes_{i \in I} \mathcal{S}_i$, quella generata dalla famiglia

$$\{p_i^{-1}(A) \mid i \in I, A \in \mathcal{S}_i\},$$

dove le p_i sono le proiezioni canoniche. Notiamo che $p_i^{-1}(A) = \prod_{j \in I} Z_j$, dove $Z_j = A$ se $j = i$ e $Z_j = S_j$ se $j \neq i$. Il lettore può verificare le analoghe proprietà viste nel caso di due spazi.

Definizione 4.4.1. Uno spazio di misura (S, \mathcal{S}, μ) viene detto σ -finito se e solo se esiste $\{A_i \in \mathcal{S}, i \in \mathbb{N}\}$ tale che

$$S = \bigcup_{i=0}^{\infty} A_i, \text{ con } \mu(A_i) < \infty \forall i.$$

Esempio 4.4.1. \mathbb{R}^n con la misura di Lebesgue è σ -finito in quanto può essere ottenuto come unione numerabile delle palle a raggio razionale, tutte di misura di Lebesgue finita.

Esempio 4.4.2. \mathbb{R} con la misura che “conta i punti”, a dire

$$\mu(S) = \begin{cases} |S| & \text{se } S \text{ è finito} \\ \infty & \text{se } S \text{ è infinito} \end{cases}$$

non è σ -finito. Infatti se $\mathbb{R} = \bigcup_{i=0}^{\infty} A_i$ con $\mu(A_i) < \infty$ (ossia A_i è finito) per ogni i allora avremmo che \mathbb{R} sarebbe al più numerabile, il che è falso.

Esempio 4.4.3. Sia $X \neq \emptyset$ insieme e $\mathcal{P}(X)$ la σ -algebra su di esso definita. Con la misura

$$\mu(A) = \begin{cases} 0 & \text{se } A = \emptyset \\ \infty & \text{se } A \neq \emptyset \end{cases}$$

X non è σ -finito. Infatti l'unico insieme di misura finita è quello vuoto, ma X non lo è e dunque non può essere unione di insiemi con misura tutti finita.

Adesso vogliamo definire una misura sullo spazio misurabile $(S \times T, \mathcal{S} \otimes \mathcal{T})$ a partire dalle misure dei fattori. Enunciamo il teorema relativo, di cui però non daremo dimostrazione:

Teorema 4.4.1. *Se (S, \mathcal{S}, μ) e (T, \mathcal{T}, ν) sono spazi di misura σ -finiti allora esiste un'unica misura (denotata con $\mu \otimes \nu$) su $(S \times T, \mathcal{S} \otimes \mathcal{T})$ tale che per ogni $A \in \mathcal{S}$ e $B \in \mathcal{T}$ si ha*

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B),$$

dove si assume la convenzione che $0 \cdot \infty = 0$.

Ci limitiamo solo ad osservare che la dimostrazione del teorema discende dal teorema di estensione di Carathéodory: per dimostrarne le ipotesi si deve solo mostrare la σ -additività. Un'altra osservazione è che senza l'ipotesi di σ -finitezza si ha comunque definita una misura prodotto, ma si viene a perderne l'unicità.

4.5 Integrali su spazi prodotto

Adesso vogliamo vedere cosa accade quando si integra su spazi prodotto, e cercare di esprimere tali integrali in funzione di integrali più semplici. Il teorema che segue è fondamentale per questo scopo, in quanto permette di esprimere gli integrali su spazi prodotto mediante gli integrali sugli spazi fattori.

Teorema 4.5.1 (di Tonelli). *Siano (S, \mathcal{S}, μ) e (T, \mathcal{T}, ν) spazi di misura σ -finiti e sia poi $f : (S \times T, \mathcal{S} \otimes \mathcal{T}) \rightarrow [0, \infty]$ funzione misurabile. Allora valgono le seguenti proprietà:*

- (1) per ogni $s \in S$ la funzione $f(s, \cdot) : T \rightarrow [0, \infty]$ è misurabile;
 (2) è ben definita e misurabile la funzione $S \rightarrow [0, \infty]$ tale che

$$s \mapsto \int_T f(s, t) d\nu;$$

(3) risulta infine che

$$\int_{S \times T} f(s, t) d(\mu \otimes \nu) = \int_S \left(\int_T f(s, t) d\nu \right) d\mu.$$

Dimostrazione. Rimandiamo la dimostrazione ad un corso più avanzato. \square

Osservazione 4.5.1. Osserviamo che nel caso di $S = T = \mathbb{N}$ (spazi discreti) si ritrova la proprietà di associatività per famiglie di numeri reali positivi, enunciata già come teorema di Tonelli:

$$\sum_{(i,j) \in \mathbb{N}^2} a_{ij} = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} a_{ij},$$

date precedentemente con I e J generici.

L'analogo teorema con l'ipotesi che $f \in {}^1(S \times T, \mathcal{S} \otimes \mathcal{T})$ è il teorema di Fubini (dove f deve essere una funzione di 1 anziché misurabile).

Osservazione 4.5.2. Per osservare le validità del teorema di Tonelli mostriamo che l'ipotesi di σ -finitzza è necessaria. Consideriamo $S = T = [0, 1] = I$ con μ misura di Lebesgue e ν misura della cardinalità. Sia poi $f = \chi_\Delta$, dove Δ è la diagonale di I^2 . Allora

$$\int_I \left(\int_I f(x, y) d\mu \right) d\nu \neq \int_I \left(\int_I f(x, y) d\nu \right) d\mu,$$

in quanto il primo è nullo e il secondo è 1.

Capitolo 5

Probabilità discreta e generale

Adesso è giunto il momento di applicare quanto visto finora alla probabilità. Vogliamo richiamare ora quanto detto sulla probabilità uniforme e condizionata.

5.1 Richiami di probabilità discreta

Sia (Ω, \mathcal{A}) uno spazio misurabile. Se Ω è finito allora possiamo definire una misura di probabilità su Ω grazie alla cardinalità nel modo seguente: per ogni $A \in \mathcal{A}$ poniamo

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

Questa è detta *probabilità uniforme*. Se consideriamo uno spazio già probabilizzato (Ω, \mathcal{A}, P) e prendiamo $A \in \mathcal{A}$ di probabilità non nulla. Si può definire dunque una nuova probabilità come segue:

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

Se B è disgiunto da A tale probabilità è nulla, e si vede dalla sua definizione che quella calcolata è la probabilità che si verifichi B condizionata dal fatto che già si è verificato A . Inoltre ricordiamo che due eventi A e B si dicono *indipendenti* se e solo se

$$P(A)P(B) = P(A \cap B),$$

ossia se e solo se $P(B|A) = P(B)$: questa scrittura traduce il fatto che il verificarsi di B non è condizionato al verificarsi di A .

Inoltre la probabilità condizionata è utile anche nel descrivere intersezioni di serie di eventi consecutivi ciascuno dei quali potenzialmente condizionato a tutti i

precedenti. Svolgendo il calcolo si mostra che se $\bigcap_{i=1}^n A_i$ non ha probabilità nulla allora

$$P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P\left(A_i \mid \bigcap_{j=1}^{i-1} A_j\right).$$

A proposito della probabilità condizionata esiste una formula molto importante che sostanzialmente scambia i condizionamenti e riguarda i sistemi di alternative, ossia partizioni misurabili di Ω . Sia $\{A_i\}_{i=1}^n$ un sistema di alternative per uno spazio di probabilità Ω . Allora valgono le seguenti formule:

$$P(B) = \sum_{i=1}^n P(A_i \cap B) \quad \text{e} \quad P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)},$$

per ogni $B \in \mathcal{A}$, e per ogni $j = 1, \dots, n$ a seconda della formula. La prima di esse è nota come *teorema della probabilità totale*, mentre l'ultima come *formula di Bayes*.

5.2 Variabili aleatorie

Premettiamo alla definizione di variabile aleatoria un esempio. Supponiamo di aver puntato alla roulette 1 euro sul numero 28 e 1 euro sul pari: possiamo domandarci qual è la probabilità di vincere più di 10 euro, oppure qual è la probabilità di perdere. Lo spazio naturale per descrivere l'esito di un giro di roulette è $\Omega = \{0, \dots, 36\}$ munito della distribuzione uniforme di probabilità su tutte le sue parti. Le domande che ci siamo posti non corrispondono direttamente a sottoinsiemi di Ω . Siamo naturalmente portati a introdurre una funzione $X : \Omega \rightarrow \mathbb{R}$ che associa ad ogni uscita la vittoria netta. Avremo dunque

$$X(\omega) = \begin{cases} 2 & \text{se } \omega = 28 \\ 0 & \text{se } \omega \text{ è pari ma } \omega \neq 28 \\ -1 & \text{se } \omega = 0 \\ -2 & \text{se } \omega \text{ è dispari} \end{cases}.$$

Allora la risposta alla prima domanda è $P\{\omega_i \mid X(\omega_i) \geq 10\} = P(X^{-1}[10, \infty)) = \frac{1}{37}$, e analogamente la risposta all'altra è $\frac{19}{37}$. Quello che abbiamo fatto è introdurre una funzione $X : \Omega \rightarrow \mathbb{R}$ e aver trasportato la probabilità dai sottoinsiemi di Ω a quelli di \mathbb{R} . Questo è ciò che si intende per variabile aleatoria:

Definizione 5.2.1. Una *variabile aleatoria discreta* è una $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{Z}$ misurabile.

Riprendiamo adesso l'osservazione 3.3.3. Sia $f : \Omega \rightarrow \mathbb{Z}$ una funzione misurabile, abbiamo visto che si può scrivere

$$f = \sum_{k \in \mathbb{Z}} k \cdot \chi_{E_k},$$

dove $E_k = \{f = k\}$ (e questi formano una partizione di Ω). Per ogni $\omega \in \Omega$ si ha che $f(\omega) = \sum_{k \in \mathbb{Z}} k \chi_{E_k}(\omega) = j$, dove j è l'intero tale che $\omega \in E_j$.

Ora consideriamo come f misurabile la funzione X variabile aleatoria. Quanto appena osservato ci dice che l'applicazione X rimane misurabile indebolendo la σ -algebra \mathcal{A} ad $\mathcal{A}_0 \subseteq \mathcal{A}$, dove $\mathcal{A}_0 = \sigma A(\mathcal{B})$ e $\mathcal{B} = \{E_k\} = \{\{X = k\} \mid k \in \mathbb{Z}\}$, che è atomica¹. Quindi considerando variabili aleatorie a valori in \mathbb{Z} (o più in generale in un insieme discreto) si può pensare Ω , senza perdere di generalità, come se fosse numerabile.

Definizione 5.2.2. Una *variabile aleatoria reale* è una $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ misurabile.

Osservazione 5.2.1. Senza perdita di generalità, per quanto detto subito prima della definizione, si potrebbe pensare \mathcal{A} atomica, o addirittura Ω numerabile (che implica che \mathcal{A} è atomica). Notare anche che se Ω è numerabile, una variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ è automaticamente discreta.

Si è visto che per una funzione $f : (\Omega, \mathcal{A}, \mu) \rightarrow (\Omega', \mathcal{A}')$ misurabile dallo spazio di misura allo spazio misurabile è definita la misura immagine $f(\mu)$ come

$$f(\mu)(B) = \mu(f^{-1}(B)).$$

per ogni $B \in \mathcal{A}'$. Inoltre si ha $(fg)(\mu)(B) = \mu(g^{-1}f^{-1}(B)) = g(\mu)(f^{-1}(B)) = f(g(\mu))$.

Osservazione 5.2.2. Se μ è una misura di probabilità, allora anche la misura immagine lo è. Sia $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}'$ una successione numerabile di insiemi a due a due disgiunti; allora

$$\begin{aligned} f(\mu) \left(\bigcup_{n \in \mathbb{N}} A_n \right) &= \mu \left(f^{-1} \left(\bigcup_{n \in \mathbb{N}} A_n \right) \right) = \mu \left(\bigcup_{n \in \mathbb{N}} f^{-1}(A_n) \right) = \\ &= \sum_{n \in \mathbb{N}} \mu(f^{-1}(A_n)) = \sum_{n \in \mathbb{N}} f(\mu)(A_n). \end{aligned}$$

Inoltre $f^{-1}(\Omega') = \Omega$ e dunque $f(\mu)(\Omega') = \mu(f^{-1}(\Omega')) = \mu(\Omega) = 1$ e dunque anche $f(\mu)$ è una misura di probabilità.

¹si noti però che già considerando una successione di variabili aleatorie $X_i : \Omega \rightarrow \mathbb{Z}$, questa semplificazione non è più possibile, perché ogni X_i avrebbe la sua speciale σ -algebra atomica che la lascia misurabile, ma in generale non si può garantire l'esistenza di una unica σ -algebra atomica che vada bene per tutte le X_i . Questo è già un buon motivo per uscire dal caso discreto.

Definizione 5.2.3. Se $X : \Omega \rightarrow E$ è una variabile aleatoria, la misura di probabilità $X(P)$ si chiama *distribuzione di probabilità* della variabile aleatoria X , o anche *legge* di X (specialmente quando E è \mathbb{R} , \mathbb{Z} oppure \mathbb{R}^n).

Teorema 5.2.1. Se una applicazione $f : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$ è misurabile, allora f è limite uniforme di una successione di funzioni $f_n : (\Omega, \mathcal{A}) \rightarrow \mathbb{R}$, misurabili e discrete.

Dimostrazione. Basta prendere per ogni $n \in \mathbb{N}_0$ la funzione $f_n = \frac{1}{n} [nf]$. Siccome per ogni $x \in \mathbb{R}$ si ha che $[x] \leq x \leq [x] + 1$ segue

$$f_n \leq f \leq f_n + \frac{1}{n},$$

d'altra parte le f_n sono discrete perché $f_n(\Omega) \subseteq n\mathbb{Z}$. \square

5.3 Speranza matematica e varianza

Adesso veniamo come al solito alle applicazioni in probabilità di quanto in astratto detto fino ad ora sugli integrali. La definizione che possiamo dare grazie al concetto di integrale è quella di speranza matematica:

Definizione 5.3.1. Sia Ω uno spazio probabilizzato con probabilità P . Se $X : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria si definisce (se esiste, cioè se $X \in {}^1(\Omega, \mathcal{A}, P)$) *speranza matematica* (o valore atteso) di X

$$E(X) = \int_{\Omega} X dP.$$

Esempio 5.3.1. Se A_1, \dots, A_n sono eventi in Ω spazio di probabilità e $X = \sum_{i=1}^n a_i \cdot \chi_{A_i}$. Allora la speranza di X è

$$E(X) = \int_{\Omega} \left(\sum_{i=1}^n a_i \cdot \chi_{A_i} \right) dP = \sum_{i=1}^n a_i P(A_i).$$

Osservazione 5.3.1. Osserviamo che anche l'impostazione soggettivista di De Finetti è coerente con questa definizione e con l'esempio precedente. Se si propone ad una persona un affare per cui il verificarsi di A_i produca un guadagno a_i allora egli ritiene equo pagare tanto quanto la speranza matematica.

Abbiamo visto nel teorema 4.3.1 che possiamo passare dall'integrale rispetto alla misura di partenza a quello rispetto alla misura di arrivo. Nel caso di una variabile aleatoria reale X questo fatto si traduce nell'equazione seguente:

$$E(X) = \int_{\Omega} X dP = \int_{\mathbb{R}} x d(X(P)).$$

Più in generale per $g : [0, \infty) \rightarrow [0, \infty]$ misurabile si ha

$$E(g(X)) = \int_0^\infty g(x) dX(P).$$

Possiamo scrivere in forma migliore il valore atteso di una variabile aleatoria X sfruttando l'integrale secondo la misura immagine.

Definizione 5.3.2. Sia P una proprietà. Denotiamo la *funzione di verità* di P con $[P]$, e tale funzione è definita 1 se P vale e 0 se P non vale.

Osservazione 5.3.2. Con la notazione appena introdotta si ha

$$x = \int_0^x 1 dt = \int_0^\infty [0 \leq t \leq x] dt,$$

e questo ci servirà nella prossima dimostrazione.

Teorema 5.3.1. Sia $g : [0, \infty) \rightarrow [0, \infty] C^1$ e crescente. Valgono le seguenti due uguaglianze:

$$E(X) = \int_0^\infty P(X \geq t) dt \quad e \quad E(g(x)) = \int_0^\infty P(X \geq t)g'(t) dt.$$

Dimostrazione. Dall'osservazione che precede l'enunciato del teorema e dal teorema di Fubini–Tonelli segue che

$$E(X) = \int_0^\infty x dX(P) = \int_0^\infty \left(\int_0^\infty [0 \leq t \leq x] dt \right) dX(P) = \int_0^\infty \int_0^\infty [0 \leq t \leq x] dX(P) dt$$

Ma inoltre $X(P)[0 \leq t \leq x] = P(X^{-1}([t, \infty))) = P(X \geq t)$, e dunque si conclude

$$E(X) = \int_0^\infty P(X \geq t) dt.$$

Se poi $g : [0, \infty) \rightarrow [0, \infty]$ è C^1 e crescente allora

$$g(x) = \int_0^x g'(t) dt = \int_0^\infty g'(t)[0 \leq t \leq x] dt,$$

e quindi con analogo calcolo si ha la tesi. \square

Definizione 5.3.3. La funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ associata alla variabile aleatoria X definita da $F_X(t) = P(X \leq t) = X(P)(-\infty, t]$, è detta *funzione di ripartizione* di X .

Osservazione 5.3.3. Conoscere F_X implica conoscere la legge X . Dire che conosciamo $P(X \leq t)$ per tutti i $t \in \mathbb{R}$ significa dire che conosciamo la misura immagine di tutti gli intervalli illimitati inferiormente del tipo $(-\infty, t]$, e questo implica conoscere la misura di tutti gli intervalli limitati semi-aperti $(t_1, t_2] = (-\infty, t_2] - (-\infty, t_1]$. Infine, con unioni e intersezioni numerabili monotone, possiamo conoscere anche la probabilità immagine di tutti gli intervalli. Per il teorema di unicità² questo implica conoscere la misura immagine su tutti i boreliani.

Lemma 5.3.1. *Siano X_1 e X_2 due variabili aleatorie indipendenti e integrabili. Allora $X_1 X_2$ è integrabile e vale $E(X_1 X_2) = E(X_1)E(X_2)$.*

Dimostrazione. Consideriamo la variabile $(X_1, X_2) : \Omega \rightarrow \mathbb{R}^2$. Data $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ funzione misurabile si ha

$$E(\varphi(X_1, X_2)) = \int_{\Omega} \varphi(X_1(\omega), X_2(\omega)) dP = \int_{\mathbb{R}^2} \varphi(x_1, x_2) d(X_1, X_2)(P)$$

Nel nostro caso prendiamo $\varphi(x_1, x_2) = x_1 x_2$ e osserviamo che dire X_1 e X_2 indipendenti equivale a dire $(X_1, X_2)(P) = X_1(P) \otimes X_2(P)$ e quindi

$$\begin{aligned} E(X_1 X_2) &= E(\varphi(X_1, X_2)) = \int_{\mathbb{R}^2} x_1 x_2 d(X_1(P) \otimes X_2(P)) = \\ &= \int_{\mathbb{R}^2} x_1 x_2 dX_1(P) dX_2(P) = \int_{\mathbb{R}} x_1 dX_1(P) \int_{\mathbb{R}} x_2 dX_2(P), \end{aligned}$$

e abbiamo concluso. \square

Definizione 5.3.4. Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria. Si definisce *varianza* di X la quantità:

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

Sostanzialmente la varianza misura di quanto una variabile aleatoria si scosta dal suo valor medio $E(X)$. In effetti la proprietà (4) delle seguenti lo illustra:

Proposizione 5.3.1. *Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria di classe¹. Valgono le seguenti proprietà:*

- (1) $\text{Var}(X) = E(X^2) - E(X)^2$;
- (2) $\text{Var}(X) < \infty$ se e solo se $E(X^2) < \infty$;
- (3) se $\text{Var}(X) < \infty$ allora $E(|X|) < \infty$;
- (4) $\text{Var}(X) = \min_{t \in \mathbb{R}} E((X - t)^2)$.

Dimostrazione. Le dimostrazioni sono tutte molto facili e quindi le lasciamo come esercizio. Ad esempio la (1) è facile conseguenza della linearità dell'integrale. \square

²il teorema afferma che se (Ω, \mathcal{A}) è uno spazio di misura con μ e ν che coincidono su una classe $D \subseteq \mathcal{A}$ che genera \mathcal{A} e stabile per intersezione finita, allora le misure coincidono su \mathcal{A} .

Quando dobbiamo correlare due variabili aleatorie è utile la nozione di covarianza:

Definizione 5.3.5. Siano $X, Y \in \mathcal{L}^2$ variabili aleatorie. Definiamo la *covarianza* di X e Y come

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

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

Osservazione 5.3.4. Osserviamo che $\text{Var}(X) = \text{Cov}(X, X)$; infatti indicheremo $\text{Cov}(X, X)$ semplicemente con $\text{Var}(X)$ per abbreviare le notazioni.

La nozione di indipendenza è una nozione più forte della scorrelazione, tant'è che vale il seguente risultato:

Proposizione 5.3.2. *Se X e Y sono variabili aleatorie indipendenti allora sono scorrelate.*

Dimostrazione. Basta calcolare $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$, che segue dall'indipendenza per quanto abbiamo mostrato nel lemma 5.3.1. \square

Per vedere che il viceversa della precedente proposizione non vale faremo un esempio:

Esempio 5.3.2 (controesempio). Sia $\Omega = \{1, 2, 3, 4\}$, e si consideri lo spazio $(\Omega, \mathcal{P}(\Omega), P)$ con P misura di probabilità uniforme. Definiamo $X, Y : \Omega \rightarrow \{-1, 0, 1\}$ come sequenze

$$\langle 1, 0, -1, 0 \rangle \quad \text{e} \quad \langle 0, 1, 0, -1 \rangle.$$

Vale che $E(X) = E(Y) = 0$ ed anche $E(XY) = 0$ (fare il calcolo). Dunque $\text{Cov}(X, Y) = 0$. Questo indica che le variabili sono scorrelate; però non sono indipendenti perché

$$P(X = Y = 1) = 0 \neq P(X = 1)P(Y = 1).$$

Questo conclude l'esempio.

Adesso mostriamo che la varianza, così come la speranza, è additiva però sotto certe ipotesi:

Proposizione 5.3.3. *Siano $X_i \in \mathcal{L}^2$ variabili aleatorie scorrelate. Allora*

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

Dimostrazione. Questo segue semplicemente osservando che vale

$$\text{Var}(\lambda X_1 + X_2)^2 = \lambda^2 \text{Var}(X_1) + 2\lambda \text{Cov}(X_1, X_2) + \text{Var}(X_2)$$

per ogni $\lambda \in \mathbb{R}$. Nell'ipotesi di scorrelazione si cancella il termine $\text{Cov}(X_1, X_2)$ e poi basta concludere per induzione. \square

5.4 Indipendenza di variabili aleatorie

La nozione di indipendenza data per eventi può essere anche data per variabili aleatorie:

Definizione 5.4.1. Sia (Ω, \mathcal{A}, P) uno spazio di probabilità e siano $X : \Omega \rightarrow S$ e $Y : \Omega \rightarrow T$ due variabili aleatorie. Queste si dicono *indipendenti* se per ogni $A \in \mathcal{S}$ e per ogni $B \in \mathcal{T}$

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Le due nozioni di indipendenza sinora date non sono scorrelate. Infatti la definizione precedente equivale a chiedere che per ogni $A \in \mathcal{S}$ e per ogni $B \in \mathcal{T}$ gli eventi $\{X \in A\}$ e $\{Y \in B\}$ sono indipendenti.

Osservazione 5.4.1. Più in generale se \mathcal{B} e \mathcal{B}' sono σ -algebre contenute in \mathcal{A} , diciamo che \mathcal{B} e \mathcal{B}' sono indipendenti se e solo se per ogni $A \in \mathcal{B}$ e per ogni $B \in \mathcal{B}'$ si ha $P(A \cap B) = P(A)P(B)$. Nel caso delle variabili aleatorie si ha che $\mathcal{B} = \{X^{-1}(A) \mid A \in \mathcal{S}\}$ e che $\mathcal{B}' = \{Y^{-1}(B) \mid B \in \mathcal{T}\}$.

Vogliamo adesso dare queste nozioni in termini degli spazi di misura prodotto. Intanto X e Y , essendo misurabili, determinano una variabile aleatoria (X, Y) misurabile.

Definizione 5.4.2. La variabile aleatoria (X, Y) è detta *blocco* di X e Y . La probabilità immagine di P secondo (X, Y) viene detta *legge congiunta* di X e Y . La legge congiunta di X e Y è una misura di probabilità su $S \times T$. Ma su $S \times T$ c'è anche la misura

$$X(P) \otimes Y(P),$$

dove $X(P)$ e $Y(P)$ sono le due probabilità immagine. Adesso, dire che X e Y sono indipendenti significa chiedere che la legge congiunta coincide con il prodotto $X(P) \otimes Y(P)$ delle leggi:

Proposizione 5.4.1. *Due variabili aleatorie X e Y sono indipendenti se e solo se $(X, Y)(P) = X(P) \otimes Y(P)$.*

Dimostrazione. I due membri sono ben definite misure su $(S \times T, \mathcal{S} \otimes \mathcal{T})$, e per l'unicità esse coincidono perché coincidono sui prodotti $A \times B$ con $A \in \mathcal{S}$ e $B \in \mathcal{T}$. Infatti per definizione

$$(X, Y)(P)(A \times B) = P((X, Y)^{-1}(A \times B)) = P(\{X \in A, Y \in B\})$$

ed anche

$$X(P) \otimes Y(P)(A \times B) = X(P)(A) \cdot Y(P)(B) = P(\{X \in A\})P(\{Y \in B\}).$$

Adesso la due variabili sono indipendenti se e solo se coincidono i secondi membri delle due uguaglianze e quindi se e solo se coincidono i primi membri. \square

È chiaro che dalla proposizione precedente derivi la stessa proprietà anche per un qualsiasi numero finito di variabili aleatorie $\{X_i\}_{i=1}^n$. Queste sono indipendenti se e solo se la loro legge congiunta coincide con la legge prodotto $\bigotimes_{i=1}^n X_i(P)$.

Le leggi marginali di n variabili aleatorie distinte si possono dedurre dalla legge congiunta, ovviamente. Siano $\{X_i\}_{i=1}^n$ variabili aleatorie $X_i : \Omega \rightarrow S_i$, allora $X_i = p_i \circ X$ – dove con p_i abbiamo denotato l' i -esima proiezione e con X la legge congiunta (X_1, \dots, X_n) . Ossia consideriamo il seguente diagramma commutativo:

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \prod_{i=1}^n S_i \\ & \searrow X_j & \downarrow p_j \\ & & S_j \end{array} .$$

Osservazione 5.4.2. Se $\{X_i\}_{i=1}^n$ sono n variabili aleatorie indipendenti, allora ogni sottofamiglia finita $\{X_i\}_{i \in J}$ è una famiglia di variabili aleatorie indipendenti. Dall'indipendenza di tutte segue che la legge congiunta $\prod_{i=1}^n X_i : \Omega \rightarrow \prod_{i=1}^n S_i$ è tale che

$$\prod_{i=1}^n X_i(P) = \bigotimes_{i=1}^n X_i(P).$$

Applicando $p_j : \prod_{i=1}^n S_i \rightarrow \prod_{i \in J} S_i$ si trova la relazione voluta, ossia che la legge congiunta delle $\{X_i\}_{i \in J}$ ha per misura immagine proprio la misura prodotto delle leggi marginali.

Osservazione 5.4.3. Dalla definizione che abbiamo dato, un evento A è indipendente da se stesso se e solo se $P(A) = 0$ o $P(A) = 1$. Quindi una sola variabile aleatoria è sempre indipendente da se stessa.

Vediamo adesso qualche modo in cui è possibile combinare variabili aleatorie indipendenti ed ottenerne delle altre indipendenti.

Proposizione 5.4.2. *Siano $X_i : \Omega \rightarrow S_i$ per $i = 1, \dots, n$ variabili aleatorie indipendenti e siano inoltre $g_i : S_i \rightarrow T_i$ funzioni misurabili. Allora le variabili aleatorie $g_i \circ X_i : \Omega \rightarrow T_i$ sono n variabili aleatorie indipendenti.*

Dimostrazione. Basta verificare che la legge congiunta delle $g_i \circ X_i$ è il prodotto delle leggi marginali. Consideriamo la mappa definita in modo naturale

$$\prod_{i=1}^n (g_i \circ X_i) : \Omega \longrightarrow \prod_{i=1}^n T_i.$$

Altrettanto naturalmente sono definite le due mappe seguenti: il blocco delle X_i definito come $\prod_{i=1}^n X_i : \Omega \rightarrow \prod_{i=1}^n S_i$ e il blocco delle g_i dato da $G : \prod_{i=1}^n S_i \rightarrow \prod_{i=1}^n T_i$. Intanto

$$\prod_{i=1}^n (g_i \circ X_i)(P) = \left(G \circ \prod_{i=1}^n X_i \right) (P) = G \left(\prod_{i=1}^n X_i(P) \right) = G \left(\bigotimes_{i=1}^n X_i(P) \right),$$

e d'altro canto

$$\bigotimes_{i=1}^n (g_i \circ X_i)(P) = \bigotimes_{i=1}^n g_i(X_i(P)).$$

Poniamo per semplicità $X_i(P) = m_i$. Se abbiamo due spazi di misura $(S_1, \mathcal{S}_1, m_1)$ e $(S_2, \mathcal{S}_2, m_2)$ e $g_i : S_i \rightarrow T_i$ funzioni misurabili per $i = 1, 2$. Allora su $(T_1 \times T_2, \mathcal{T}_1 \otimes \mathcal{T}_2)$ coincidono le due misure seguenti: la misura prodotto delle misure immagini e l'immagine del prodotto $m_1 \otimes m_2$ tramite l'applicazione prodotto $S_1 \times S_2 \rightarrow T_1 \times T_2$. Infatti queste due misure coincidono sulla base dei rettangoli $A_1 \times A_2 \subseteq T_1 \times T_2$ poiché

$$g_1(m_1) \otimes g_2(m_2)(A_1 \times A_2) = g_1(m_1)(A_1)g_2(m_2)(A_2) = m_1(g_1^{-1}(A_1))m_2(g_2^{-1}(A_2))$$

e inoltre

$$\begin{aligned} ((g_1 \times g_2)(m_1 \otimes m_2))(A_1 \times A_2) &= (m_1 \otimes m_2)(g_1 \times g_2)^{-1}(A_1 \times A_2) = \\ &= (m_1 \otimes m_2)(g_1^{-1}(A_1) \times g_2^{-1}(A_2)). \end{aligned}$$

Questo conclude la dimostrazione del teorema perché si può estendere questo fatto al caso di n spazi di misura. \square

Più in generale date $\{X_i\}_{i \in I}$ variabili aleatorie con I insieme finito e data una partizione $\{I_\lambda\}_{\lambda \in \Lambda}$, si può considerare la variabile aleatoria blocco delle $\{X_i\}_{i \in I_\lambda}$, cioè

$$\tilde{X}_\lambda = \prod_{i \in I_\lambda} X_i : \Omega \longrightarrow \prod_{i \in I_\lambda} S_i,$$

e risulta

Teorema 5.4.1. *La famiglia $\{X_i\}_{i \in I}$ è indipendente se e solo se valgono le seguenti due condizioni:*

- (1) *la famiglia dei blocchi $\tilde{X}_\lambda : \Omega \rightarrow \prod_{i \in I_\lambda} S_i$ è fatta di variabili indipendenti;*
- (2) *per ogni λ la famiglia degli $\{X_i\}_{i \in I_\lambda}$ è fatta di variabili aleatorie indipendenti.*

In realtà tutto ciò vale anche per insiemi I che non siano necessariamente finiti, basta dare la seguente:

Definizione 5.4.3. Una famiglia $\{X_i\}_{i=1}^\infty$ di variabili aleatorie, si dice essere una *famiglia indipendente* se per ogni insieme finito $J \subseteq I$ si ha che $\{X_i\}_{i \in J}$ è un insieme di variabili aleatorie indipendenti.

5.5 Variabili aleatorie discrete

Intanto un'osservazione che a tal punto si rivela necessaria: finora abbiamo fatto la nostra trattazione nel caso di insiemi come \mathbb{R} in cui le variabili aleatorie prendono i valori, e ciò produceva la nozione di integrale. Tale nozione, però, nel caso di variabili aleatorie discrete passa a diventare la nozione di serie.

Nel caso di funzioni misurabili definite su spazi numerabili in effetti il passaggio all'integrale è superfluo: la definizione data di integrale va sicuramente bene, ma nel caso di tali funzioni tutto si riduce alle somme di famiglie di numeri reali. Sia dunque $\Omega = \{\omega_i\}_{i \in \mathbb{N}}$ uno spazio di misura con μ misura tale che $\mu(\omega_i) < \infty$ per ogni $i \in \mathbb{N}$. In questo caso si dice che f è integrabile se $\sum_{\omega_i \in \Omega} |f(\omega_i)|\mu(\omega_i) < \infty$ ed in tal caso si chiama integrale di f la quantità

$$\int_{\Omega} f d\mu = \sum_{\omega_i \in \Omega} f(\omega_i)\mu(\omega_i).$$

Richiedere la assoluta sommabilità della famiglia $\{f(\omega_i)\mu(\omega_i)\}$ è necessario perché sennò, come è noto dall'analisi, si potrebbe cambiare l'ordine agli addendi cambiando però il valore della somma.

Adesso vogliamo capire come è fatta la legge di probabilità di una variabile aleatoria discreta. Considereremo dunque $X : \Omega \rightarrow \mathbb{Z}$ variabile aleatoria discreta, dove prendiamo Ω spazio probabilizzato con P e numerabile (tanto, come visto, possiamo sempre ridurci a questo caso). Visto che Ω è numerabile la sua immagine secondo X è finita o numerabile in \mathbb{Z} : dunque scriveremo

$$X(\Omega) = \{x_i \mid i \in I, I \subseteq \mathbb{N}\},$$

dove chiaramente si intende I sottoinsieme finito o numerabile, come appena detto. Poniamo poi $p_i = P(\{X = x_i\}) = P(X^{-1}(x_i)) = X(P)(x_i)$. Per ogni $A \subseteq X(\Omega)$, potendosi scrivere

$$X^{-1}(A) = \bigcup_{x_i \in A} \{X = x_i\},$$

per additività della probabilità otteniamo la seguente formula:

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

In questo modo si capisce che per assegnare la probabilità su $X(\Omega)$ basta definirla su ogni singoletto e poi si estende per somma.

Definizione 5.5.1. La funzione $x \mapsto p(x) = X(P)(x) = P(\{X = x\})$, viene detta *funzione di probabilità* o anche *densità di probabilità*.

Naturalmente i numeri $p_i = X(P)(x_i)$ sono positivi e vale

$$\sum_{i \in I} p_i = \sum_{x \in X(\Omega)} X(P)(x) = X(P)(X(\Omega)) = P(\Omega) = 1. \quad (5.1)$$

Adesso vedremo le principali densità su uno spazio numerabile, in particolare su \mathbb{N} . Indicheremo le densità di probabilità come funzioni $p : X(\Omega) \rightarrow [0, 1]$, dove p assegna ad ogni $x \in X(\Omega)$ la probabilità che la variabile X assuma il valore x .

5.5.1 Legge binomiale e di Bernoulli

Scegliamo di partire dalla legge di Bernoulli perché concettualmente più semplice della legge binomiale, benché la prima sia un caso particolare di questa. Sia $p \in [0, 1]$ un parametro fissato, e poniamo

$$p(0) = 1 - p \quad \text{e} \quad p(1) = p.$$

Questa legge modella un esperimento con due esiti possibili (il successo e il fallimento, di solito indicati con 1 e 0 rispettivamente), di cui uno ha probabilità p . La *variabile di Bernoulli* è sostanzialmente una variabile aleatoria $X : \Omega \rightarrow \{0, 1\}$.

Esempio 5.5.1. Il lancio di una moneta è il classico esempio di esperimento bernoulliano: in quel caso 0 e 1 rappresentano testa e croce, e $p = \frac{1}{2}$.

Osserviamo che se una variabile aleatoria $X : \Omega \rightarrow \{0, 1\}$ assume valori 0 o 1 allora necessariamente deve soddisfare una legge di Bernoulli.

Adesso passiamo alla variabile binomiale, prima dandone la descrizione formale e poi vedendone le applicazioni. Sia nuovamente $p \in [0, 1]$ un parametro fissato e $n \geq 1$. La funzione di densità in questo caso è data da

$$p(k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{se } k \leq n,$$

e volendo per gli altri k si può porre $p(k) = 0$; il motivo sarà chiaro subito dopo il teorema che segue, che dà l'interpretazione pratica della legge binomiale. Molto spesso si trova indicata la distribuzione binomiale di n tentativi con probabilità di successo p con $B(n, p)$, che è una funzione di $k \in \mathbb{N}$. L'interpretazione della distribuzione binomiale è la seguente: la variabile binomiale considera n ripetizioni (in condizioni di indipendenza) di un esperimento bernoulliano con probabilità di successo p .

Teorema 5.5.1. *La funzione $B(n, p)(k)$ con $k \leq n$ esprime la probabilità di avere k successi nella ripetizione con indipendenza di n esperimenti di Bernoulli di probabilità di successo p .*

Dimostrazione. Siano X_i con $i = 1, \dots, n$ le variabili aleatorie indipendenti tutte con densità di Bernoulli di parametro p : si può supporre quindi che

$$X_i : \Omega \rightarrow \{0, 1\}.$$

Inoltre le X_i devono essere variabili indipendenti. Allora, detta $S_n = \sum_{i=1}^n X_i$, si hanno k successi su n ripetizioni se e solo se $S_n = k$ (abbiamo sommato k uno e $n - k$ zero). Si può scrivere

$$\begin{aligned} \{S_n = k\} &= \{I \subseteq \{1, \dots, n\} \mid |I| = k, X_i = 1 \iff i \in I\} = \\ &= \bigcup_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} \left(\bigcap_{i \in I} \{X_i = 1\} \cap \bigcap_{i \notin I} \{X_i = 0\} \right). \end{aligned}$$

Quindi, ricordando anche che le variabili X_i sono indipendenti, si ha

$$\begin{aligned} P(S_n = k) &= \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} P \left(\bigcap_{i \in I} \{X_i = 1\} \cap \bigcap_{i \notin I} \{X_i = 0\} \right) = \\ &= \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} \prod_{i \in I} P(\{X_i = 1\}) \prod_{i \notin I} P(\{X_i = 0\}) = \\ &= \sum_{\substack{I \subseteq \{1, \dots, n\} \\ |I|=k}} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

in quanto si possono scegliere $\binom{n}{k}$ sottinsiemi di k elementi da un insieme di n . \square

Osservazione 5.5.1. Notiamo che se $n = 1$ allora ritroviamo la legge di Bernoulli che già avevamo dato.

Osservazione 5.5.2. Vogliamo anche far notare che in effetti vale la formula (5.1). Infatti:

$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^n p(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1.$$

Visto che abbiamo introdotto il concetto di indipendenza di più variabili aleatorie in maniera formale, possiamo ricavare senza problemi la definizione di legge binomiale passando però attraverso la somma di variabili aleatorie.

Proposizione 5.5.1. Sia $\{X_i\}_{i=1}^n$ una successione di variabili aleatorie indipendenti di tipo bernoulliano di parametro p . Definita la variabile aleatoria

$$S_n = X_1 + \cdots + X_n$$

si ha che la sua legge è la legge binomiale $B(n, p)$.

Dimostrazione. La legge prodotto di n copie della misura di Bernoulli con parametro p è la misura su $\{0, 1\}^n$ tale che dà all'elemento $(x_1, \dots, x_n) \in \{0, 1\}^n$ la misura

$$p^k(1-p)^{n-k},$$

dove $k = \sum_{i=1}^n x_i = |\{j \leq n \mid x_j = 1\}|$ è il numero di successi. Dire allora che $\{X_i\}_{i=1}^n$ sono n variabili aleatorie bernoulliane di parametro p e indipendenti significa dire che la loro legge congiunta è la legge prodotto appena detta, ossia

$$P(X_i = \varepsilon_i, i = 1, \dots, n) = P((X_1, \dots, X_n) = (\varepsilon_1, \dots, \varepsilon_n)) = p^k(1-p)^{n-k}.$$

Adesso vogliamo $P(S_n = k)$, ma osserviamo che

$$\{S_n = k\} = \bigcup_{\sum \varepsilon_i = k} \{(X_1, \dots, X_n) = (\varepsilon_1, \dots, \varepsilon_n)\}$$

e allora segue

$$\begin{aligned} P(S_n = k) &= P\left(\bigcup_{\sum \varepsilon_i = k} \{(X_1, \dots, X_n) = (\varepsilon_1, \dots, \varepsilon_n)\}\right) = \\ &= \sum_{\sum \varepsilon_i = k} P((X_1, \dots, X_n) = (\varepsilon_1, \dots, \varepsilon_n)) = \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

e abbiamo così concluso. \square

Esempio 5.5.2. Si calcoli la probabilità che lanciando un dado 5 volte si ottenga il numero 4 esattamente tre volte.

Se non volessimo usare la legge di Bernoulli dovremmo ragionare come segue. La probabilità che esca il numero 4 nel lancio di un dado è $p = \frac{1}{6}$, quindi se vogliamo che tale numero esca esattamente tre volte significa che su cinque lanci tre devono avere il numero 4 come uscita e gli altri due no. Adesso fissiamo l'ordine di uscita e supponiamo di voler calcolare la probabilità che esca 4 nei primi tre lanci e che non esca negli altri due: questa è data da

$$\left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2.$$

Poi osservando che non importa in quale ordine le cinque uscite escano dobbiamo riassortirle in tutti i modi possibili, che sono chiaramente $\binom{5}{3}$: infatti dobbiamo fissare su cinque uscite l'insieme di tre elementi delle tre uscite "buone". Questo è il coefficiente con cui moltiplicare la probabilità precedente per ottenere quella cercata.

Con la distribuzione di Bernoulli tutto è molto più semplice. Il lancio di un dado può essere visto come un esperimento bernoulliano di parametro di successo $p = \frac{1}{6}$. Vogliamo che su $n = 5$ lanci si abbiano $k = 3$ successi, e dunque

$$P(S_5 = 3) = \binom{5}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^2 \approx 0,0032,$$

che è quanto avevamo ottenuto.

5.5.2 Legge geometrica

Sulla legge geometrica esiste in probabilità una sorta di ambiguità. Con il nome di legge geometrica si indicano due densità di probabilità:

- (1) si può indicare il numero di ripetizioni di un esperimento di Bernoulli necessario per ottenere il primo successo;
- (2) il numero di fallimenti prima del primo successo.

Chiaramente secondo la prima interpretazione la funzione ha valori in $\{1, 2, \dots\}$ per calcolarne la probabilità, mentre l'altra li ha in $\{0, 1, \dots\}$. Quale delle due sia chiamata *la* distribuzione geometrica è questione di convenzione e convenienza. Queste due differenti distribuzioni geometriche non devono essere confuse una con l'altra. Spesso, il nome di distribuzione geometrica modificata viene adottato per il caso (1) che abbiamo presentato; tuttavia, per eliminare ambiguità, è considerata buona norma esprimere sempre l'insieme dei valori accettati dalla legge.

Intanto presentiamo le due leggi di distribuzione, poi come prima vedremo un modello che la segue. Sia $p \in (0, 1)$ parametro fissato, allora le due leggi sono

$$p(n) = p(1-p)^{n-1} \quad \forall n \geq 1 \quad \text{e} \quad p(n) = p(1-p)^n \quad \forall n \geq 0.$$

Per poter descrivere la variabile geometrica abbiamo bisogno della nozione di indipendenza per un insieme finito di eventi.

Teorema 5.5.2. *La funzione $p(n) = p(1-p)^{n-1}$ per $n \geq 1$ esprime la probabilità di avere un successo all' n -esima ripetizione con indipendenza di un esperimento di tipo bernoulliano di probabilità di successo p .*

Dimostrazione. Siano $\{X_i\}_{i \in \mathbb{N}}$ una quantità numerabile di variabili aleatorie indipendenti (ossia per ogni n le prime n lo sono). Si suppongono le X_i variabili

aleatorie di tipo bernoulliano di parametro di successo $p \in (0, 1)$. Il tempo al quale si ottiene il primo successo è

$$T(\omega) = \min\{i \geq 1 \mid X_i(\omega) = 1\} \cup \{\infty\},$$

dove ω è un elemento di uno spazio Ω .³ Ci chiediamo quale sia $P(T = n)$. Questo non è difficile, in quanto si ha semplicemente

$$P(T = n) = P\left(\bigcap_{i=1}^{n-1} \{X_i = 0\} \cap \{X_n = 1\}\right) = \prod_{i=1}^{n-1} P(X_i = 0) \cdot P(X_n = 1),$$

e ciò è $p(1-p)^{n-1}$, ed abbiamo concluso. \square

Come annunciato, adesso abbiamo anche l'analogo del teorema precedente però nella seconda impostazione del problema della variabile geometrica, ossia di descrivere il numero di fallimenti prima del primo successo.

Teorema 5.5.3. *La funzione $p(n) = p(1-p)^n$ per $n \geq 0$ esprime la probabilità di avere n fallimenti prima del primo successo nella ripetizione con indipendenza di un esperimento di tipo bernoulliano di probabilità di successo p .*

Dimostrazione. Rimettiamoci nelle ipotesi del precedente teorema. Ciò che stiamo cercando stavolta è invece $P(T = n + 1)$ e quindi è ovvia la tesi. \square

Osservazione 5.5.3. Supponiamo di voler calcolare la probabilità che si abbia il primo successo almeno dopo n fallimenti. Avremo

$$P(T > n) = P\left(\bigcap_{i=1}^n \{X_i = 0\}\right) = (1-p)^n.$$

Visto che $\{T > n\} \supseteq \{T > n + 1\}$ e tale successione converge a $\{T = \infty\}$ si ha

$$P(T = \infty) = \lim_{n \rightarrow \infty} P(T > n) = \lim_{n \rightarrow \infty} (1-p)^n = 0.$$

Questo fatto esprime in maniera intuitiva che prima o poi, in una successione di esperimenti, dovremo ottenere un successo. Inoltre questo è un vantaggio dal punto di vista formale perché si può pensare la variabile aleatoria come a valori in numeri naturali.

Osservazione 5.5.4. Anche in questo caso vale la (5.1). Infatti

$$\sum_{n=0}^{\infty} p(n) = \sum_{n=0}^{\infty} p(1-p)^n = p \frac{1}{1-(1-p)} = 1.$$

³non descriveremo, perché non serve, lo spazio Ω ; basterà considerare uno spazio abbastanza grande da contenere la possibilità di lanciare infinite volte.

Adesso presentiamo una proprietà che caratterizza esattamente le variabili geometriche:

Proposizione 5.5.2 (assenza di usura). *Sia $X : \Omega \rightarrow \mathbb{N}$ una variabile aleatoria e sia $t \in \mathbb{N}$. Considerato $A \subseteq \mathbb{N}$ e posto $\tilde{X} = X - t$ si ha che X è geometrica se e solo se*

$$P(X \in A) = P(\tilde{X} \in A \mid X \geq t).$$

Dimostrazione. (\implies) Sia p il parametro della variabile aleatoria geometrica X . Mostriamo l'uguaglianza intanto per ogni $a \in A$. Sappiamo che

$$P(X = a) = p(1 - p)^a \quad \text{e} \quad P(X \geq t) = (1 - p)^t$$

per ogni $a \in A \subseteq \mathbb{N}$ e per ogni $t \in \mathbb{N}$. Ma allora

$$P(X = a + t \mid X \geq t) = \frac{P(X = a + t)}{P(X \geq t)} = \frac{p(1 - p)^{a+t}}{(1 - p)^t} = P(X = a).$$

A questo punto, dimostrata l'uguaglianza per ogni singoletto $a \in A$, essendo la variabile discreta, basta osservare che

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

e la stessa cosa anche per il secondo membro.

(\implies) Sia $p(n) = P(X = n)$ la legge di X . Posto $q = \sum_{n=1}^{\infty} p(n)$ e ponendo $t = 1$ nella legge di assenza di usura si ottiene

$$P(X = a + 1) = p(a + 1) = q \cdot p(a).$$

Iterando si ottiene $p(n) = q^n p(0)$, e ricordando che $\sum_{n=0}^{\infty} p(n) = 1$ si ha $q = 1 - p(0)$, da cui segue allora che

$$p(n) = p(0)(1 - p(0))^n,$$

il che mostra che la densità è di tipo geometrico. \square

5.5.3 Legge di Poisson

Fissato $\lambda \in (0, 1)$ la distribuzione di Poisson è descritta dalla seguente distribuzione:

$$p(n) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

Prima dell'interpretazione vogliamo proporre un lemma che ci servirà:

Lemma 5.5.1 (degli eventi rari). *Siano $N_h \geq 1$ e $p_h \in (0, 1)$ tali che $\lim_{h \rightarrow \infty} N_h = \infty$, $\lim_{h \rightarrow \infty} p_h = 0$ e $\lim_{h \rightarrow \infty} N_h p_h = \lambda > 0$. Allora*

$$\lim_{h \rightarrow \infty} \binom{N_h}{n} p_h^n (1 - p_h)^{N_h - n} = \frac{\lambda^n}{n!} e^{-\lambda}.$$

Dimostrazione. Sviluppando il coefficiente binomiale si può scrivere:

$$\binom{N_h}{n} p_h^n (1 - p_h)^{N_h - n} = \frac{N_h p_h (N_h - 1) p_h \cdots (N_h - n + 1) p_h}{n!} (1 - p_h)^{N_h} \frac{1}{(1 - p_h)^n}.$$

A questo punto si può scrivere per il numeratore della frazione

$$\lim_{h \rightarrow \infty} \prod_{i=0}^{n-1} (N_h - i) p_h = \prod_{i=0}^{n-1} \lim_{h \rightarrow \infty} (N_h p_h - i p_h) = \lambda^n.$$

Inoltre

$$(1 - p_h)^{N_h} = \left((1 - p_h)^{\frac{1}{p_h}} \right)^{N_h p_h} \rightarrow e^{-\lambda},$$

e infine $\frac{1}{(1 - p_h)^n} \rightarrow 1$ se $h \rightarrow \infty$. \square

Adesso possiamo spiegare il senso del lemma appena dimostrato. Nel lemma abbiamo dimostrato che la distribuzione di Poisson si ottiene come limite all'infinito della distribuzione binomiale con parametri p_h e N_h :

$$p(n) = \binom{N_h}{n} p_h^n (1 - p_h)^{N_h - n}.$$

Le prime due ipotesi per poter passare a tale limite sono $N_h \rightarrow \infty$ e $p_h \rightarrow 0$: questo esprime il fatto che il numero di eventi che si considera è molto grande, ma che ciascuno di essi ha probabilità infinitesima; questo è il motivo per cui la legge di Poisson si chiama anche *legge degli eventi rari*. Il fatto che $N_h p_h \rightarrow \lambda$ significa che “in media” avvengono λ eventi.

La legge di Poisson può essere anche utilizzata, e in statistica in certi casi lo è, per approssimare una legge binomiale con n grande e p piccolo: generalmente di fa l'approssimazione quando $n > 100$ e $p < 10^{-1}$. Vediamo un esempio:

Esempio 5.5.3. Prendiamo la quantità di uranio radioattivo che contiene 10^{24} atomi. Sia $p = 10^{-20}$ la probabilità che in un secondo un atomo di uranio decada. Per sapere la probabilità che decadano k atomi potremmo usare la legge binomiale, ma essendo il numero degli eventi elevato e ciascuno di essi raro si può utilizzare la legge di Poisson con $\lambda = Np = 10^{-4}$.

5.5.4 Legge ipergeometrica

Adesso presentiamo l'ultima legge discreta che vogliamo considerare, la legge ipergeometrica. Questa descrive il numero di successi in una sequenza di n estrazioni da un insieme finito e senza reinserimento; infatti è la variabile binomiale che descrive questo caso ma con il reinserimento. Però la via più facile per capire questa distribuzione è nei termini di modelli con le urne. Partiamo dunque dalla presentazione in pratica della legge, per poi ricavarne l'espressione.

Abbiamo un'urna con N palline, di cui r rosse e $b = N - r$ blu. Estraiamo senza reinserimento n palline e sia R il numero di palline rosse estratte. Ci chiediamo quant'è $P(R = k)$, ossia qual è la probabilità che delle n palline estratte esattamente k siano rosse. Le palline rosse sono distribuite tra quelle estratte (e sono k) e quelle non estratte (e sono $r - k$). Vale dunque

$$p(k) = P(R = k) = \frac{\binom{n}{k} \binom{N-n}{r-k}}{\binom{N}{r}} = \frac{r!(N-r)!n!(N-n)!}{N!k!(r-k)!(n-k)!(N+k-r-n)!}$$

Si osserva che vale $P(R = k) > 0$ se e solo se $k \geq 0$, $k \geq r + n - T$, $k \leq n$ e $k \leq r$. Se vogliamo rendere il ragionamento precedente rigoroso, e costruire un modello di probabilità, possiamo scegliere come Ω lo spazio dei possibili sottoinsiemi ω di n elementi in un insieme $I = \{1, \dots, N\}$ di N elementi; ω rappresenta l'estrazione. Aggiungiamo la probabilità P uniforme su Ω e decidiamo, senza perdita di generalità, che i primi r elementi di I sono le palline rosse: la variabile aleatoria

$$R(\omega) = |\{k \mid k \leq r, k \in \omega\}|$$

conta quanti elementi rossi sono in ω . In questo caso, in modo simile a quanto fatto finora, si ha

$$p(k) = P(R = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}},$$

ed un semplice calcolo mostra che questa e la precedente coincidono. Indicheremo la variabile ipergeometrica dell'estrazione tra N palline di cui r rosse e n palline di cui k rosse con $H(N, r, n)$.

Osservazione 5.5.5. Una prima semplice osservazione che possiamo fare deriva dal fatto che possiamo decidere di contare le palline rosse non estratte invece delle rosse estratte. A dire che vale

$$H(N, r, n)(k) = H(N, r, N - n)(r - k).$$

Una simmetria più sottile è quella che segue, in cui invertiamo il ruolo delle palline "rosse" e delle palline "estratte", ossia vale

$$H(N, r, n)(k) = H(N, n, r)(k).$$

Osservazione 5.5.6. Adesso riscriveremo la legge ipergeometrica in una forma più mnemonica; tale forma è dovuta a Fischer. Posti A , B , C e D rispettivamente il numero di blu estratte, il numero di rosse estratte, il numero di rosse non estratte e il numero di blu non estratte si ha

$$H(N, r, n)(k) = \frac{(A+B)!(B+C)!(C+D)!(D+A)!}{N!A!B!C!D!}.$$

Adesso vogliamo presentare un altro tipo di problema in cui interviene la legge ipergeometrica. Supponiamo che le palline siano numerate, e che il loro colore sia scelto casualmente con variabili aleatorie $\{X_i\}_{i=1}^N$ nel senso che $\{X_i = 1\}$ se l' i -esima pallina è rossa; scegliamo le variabili aleatorie indipendenti e bernoulliane di parametro $p \in (0, 1)$. Poniamo poi $S_n = \sum_{i=1}^n X_i$. Possiamo creare un modello canonico per il precedente ragionamento, dove $\Omega = \{0, 1\}^N$ e P è il prodotto di N copie della legge di Bernoulli di parametro p ; le variabili aleatorie X_i sono le proiezioni canoniche.

Possiamo decidere senza perdita di generalità che le palline estratte sono le prime n palline; aggiungiamo il requisito che il numero totale di palline rosse sia r , cosicché la probabilità cercata è $P(S_n = k \mid S_N = r)$ che possiamo calcolare intanto come

$$P(S_n = k \mid S_N = r) = \frac{P(S_n = k \cap S_N = r)}{P(S_N = r)}.$$

Il denominatore è piuttosto semplice, in quanto è descritto dalla legge binomiale seguente: $\binom{N}{r} p^r (1-p)^{N-r}$. Invece per il numeratore occorrono dei passaggi in più, che riportiamo di seguito tanto sono piuttosto semplici:

$$\begin{aligned} P(S_n = k \mid S_N = r) &= P(X_1 + \dots + X_n = k \cap X_{n+1} + \dots + X_N = r - k) = \\ &= P(X_1 + \dots + X_n = k) P(X_{n+1} + \dots + X_N = r - k). \end{aligned}$$

A questo punto i calcoli mostrano che la probabilità risulta effettivamente $H(N, r, n)$.

Proposizione 5.5.3. *Detta R la variabile aleatoria con legge ipergeometrica si ha*

$$E(R) = \frac{nr}{N} \quad e \quad Var(R) = \frac{rn(N-n)(N-r)}{N^2(N-1)}.$$

Dimostrazione. Per abbreviare sia $Q(\cdot) = P(\cdot \mid S_N = r)$ la probabilità condizionale indicata; per quanto appena mostrato si ha $S_n(Q) = H(N, r, n)$. Dunque $E_Q[S_n]$ è la speranza della variabile aleatoria ipergeometrica:

$$E_Q[S_n] = \sum_{i=1}^n E_Q[X_i] = \sum_{i=1}^n Q(X_i = 1),$$

ma allora

$$\begin{aligned} Q(X_i = 1) &= P(X_i = 1 \mid S_N = r) = \frac{P(X_i = 1 \cap S_N = r)}{P(S_N = r)} = \\ &= \frac{P(X_i)P(X_1 + \dots + X_{i-1} + X_{i+1} + \dots + X_n = r - 1)}{P(S_N = r)} = \frac{\binom{N-1}{r-1}}{\frac{N}{r}} = \frac{r}{N}. \end{aligned}$$

Dunque $E(R) = E_Q[S_n] = \frac{nr}{N}$. Per la varianza si ragiona in modo del tutto analogo. \square

5.6 Funzione generatrice di probabilità

Per i calcoli con le variabili aleatorie discrete (a valori in \mathbb{N}) è utile la nozione di funzione generatrice della legge di X . La legge di $X : \Omega \rightarrow \mathbb{N}$ è una misura di probabilità su \mathbb{N} ed è determinata univocamente dalla sua densità $p(n) = P(X = n)$.

Definizione 5.6.1. La *funzione generatrice* della legge di X è per definizione la somma di una serie, più precisamente

$$F_X(t) = \sum_{n=0}^{\infty} p(n) \cdot t^n$$

Osservazione 5.6.1. Poiché $\sum_{n=0}^{\infty} p(n) = 1$, il raggio di convergenza di $F_X(t)$ è maggiore è uguale a 1.

Con i prossimi risultati sarà chiara l'utilità della funzione generatrice, essa infatti permette di calcolare abbastanza agevolmente speranza e varianza di una variabile aleatoria.

Proposizione 5.6.1. Sia X una variabile aleatoria a valori in \mathbb{N} . Allora

$$E(X) = \lim_{t \rightarrow 1^-} F'_X(t).$$

Dimostrazione. Vale che

$$E(X) = \sum_{n=0}^{\infty} np(n) = \lim_{t \rightarrow 1^-} \sum_{n=0}^{\infty} np(n)t^{n-1} = \lim_{t \rightarrow 1^-} F'_X(t).$$

Dobbiamo giustificare il secondo uguale che abbiamo scritto: se la somma $\sum_{n=0}^{\infty} np(n)$ è infinita è ovvio, mentre nel caso finito si deve usare il teorema di convergenza dominata. \square

Proposizione 5.6.2. *Sia X una variabile aleatoria a valori in \mathbb{N} . Allora*

$$\text{Var}(X) = F_X''(1) + F_X'(1) - F_X'^2(1).$$

Dimostrazione. Mostriamo dapprima che $F_X''(t) \rightarrow E(X^2) - E(X)$ se $t \rightarrow 1^-$. Sia $0 < t < 1$, allora la serie che definisce la funzione generatrice converge e possiamo scrivere le due derivate

$$F_X'(t) = \sum_{n=1}^{\infty} np(n)t^{n-1} \quad \text{e} \quad F_X''(t) = \sum_{n=2}^{\infty} n(n-1)p(n)t^{n-2}.$$

Allora

$$F_X''(t) = \sum_{n=0}^{\infty} n^2 p(n)t^{n-2} - \sum_{n=0}^{\infty} np(n)t^{n-2}$$

e passando al limite con $t \rightarrow 1^-$ (utilizzando Beppo Levi) si ha la tesi. Unendo il risultato appena conseguito con la precedente proposizione si ottiene la formula per la varianza. \square

La funzione generatrice ha un buon comportamento rispetto a variabili aleatorie indipendenti, vale infatti il seguente:

Lemma 5.6.1. *Se X e Y sono due variabili aleatorie indipendenti a valori in \mathbb{N} allora si ha che $F_{X+Y}(t) = F_X(t)F_Y(t)$.*

Dimostrazione. Il coefficiente di t^n in F_{X+Y} è

$$P(X+Y = n) = P\left(\bigcup_{k=0}^n (\{X = k\} \cap \{Y = n - k\})\right) = \sum P(X = k)P(Y = n-k),$$

che è anche il coefficiente di t^n in $F_X(t)F_Y(t)$, ricordando che il prodotto di serie si fa alla Cauchy. \square

5.7 Legge dei grandi numeri e disuguaglianza di Chebyshev

La legge dei grandi numeri, detta anche legge empirica del caso, descrive il comportamento della media di una sequenza di n variabili aleatorie indipendenti e caratterizzate dalla stessa distribuzione di probabilità (n misure della stessa grandezza, n lanci della stessa moneta, eccetera) al tendere all'infinito della numerosità della sequenza stessa, ossia di n . In altre parole, grazie alla legge dei grandi numeri, possiamo "fidarci" che la media che calcoliamo a partire da un numero sufficiente di campioni sia sufficientemente vicina alla media vera.

In termini generici e ancora informali, per la legge dei grandi numeri si può dire:

- che la media della sequenza è una approssimazione, che migliora al crescere di n , della media della distribuzione;
- che, viceversa, si può prevedere che sequenze siffatte mostreranno una media tanto più spesso e tanto più precisamente prossima alla media della distribuzione quanto più grande sarà n .

Un caso particolare di applicazione della legge dei grandi numeri è la previsione probabilistica della proporzione di successi in una sequenza di n realizzazioni indipendenti di un evento E : per n che tende all'infinito, la proporzione di successi converge alla probabilità di E . Adesso formalizziamo tutto quanto detto e iniziamo con una disuguaglianza molto importante:

Proposizione 5.7.1 (disuguaglianza di Markov). *Sia $f : \Omega \rightarrow \mathbb{R}$ una funzione misurabile su $(\Omega, \mathcal{S}, \mu)$. Allora per ogni $\varepsilon > 0$ vale*

$$\mu(|f| \geq \varepsilon) \leq \frac{1}{\varepsilon} \|f\|_1.$$

Dimostrazione. Basta osservare che per ogni $x \in \Omega$ vale

$$|f(x)| \geq \varepsilon \cdot \chi_{\{|f| \geq \varepsilon\}}(x);$$

infatti $\chi_{\{|f| \geq \varepsilon\}}(x) = 1$ se e solo se $|f(x)| \geq \varepsilon$. Se si integra tale disuguaglianza si ottiene allora $\|f\|_1 \geq \varepsilon \cdot \mu(|f| \geq \varepsilon)$ e abbiamo concluso. \square

Corollario 5.7.1 (disuguaglianza di Chebyshev). *Sia X una variabile aleatoria sullo spazio di probabilità (Ω, \mathcal{S}, P) e integrabile. Allora*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

Dimostrazione. Il fatto che X sia integrabile ci garantisce l'esistenza della speranza di X . Sia dunque $f = X - E(X)$, allora

$$P(|X - E(X)| \geq \varepsilon) = P(|X - E(X)|^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} E(|X - E(X)|^2) = \frac{\text{Var}(X)}{\varepsilon^2},$$

dove nella prima maggiorazione abbiamo utilizzato la disuguaglianza di Chebyshev. \square

Sia $\{X_i\}_{i=1}^\infty$ una successione di variabili aleatorie indipendenti con legge di Bernoulli di parametro p . Sia poi $S_n = \sum_{i=1}^n X_i$, allora

$$\frac{1}{n} S_n = \frac{\sum_{i=1}^n X_i}{n}$$

è la frequenza dei successi. Quello che la legge dei grandi numeri dimostra è che tale frequenza, al crescere di n , tende ad assumere il valore p ; tutto ciò è espresso nel seguente:

Teorema 5.7.1 (legge dei grandi numeri). Per ogni $\varepsilon > 0$ vale che

$$P\left(\left|\frac{1}{n}S_n - p\right| \geq \varepsilon\right) \rightarrow 0$$

Dimostrazione. Ricordiamo che la variabile aleatoria S_n ha una legge binomiale, ed ha funzione generatrice

$$F_{S_n}(t) = \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} t^k = \sum_{k=0}^n \binom{n}{k} (tp)^k (1-p)^{n-k} = (1-p+tp)^n.$$

Tale funzione è derivabile con continuità infinite volte, quindi sappiamo che $E(S_n) = F'_{S_n}(1) = np$ e $Var(S_n) = F''_{S_n}(1) + F'_{S_n}(1) - F'^2_{S_n}(1) = np(1-p)$. Quindi

$$E\left(\frac{1}{n}S_n\right) = p \quad \text{e} \quad Var\left(\frac{1}{n}S_n\right) = \frac{p(1-p)}{n}.$$

Applicando il corollario 5.7.1 alla variabile aleatoria $X = \frac{1}{n}S_n$ si ottiene

$$P\left(\left|\frac{1}{n}S_n - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2},$$

e ciò conclude. \square

Osservazione 5.7.1. Si può anche dire che la legge di X si *concentra* su p : infatti nel complementare di un qualsiasi intervallo di centro p la misura di probabilità è nulla. In effetti potremmo anche esprimere questo fatto in termini di convergenza:

Definizione 5.7.1. Una successione $\{X_i : \Omega \rightarrow \mathbb{R}\}$ di variabili aleatorie *converge in probabilità* alla variabile aleatoria X se e solo se per ogni $\varepsilon > 0$ si ha

$$P(|X_n - X| \geq \varepsilon) \rightarrow 0.$$

In effetti quella data è la nozione di convergenza è la convergenza rispetto a una opportuna metrica sullo spazio vettoriale di tutte le funzioni $X : \Omega \rightarrow \mathbb{R}$ misurabili. Inoltre questa distanza su tale spazio rende continue le operazioni di spazio vettoriale.

In effetti possiamo anche mostrare che la convergenza nella legge dei grandi numeri è esponenziale:

Lemma 5.7.1. Per ogni $\varepsilon > 0$ e per ogni $p \in [0, 1]$ esiste una costante C che dipende solo da ε e da p tale che

$$P\left(\left|\frac{1}{n}S_n - p\right| \geq \varepsilon\right) \leq C^n.$$

Dimostrazione. Introduciamo un parametro ausiliario $t > 1$. Sia α tale che $p \leq \alpha \leq 1$; si ha allora

$$\begin{aligned} P\left(\frac{1}{n}S_n \geq \alpha\right) &= P(S_n - n\alpha \geq 0) = P(t^{S_n - n\alpha} \geq 1) \leq E(t^{S_n - n\alpha}) = \\ &= E\left(\prod_{i=1}^n t^{X_i - \alpha}\right) = \prod_{i=1}^n E(t^{X_i - \alpha}) = ((1-p)t^{-\alpha} + pt^{1-\alpha})^n \end{aligned}$$

dove il minore o uguale deriva dalla disuguaglianza di Chebyshev e il prodotto passa fuori dalla speranza grazie all'indipendenza delle $t^{X_i - \alpha}$. Adesso cerchiamo $t > 1$ tale che

$$g(t) = (1-p)t^{-\alpha} + pt^{1-\alpha} < 1.$$

In effetti basta calcolare il minimo della funzione $g(t)$: osservando che $g(1) = 1$ e $g'(1) = p - \alpha < 0$ abbiamo che la funzione è decrescente in $t = 1$; poi visto che se $t \rightarrow \infty$ allora $g(t) \rightarrow +\infty$ segue che ci deve essere un minimo e che questo è più piccolo di 1. La derivata è

$$g'(t) = -(1-p)\alpha + p(1-\alpha)t^{-1-\alpha},$$

e tale minimo è l'unico punto in cui si annulla la derivata, ossia

$$t = \frac{\alpha(1-p)}{p(1-\alpha)},$$

da cui $\min_{t>1} g(t) = \left(\frac{p}{\alpha}\right)^\alpha \left(\frac{1-p}{1-\alpha}\right)^{1-\alpha} < 1$. Inoltre per $0 \leq \alpha \leq p$ si ha analogamente

$$P\left(\frac{1}{n}S_n \leq \alpha\right) = P\left(1 - \frac{1}{n}S_n \geq 1 - \alpha\right) \leq \left(\frac{p}{\alpha}\right)^\alpha \left(\frac{1-p}{1-\alpha}\right)^{1-\alpha};$$

infatti $1 - \frac{1}{n}S_n$ è la frequenza dei successi rispetto alla successione di variabili aleatorie (sempre indipendenti) $1 - X_i$, con logge di Bernoulli di parametro $1 - p$. In definitiva ponendo $\alpha = \varepsilon + p$ si ottiene

$$P\left(\left|\frac{1}{n}S_n - p\right| \geq \varepsilon\right) \leq \left(\left(\frac{p}{p+\varepsilon}\right)^{p+\varepsilon} \left(\frac{1-p}{1-p+\varepsilon}\right)^{1-p+\varepsilon}\right)^n,$$

ed abbiamo concluso. \square

5.7.1 La formula di Stirling

Molto utile in probabilità, specialmente nelle applicazioni, è la *formula di Stirling*, che forse il lettore già conosce. Decidiamo di dare anche una dimostrazione:

Teorema 5.7.2 (formula di Stirling). *Per $n \rightarrow \infty$ vale*

$$n! = \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} (1 - o(1)).$$

Dimostrazione. Dal fatto che $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ segue $e^n \geq \frac{n^n}{n!}$, da cui

$$n! \geq n^n e^{-n}.$$

Visto che $y = \log x$ è crescente, la somma di Riemann inferiore per l'intervallo $[1, n]$ rispetto alla partizione $\sigma = \{1, 2, \dots, n\}$ è data da

$$S(\sigma) = \sum_{k=1}^n \log k = \log n! \leq \int_1^n \log x \, dx = n \log n - n + 1.$$

Da ciò segue $n! \leq n^{n+1} e^{-n+1}$. In definitiva

$$n^n e^{-n} \leq n! \leq n^{n+1} e^{-n+1}.$$

Ciò suggerisce un comportamento del fattoriale del tipo $n! = cn^{n+\alpha} e^{-n}$. Dato $\alpha \in [0, 1]$ si considera

$$\rho_n = \frac{n^{n+\alpha} e^{-n}}{n!},$$

allora

$$\frac{\rho_{n+1}}{\rho_n} = \frac{1}{e} \left(1 + \frac{1}{n}\right)^{n+\alpha} = e^{(n+\alpha) \log\left(1 + \frac{1}{n}\right) - 1}.$$

Dallo sviluppo di $\log(1+x)$ con $x = \frac{1}{n}$ si ottiene

$$\log \frac{\rho_{n+1}}{\rho_n} = \left(\alpha - \frac{1}{2}\right) \frac{1}{n} + \left(\frac{1}{3} - \frac{\alpha}{2}\right) \frac{1}{n^2} + o(n^{-2}).$$

per $n \rightarrow \infty$. Scelto $\alpha = \frac{1}{2}$ si ha $\log \frac{\rho_{n+1}}{\rho_n} = \frac{1}{12n^2} (1 + o(1))$ e esponenziando da entrambe le parti si ottiene

$$\frac{\rho_{n+1}}{\rho_n} = e^{\frac{1}{12n^2}} (1 + o(1)).$$

Dunque ρ_n è decrescente e limitata, quindi ammette limite k e $c = \frac{1}{k}$. Adesso il teorema è concluso non appena calcolato il limite di ρ_n , parte semplice che lasciamo al lettore. \square

5.7.2 Alcune osservazioni

Adesso presentiamo una variazione della disuguaglianza di Chebyshev che adotta notazioni più probabilistiche. Il risultato è il seguente:

Proposizione 5.7.2 (disuguaglianza di Markov). *Sia $X : \Omega \rightarrow [0, \infty)$ una variabile aleatoria integrabile (ossia 1) e sia $\varepsilon > 0$. Allora*

$$P(X > \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Dimostrazione. Basta considerare l'ovvia disuguaglianza $\varepsilon \cdot \chi_{\{X > \varepsilon\}}(x) \leq X(x)$ e integrarla. \square

Corollario 5.7.2. *Se $X \geq 0$ e $E(X) = 0$ allora $X = 0$ quasi ovunque.*

Dimostrazione. Dalla disuguaglianza di Markov segue che $P(X > \frac{1}{n}) = 0$ per ogni $n \in \mathbb{N}^+$. Visto che la famiglia $\{X > \frac{1}{n}\}$ è crescente possiamo passare al limite e portare il limite dentro la funzione di probabilità e ottenere dunque $P(X > 0) = 0$. \square

Corollario 5.7.3. *Se $f, g \in {}^1(\mathbb{R})$ e se $\int_I f(x) dx = \int_I g(x) dx$ per ogni intervallo $I \subseteq \mathbb{R}$, allora $f = g$ quasi ovunque.*

Dimostrazione. Sappiamo che $\int_I f(x) dx = \int_I g(x) dx$ per ogni I , ossia $\int_{I^+} (f(x) - g(x)) dx = 0$ e $\int_{I^-} (g(x) - f(x)) dx = 0$, dove $I^+ = \{f > g\}$ e $I^- = \{g \geq f\}$. La differenza tra i due integrali è ovviamente l'integrale del modulo, che essendo nullo, fa sì che la funzione $|f(x) - g(x)|$ sia quasi ovunque nulla su I . \square

5.8 Il teorema del limite centrale

Adesso presentiamo un risultato molto importante in probabilità, che in un certo senso precisa quanto già detto dalla legge dei grandi numeri, e che va sotto il nome di *teorema del limite centrale*, dovuto a De Moivre e Laplace. La legge dei grandi numeri può essere anche riscritta come segue:

$$P(np - \varepsilon \leq S_n \leq np + \varepsilon n) \longrightarrow 1$$

quando $n \rightarrow \infty$. Visto che l'intervallo intorno alla media np in cui è concentrata S_n cresce linearmente con n ci possiamo chiedere se c'è una limitazione a tale crescita. Spieghiamo meglio: vogliamo vedere se c'è un modo più lento di far crescere l'intervallo ma in modo tale che si abbia ancora convergenza in probabilità. Questo è il contenuto del teorema del limite centrale. Però prima dobbiamo mostrare un lemma – che però assurgeremo a teorema – che ci servirà per il vero e proprio teorema:

Teorema 5.8.1 (del limite locale). *Vale*

$$P(S_n = k) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}} \left(1 + O\left(n^{-\frac{1}{2}}\right)\right)$$

per $n \rightarrow \infty$, dove $k = np + x\sqrt{npq}$ e uniformemente per x nei limitati.

Dimostrazione. Osserviamo che $n - k = nq - x\sqrt{npq}$ e anche che

$$\frac{k}{n} = p + O\left(n^{-\frac{1}{2}}\right) \quad \text{e} \quad \frac{n-k}{n} = q + O\left(n^{-\frac{1}{2}}\right)$$

per $n \rightarrow \infty$. Inoltre $k = np + O\left(n^{\frac{1}{2}}\right)$ e $n - k = nq + O\left(n^{\frac{1}{2}}\right)$ sempre per $n \rightarrow \infty$. Sappiamo poi che S_n ha legge binomiale

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}.$$

Viste le stime asintotiche precedenti possiamo, per $n \rightarrow \infty$ scrivere i vari fattoriali grazie alla formula di Stirling, e quindi

$$\begin{aligned} P(S_n = k) &= \frac{\sqrt{2\pi n} n^n e^{-n}}{\sqrt{2\pi k} k^k e^{-k} \cdot \sqrt{2\pi(n-k)} (n-k)^{(n-k)} e^{-(n-k)}} p^k q^{n-k} (1 + o(1)) = \\ &= \frac{1}{\sqrt{2\pi n \frac{k}{n} \frac{n-k}{n}}} \left(\frac{pn}{k}\right)^k \left(\frac{qn}{n-k}\right)^{n-k} (1 + o(1)) = \\ &= \frac{1}{\sqrt{2\pi npq}} \cdot e^{-k \log \frac{k}{pn} - (n-k) \log \frac{n-k}{qn}} (1 + o(1)). \end{aligned}$$

Chiamiamo E l'esponente di e nell'espressione precedente, vogliamo scriverlo in funzione di x e svilupparlo grazie alle serie di Taylor dei logaritmi fino al secondo ordine (ricordiamo che x è in un limitato, quindi gli sviluppi hanno senso):

$$\begin{aligned} E &= -k \log \left(1 + x\sqrt{\frac{q}{np}}\right) - (n-k) \log \left(1 - x\sqrt{\frac{p}{nq}}\right) \\ &= -(np + x\sqrt{npq}) \left(x\sqrt{\frac{q}{np}} - \frac{x^2}{2} \frac{q}{np} + O\left(n^{-\frac{3}{2}}\right)\right) - \\ &\quad -(nq - x\sqrt{npq}) \left(-x\sqrt{\frac{p}{nq}} - \frac{x^2}{2} \frac{p}{nq} + O\left(n^{-\frac{3}{2}}\right)\right) = \\ &= -x\sqrt{npq} + x\sqrt{npq} - x^2 q - x^2 p + \frac{x^2}{2} q + \frac{x^2}{2} p + O\left(n^{-\frac{1}{2}}\right) = -\frac{x^2}{2} + O\left(n^{-\frac{1}{2}}\right). \end{aligned}$$

Quindi, per x in un limitato, vale

$$P(S_n = k) = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}} \left(1 + O\left(n^{-\frac{1}{2}}\right)\right),$$

ed abbiamo concluso. \square

Teorema 5.8.2 (del limite centrale). *Con le stesse convenzioni dei precedenti teoremi, con $q = 1 - p$ e per $a, b \in \mathbb{R} \cup \{\pm\infty\}$ costanti, si ha*

$$P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx \cdot \left(1 + O\left(n^{-\frac{1}{2}}\right)\right).$$

Dimostrazione. Siano $a < b$ numeri reali, allora basta osservare che

$$\begin{aligned} P(np + a\sqrt{npq} \leq S_n \leq np + b\sqrt{npq}) &= \sum_{np+a\sqrt{npq} \leq k \leq np+b\sqrt{npq}} P(S_n = k) = \\ &= \frac{1}{\sqrt{2\pi}} \left(\sum_k \frac{1}{\sqrt{npq}} e^{-\frac{x_k^2}{2}} \right) (1 + o(1)), \end{aligned}$$

dove $a \leq x_k = \frac{k-np}{\sqrt{npq}} \leq b$. Se si osserva, la quantità $\frac{1}{\sqrt{npq}}$ è la lunghezza dell'intervallo $x_{k+1} - x_k$ e quindi la sommatoria scritta è proprio una somma di Riemann per l'integrale di $e^{-\frac{x^2}{2}}$ sull'intervallo $[a, b]$. Mandando $n \rightarrow \infty$ si ottiene esattamente la tesi.

Trattiamo adesso il caso in cui uno dei due, o entrambi, tra a e b sono infiniti. Per simmetria è sufficiente vedere il caso $a = -\infty$, e per fare ciò andiamo a stimare la probabilità

$$P = P(S_n \leq np + b\sqrt{npq}).$$

Introduciamo un parametro ausiliario $\lambda > 0$ e osserviamo che

$$\begin{aligned} \left| P - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{x^2}{2}} dx \right| &\leq P(S_n < np - \lambda\sqrt{npq}) + \\ &+ \left| P(np - \lambda\sqrt{npq} \leq S_n < np + b\sqrt{npq}) - \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^b e^{-\frac{x^2}{2}} dx \right| + \\ &+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\lambda} e^{-\frac{x^2}{2}} dx. \end{aligned}$$

Ma adesso occupiamoci del primo addendo

$$\begin{aligned} P(S_n < np - \lambda\sqrt{npq}) &\leq P(S_n < nq - \lambda\sqrt{npq}) + P(np + \lambda\sqrt{npq} < S_n) = \\ &= 1 - P(np - \lambda\sqrt{npq} \leq S_n \leq np + \lambda\sqrt{npq}) = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-\frac{x^2}{2}} dx + o(1) = \frac{2}{\sqrt{2\pi}} \int_{\lambda}^{+\infty} e^{-\frac{x^2}{2}} dx + o(1). \end{aligned}$$

Il secondo addendo va a zero per quanto già mostrato nella prima parte, e quindi, considerando solo il primo e il terzo avremo che, per ogni $\lambda > 0$ si ha

$$\limsup_{n \rightarrow \infty} \left| P(S_n \leq np + b\sqrt{npq}) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{x^2}{2}} dx \right| \leq \frac{3}{\sqrt{2\pi}} \int_{\lambda}^{+\infty} e^{-\frac{x^2}{2}} dx.$$

Si conclude mandando $\lambda \rightarrow +\infty$. \square

Osservazione 5.8.1. Se $b = 0$ troviamo $P(S_n \leq np) = \frac{1}{2} + o(1)$ se $n \rightarrow \infty$. Questo è un fatto molto intuitivo, ma non è per niente ovvio dimostrarlo, come visto sopra.

Osservazione 5.8.2. Si osservi che \sqrt{npq} è la radice della varianza di S_n . Abbiamo calcolato in precedenza che in un processo di Bernoulli vale $Var(S_n) = np(1-p)$. In generale alla radice quadrata della varianza si dà il nome di *deviazione standard*.

Osservazione 5.8.3. Se $a = -\infty$ e $b = +\infty$ si trova che $P(S_n \in \bar{\mathbb{R}}) = 1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx$. Da cui, se π era solo definito dalla formula di Stirling, e considerando che in questa formula π è il pi greco che tutti conoscono e già definito, allora segue che anche la costante della formula di Stirling deve essere $\sqrt{2\pi}$.

Esempio 5.8.1 (teorema di Stone–Weierstrass). Adesso vediamo un’applicazione della legge dei grandi numeri, che dà una dimostrazione probabilistica del teorema di Stone–Weierstrass, che riguarda l’approssimazione di funzioni continue su $[0, 1]$. L’idea è la seguente: sia $\{X_n\}_{n \geq 1}$ uno schema di Bernoulli con parametro $\lambda \in [0, 1]$. Vogliamo approssimare una funzione continua $f(x)$ con dei polinomi nel punto $\lambda \in [0, 1]$. Consideriamo poi $f_n(\lambda) = E\left(f\left(\frac{1}{n}S_n^{(\lambda)}\right)\right)$. I fatti sono due, questa speranza è un polinomio e converge uniformemente a f . In effetti intanto vale

$$f_n(\lambda) = E\left(f\left(\frac{1}{n}S_n^{(\lambda)}\right)\right) = \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} f\left(\frac{k}{n}\right),$$

ossia è un polinomio di grado $\leq n$ in λ . Introduciamo un parametro ausiliario $\delta > 0$ e poniamo $\omega_f(\delta) = \sup_{|x-\lambda| \leq \delta} |f(\lambda) - f(x)|$ ⁴ si ha

$$\begin{aligned} |f_n(\lambda) - f(\lambda)| &= \left| E\left(f\left(\frac{1}{n}S_n^{(\lambda)}\right)\right) - f(\lambda)E(1) \right| = \left| E\left(f\left(\frac{1}{n}S_n^{(\lambda)}\right) - f(\lambda)\right) \right| \leq \\ &\leq E\left(\left|f\left(\frac{1}{n}S_n^{(\lambda)}\right) - f(\lambda)\right|\right) \leq \sum_{k=0}^n \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \left|f(\lambda) - f\left(\frac{k}{n}\right)\right| \leq \\ &\leq \left(\sum_{|\lambda - \frac{k}{n}| \leq \delta} + \sum_{|\lambda - \frac{k}{n}| > \delta} \right) \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \left|f(\lambda) - f\left(\frac{k}{n}\right)\right| \leq \\ &\leq \omega_f(\delta) + 2\|f\|_\infty \sum_{|\lambda - \frac{k}{n}| > \delta} \binom{n}{k} \lambda^k (1-\lambda)^{n-k} \leq \\ &\leq \omega_f(\delta) + 2\|f\|_\infty P\left(\left|\frac{1}{n}S_n^{(\lambda)}\right| > \delta\right) \leq \\ &\leq \omega_f(\delta) + 2\|f\|_\infty \frac{\lambda(1-\lambda)}{\delta^2 n} \leq \omega_f(\delta) + \frac{\|f\|_\infty}{2n\delta^2}, \end{aligned}$$

⁴tale numero è chiamato *modulo di continuità* di f e f è uniformemente continua se e solo se $\omega_f(\delta)$ è infinitesimo per $\delta \rightarrow \infty$.

dove nella penultima maggiorazione abbiamo utilizzato la legge dei grandi numeri. Passando agli estremi superiori in λ si ottiene dunque, per ogni $\delta > 0$

$$\|f - f_n\|_\infty \leq \omega_f(\delta) + \frac{\|f\|_\infty}{2\delta^2 n},$$

e passando al limite superiore su n si ha

$$\limsup_{n \rightarrow \infty} \|f - f_n\|_\infty \leq \sup_{|x-\lambda| \leq \delta} |f(\lambda) - f(x)| \rightarrow 0.$$

e con ciò abbiamo finito.

5.9 Leggi di probabilità continue

Ricordiamo dall'analisi il *lemma di Fatou*: se f_n sono funzioni misurabili non negative allora

$$\liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \geq \int_{\Omega} \liminf_{n \rightarrow \infty} f_n d\mu.$$

Inoltre vale anche un analogo del *teorema di convergenza dominata* che abbiamo dato per serie, che nel caso degli integrali si enuncia come segue: se f_n sono funzioni misurabili su D , se vale $|f_n(x)| \leq g(x)$ quasi ovunque in D con g sommabile e se esiste $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ quasi ovunque in D allora

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n(x) d\mu = \int_{\Omega} f(x) d\mu.$$

Teorema 5.9.1 (formula di Stirling). Per $n \rightarrow \infty$ vale $n! = \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n} (1 - o(1))$.

Dimostrazione. Consideriamo lo sviluppo in serie $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$. Scritto z in forma polare $z = \rho e^{it}$ abbiamo allora $e^z = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} e^{ikt}$, ie di funzioni che converge uniformemente in t una volta fissato ρ . Moltiplicando per e^{-int} e integrando su $[-\pi, \pi]$ si ha

$$\int_{-\pi}^{\pi} e^z e^{-int} dt = \int_{-\pi}^{\pi} \left(\sum_{k=0}^{\infty} \frac{\rho^k}{k!} e^{i(n-k)t} \right) dt = \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \int_{-\pi}^{\pi} e^{i(n-k)t} dt = 2\pi \frac{\rho^n}{n!},$$

in quanto se $k \neq n$ gli integrali sono tutti nulli. Adesso scegliamo $\rho = n$ e dividiamo per $2\pi e^n$, si ha dunque

$$\frac{n^n}{e^n n!} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{n(e^{it}-1-it)} dt.$$

Per studiare il comportamento dell'integrale cambiamo variabile ponendo $s = t\sqrt{n}$, e inoltre ricordiamo che $e^{it} = 1 + it - \frac{t^2}{2} + o(t^2)$ per $t \rightarrow 0$, da cui

$$\int_{-\pi}^{\pi} e^{n(e^{it}-1-it)} dt = \frac{1}{\sqrt{n}} \int_{-\sqrt{n}\pi}^{\sqrt{n}\pi} e^{-\frac{s^2}{2} + nR\left(\frac{s}{\sqrt{n}}\right)} ds = \frac{1}{\sqrt{n}} \int_{-\infty}^{+\infty} f_n(s) ds,$$

dove abbiamo posto

$$f_n(s) = e^{-\frac{s^2}{2} + nR\left(\frac{s}{\sqrt{n}}\right)} \chi_{[-\pi\sqrt{n}, \pi\sqrt{n}]}$$

Per s fissato e per $n \rightarrow \infty$ si ha $nR\left(\frac{s}{\sqrt{n}}\right) = n \cdot o\left(\frac{1}{n}\right) = o(1)$ e inoltre la caratteristica è 1 definitivamente. Quindi $\lim_{n \rightarrow \infty} f_n(s) = e^{-\frac{s^2}{2}}$. Adesso vogliamo fare in modo che il limite passi dentro l'integrale per poi ottenere la tesi. Osserviamo che

$$R(t) = e^{it} - \left(1 + it - \frac{t^2}{2}\right),$$

e visto che $|e^z| = e^{\operatorname{Re}z}$, calcoliamo il modulo

$$|\operatorname{Re}R(t)| = \left| \cos t - 1 + \frac{t^2}{2} \right| \leq \frac{t^4}{4!} |\cos \xi| \leq \frac{t^4}{24}.$$

Dunque $|f_n(s)| = e^{-\frac{s^2}{2} + n\operatorname{Re}R\left(\frac{s}{\sqrt{n}}\right)} \leq e^{-\frac{s^2}{2} + \frac{s^4}{24n}} \chi(s) \leq e^{\left(-\frac{1}{2} + \frac{\pi^2}{24}\right)s^2}$, in quanto $\frac{s^2}{n} \leq \pi^2$. A questo punto basta utilizzare la convergenza dominata e passare il limite nell'integrale. \square

Grazie al teorema precedente abbiamo così visto una applicazione del teorema di convergenza dominata. Ora veniamo a questioni probabilistiche: quando si parla di variabili aleatorie reali $X : \Omega \rightarrow \mathbb{R}$ è consuetudine nella pratica utilizzare l'algebra dei boreliani $B(\mathbb{R})$. Come già abbiamo detto la misura di probabilità P presente su Ω viene trasportata da X nella misura immagine $X(P)$ tale che

$$X(P)(A) = P(X \in A),$$

per ogni $A \in B(\mathbb{R})$. Tale misura immagine viene detta *legge* della variabile aleatoria X . Nel caso discreto a questo punto abbiamo introdotto la nozione di densità: se $X : \Omega \rightarrow \mathbb{N}$ è una variabile discreta dicevamo densità della legge di X la probabilità $p(n) = X(P)(n) = P(X = n)$. Osserviamo che nel caso reale ciò non è possibile in quanto se, fissato $x \in \mathbb{R}$, scrivessimo $P(X = x)$ avremmo che tale probabilità è nulla in quanto $\{x\}$ è un insieme di misura nulla in \mathbb{R} (secondo la misura di Lebesgue). Allora la definizione che si dà è un'altra, ma simile:

Definizione 5.9.1. Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria reale sullo spazio di probabilità $(\Omega, B(\mathbb{R}), P)$. Diremo che $f(x)$ è una *densità* per la legge di X se per ogni $A \in B(\mathbb{R})$

$$X(P)(A) = P(X \in A) = \int_A f(x) dx = \int_{\mathbb{R}} \chi_A(x) f(x) dx,$$

se f è quasi ovunque non negativa e se $\int_{\mathbb{R}} \rho_X(x) dx = 1$.

5.10 Convoluzione di funzioni misurabili su \mathbb{R}^n

Abbiamo già incontrato nel caso discreto la nozione di convoluzione di due variabili aleatorie discrete. Adesso vediamo in generale cosa significa:

Definizione 5.10.1. Se f e g sono funzioni su \mathbb{R}^n , la loro *convoluzione* è – purché esista – la funzione $f * g$ su \mathbb{R}^n data da

$$x \mapsto \int_{\mathbb{R}^n} f(x-y)g(y) dy.$$

Osservazione 5.10.1. Questa operazione è l'analogo continuo della convoluzione discreta. Se u e v sono due successioni da \mathbb{Z} in \mathbb{R} allora $u * v : \mathbb{Z} \rightarrow \mathbb{R}$ è la successione tale che

$$(u * v)(n) = \sum_{k \in \mathbb{Z}} u(n-k)v(k).$$

Osserviamo che se u e v sono nulle sugli interi negativi allora la somma di convoluzione si riduce ad una somma finita, e quindi la serie converge certamente per ogni n . Osserviamo per esempio che è ben definita se $\|u\|_1 < \infty$ (ossia u è assolutamente sommabile) e se $\|v\|_\infty < \infty$ (ossia se v è limitata).

Si osservi che se u e v sono due variabili aleatorie discrete allora la convoluzione $(u * v)(n)$ esprime la legge della somma delle due variabili.

Proposizione 5.10.1. Se $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ sono misurabili e non negative (o integrabili), allora la convoluzione $f * g$ è ben definita ed è misurabile non negativa (rispettivamente integrabile). Nel caso in cui la convoluzione sia integrabile si ha

$$\int_{\mathbb{R}^n} (f * g) dx = \int_{\mathbb{R}^n} f dx \cdot \int_{\mathbb{R}^n} g dx$$

e anche $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$ (anzi, nel primo caso vale l'uguaglianza).

Dimostrazione. Intanto osserviamo che

$$(x, y) \longmapsto f(x - y)g(y)$$

è una mappa misurabile perché prodotto di misurabili (entrambe perché composizione di funzioni misurabili).

Supponiamo valga l'ipotesi (a), considerato $\int_{\mathbb{R}^n \times \mathbb{R}^n} f(x - y)g(y) d(x, y)$ allora per il teorema di Tonelli si ha

$$\int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x - y)g(y) dy \right) dx = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x - y)g(y) dx \right) dy.$$

In effetti il primo integrale scritto è proprio l'integrale della convoluzione $(f * g)(x)$, mentre il secondo membro è

$$\begin{aligned} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(x - y)g(y) dx \right) dy &= \int_{\mathbb{R}^n} g(y) \left(\int_{\mathbb{R}^n} f(x - y) dx \right) dy = \\ &= \int_{\mathbb{R}^n} g(y) \left(\int_{\mathbb{R}^n} f(x) dx \right) dy; \end{aligned}$$

a questo punto possiamo separare i due integrali ed ottenere la tesi. Per l'uguaglianza tra le norme basta osservare che tutte le funzioni integrande sono positive, quindi la norma coincide con il rispettivo integrale. Adesso passiamo al caso (b), con $f, g \in {}^1$. Per mostrare la tesi basta mostrare che la funzione $|f(x - y)g(y)|$ è integrabile su $\mathbb{R}^n \times \mathbb{R}^n$; infatti poi basterà applicare il teorema di Fubini. In effetti si ha

$$\begin{aligned} \int_{\mathbb{R}^n \times \mathbb{R}^n} |f(x - y)g(y)| d(x, y) &= \int_{\mathbb{R}^n \times \mathbb{R}^n} |f(x - y)||g(y)| d(x, y) = \\ &= \int_{\mathbb{R}^n} |f| * |g| dx = \\ &= \int_{\mathbb{R}^n} |f(x)| dx \cdot \int_{\mathbb{R}^n} |g(y)| dy = \|f\|_1 \|g\|_1, \end{aligned}$$

che è finita perché $f, g \in {}^1$. Quindi si può applicare il teorema di Fubini alla funzione $f(x - y)g(y)$ su $\mathbb{R}^n \times \mathbb{R}^n$ e ripetendo il calcolo già svolto si ha la prima tesi anche in questo caso. Infine

$$\begin{aligned} \|f * g\|_1 &= \int_{\mathbb{R}^n} |(f * g)(x)| dx = \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} f(x - y)g(y) dy \right| dx \leq \\ &\leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x - y)g(y)| dy \right) dx = \|f\|_1 \|g\|_1, \end{aligned}$$

e abbiamo finito. \square

Vediamo il significato di $f * g$ con una funzione non negativa e con integrale unitario, come può essere una densità. Sia $f_\varepsilon(x) = \frac{1}{2\varepsilon} \chi_{[-\varepsilon, \varepsilon]}(x)$, allora

$$(f_\varepsilon * g)(x) = \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} g(y) dy,$$

e quindi è la media integrale di g sull'intervallo $[x - \varepsilon, x + \varepsilon]$. Si noti che (il lettore provi a dimostrarlo) che $f_\varepsilon * g \rightarrow g$ in L^1 se $\varepsilon \rightarrow 0$.

Adesso accenniamo alla nozione di *convoluzione unilatera* su ${}^1(\mathbb{R}^+) \subseteq {}^1(\mathbb{R})$: l'inclusione si ha per immersione, in quanto si può estendere ogni funzione a zero su \mathbb{R}^- . Per $u, v \in {}^1(\mathbb{R}^+)$ allora

$$(u * v)(x) = \int_{\mathbb{R}} u(y)v(x-y) dy = \int_0^\infty u(y)v(x-y) dy,$$

ed è l'analogo continuo del prodotto di Cauchy per serie di potenze.

5.11 Misure su \mathbb{R}^n dotate di densità ¹

Se $f \in {}^1(\mathbb{R})$ è una funzione non negativa allora è definita una misura μ_f sui boreliani ponendo per ogni $A \in B(\mathbb{R})$

$$\mu_f(A) = \int_A f(x) dx.$$

Lasciamo come esercizio al lettore la verifica del fatto che μ_f è una misura finita. Se A è un boreliano di misura nulla secondo Lebesgue allora è di misura nulla anche per μ_f , in quanto $\mu_f(A) = \int_A f(x) dx = 0$.

Definizione 5.11.1. Se μ e ν sono misure su (Ω, \mathcal{A}) si dice che μ è *assolutamente continua* rispetto a ν , e si scrive $\mu \ll \nu$, se

$$\{A \in \mathcal{A} \mid \mu(A) = 0\} \subseteq \{A \in \mathcal{A} \mid \nu(A) = 0\}.$$

Esempio 5.11.1. La misura di Dirac non è assolutamente continua rispetto alla misura di Lebesgue.

Quindi il fatto enunciato prima della definizione può essere enunciato dicendo che per ogni $f \in {}^1(\mathbb{R})$ si ha che μ_f è assolutamente continua rispetto alla misura di Lebesgue. Come informazione riportiamo che vale il teorema di Radon Nikodym che caratterizza tutte le misure assolutamente continue rispetto ad una data: infatti si ha che per ogni misura μ assolutamente continua rispetto a ν esiste una funzione f misurabile non negativa tale che $\mu = \mu_f$.

5.12 Variabili aleatorie con densità

Quando una variabile aleatoria ha densità possiamo, da essa, determinare i parametri che più ci interessano, come la speranza, la funzione di ripartizione, la densità di diffeomorfismi applicati alla variabile, la densità congiunta di un blocco. Vediamo come fare.

Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria reale e $f(x)$ la sua densità. È immediato dunque che la *funzione di ripartizione* si possa scrivere come

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]) = \int_{-\infty}^x f(t) dt.$$

Da ciò segue anche la funzione di ripartizione è continua, ed il viceversa di questa affermazione è falso in generale.

Osservazione 5.12.1. Se alteriamo la densità f su un insieme di misura di Lebesgue nulla abbiamo che l'integrale non cambia. Per questo la densità non è tanto una funzione, bensì è una classe di funzioni, quelle che differiscono da f solo su un insieme di misura nulla.

Teorema 5.12.1. *Sia $F : \mathbb{R} \rightarrow [0, 1]$ una funzione continua, monotona, con $\lim_{x \rightarrow \infty} F(x) = 1$ e $\lim_{x \rightarrow -\infty} F(x) = 0$ e di classe C^1 salvo in un numero finito di punti, ossia esistono punti $t_0 < t_1 < \dots < t_n < t_{n+1}$ tali che $F|_{(t_i, t_{i+1})}$ è di classe C^1 per ogni i . Allora posto*

$$f(x) = \begin{cases} F'(x) & \text{se } x \neq t_i \\ 0 & \text{altrimenti} \end{cases}$$

abbiamo che f è densità dell'unica probabilità che ha F come funzione di ripartizione.

Dimostrazione. Si applica il teorema fondamentale del calcolo integrale in ogni intervallo e poi si raccorda per continuità. Il lettore completi i dettagli. \square

Osserviamo solo che la densità f poteva essere definita in modo qualsiasi su $\{t_0, \dots, t_n\}$, tanto questo è un insieme trascurabile, ed intervenendo le densità sempre dentro integrali avremo che tali integrali non dipendono dal valore della funzione su un insieme di misura nulla.

Teorema 5.12.2. *Sia X una variabile aleatoria reale. X ha densità f se e solo se per ogni $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ misurabile e limitata si ha*

$$E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx.$$

Dimostrazione. (\implies) Supponendo che X abbia densità f si ha

$$E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) dX(P)(x) = \int_{-\infty}^{\infty} \varphi(x)f(x) dx.$$

L'ultima uguaglianza si prova per ogni φ boreliana e limitata⁵.

(\impliedby) Preso A misurabile e considerata $\varphi = \chi_A$ si ha

$$P(X \in A) = E(\chi_A \circ X) = \int_{-\infty}^{\infty} \chi_A(x)f(x) dx = \int_A f(x), dx,$$

e ciò basta a dire che $f(x)$ è densità per X . \square

5.12.1 Caso del blocco di variabili aleatorie

Intanto iniziamo osservando che anche nel caso delle variabili aleatorie reali si può parlare di *modello canonico* di probabilità. Si considera lo spazio di probabilità $(\mathbb{R}^2, B(\mathbb{R}^2), P)$ e si prendono come variabili aleatorie le due proiezioni canoniche. In modo analogo a quanto ora detto possiamo anche definire la densità di variabili aleatorie vettoriali ed estendere il teorema appena mostrato al caso n -dimensionale (enunciarla). Consideriamo infatti il blocco (X, Y) di due variabili aleatorie e supponiamo che abbia densità $f(x, y)$; ciò significa che per ogni $C \in B(\mathbb{R}^2)$ si ha

$$(X_1, X_2)(P)(C) = \iint_C f(x, y) dx dy.$$

Vogliamo vedere se si possono conoscere le densità delle due variabili aleatorie. In effetti ciò è possibile grazie al seguente risultato, già incontrato nel caso discreto:

Lemma 5.12.1. *Siano X_1 e X_2 variabili aleatorie il cui blocco (X_1, X_2) ha densità $f(x, y)$. Allora*

$$f_1(x) = \int_{\mathbb{R}} f(x, y) dy \quad e \quad f_2(y) = \int_{\mathbb{R}} f(x, y) dx.$$

Dimostrazione. Vale in effetti

$$\begin{aligned} X_1(P)(A) &= (X_1, X_2)(P)(A \times \mathbb{R}) = P((X_1, X_2) \in A \times \mathbb{R}) = \\ &= \iint_{A \times \mathbb{R}} f(x, y) dx dy = \int_A \left(\int_{\mathbb{R}} f(x, y) dy \right) dx, \end{aligned}$$

e da questa vediamo che l'integrale interno soddisfa la definizione di densità per $X_1(P)$. \square

⁵si prenda dapprima $\varphi = \chi_A$ per un certo A misurabile e si dimostri in questo caso; fatto ciò abbiamo la tesi per le funzioni semplici. Ma allora poi bisogna prendere φ misurabile e positiva e dimostrare la tesi anche in questo caso e poi seguirà immediatamente il caso di φ misurabile generica.

Teorema 5.12.3. *Siano X_1 e X_2 due variabili aleatorie indipendenti. Allora hanno densità $f_1(x)$ e $f_2(y)$ se e solo se il blocco (X_1, X_2) ha densità $f(x, y) = f_1(x)f_2(y)$ quasi ovunque.*

Dimostrazione. (\implies) Intanto osserviamo che per ogni $A, B \in B(\mathbb{R})$ vale

$$\begin{aligned} (X_1, X_2)(P)(A \times B) &= P(X_1 \in A, X_2 \in B) = P(X_1 \in A)P(X_2 \in B) = \\ &= \int_A f_1(x) dx \cdot \int_B f_2(y) dy = \iint_{A \times B} f_1(x)f_2(y) dx dy. \end{aligned}$$

Questo mostra che la tesi è vera per la famiglia $D = \{A \times B \mid A, B \in B(\mathbb{R})\}$; ma questa famiglia soddisfa le ipotesi del teorema di unicità e dunque la tesi è vera quasi ovunque sui boreliani $B(\mathbb{R}^2)$.

(\impliedby) La seconda parte è del tutto analoga. \square

Osservazione 5.12.2. Osserviamo che se vogliamo due variabili aleatorie X_1 e X_2 indipendenti e con densità di legge f_1 e f_2 allora, per i risultati visti prima, dobbiamo prendere P con densità della legge $f_1(x)f_2(y)$.

Osservazione 5.12.3. Nelle ipotesi precedenti si ha che $f > 0$ se e solo se $f_1 > 0$ e $f_2 > 0$ (quasi certamente): allora $\{f > 0\}$ deve essere un rettangolo. Questa è una condizione necessaria all'indipendenza.

5.12.2 Somma di variabili aleatorie

Abbiamo ricordato parlando di convoluzione di funzioni misurabili che nel caso discreto la legge della somma di variabili aleatorie è data dalla convoluzione. In effetti ciò vale anche nel caso continuo, solo che parleremo di densità, come mostra il seguente:

Teorema 5.12.4. *Siano X e Y due variabili aleatorie indipendenti con densità rispettivamente $f_1(x)$ e $f_2(y)$. La somma $X + Y$ ha densità $g(x)$ data da*

$$g(x) = \int_{-\infty}^{\infty} f_1(x-y)f_2(y) dy = (f_1 * f_2)(x).$$

Dimostrazione. Utilizziamo ancora il teorema 5.12.2: sia dunque $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ un'applicazione misurabile limitata. Allora per $E = E(\varphi(X + Y))$ si ha

$$\begin{aligned} E &= \iint_{\mathbb{R}^2} \varphi(x+y)f_1(x)f_2(y) dx dy = \int_{-\infty}^{\infty} f_2(y) dy \int_{-\infty}^{\infty} \varphi(x+y)f_1(x) dx = \\ &= \int_{-\infty}^{\infty} f_2(y) dy \int_{-\infty}^{\infty} \varphi(t)f_1(t-y) dt = \int_{-\infty}^{\infty} \varphi(t) \left(\int_{-\infty}^{\infty} f_1(t-y)f_2(y) dy \right) dt, \end{aligned}$$

e ciò mostra la tesi. \square

5.12.3 Applicazione di diffeomorfismi

Ora vogliamo presentare un fatto molto utile negli esercizi e che coinvolge la funzione di ripartizione di una probabilità. Sia $X : \Omega \rightarrow I$ con I intervallo e X di densità f_X , e sia $\varphi : I \rightarrow J$ un diffeomorfismo crescente. Posto $Y = \varphi \circ X$ si ha

$$F_Y(t) = P(Y \leq t) = P(\varphi \circ X \leq t) = P(X \leq \varphi^{-1}(t)) = F_X(\varphi^{-1}(t)),$$

e ciò è vero solo supponendo φ invertibile crescente. Avendo anche densità, ed essendo φ differenziabile con inversa differenziabile, abbiamo

$$\rho_Y(t) = F'_Y(t) = F'_X(\varphi^{-1})(\varphi^{-1})'(t) = \rho_X(\varphi^{-1}(t))(\varphi^{-1})'(t).$$

Questa regola per ricavare la densità della variabile immagine $Y = \varphi \circ X$ si può estendere anche al caso bidimensionale e il teorema relativo è il seguente:

Teorema 5.12.5. *Sia $X : \Omega \rightarrow A$ con $A \subseteq \mathbb{R}^2$ aperto X di densità ρ_X , e sia $\varphi : A \rightarrow B$ un diffeomorfismo, ψ il suo inverso e B un aperto. Posto $Y = \varphi \circ X$ si ha*

$$\rho_Y(x, y) = \begin{cases} \rho_X(\psi(x, y))|J_\psi(x, y)| & \text{se } (x, y) \in B \\ 0 & \text{se } (x, y) \notin B \end{cases}.$$

Dimostrazione. La dimostrazione consiste in un semplice cambio di variabili nell'integrale che dà la densità di Y . \square

5.13 Densità continue notevoli

Adesso, come nel caso delle variabili aleatorie discrete, vedremo alcuni esempi importanti di densità di probabilità su \mathbb{R} .

5.13.1 Densità uniforme

La distribuzione continua uniforme è una distribuzione di probabilità continua che è uniforme su un insieme misurabile, ovvero che attribuisce la stessa probabilità a due intervalli della stessa lunghezza contenuti nell'insieme. In altre parole se $[a, b]$ è un intervallo, basta definire la *densità uniforme* come

$$\rho(x) = \chi_{[a,b]}(x) \frac{1}{b-a}.$$

Adesso possiamo calcolare speranza e varianza di questa variabile; un rapidissimo calcolo che non stiamo a riportare ci dice che:

$$E(X) = \frac{b+a}{2} \quad \text{e} \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

5.13.2 Legge esponenziale

Sia λ un numero reale positivo. La *densità esponenziale* $\mathcal{E}(\lambda)$ è data dall'equazione:

$$\mathcal{E}(\lambda)(x) = \chi_{[0, \infty)} \cdot \lambda e^{-\lambda x}.$$

Anche in questo caso apprestiamoci a calcolare speranza e varianza di questa variabile aleatoria. Intanto

$$\begin{aligned} E(X) &= \int_0^\infty P(X \geq t) dt = \int_0^\infty \int_t^\infty \mathcal{E}_\lambda(x) dx dt = \int_0^\infty \int_t^\infty \lambda \cdot e^{-\lambda x} dx dt = \\ &= \int_0^\infty e^{-\lambda t} dt = -\frac{1}{\lambda} e^{-\lambda t} \Big|_0^\infty = \frac{1}{\lambda}. \end{aligned}$$

Analogamente si può calcolare che $Var(X) = \frac{1}{\lambda^2}$. Come nel caso discreto, possiamo anche dare una sorta di proprietà di assenza di usura:

Proposizione 5.13.1. *Sia $X : \Omega \rightarrow \mathbb{R}$ una variabile aleatoria. Dimostrare che le seguenti sono equivalenti:*

- (1) *la legge di X è esponenziale di parametro λ_1 ;*
- (2) *$P(X - t \geq x \mid X \geq t) = P(X \geq x)$ per ogni $t \geq 0$ e per ogni $x \geq 0$;*
- (3) *$Q_t(\cdot) = P(\cdot \mid X \geq t)$, esiste λ_3 tale che $E_{Q_t}(X - t) = \lambda_3$.*

Dimostrazione. ((1) \implies (2)) Si ha per ipotesi $P(X \geq x) = e^{-\lambda_1 x}$, e dunque basta scrivere le definizioni:

$$P(X - t \geq x \mid X \geq t) = \frac{P(X - t \geq x)}{P(X \geq t)} = \frac{e^{-\lambda_1(x+t)}}{e^{-\lambda_1 t}} = e^{-\lambda_1 x} = P(X \geq x),$$

e la prima parte è provata.

((2) \implies (3)) Vale $Q_t(X - t) = \int_0^\infty Q_t(X - t \geq x) dx = E_P(X)$, e non dipende da t . Chiamata $E_P(X) = \lambda_3$ abbiamo la tesi.

((3) \implies (1)) Abbiamo una sola ipotesi, quindi esprimiamo la costante λ_3 come speranza:

$$\lambda_3 = \int_0^\infty P(X \geq t + x \mid X \geq t) dx = \frac{\int_0^\infty P(X \geq t + x) dx}{P(X \geq t)}.$$

Sia $G(s) = P(X \geq s)$, quindi

$$\lambda_3 G(t) = \int_0^\infty G(t + x) dx = \int_0^\infty G(x) dx.$$

L'equazione integrale permette di concludere informazioni molto importanti sulla derivabilità della funzione $G(t)$; quello che diciamo ora è un procedimento standard: G è continua decrescente e quindi è integrabile secondo Riemann in $[t, \infty)$,

dunque l'integrale a secondo membro definisce una funzione continua, e quindi $G(t)$ è continua, e analogamente si va avanti indefinitamente concludendo che $G(t)$, se deve risolvere l'equazione integrale, deve essere di classe C^∞ . Derivando si ha dunque

$$\lambda_3 G'(t) = -G(t).$$

Ma allora $G(t) = c \cdot e^{-\frac{1}{\lambda_3}t}$ possiamo ricavare che $c = 1$ solo osservando che $G(0) = 1$. A questo punto basta prendere $F(t) = 1 - G(t)$ e osservare che è la funzione di ripartizione che definisce una legge esponenziale. \square

5.13.3 Legge gamma

Definiamo la *funzione gamma*, denotata con $\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}$, come

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Osservazione 5.13.1. Artin ha anche dimostrato che la funzione gamma è l'unica funzione da \mathbb{R}^+ a valori reali tale che

$$\Gamma(1) = 1, \quad \Gamma(x+1) = x\Gamma(x) \quad \text{e} \quad \log \Gamma(x) \text{ è convesso.}$$

Le prime due proprietà non bastano perché basterebbe prendere una funzione periodica di periodo unitario e tale che $\Gamma(1) = 1$. Si noti che se $n \in \mathbb{N}$ allora la seconda proprietà ci dice che $\Gamma(n+1) = n!$.

Quella precedente era solo una definizione di una funzione che però interviene in problemi di probabilità perché con essa possiamo definire la cosiddetta *densità gamma*. Questa è

$$\gamma_x(t) = \frac{t^{x-1} e^{-t}}{\Gamma(x)} \chi_{\mathbb{R}^+}(t).$$

Osservazione 5.13.2. Potremmo anche dare una definizione lievemente più generale della densità gamma, che dipenda anche da un parametro reale $\lambda > 0$. La densità

$$\gamma_{x,\lambda}(t) = \frac{\lambda^x t^{x-1} e^{-\lambda t}}{\Gamma(x)} \chi_{\mathbb{R}^+}(t).$$

Si può verificare che in effetti γ_x (così come $\gamma_{x,\lambda}$) definisce una densità, nel senso che il suo integrale è 1. Si noti che $\Gamma(1, \lambda) = \mathcal{E}(\lambda)$.

Se X ha per densità $\gamma_{x,\lambda}$ allora

$$E(X) = \frac{r}{\lambda} \quad \text{e} \quad \text{Var}(X) = \frac{r}{\lambda^2},$$

come un semplice calcolo dimostra.

5.13.4 Legge normale (o gaussiana)

La variabile aleatoria normale è una variabile aleatoria continua con due parametri, indicata tradizionalmente con $N(m, \sigma^2)$. Si tratta di una delle più importanti variabili casuali, soprattutto continue, in quanto è la base di partenza per altre variabili aleatorie o comunque un metodo per approssimarle in certe situazioni limite.

Siano dunque $m \in \mathbb{R}$ e $\sigma > 0$, definiamo la *densità gaussiana* come

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Il grafico di questa funzione, detta anche *gaussiana*, ha la classica e nota forma a campana. Il picco di massimo del grafico si ha in $x = m$ e la larghezza della “campana” è determinata da σ . Un veloce calcolo mostra che

$$E(X) = m \quad \text{e} \quad \text{Var}(X) = \sigma^2.$$

Capitolo 6

Statistica

Introduciamo le idee fondamentali dell'inferenza statistica con un esempio, che d'ora innanzi chiameremo "controllo di qualità": è probabilmente il più semplice che si possa immaginare, ma sufficiente per presentare le idee fondamentali. Vogliamo controllare la percentuale sconosciuta di pezzi difettosi in un insieme (ad esempio un grosso acquisto di certi componenti elettronici dall'estero), insieme che in statistica è usualmente denominato *popolazione*: per fare questo (non potendo verificare tutti i pezzi, per mancanza di tempo o altri motivi) estraiamo un campione di n pezzi che vengono verificati. I risultati di questa verifica saranno n variabili aleatorie X_1, \dots, X_n indipendenti, con legge di Bernoulli di parametro ϑ , con $0 < \vartheta < 1$ (la variabile X_i prende il valore 1 se l' i -esimo pezzo risulta difettoso, altrimenti prende il valore 0). Possiamo formalizzare la situazione in questo modo. Consideriamo sullo spazio $\Omega = \{0, 1\}^n$ (munito della σ -algebra di tutte le parti) la famiglia di probabilità $\{P^\vartheta \mid \vartheta \in (0, 1)\}$ definite da

$$P^\vartheta(k_1, \dots, k_n) = \vartheta^{k_1 + \dots + k_n} (1 - \vartheta)^{n - (k_1 + \dots + k_n)}.$$

Definiamo poi $X_i(k_1, \dots, k_n) = k_i$, ossia la i -esima proiezione canonica. È immediato verificare che, se si considera su Ω la probabilità P^ϑ , le variabili X_i risultano indipendenti e con legge di Bernoulli di parametro ϑ .

6.1 Introduzione all'inferenza statistica

Sia (Ω, \mathcal{A}) uno spazio di misura numerabile e Θ è un insieme di parametri, usualmente (ma non necessariamente) intervalli reali; inoltre sia $\{P^\vartheta\}_{\vartheta \in \Theta}$ una famiglia di misure di probabilità su Ω . Consideriamo ora lo spazio

$$\left(\Omega, \mathcal{A}, \{P^\vartheta\}_{\vartheta \in \Theta} \right),$$

che chiameremo *modello statistico*.

Osservazione 6.1.1. Osserviamo che nel caso in cui Ω sia numerabile le probabilità su Ω sono determinate univocamente dalle loro densità discrete.

Scopo dell'inferenza statistica è partire dall'esperienza (l'osservazione del campione) per risalire a informazioni sulla legge di probabilità che meglio si adatta a descrivere il modello, e per ottenere questo i metodi dell'inferenza statistica sono essenzialmente tre:

- (1) la stima statistica;
- (2) gli intervalli di fiducia;
- (3) i test statistici.

Noi ci occuperemo solo del secondo metodo, me lo vedremo in dettagli fra poco, cerchiamo ora di introdurre questi concetti a livello intuitivo, sempre riferendoci all'esempio del controllo di qualità. Indichiamo con $\bar{X}(\omega)$ la media aritmetica (chiamata anche media empirica) delle n variabili X_i (percentuale di pezzi difettosi riscontrati nell'indagine statistica), ed è importante ribadire che si tratta di una variabile aleatoria, cioè il risultato di questa indagine statistica dipende dal caso. Non avendo per il momento risultati teorici più precisi, sembra opportuno considerare proprio $\bar{X}(\omega)$ come stima del parametro ϑ . Quanto all'*intervallo di fiducia*, appare evidente che una maggiore ampiezza del campione permettere di rafforzare l'attendibilità dell'informazione: per spiegarci meglio, 2 pezzi difettosi su 10 oppure 200 su 1000 portano alla stessa stima (in entrambi i casi ϑ viene stimato 0,2), ma è evidente che il secondo risultato è molto più rassicurante. Come si può misurare questa sicurezza? È interessante osservare che nella vita pratica si incontrano più volte gli intervalli di fiducia, senza rendersene conto, ad esempio quando vengono trasmesse le proiezioni sui risultati delle elezioni. Le prime proiezioni danno per il partito una percentuale con un'oscillazione ad esempio di 2 punti percentuali (in più o in meno), dopo due ore la percentuale è cambiata (magari di poco) ma l'oscillazione è stata ridotta a 0,5 punti, e così via.

Effettuare un *test statistico* significa invece formulare un'ipotesi e pianificare un'esperienza per decidere se accettare o rifiutare l'ipotesi: ad esempio nel caso del controllo di qualità l'ipotesi potrebbe essere "la ditta fornitrice garantisce che la percentuale di pezzi difettosi non supera il 5%" (cioè $\vartheta \leq 0,05$). È evidente che l'ipotesi viene accettata se si osserva $\bar{X}(\omega) = 0,036$ e rifiutata se $\bar{X}(\omega) = 0,09$; ma che fare se $\bar{X}(\omega) = 0,049$ oppure $0,052$?

6.2 Intervalli di fiducia

Scegliamo convenzionalmente un numero positivo α positivo con $0 < \alpha < 1$; i valori tipici sono 0,1, 0,05 e 0,01. Supponiamo che per ogni ϑ esista un evento A_ϑ con $P^\vartheta(A_\vartheta) \geq 1 - \alpha$. Possiamo "assemblare" gli A_ϑ in come sezioni di un certo

$A \subseteq \Theta \times \Omega$, che precisamente sarà

$$A = \{(\vartheta, \omega) \in \Theta \times \Omega \mid \omega \in A_\vartheta\} = \bigcup_{\vartheta \in \Theta} \{\vartheta\} \times A_\vartheta.$$

Adesso però possiamo considerare le sezioni orizzontali di A , che saranno dunque una famiglia $\{C_\omega\}_{\omega \in \Omega}$ di sottoinsiemi di Θ tali che

$$A = \bigcup_{\omega \in \Omega} C_\omega \times \{\omega\}.$$

Definizione 6.2.1. Gli insiemi C_ω si chiamano *regioni di fiducia* per ϑ al livello $1 - \alpha$ e sono caratterizzate dalla proprietà

$$P^\vartheta(\vartheta \in C_\omega) \geq 1 - \alpha \quad \text{per ogni } \vartheta \in \Theta.$$

Adesso vogliamo riprendere quanto dicevamo prima della definizione, che è un metodo *pratico* per individuare una regione di fiducia. Fissato un parametro ϑ si determina un evento $A_\vartheta \subseteq \Omega$ tale che si abbia

$$P^\vartheta(A_\vartheta) \geq 1 - \alpha,$$

e che sia il più piccolo possibile. Si pone poi $C_\omega = \{\vartheta \mid \omega \in A_\vartheta\}$; osservando che $\vartheta \in C_\omega$ se e solo se $\omega \in A_\vartheta$ si ha che $A_\vartheta = \{\omega \mid \vartheta \in C_\omega\}$. Tutto ciò sarà chiaro nei prossimi esempi.

6.2.1 Utilizzo della legge dei grandi numeri

Vediamo come costruire regioni di fiducia (intervalli) usando la legge dei grandi numeri. Consideriamo $\Theta = [0, 1]$ e siano $\{X_i\}_{i=1}^\infty$ una successione di variabili aleatorie indipendenti con legge di Bernoulli di parametro ϑ . Detto

$$\overline{X}_n = \frac{1}{n} S_n = \frac{1}{n} (X_1 + \dots + X_n)$$

sappiamo che per ogni $\vartheta \in [0, 1]$

$$P(|\overline{X}_n - \vartheta| \geq \varepsilon) \leq \frac{\vartheta(1 - \vartheta)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n}$$

grazie alla disuguaglianza di Markov. Posto $\alpha = \frac{1}{4\varepsilon^2 n}$ si ha $\varepsilon = \frac{1}{2\sqrt{n\alpha}}$ si ottiene

$$P(|\overline{X}_n - \vartheta| < \varepsilon) \geq 1 - \alpha$$

e dunque gli insiemi

$$C_\omega = \left(\overline{X}_n(\omega) - \frac{1}{2\sqrt{n\alpha}}, \overline{X}_n(\omega) + \frac{1}{2\sqrt{n\alpha}} \right)$$

sono intervalli di fiducia (con la scelta di α) e per n grande si restringono sempre più.

Esempio 6.2.1. Valutare l'ipotesi che una moneta sia equilibrata (ossia l'esito di un lancio è testa oppure croce con probabilità uguali) sapendo che $\bar{X}_n = \frac{65}{100}$, cioè sono uscite 65 teste contro 35 croci.

Sia ora $\alpha = \frac{1}{10}$: vogliamo calcolare qual è l'intervallo di fiducia corrispondente. Per quanto visto prima sarà

$$I = \left[0,65 - \frac{1}{2\sqrt{10}}, 0,65 + \frac{1}{2\sqrt{10}} \right].$$

Visto che $\frac{1}{2} \in I$ possiamo dire che siamo nell'intervallo di fiducia, ossia si crede che la moneta possa essere equilibrata.

6.2.2 Utilizzo del teorema di de Moivre–Laplace

L'intervallo di fiducia che abbiamo determinato sopra in realtà non è molto buono (cioè non è molto stretto) perché è basato sulla disuguaglianza di Chebyshev, che in genere fa perdere qualcosa rispetto ai calcoli precisi; per n grande si ha una stima migliore utilizzando il teorema di de Moivre–Laplace.

Prima però premettiamo una definizione:

Definizione 6.2.2. La *funzione di errore*, denotata con $\Phi(x)$, è la funzione

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Per $0 < \alpha < 1$ si chiama *quantile* di α , e lo si indica con q_α , quel numero tale che $\Phi(q_\alpha) = \alpha$.

Osservazione 6.2.1. Esistono delle tavole statistiche che servono principalmente ai seguenti due scopi: dato x determinare il valore $\Phi(x)$, ossia il valore in x della funzione di ripartizione di una variabile aleatoria con densità gaussiana $N(0, 1)$; oppure dato α determinare q_α .

Nella stessa situazione sappiamo che per ogni x

$$P(n\vartheta - x\sqrt{n\vartheta(1-\vartheta)} \leq S_n \leq n\vartheta + x\sqrt{n\vartheta(1-\vartheta)}) = \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{t^2}{2}} dt + o(1)$$

per $n \rightarrow \infty$. Scriveremo quindi

$$P\left(\sqrt{n} \frac{|\bar{X}_n - \vartheta|}{\sqrt{\vartheta(1-\vartheta)}} \leq x\right) \approx \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-\frac{t^2}{2}} dt = \Phi(x) - \Phi(-x),$$

da cui

$$P\left(\sqrt{n}\frac{|\bar{X}_n - \vartheta|}{\sqrt{\vartheta(1-\vartheta)}} > x\right) \approx 1 - \Phi(x) + \Phi(-x) = \Phi(-x) + \Phi(-x) = 2\Phi(-x).$$

Posto come sopra $2\Phi(-x) = \alpha$ si ha $x = -q_{\frac{\alpha}{2}} = q_{1-\frac{\alpha}{2}}$, e quindi

$$P\left(\sqrt{n}\frac{|\bar{X}_n - \vartheta|}{\sqrt{\vartheta(1-\vartheta)}} > q_{1-\frac{\alpha}{2}}\right) \approx \alpha.$$

Visto che il massimo di $\sqrt{\vartheta(1-\vartheta)}$ è $\frac{1}{2}$ possiamo dire che la probabilità precedente è maggiorata dalla probabilità su un intervallo che non dipende da ϑ : tale intervallo è l'intervallo di fiducia, e in questo caso risulta

$$\left[\bar{X}_n - \frac{q_{1-\frac{\alpha}{2}}}{2\sqrt{n}}, \bar{X}_n + \frac{q_{1-\frac{\alpha}{2}}}{2\sqrt{n}}\right].$$

6.3 Cenni sulla teoria della stima

La funzione

$$\begin{aligned} \Theta \times \Omega &\longrightarrow [0, 1] \\ (\vartheta, \omega) &\longmapsto P^\vartheta(\omega) \end{aligned}$$

è detta *funzione di verosimiglianza*. Più in generale la funzione di verosimiglianza è definita se le P^ϑ hanno tutte densità rispetto a una fissata misura su Ω .

Definizione 6.3.1. Un *campionamento* (di numerosità n) per un modello statistico è una n -upla (X_1, \dots, X_n) di variabili aleatorie reali indipendenti, e con uguale legge rispetto ad ogni $\vartheta \in \Theta$.

Definizione 6.3.2. Lo *stimatore* di un modello statistico è una coppia di funzioni misurabili $U : \Omega \rightarrow \mathbb{R}$ e $g : \Theta \rightarrow \mathbb{R}$. Diremo che lo stimatore è *corretto* se e solo se per ogni $\vartheta \in \Theta$ vale

$$E^\vartheta(U) = g(\vartheta).$$

Esempio 6.3.1. Consideriamo X_1, \dots, X_n , campione su $\Omega \rightarrow \{0, 1\}$ con legge di Bernoulli di parametro ϑ . Detta

$$U_n = \frac{1}{n} \sum_{i=1}^n X_i$$

abbiamo che U_n è corretto da $g = id_\Theta$, visto che $E^\vartheta(U_n) = \vartheta$ per ogni ϑ .

Definizione 6.3.3. Una successione (U_n, g) di stimatori si dice *consistente* se e solo se

$$\forall \vartheta \in \Theta, \forall \varepsilon > 0 \quad \text{tale che} \quad P(|U_n - g(\vartheta)| \geq \varepsilon) = o(1)$$

per $n \rightarrow \infty$.

Esempio 6.3.2. Nell'esempio precedente gli $(U_n, g)_{n \geq 1}$ sono una successione consistente.

Appendice A

Esercizi risolti

In quest'ultima parte presentiamo degli esercizi su tutti gli argomenti svolti nel testo, suddividendoli in sezioni molto simili a quelle della trattazione stessa. Proporremo sia esercizi applicativi e quindi con uno scarso livello di astrazione, benché presuppongano le conoscenze teoriche, ed altri invece più concettuali, specialmente sulla prima parte più astratta.

A.1 Algebre e spazi di misura

Definizione A.1.1. Sia $x \in \Omega$ e $A \in \mathcal{A}$ elemento di un'algebra su Ω . Definiamo *delta di Dirac* la funzione $\delta_x : \mathcal{A} \rightarrow \{0, 1\}$ tale che $\delta_x(A) = 1$ se e solo se $x \in A$.

Esercizio A.1.1. Mostrare che la terna $(\Omega, \mathcal{P}(\Omega), \delta_x)$ è uno spazio di probabilità. *Soluzione.* Ovviamente $\mathcal{P}(\Omega)$ è un'algebra su Ω . Dobbiamo quindi solo verificare che δ_x è una misura su tale spazio. Intanto vale che

$$\delta_x(\emptyset) = 0 \quad \text{e} \quad \delta_x(\Omega) = 1.$$

Inoltre se $A, B \in \mathcal{P}(\Omega)$ e $A \cap B = \emptyset$ allora $\delta_x(A \cup B) = \delta_x(A) + \delta_x(B)$, come segue subito dalla definizione. Adesso passiamo alle unioni numerabili: siano $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{P}(\Omega)$ disgiunti allora

$$\delta_x \left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \delta_x(A_i),$$

e dunque abbiamo concluso.

Esercizio A.1.2. Sia $\Omega = \mathbb{N}$ e $\mathcal{A} = \{A \subseteq \mathbb{N} \mid A \text{ o } A^C \text{ è finito}\}$. Definiamo una funzione $\mu : \mathcal{A} \rightarrow [0, 1]$ tale che

$$\mu(A) = \begin{cases} 0 & \text{se } A \text{ è finito} \\ 1 & \text{se } A^C \text{ è finito} \end{cases}.$$

La famiglia \mathcal{A} è un'algebra? Di che tipo di additività è dotata la funzione μ ?

Soluzione. In effetti \mathcal{A} è un'algebra. Infatti $\emptyset \in \mathcal{A}$ e la famiglia è ovviamente chiusa per complementazione. Siano $A, B \in \mathcal{A}$ mostriamo che $A \cap B \in \mathcal{A}$. Se A è finito o B è finito allora la loro intersezione è finita; se entrambi sono infiniti si ha che $(A \cap B)^C = (A^C \cup B^C)$ e quindi ha complementare finito. Questo mostra che la famiglia \mathcal{A} è un'algebra. Siano $A, B \in \mathcal{A}$ disgiunti, mostriamo che μ è additiva. Se A e B sono finiti segue che sia la misura dell'unione che la somma delle misure sono entrambi 0, e quindi in particolare coincidono. Supponiamo ora che A sia finito e che B^C sia finito, allora si ha che le misure fanno entrambe 1. Se sia A^C che B^C sono finiti vediamo che tale caso non può darsi: infatti si può scrivere $(A \cap B)^C$, che è finito, e dunque $(A \cap B)^C \neq \Omega$ ma allora $A \cap B \neq \emptyset$ e non erano dunque insiemi disgiunti.

Ora mostriamo che \mathcal{A} non è una σ -algebra. Infatti fissato $n \in \mathbb{N}$ l'insieme $\{2n\} \in \mathcal{A}$ ma $\{2n \mid n \in \mathbb{N}\} \notin \mathcal{A}$. Inoltre non vale neanche la σ -additività di μ in quanto

$$1 = \mu(\mathbb{N}) \neq \sum_n \mu(\{n\}) = 0.$$

Esercizio A.1.3. Sia $\Omega = \mathbb{R}$ e $\mathcal{A} = \{A \subseteq \mathbb{N} \mid A \text{ o } A^C \text{ è finito o numerabile}\}$. La famiglia \mathcal{A} è un'algebra? Di che tipo di additività è dotata la funzione μ ?

Soluzione. Analogamente a prima si dimostra che \mathcal{A} è una σ -algebra. L'unico punto diverso è mostrare la chiusura per unione numerabile. Sia $\{A_n\} \subseteq \mathcal{A}$, mostriamo che $\bigcup_n A_n \in \mathcal{A}$. Se per ogni n si ha che A_n è finito o numerabile allora la loro unione è finita o numerabile e quindi sta in \mathcal{A} . Se esiste un \bar{n} tale che $A_{\bar{n}}^C$ è finito o numerabile allora

$$\left(\bigcup_n A_n \right)^C = \bigcap_n A_n^C \subseteq A_{\bar{n}}^C,$$

e dunque il complementare di $\bigcup_n A_n$ è finito o numerabile. Se per ogni n vale che A_n^C è finito o numerabile si ha che

$$\bigcap_n A_n^C \text{ è finita o numerabile} \implies \left(\bigcup_n A_n \right)^C \in \mathcal{A} \implies \bigcup_n A_n \in \mathcal{A}$$

ed abbiamo concluso.

Esercizio A.1.4. Sia $\Omega = \mathbb{R}$ e $D = \{(a, b) \mid a < b\}$. Mostrare che

$$\mathcal{A}(D) = \left\{ \bigcup_{i=1}^n I_i \mid I_i \in D \right\}.$$

Soluzione. Basta ricordare che

$$\mathcal{A}(D) = \left\{ \bigcup_{i=1}^n \bigcap_{j=1}^m A_{ij} \mid A_{ij} \in D \vee A_{ij}^C \in D \right\}$$

e osservare che l'intersezione e il complementare di due elementi di D sono ancora elementi di D o loro unioni.

Esercizio A.1.5. Sia $\Omega = \mathbb{R}$ e $D = \{I \mid I \text{ intervalli}\}$. Mostrare che

$$\mathcal{A}(D) = \left\{ \bigcup_{i=1}^n I_i \mid I_i \in D \right\}.$$

Soluzione. Analogo al precedente.

A.2 Combinatoria

Ricordiamo qualche risultato elementare di combinatoria. Siano A e B insiemi finiti con $|A| = k$ e $|B| = n$, allora

- (1) $|\{f : A \rightarrow B\}| = n^k$;
- (2) se $k < n$ si ha $|\{f : A \rightarrow B \mid f \text{ è iniettiva}\}| = \frac{n!}{(n-k)!}$;
- (3) $|\{C \subseteq B \mid |C| = k\}| = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ ¹;
- (4) consideriamo le funzioni $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ tali che per ogni i vale $f(i) < f(i+1)$. Allora questo è analogo al punto precedente perché dobbiamo scegliere dall'insieme di arrivo un suo sottoinsieme di k elementi. Quindi il numero di queste funzioni è ancora dato da $\binom{n}{k}$;
- (5) adesso consideriamo sempre le funzioni $f : \{1, \dots, k\} \rightarrow \{1, \dots, n\}$ tali che per ogni i valga stavolta $f(i) \leq f(i+1)$. In questo caso abbiamo che queste sono $\binom{n+k-1}{k}$ ².

Un concetto che torna molto utile in combinatoria è quello di multiinsieme:

Definizione A.2.1. Un *multiinsieme* M è definito come una coppia $M = (A, m)$, dove A è un insieme e m è una funzione a valori naturali positivi. Gli elementi di A vengono detti elementi del multiinsieme ed m la molteplicità del multiinsieme.

¹si può ottenere dalla precedente osservando che ci sono $k!$ funzioni distinte che mandano A in C .

²basta definire $g(i) = f(i) + i - 1$ e osservare che sono verificate le ipotesi del punto (4) con $n + k - 1$ al posto di n .

Si può dire che la funzione molteplicità associa ad ogni elemento del multiinsieme il numero di volte che viene ripetuto. Informalmente un multinsieme è una collezione in cui non conta l'ordine ma in cui ci possono essere ripetizioni. Adesso vogliamo contare il numero di multiinsiemi di cardinalità k che si possono costruire da un insieme di n elementi. Una deduzione errata sarebbe dire che questi sono $\frac{n^k}{k!}$: potremmo ragionare dicendo che mettiamo k volte n scelte e non contando l'ordine possiamo riarrangiare gli elementi in $k!$ modi. Ma ciò è falso perché ad esempio il multiinsieme $[2, 2, 3, 3]$ non può essere riordinato in $4!$ modi distinti, ma in meno. Adesso daremo la risposta a queste domanda.

Consideriamo di dover effettuare l'estrazione di k oggetti su un insieme di n oggetti (dove $n \geq k$). Le estrazioni possono essere con reinserimento dell'oggetto nell'urna da cui si estrae o no, e di un'estrazione possiamo o meno tener conto dell'ordine degli oggetti estratti. I quattro casi possibili sono rappresentati nella tabella seguente:

	<i>estrazioni ordinate</i>	<i>estrazioni non ordinate</i>
<i>senza reinserimento</i>	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$
<i>con reinserimento</i>	n^k	$\binom{n+k-1}{k}$

Le estrazioni ordinate vengono naturalmente rappresentate tramite un vettore; per quelle senza reinserimento ma non ordinate invece rappresentiamo con un insieme; infine le estrazioni non ordinate con reinserimento sono rappresentabili mediante un multiinsieme.

Osserviamo che gli stessi calcoli possono essere anche fatti in altri tipi di problemi, come ad esempio nel problema di disporre n oggetti in k contenitori distinti in modo che siano soddisfatte certe proprietà (ad esempio possibilità di ripetere eccetera).

Esercizio A.2.1 (terno al lotto). Calcolare la probabilità di un terno al lotto.

Soluzione. Sia Ω lo spazio degli eventi, nel nostro caso estrazioni del lotto. Avremo allora che

$$\Omega = \{\{a_1, a_2, a_3, a_4, a_5\} \mid a_i \in \{1, \dots, 90\} \wedge \text{senza ripetizioni}\}.$$

L'evento che ci interessa, ossia un terno fissato, supponiamo per esempio $1, 2, 3$, è $A = \{\{1, 2, 3, a_4, a_5\} \mid a_i \in \{1, \dots, 90\} \wedge \text{senza ripetizioni}\}$. La sua probabilità risulta dunque

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\binom{87}{2}}{\binom{90}{5}} \approx 0,000084,$$

ed abbiamo concluso.

Coefficiente multinomiale

Adesso quello che vogliamo fare è calcolare la potenza n -esima di un polinomio a r termini. Questo è un'estensione del caso di un binomio, in cui compare il coefficiente binomiale. Come dimostra la formula di Newton, nel caso $r = 2$ si ha

$$(x_1 + x_2)^n = \sum_{k=1}^n \binom{n}{k} x_1^k x_2^{n-k}.$$

Anche il caso generale ammette uno sviluppo in monomi con opportuni coefficienti: ebbene, tali coefficienti sono espressi dal cosiddetto coefficiente multinomiale. Vale il seguente risultato:

Teorema A.2.1. *Lo sviluppo della potenza n -esima di un polinomio di r termini è*

$$(x_1 + \dots + x_r)^n = \sum_{\substack{k_i \geq 0 \\ k_1 + \dots + k_r = n}} \frac{n!}{k_1! \cdot \dots \cdot k_r!} \cdot x_1^{k_1} \cdot \dots \cdot x_r^{k_r}.$$

Dimostrazione. Ogni addendo dello sviluppo è un prodotto i cui fattori sono n e vengono scelti fra x_1, \dots, x_r . Gli addendi sono tanti quante sono le disposizioni con ripetizione di r oggetti in n posti, cioè r^n . Possiamo raccogliere alcuni addendi, più precisamente si raccolgono fra loro tutti quelli che coincidono a meno dell'ordine dei fattori. Questo conduce alla tesi, poiché il fattore $x_1^{k_1} \cdot \dots \cdot x_r^{k_r}$ compare tante volte quante sono i modi di scegliere r sottoinsiemi A_i di $\{1, \dots, n\}$ in modo che $|A_i| = k_i$ e ovviamente $\sum_{i=1}^r |A_i| = n$. Il numero di queste scelte è ovviamente

$$\binom{n}{k_1} \binom{n-k_1}{k_2} \cdot \dots \cdot \binom{n-\sum_{i=1}^{r-2} k_i}{k_{r-1}} = \prod_{i=1}^{r-2} \frac{(n-\sum_{j=1}^i k_j)!}{k_{i+1}! (n-\sum_{j=1}^{i+1} k_j)!} = \frac{n!}{k_1! \cdot \dots \cdot k_r!},$$

e quindi abbiamo concluso. \square

A.3 Probabilità condizionata

Ricordiamo che due eventi A e B in \mathcal{A} algebra di uno spazio di probabilità (Ω, \mathcal{A}, P) si dicono *indipendenti* se $P(A \cap B) = P(A)P(B)$, il che corrisponde anche alla nostra intuizione. Si mostrano facilmente le seguenti proprietà:

- (1) se $P(A) = 0$ allora per ogni $B \in \mathcal{A}$ si ha che A e B sono indipendenti;
- (2) se $P(A) > 0$ allora A e B sono indipendenti se e solo se $P(B|A) = P(B)$;
- (3) A e B sono indipendenti se e solo se A e B^C lo sono.

Notiamo che la proprietà (2) esprime il fatto che due eventi sono indipendenti quando la probabilità di uno non dipende dal fatto che si sia già verificato l'altro. Abbiamo dato anche la definizione di n eventi indipendenti: $\{A_i\}_{i=1}^n \subseteq \mathcal{A}$ sono indipendenti se per ogni $I \subseteq \{1, \dots, n\}$ si ha $P(\bigcap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$.

Osservazione A.3.1. Questo fatto è equivalente a chiedere che per ogni $X_j \in \{A_j, A_j^C\}$ si ha

$$P\left(\bigcap_{j=1}^n X_j\right) = \prod_{j=1}^n P(X_j).$$

La dimostrazione dell'equivalenza è piuttosto semplice, quindi la lasciamo come esercizio.

Esercizio A.3.1. Un viaggiatore deve compiere un viaggio che si compone di tre tratte e vorrebbe, una volta consegnata la sua valigia all'inizio, riaverla alla fine. Consideriamo gli eventi

$A = \text{"la valigia si perde nella prima tratta"}$

e analogha definizione per B e C . Si suppongano i tre eventi indipendenti e sia $P(A) = 20\%$, $P(B) = 10\%$ e $P(C) = 25\%$.

- (1) Qual è la probabilità che la valigia arrivi a destinazione?
- (2) Nell'ipotesi in cui la valigia non arrivi a destinazione, qual è la probabilità che la valigia sia stata persa nella prima tratta? Analogo con la seconda e con la terza tratta.
- (3) Qual è la probabilità che la valigia sia stata persa nella prima tratta sapendo che si è persa? Analogo con la seconda e con la terza tratta.

Soluzione. (1) In effetti l'evento di cui bisogna calcolare la probabilità è $A^C \cap B^C \cap C^C$. Dunque, ricordando l'indipendenza di questi eventi:

$$P(A^C \cap B^C \cap C^C) = P(A^C) P(B^C) P(C^C) \approx 0,54.$$

(2) L'evento $A_1 = \text{"la valigia si è persa nella prima tratta"}$ ha probabilità $P(A_1) = P(A) = 20\%$. L'evento $A_2 = \text{"la valigia si è persa nella seconda tratta"}$ deve essere interpretato come "la valigia si è persa nella seconda ma non nella prima", e quindi $A_2 = B \cap A^C$. Da ciò

$$P(A_2) = P(B \cap A^C) = P(B)P(A^C) = P(B)(1 - P(A)) = 0,08.$$

Infine l'evento $A_3 = \text{"la valigia si è persa nella terza tratta"}$ può essere scritto analogamente come $A_3 = B \cap A^C \cap B^C$ e dunque $P(A_3) = P(B \cap A^C \cap B^C) = P(B)P(A^C)P(B^C) = 0,18$.

(3) La probabilità che la valigia si sia persa è $A \cup B \cup C$, quindi la probabilità da calcolare è

$$P(A|A \cup B \cup C) = \frac{P(A)}{P(A \cup B \cup C)} = \frac{P(A)}{1 - P(A^C \cap B^C \cap C^C)} \approx 0,435.$$

Analogamente $P(B|A \cup B \cup C) \approx 0,174$ e $P(C|A \cup B \cup C) = 0,391$.

Esercizio A.3.2. Si lanciano due dadi equilibrati e si considerano gli eventi A : “la somma dei due numeri usciti è 7” e B : “il numero uscito sul primo dado è 4”.

- (1) Dimostrare che i due eventi sono indipendenti;
 (2) più in generale, indicando rispettivamente con le variabili aleatorie X ed Y il numero uscito sul primo e sul secondo dado, si considerino gli eventi $\{X + Y = n\}$ e $\{X = k\}$. Dire per quali valori di n e k questi due eventi sono indipendenti.

Soluzione. (1) Prendiamo le notazioni con X e Y descritte nel punto (2). Allora possiamo scrivere $A = \{X + Y = 7\}$, e quindi

$$A = \bigcup_{i=1}^6 (\{X = i\} \cap \{Y = 7 - i\}).$$

Essendo gli eventi disgiunti e indipendenti si può scrivere

$$P(A) = \sum_{i=1}^6 P(\{X = i\} \cap \{Y = 7 - i\}) = \sum_{i=1}^6 P(X = i) \cdot P(Y = 7 - i) = \frac{6}{36} = \frac{1}{6}.$$

Chiaramente $P(B) = P(X = 4) = \frac{1}{6}$. Inoltre $A \cap B = \{X + Y = 7, X = 4\} = \{X = 4, Y = 3\}$ e dunque $P(A \cap B) = \frac{1}{36}$: quindi i due eventi sono indipendenti.

(2) In questo caso sicuramente $2 \leq n \leq 12$ e $1 \leq k \leq 6$. Possiamo scrivere

$$\{X + Y = n\} = \bigcup_{i=\max(1, n-6)}^{\min(6, n-1)} (\{X = i\} \cap \{Y = n - i\}),$$

dove le limitazioni per l'unione tengono di conto del fatto che se fissiamo $2 \leq n \leq 6$ non possono presentarsi tutte le possibili uscite sui due dadi, cosa che invece accade per $7 \leq n \leq 12$. Quindi $P(X + Y = n) = (\min(6, n - 1) - \max(1, n - 6) + 1) \cdot \frac{1}{36}$. Ovviamente, fissato k come detto, si ha $P(X = k) = \frac{1}{6}$ e

$$P(X + Y = n, Y = k) = \begin{cases} \frac{1}{36} & \text{se } n \geq 7 \text{ e } 1 \leq k \leq 6 \text{ oppure se } n < 7 \text{ e } k \leq n - 1 \\ 0 & \text{se } n < 7 \text{ e } k > n - 1 \end{cases}.$$

Non essendo nessuno dei due eventi di probabilità nulla chiaramente se $n < 7$ e $k > n - 1$ gli eventi non possono essere indipendenti; gli unici casi sono quelli tali per cui

$$(\min(6, n - 1) - \max(1, n - 6) + 1) \cdot \frac{1}{36} \cdot \frac{1}{6} = \frac{1}{36},$$

ossia $\min(6, n - 1) - \max(1, n - 6) = 5$. Una semplice verifica mostra che ciò è possibile se e solo se $n = 7$ e anche per ogni $k < n$ quindi³.

³Potevamo anche utilizzare un argomento di probabilità condizionata, ricordando che A e B sono indipendenti se e solo se $P(A|B) = P(A)$. Qui la probabilità condizionata $P(X + Y = n | Y = k)$ corrisponde alla probabilità di ottenere $X + Y = n$ noto che sia $Y = k$. Dunque (essendo X e Y variabili aleatorie indipendenti) essa vale $P(X = n - k) = \frac{1}{6}$ se $k < n$, e 0 altrimenti. Questa quindi coincide con $P(X + Y = n)$ se e solo se $k < n = 7$.

Esercizio A.3.3. Una compagnia di assicurazione vuole confrontare la probabilità di fare incidenti supponendo che il conducente dell'automobile sia in stato di ebbrezza. Siano A , B_1 e B_2 gli eventi "avviene l'incidente", "il conducente è ubriaco" e "il conducente non è ubriaco" rispettivamente. Da indagini risulta inoltre che $P(B_1) = 5\%$ e $P(B_1|A) = 30\%$. Calcolare il rapporto tra la probabilità che si faccia un incidente sapendo il conducente ubriaco con quella sapendo il conducente non ubriaco.

Soluzione. Vogliamo calcolare $\frac{P(A|B_1)}{P(A|B_2)}$. Per le due probabilità condizionate basta applicare la definizione:

$$\frac{P(A|B_1)}{P(A|B_2)} = \frac{P(A \cap B_1)P(B_2)}{P(A \cap B_2)P(B_1)} = \frac{P(B_1|A) P(B_2)}{P(B_2|A) P(B_1)},$$

dove nell'ultimo passaggio abbiamo diviso entrambi i termini della frazione per $P(A)$. Nella formula finale ottenuta conosciamo ogni termine, quindi possiamo calcolare il rapporto richiesto, che risulta essere circa 8,1.

Esercizio A.3.4. Abbiamo due sacchetti indicati con A e B : il primo contiene 3 palle rosse e 2 verdi, mentre il secondo ne contiene 2 rosse e 8 verdi. Per decidere da quale urna pescare si lancia una moneta e si pesca dalla prima urna se e solo se esce testa. Sappiamo che è stata estratta una pallina rossa; calcolare la probabilità che questa sia stata pescata da A .

Soluzione. Questa è una semplice applicazione della formula di Bayes. Sia A l'evento "la palla è stata pescata da A " e B l'analogo per l'altra urna. Essendo determinati dal lancio di una moneta gli eventi A e B hanno probabilità entrambi $\frac{1}{2}$. Indicato con R l'evento "la palla estratta è rossa" si ha

$$P(A|R) = \frac{P(R|A)P(A)}{P(R|A)P(A) + P(R|B)P(B)} = \frac{\frac{3}{5} \cdot \frac{1}{2}}{\frac{3}{5} \cdot \frac{1}{2} + \frac{1}{5} \cdot \frac{1}{2}} = \frac{3}{4},$$

ed abbiamo finito.

Esercizio A.3.5. In una gabbia di 100 conigli ve ne sono 10 malati. Qual è la probabilità che scegliendo a caso 4 conigli uno solo sia malato?

Soluzione. Consideriamo quattro scelte di conigli "senza reinserimento" e sia M_i l'evento "l' i -esimo coniglio è malato", per $i = 1, \dots, 4$. Sia invece S_i l'evento "tutti i conigli tranne l' i -esimo" non sono malati. Avremo allora che il nostro evento è $E = \bigcup_{i=1}^4 (M_i \cap S_i)$. Quindi $P(E) = 4 \cdot P(M_1 \cap S_1)$; adesso poniamo $S_1 = R_2 \cap R_3 \cap R_4$, dove R_j è l'evento "il j -esimo coniglio non è malato", per $j = 2, 3, 4$. Quindi

$$\begin{aligned} P(E) &= 4 \cdot P(M_1 \cap S_1) = 4 \cdot P(M_1 \cap R_2 \cap R_3 \cap R_4) \\ &= 4 \cdot P(M_1)P(R_2|M_1)P(R_3|M_1 \cap R_2)P(R_4|M_1 \cap R_2 \cap R_3) = \\ &= 4 \cdot \frac{10}{100} \cdot \frac{90}{99} \cdot \frac{89}{98} \cdot \frac{88}{97} \approx 0,30. \end{aligned}$$

Esercizio A.3.6. Un'urna contiene 3 palle rosse, 4 verdi e 5 blu. Si estraggono 3 palle senza reinserimento; determinare la probabilità degli eventi:

- (1) siano tutte di colore diverso;
- (2) ve ne siano 2 blu e 1 verde;

Soluzione. (1) Fissiamo gli eventi A , B e C rispettivamente “la prima palla estratta è rossa”, “la seconda palla estratta è verde” e “la terza palla estratta è blu”. Intanto vale ovviamente

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B) = \frac{3}{12} \cdot \frac{4}{11} \cdot \frac{5}{10} = \frac{1}{22}.$$

Ora, quella calcolata era la probabilità dell'evento cercato ma una volta fissato un particolare ordine. Visto che gli ordini possibili con cui scegliere le tre palle di colore diverso sono $3! = 6$ otteniamo che la probabilità cercata è $\frac{1}{22} \cdot 3! = \frac{3}{11}$. Una possibile obiezione che si potrebbe muovere al ragionamento sull'ordine è la seguente: visto che le probabilità sono condizionate una a tutte le precedenti non sembra uguale considerare ordini diversi. In realtà a livello di calcolo lo è, in quanto si hanno $3!$ casi in cui i calcoli sono prodotti delle medesime frazioni con i numeratori permutati.

(2) Adesso consideriamo gli eventi A e B rispettivamente “le prime due palle estratte sono blu” e “la terza estratta è rossa”. Si ha certamente

$$P(A \cap B) = P(A)P(B|A) = \frac{5}{12} \cdot \frac{4}{11} \cdot \frac{4}{10} = \frac{2}{33}.$$

I possibili modi di fissare l'uscita della palla rossa sono tre (e una volta fatto ciò le altre due sono fissate) quindi la probabilità cercata è $\frac{2}{11}$.

Esercizio A.3.7. Una Società di Assicurazioni ha rilevato che la probabilità che un guidatore maschio abbia almeno un incidente in un anno è α , indipendentemente da quello che succede negli altri anni; l'analogha probabilità per un guidatore femmina è β . Tra gli assicurati presso quella compagnia, il 55% sono maschi.

- (1) Qual è la probabilità che un assicurato abbia incidenti in due anni consecutivi?
- (2) Se un assicurato ha avuto incidenti in due anni consecutivi, qual è la probabilità che sia femmina?

Soluzione. Sia A l'evento “un maschio ha almeno un incidente in un anno”, sia B “una femmina ha almeno un incidente in un anno”, sia M “l'assicurato è maschio” e infine F “l'assicurato è femmina”. Dalle ipotesi si ha

$$P(A) = \alpha, \quad P(B) = \beta, \quad P(M) = \frac{11}{20} \quad \text{e} \quad P(F) = \frac{9}{20}.$$

(1) Sia E l'evento “un assicurato preso a caso ha avuto incidenti in due anni consecutivi”. Essendo $\{M, F\}$ un sistema di alternative si ha

$$P(E) = P(E|M)P(M) + P(E|F)P(F) = \frac{11}{20}\alpha^2 + \frac{9}{20}\beta^2.$$

(2) Si vuole calcolare $P(F|E)$, quindi si può calcolare grazie alla formula di Bayes

$$P(F|E) = \frac{P(E|F)P(F)}{P(E|M)P(M) + P(E|F)P(F)} = \frac{\frac{9}{20}\beta^2}{\frac{11}{20}\alpha^2 + \frac{9}{20}\beta^2}.$$

Esercizio A.3.8. Da un mazzo di 52 carte se ne estraggono cinque. Qual è la probabilità di fare poker?

Soluzione. Estrarre cinque carte corrisponde a una successione di cinque estrazioni senza reinserimento. Su cinque carte avremo che quattro concorreranno a formare il poker e la rimanente può essere qualsiasi. Fissiamo che la carta rimanente sia l'ultima, fissiamo che il poker debba essere di assi e fissiamo che i semi siano ordinati come cuori, quadri, picche e fiori. Allora la probabilità di tale evento E è

$$P(E) = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49}.$$

A questo punto, non contando la posizione della carta rimanente, che può assumere cinque diverse, non contando quale carta viene ripetuta, e queste sono 13, e non contando l'ordine, che può assumere quattro coinfigurazioni, si ha che la probabilità cercata è

$$5 \cdot 13 \cdot 4 \cdot P(E) = 624 \cdot \frac{5! \cdot 47!}{52!},$$

ed abbiamo concluso.

A.4 Variabili aleatorie discrete

Esercizio A.4.1. Un dado equilibrato viene lanciato n volte. Si determini la probabilità che siano usciti esattamente due 6. Per quale n tale probabilità è massima?

Soluzione. Il processo può essere descritto da una variabile binomiale in quanto vogliamo 2 successi su n lanci, ciascuno dei quali ha probabilità di successo $p = \frac{1}{6}$. La probabilità cercata è dunque

$$p_n(2) = \binom{n}{2} p^2 (1-p)^{n-2} = \frac{n(n-1)}{50} \frac{5^n}{6^n}.$$

Svolgendo i calcoli si trova che $f'(\xi) = 0$ in un punto $\xi \in (11, 12)$. Inoltre $p_{11}(2) = p_{12}(2) \approx 0,249$ e quindi il massimo è assunto sia in $n = 11$ che in $n = 12$ ⁴.

⁴si poteva osservare a livello di successione che $p_{n+1}(2) \geq p_n(2)$ se e solo se $n \leq 11$.

Esercizio A.4.2. Quante volte almeno si deve lanciare un dado perché la probabilità che esca almeno un 6 sia almeno 0,99.

Soluzione. Chiamiamo n il numero di lanci cercato: se B_n è l'evento di cui vogliamo la probabilità, ossia "è uscito almeno un 6 nei primi n lanci, è più semplice osservare che B_n^C è l'evento "non è uscito nessun 6 nei primi n lanci. Dunque

$$P(B_n^C) = \left(\frac{5}{6}\right)^n.$$

Bastava osservare che il fenomeno si poteva descrivere con una variabile geometrica che conta i fallimenti prima del primo successo, e imporre che, detto T tale numero di fallimenti, si abbia $T > n$. Comq già calcolato nella sezione di teoria abbiamo che

$$P(T > n) = (1 - p)^n.$$

Imponendo che tale probabilità sia più grande di 0,99 si ottiene $n > 26$.

Esercizio A.4.3. Sono date $n+2$ variabili aleatorie indipendenti Y, Z, X_1, \dots, X_n , dove Y e Z hanno distribuzione uniforme su $\{0, \dots, n\}$ e X_1, \dots, X_n sono variabili aleatorie con legge di Poisson di parametro λ . Sia ora $V = X_Y$ e $W = X_Z$: con questo intendiamo che

$$V(\omega) = X(\omega)_{Y(\omega)} \quad \text{e} \quad W(\omega) = X(\omega)_{Z(\omega)}.$$

(1) Calcolare le densità delle leggi di V e W .

(2) V e W sono indipendenti?

Soluzione. (1) Dobbiamo calcolare $P(V = k)$; osservato che $\{\{Y = i\} \mid i = 1, \dots, n\}$ è una partizione dello spazio degli eventi si ha

$$\begin{aligned} P(V = k) &= \sum_{i=1}^n P(V = k \mid Y = i)P(Y = i) = \\ &= \sum_{i=1}^n P(X_i = k \mid Y = i)P(Y = i) = \\ &= \sum_{i=1}^n P(X_i = k)P(Y = i) = P(X_i = k) \sum_{i=1}^n P(Y = i) = \\ &= e^{-\lambda} \frac{\lambda^k}{k!} \cdot 1 = e^{-\lambda} \frac{\lambda^k}{k!}. \end{aligned}$$

Per W poi il calcolo è analogo.

(2) In effetti V e W non sono indipendenti. Infatti si ha $P(V = 0) = P(W = 0) = e^{-\lambda}$, ma invece $P(V = 0, W = 0) \neq e^{-\lambda}e^{-\lambda}$. Infatti come prima

$$\begin{aligned}
 P(V = 0, W = 0) &= \sum_{i=1}^n \sum_{j=1}^n P(V = 0, W = 0 \mid Y = i, Z = j)P(Y = i, Z = j) = \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n P(V = 0, W = 0 \mid Y = i, Z = j) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n P(X_i = 0, X_j = 0 \mid Y = i, Z = j) = \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n P(X_i = 0, X_j = 0) = \\
 &= \frac{1}{n^2} \left(\sum_{i=1}^n P(X_i = 0) + \sum_{i \neq j} P(X_i = 0, X_j = 0) \right) = \\
 &= \frac{1}{n^2} \left(ne^{-\lambda} + n(n-1)e^{-2\lambda} \right) = \frac{1}{n} \left(e^{-\lambda} + (n-1)e^{-2\lambda} \right).
 \end{aligned}$$

Poi si conclude osservando che in effetti sono diversi.

Esercizio A.4.4. Siano X_1 e X_2 variabili aleatorie indipendenti con distribuzione geometrica di parametri p_1 e p_2 in $(0, 1)$. Calcolare $P(X_1 = X_2)$ e $P(X_1 < X_2)$.

Soluzione. Per la prima richiesta basta calcolare

$$\begin{aligned}
 P(X_1 = X_2) &= \sum_{k=0}^{\infty} P(X_1 = k, X_2 = k) = \sum_{k=0}^{\infty} P(X_1 = k)P(X_2 = k) = \\
 &= p_1 p_2 \sum_{k=0}^{\infty} (1-p_1)^k (1-p_2)^k = \frac{p_1 p_2}{1 - (1-p_1)(1-p_2)}.
 \end{aligned}$$

La seconda usa sostanzialmente le stesse idee, infatti

$$\begin{aligned}
 P(X_1 < X_2) &= \sum_{\substack{k, h=0 \\ h < k}}^{\infty} P(X_1 = h, X_2 = k) = \sum_{\substack{k, h=0 \\ h < k}}^{\infty} p_1 (1-p_1)^h p_2 (1-p_2)^k = \\
 &= \sum_{h=0}^{\infty} p_1 p_2 (1-p_1)^h (1-p_2)^{h+1} \sum_{k=h+1}^{\infty} (1-p_2)^{k-h-1} = \\
 &= \sum_{h=0}^{\infty} p_1 p_2 (1-p_1)^h (1-p_2)^{h+1} \sum_{k=0}^{\infty} (1-p_2)^k = \frac{p_1 (1-p_2)}{1 - (1-p_1)(1-p_2)}.
 \end{aligned}$$

L'esercizio è concluso. Segnaliamo solamente come spunto che questo può essere generalizzato al caso di n variabili aleatorie.

Osservazione A.4.1. Siano $X_1 : \Omega \rightarrow C_1$ e $X_2 : \Omega \rightarrow C_2$ variabili aleatorie discrete, e poniamo $f_i(n) = P(X_i = n)$ per le densità delle due leggi. Il blocco $(X_1, X_2) : \Omega \rightarrow C_1 \times C_2$ è ancora una variabile aleatoria discreta e ha densità della legge $f(n, m) = P((X_1, X_2) = (n, m))$. Allora vale che

$$f_1(n) = \sum_{m \in C_2} f(n, m). \quad (\text{A.1})$$

Questo perché semplicemente $P(X_1 = n) = P(X_1 = n, X_2 \in C_2)$; analogamente si può scrivere per f_2 . È altrettanto chiaro che X_1 e X_2 sono indipendenti se e solo se $f(n, m) = f_1(n)f_2(m)$.

Vogliamo esemplificare quanto appena detto in astratto. Consideriamo come evento il lancio di due dadi, allora avremo $C_1 = C_2 = \{1, 2, \dots, 6\}$ e lo spazio prodotto sarà $C_1 \times C_2$; siano μ_1 e μ_2 le leggi marginali delle due variabili aleatorie. Se X_1 e X_2 sono indipendenti e μ_1 e μ_2 sono leggi uniformi allora anche la legge del blocco μ è una densità uniforme. Precisiamo che se togliamo la condizione di indipendenza ciò non è più vero.

Esercizio A.4.5. Consideriamo un'urna rossa contenente T_1 palline rosse e un'urna nera contenente T_2 palline nere, con $1 \leq T_1 \leq T_2$ e le palline sono numerate da 1 a T_i . Si estraggano r_1 palline dall'urna rossa e r_2 palline da quella nera, e sia R il numero di coppie estratte (ossia palline con lo stesso numero).

(1) Costruire un modello di probabilità.

(2) Calcolare le legge di R .

Soluzione. (1) Poniamo per $i = 1, 2$

$$C_i = \{\omega_i \mid \omega_i \subseteq \{1, \dots, T_i\}, |\omega_i| = r_i\}.$$

Poniamo $\Omega = C_1 \times C_2$ e sia $\omega = (\omega_1, \omega_2)$; le variabili aleatorie X_i sono le due proiezioni canoniche. Per quanto richiesto si ha

$$R(\omega) = |\omega_1 \cap \omega_2| = |X_1(\omega) \cap X_2(\omega)|.$$

Non essendo alcuna estrazione facilitata rispetto alle altre mettiamo sull'algebra della parti $\mathcal{P}(\Omega)$ una misura di probabilità uniforme.

(2) Vediamo due modi per calcolare la legge di R . Dalla relazione A.1 segue

$$P(R = j) = \sum_{\omega_1 \in C_1} P(R = j \mid X_1 = \omega_1)P(X_1 = \omega_1).$$

Visto che $P(X_1 = \omega_1) = \binom{T_1}{r_1}^{-1}$ possiamo scrivere

$$P(R = j) = \frac{1}{\binom{T_1}{r_1}} \sum_{\omega_1 \in C_1} P(R = j \mid X_1 = \omega_1) = \frac{1}{\binom{T_1}{r_1}} \sum_{\omega_1 \in C_1} H(T_2, r_1, r_2)(j) = H(T_2, r_1, r_2).$$

L'altro modo è quello di calcolare la densità direttamente:

$$P(R = j) = \frac{|\{R = j\}|}{\binom{T_1}{r_1} \binom{T_2}{r_2}}$$

Sappiamo che vi sono state j coppie: possiamo pensare che dall'urna rossa siano state estratte j palline rosse numerate che sono in corrispondenza con le j palline dall'urna nera che hanno gli stessi numeri. Essendoci le altre $r_1 - j$ estratte dalla prima urna, bisogna fare in modo che le $r_1 - j$ con lo stesso numero nell'urna nera non vengano estratte nelle r_2 estratte dall'urna nera. Quindi per le $r_2 - j$ palline nere ancora da scegliere nella seconda urna abbiamo le scelte limitate solo alle $T_2 - r_1$ palline rimaste. Quindi

$$|\{R = j\}| = \binom{T_1}{j} \binom{T_1 - j}{r_1 - j} \binom{T_2 - r_1}{r_2 - j},$$

e allora sostituendo si ottiene la tesi.

Esercizio A.4.6. Siano A , B e C eventi e definiamo $D = B - A$ e $E = C - B$. Supponiamo che (A, B) , (B, C) , (C, D) e (A, E) siano coppie di eventi indipendenti. Mostrare che A , B e C sono una terna di eventi indipendenti.

Soluzione. Abbiamo per ipotesi

$$P(A)P(B) = P(A \cap B), \quad P(B)P(C) = P(B \cap C)$$

$$P(C)P(B \cap A^C) = P(A^C \cap B \cap C) \quad \text{e} \quad P(A)P(C \cap B^C) = P(A \cap B^C \cap C).$$

Ricordiamo che, dati A_i per $i = 1, 2, 3$ eventi, questi sono indipendenti se e solo se per ogni $J \subseteq \{1, 2, 3\}$ vale

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

Visto che A e B^C sono indipendenti si ha

$$\begin{aligned} P(A \cap B \cap C) &= P(B \cap C) - P(A^C \cap B \cap C) = P(B)P(C) - P(C)P(B)P(A^C) = \\ &= P(B)P(C)(1 - P(A^C)) = P(A)P(B)P(C). \end{aligned}$$

Analogamente

$$\begin{aligned} P(A \cap C) &= P(A \cap C \cap B^C) + P(A \cap B \cap C) = P(A)P(C)P(B^C) + P(A)P(B)P(C) = \\ &= P(A)P(C)(P(B) + P(B^C)) = P(A)P(C). \end{aligned}$$

Esercizio A.4.7. Sia X una variabile aleatoria discreta. Se X è indipendente da X allora è quasi certamente costante.

Soluzione. Sia $p(n) = P(X = n)$ la densità della legge di X . Per l'ipotesi che X sia indipendente da se stessa possiamo scrivere che

$$p(n) = P(X = n, X = n) = P(X = n)P(X = n) = p^2(n),$$

e allora $p(n) = 0$ o $p(n) = 1$ per ogni n . Ma considerando che $\sum_n p(n) = 1$ deve esistere un \bar{n} tale che $p(\bar{n}) \neq 0$; questo però implica che $p(\bar{n}) = 1$ e dunque $p(n) = 0$ per ogni $n \neq \bar{n}$. Da ciò segue $P(X \neq \bar{n}) = 0$, che è la definizione di X quasi certamente costante.

Esercizio A.4.8. Siano $\{X_i\}_{i=1}^m$ variabili aleatorie geometriche di parametri $\{p_i\}_{i=1}^m$ e indipendenti. Studiare la legge di $M = \min\{X_1, \dots, X_m\}$.

Soluzione. Poniamo $q_i = 1 - p_i$ per ogni i . Vale

$$\{M \geq n\} = \bigcap_{i=1}^m \{X_i \geq n\}$$

e avendo supposto l'indipendenza si ottiene $P(M \geq n) = (\prod_{i=1}^m q_i)^n$. Con questa legge si ha che la variabile M è geometrica, infatti

$$P(M = n) = P(M \geq n) - P(M \geq n + 1) = \left(1 - \prod_{i=1}^m q_i\right)^n \prod_{i=1}^m q_i,$$

e l'esercizio è concluso

Esercizio A.4.9. Siano X, Y e Z variabili aleatorie discrete integrabili due volte e indipendenti. Valga anche $|X(\omega)| \leq |Y(\omega)|$ per ogni $\omega \in \Omega$ e $E(Z) = 0$. Definiti $R = XY$ e $S = YZ$:

(1) è vero che $R, S \in L^2$?

(2) dimostrare che $Var(R) \leq Var(S)$.

Soluzione. (1) La risposta è affermativa, e dimostrarlo non è difficile. Infatti

$$E(R^2) = E(X^2 Z^2) = E(X^2)E(Z^2),$$

e questa speranza è finita perché lo sono i fattori che la esprimono. In modo analogo si deve procedere per S .

(2) Sviluppando come abbiamo fatto sopra si ottiene

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

ed anche $Var(S) = E(Y^2)E(Z^2)$; si conclude con la maggiorazione data per ipotesi.

A.5 Variabili aleatorie reali

Esercizio A.5.1. Siano X, Y e Z variabili aleatorie indipendenti e uniformemente distribuite su $[0, 1]$. Calcolare la probabilità di $\lfloor X + Y + Z \rfloor = 1$.

Soluzione. Dire che una variabile aleatoria X è distribuita uniformemente su $[0, 1]$ significa che la sua legge è la misura di Lebesgue unidimensionale m_1 su $[0, 1]$, e ciò ancora equivale a dire che $P(X \in A) = m_1(A)$ per ogni boreliano $A \subseteq [0, 1]$. Ora poi ricordiamo che dire che la loro legge congiunta è la misura prodotto delle leggi, e quindi in questo caso equivale a dire che $P((X, Y, Z) \in A) = m_3(A)$ per ogni boreliano A del cubo $[0, 1] \times [0, 1] \times [0, 1]$.

Siccome ci interessa la probabilità che la parte intera $\lfloor X + Y + Z \rfloor$ sia uguale a uno, basta esprimere ciò in termini di appartenenza della terna di punti del cubo (X, Y, Z) a un opportuno sottoinsieme A del cubo stesso. Adesso si tratta di determinare l'insieme A detto:

$$\lfloor X + Y + Z \rfloor = 1 \iff 1 \leq x + y + z < 2,$$

da cui $A = \{(x, y, z) \in [0, 1]^3 \mid z < 2 - x - y, z \geq 1 - x - y\}$. Determinato l'insieme A si tratta di calcolarne la misura di Lebesgue tridimensionale: in questo caso l'insieme è molto semplice e si tratta del suo volume elementare. Quindi

$$m_3(A) = 1 - m_3(A^C) = 1 - 2m_3(P),$$

dove P è la piramide $P = \{(x, y, z) \in [0, 1]^3 \mid z \geq 2 - x - y\}$ che ha per basi un triangolo equilatero di lato $\sqrt{2}$ e per spigolo un segmento di lunghezza 1.

Senza calcolare integrali ci ricordiamo che il volume tridimensionale di una piramide si calcola come area di base per altezza: elementari calcoli mostrano che $m_3(P) = A_b \cdot h = \frac{1}{2} \cdot 1 \cdot \frac{1}{3} = \frac{1}{6}$. A questo punto quindi la probabilità richiesta è $m_3(A) = 1 - 2m_3(P) = 1 - 2\frac{1}{6} = \frac{2}{3}$.

Esercizio A.5.2. Siano X_1 e X_2 variabili aleatorie con densità del blocco

$$\rho(x, y) = \begin{cases} \frac{1}{\pi} & \text{se } X_1^2 + X_2^2 \leq 1 \\ 0 & \text{altrimenti} \end{cases}.$$

Dire se sono indipendenti.

Soluzione. Intanto

$$\rho_1(x) = \int_{\mathbb{R}} \rho(x, y) dx dy = \begin{cases} \frac{2}{\pi} \sqrt{1 - X_1^2} & \text{se } |X_1| \leq 1 \\ 0 & \text{altrimenti} \end{cases}$$

e analogamente la densità di X_2 . Facendone il prodotto si vede che la densità non è ρ , neanche quasi ovunque. Ciò mostra che non sono indipendenti.

Se non volevamo fare il calcolo esplicito delle due densità era sufficiente osservare che il dominio su cui ρ è non negativa non è un rettangolo.

Esercizio A.5.3. Siano X_n variabili aleatorie indipendenti con legge uniforme su $[0, 1]$. Posto $S_n = \max_i X_i$ studiare la densità della variabile S_n .

Soluzione. Visto che la probabilità è uniforme su $[0, 1]$ si ha

$$F(t) = P(X_i \leq t) = \begin{cases} 0 & \text{se } t < 0 \\ t & 0 \leq t \leq 1 \\ 1 & \text{se } t > 1 \end{cases} .$$

Dunque, vista l'indipendenza

$$P(S_n \leq t) = \prod_{i=1}^n P(X_i \leq t) = \begin{cases} 0 & \text{se } t < 0 \\ t^n & 0 \leq t \leq 1 \\ 1 & \text{se } t > 1 \end{cases} ,$$

e questa è continua e derivabile a tratti; dunque per un teorema visto a lezione, S_n ammette densità e questa è proprio

$$\rho(x) = \begin{cases} 0 & \text{se } t \leq 0 \\ nt^{n-1} & 0 < t < 1 \\ 0 & \text{se } t \geq 1 \end{cases} .$$

Osservazione A.5.1. Vale $S_n \rightarrow Z$ in legge se e solo se $F_{S_n}(t) \rightarrow F_Z(t)$ per ogni t salvo al più numerabili casi. Questi casi sono le eventuali discontinuità a salto della funzione $F_Z(t)$.

Nell'esercizio precedente abbiamo

$$F_{S_n}(t) \rightarrow \begin{cases} 0 & \text{se } t < 1 \\ 1 & \text{se } t \geq 1 \end{cases}$$

Ci chiediamo che variabile aleatoria ha questa funzione di ripartizione. In effetti

$$P(Z \leq 1 + \varepsilon) = F_Z(1 + \varepsilon) = 0 \quad \text{e} \quad P(Z > 1 + \varepsilon) = 1 - F_Z(1 + \varepsilon) = 0,$$

e quindi $P(Z \in (1 - \varepsilon, 1 + \varepsilon)) = 1$. Allora $Z = 1$ quasi certamente.

Esercizio A.5.4. Sia (X, Y) il blocco di due variabili aleatorie, la cui densità è

$$f(x, y) = \begin{cases} (y - x)e^{-(y-x)} & \text{se } 0 < x < 1 \text{ e } y > x \\ 0 & \text{altrimenti} \end{cases} .$$

- (1) Le variabili aleatorie X e Y sono indipendenti? E X e $Y - X$?
- (2) Qual è la densità di $X + Y$?

Soluzione. (1) Le variabili X e Y non possono essere indipendenti in quanto la densità $f(x, y)$ non può essere spezzata come prodotto di una funzione di x con una funzione di y . Adesso possiamo considerare il blocco $(U, V) = (X, Y - X)$ e l'applicazione

$$h : (x, y) \mapsto (u, v) = (x, y - x).$$

Detto $D = \{(x, y) \in \mathbb{R}^2 \mid 0 < x < 1 \text{ e } y > x\}$, abbiamo che h è un diffeomorfismo da D a $h(D) = \{(u, v) \in \mathbb{R}^2 \mid 0 < u < 1 \text{ e } v > 0\}$. Il determinante dello jacobiano di h^{-1} ha modulo 1 e dunque

$$g(u, v) = \begin{cases} ve^{-v} & \text{se } 0 < u < 1 \text{ e } v > 0 \\ 0 & \text{altrimenti} \end{cases}.$$

è la densità del blocco (U, V) . In questo caso si nota che U e V sono indipendenti in quanto la densità congiunta può essere scritta come prodotto di 1 per ve^{-v} : possiamo dire di più, ossia che U ha densità uniforme sull'intervallo $[0, 1]$ ⁵, mentre V ha densità esponenziale di parametro 1.

(2) Osserviamo che $X + Y = 2U + V$, e ciò ci conviene perché le espressioni delle variabili U e V sono più semplici delle altre. Visto che $2U$ e V sono indipendenti avremo che la somma $2U + V$ ha densità $g(x)$ data da

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f_{2U}(y-x)f_V(y) dy = \int_{-\infty}^{\infty} \frac{1}{2} \cdot \chi_{\{0 \leq x-y \leq 2\}} \cdot ye^{-y} \cdot \chi_{\{y \geq 0\}} dy = \\ &= \begin{cases} 0 & \text{se } x \leq 0 \\ \frac{1}{2} \int_0^x ye^{-y} dy & \text{se } 0 < x < 2 \\ \frac{1}{2} \int_{x-2}^x ye^{-y} dy & \text{se } x \geq 2 \end{cases}, \end{aligned}$$

dove omettiamo i calcoli finali.

Esercizio A.5.5. Si consideri il blocco (X, Y) , la cui densità è data da

$$f(x, y) = \begin{cases} kxy & \text{se } x \geq 0, y \geq 0 \text{ e } x + y \leq 1 \\ 0 & \text{altrimenti} \end{cases}.$$

- (1) Calcolare k in modo che f risulti una densità;
- (2) calcolare la densità di $X + Y$;
- (3) le variabili $U = X + Y$ e $V = X - Y$ sono indipendenti?

⁵e dunque $2U$ ha densità uniforme su $[0, 2]$, in cui la densità vale chiaramente $\frac{1}{2}$, ci servirà dopo.

Soluzione. (1) Affinché f sia una densità è necessario che il suo integrale sia 1. Sia D l'insieme su cui $f(x, y)$ è non nulla: un rapido sguardo mostra che tale insieme è il triangolo delimitato dagli assi e dalla retta $y = -x + 1$ contenuto nel primo quadrante. Dunque

$$1 = \iint_{\mathbb{R}^2} f(x, y) dx dy = \int_0^1 \left(\int_0^{1-x} kxy dy \right) dx = \frac{k}{24},$$

quindi $k = 24$.

(3) Per rispondere alla domanda (2), vista la (3), ci conviene dapprima calcolare la legge congiunta del blocco (U, V) . Considerato il diffeomorfismo $h(x, y) = (x + y, x - y)$ si ha che

$$h(D) = \{(u, v) \in \mathbb{R}^2 \mid 0 < u \leq 1 \text{ e } -u < v < u\}$$

e che $|\det J_{h^{-1}}| = \frac{1}{2}$. Allora avremo che la densità congiunta è $g(u, v) = 3(u^2 - v^2) \cdot \chi_{h(D)}$. Visto che $h(D)$ (il supporto della densità g) non è un rettangolo abbiamo che U e V non possono essere indipendenti⁶.

(2) A questo punto si ha semplicemente

$$f_{X+Y}(u) = \int_{-\infty}^{\infty} g(u, v) dv = \int_{-u}^u 3(u^2 - v^2) dv \cdot \chi_{(0,1)} = 4u^3 \cdot \chi_{(0,1)}.$$

Esercizio A.5.6. Sia X una variabile aleatoria reale che ammette densità f . Provare che anche la variabile aleatoria $|X|$ ammette densità e calcolarla.

Supponiamo adesso che f sia una funzione pari e che la variabile X ammetta momento del terzo ordine. Dimostrare che le variabili X e X^2 sono scorrelate.

Soluzione. (1) Abbiamo che

$$\begin{aligned} E(|X|) &= \int_{-\infty}^{\infty} |x|f(x) dx = \int_0^{\infty} xf(x) dx + \int_{-\infty}^0 |x|f(x) dx = \\ &= \int_0^{\infty} xf(x) dx + \int_{\infty}^0 -|x|f(-x) dx = \int_0^{\infty} xf(x) dx + \int_0^{\infty} xf(-x) dx, \end{aligned}$$

dove nel terzo passaggio abbiamo utilizzato un cambio di variabile. Ma allora osserviamo che $|X|$ ha densità, ed anzi tale densità è $f(x) + f(-x)$.

(2) Mostriamo che $Cov(X, X^2) = 0$; abbiamo che $Cov(X, X^2) = E(X^3) - E(X)E(X^2)$.

Ma

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx = 0 \quad \text{e} \quad E(X^3) = \int_{-\infty}^{\infty} x^3f(x) dx = 0$$

in quanto le due funzioni integrande sono dispari.

⁶si poteva anche dire osservando che non si può spezzare $g(u, v)$ nel prodotto di due funzioni, una di u e l'altra di v .

Esercizio A.5.7. Sia X una variabile aleatoria reale con densità esponenziale di parametro λ e sia $Y = \lfloor X \rfloor$. Qual è la legge della variabile aleatoria Y ?

Si consideri adesso per ogni intero $n > 0$ una variabile aleatoria geometrica di parametro $\frac{1}{n}$ e sia $Y_n = \frac{X_n}{n}$. Esaminare se la successione $\{Y_n\}_{n \geq 1}$ converge in probabilità, ed eventualmente a quale limite.

Soluzione. (1) Intanto osserviamo che Y è discreta, e dunque per $k \geq 0$

$$P(Y = k) = P(X \in [k, k+1)) = \int_k^{k+1} \lambda e^{-\lambda x} dx = (e^{-\lambda})^k (1 - e^{-\lambda}).$$

(2) Calcoliamo la funzione di ripartizione di Y_n al variare di $n \geq 1$. Visto che Y_n assume valori positivi avremo che per $t \geq 0$

$$P(Y_n \leq t) = P(X_n \leq nt) = P(X_n \in \{1, \dots, \lfloor nt \rfloor\}) = \sum_{k=1}^{\lfloor nt \rfloor} \frac{1}{n} \left(1 - \frac{1}{n}\right)^{k-1} = 1 - \left(1 - \frac{1}{n}\right)^{\lfloor nt \rfloor}.$$

Se $n \rightarrow \infty$ abbiamo che $P(Y_n \leq t) \rightarrow 1 - e^{-t}$, che è la funzione di ripartizione della variabile esponenziale. Dunque Y_n converge in probabilità ad una variabile esponenziale.

Esercizio A.5.8. Siano X_1, \dots, X_n variabili aleatorie di Bernoulli con parametro $p \in (0, 1)$. Sia poi $a > 0$ fissato e sia A_n l'evento

$$A_n = \{X_1 + \dots + X_n \leq a\sqrt{n}\}.$$

Si dimostri che $P(A_n) \rightarrow 0$ se $n \rightarrow \infty$.

Soluzione. Detta $S_n = \sum_{i=1}^n A_i$ abbiamo che

$$P(A_n) = P(S_n - np \leq a\sqrt{n} - np) \leq P(|S_n - np| \geq np - a\sqrt{n}) \leq \frac{\text{Var}(S_n)}{(np - a\sqrt{n})^2},$$

e osservando che $\text{Var}(S_n) = p(1-p)n$ si conclude.

Concludiamo con un lemma che ci sarà utile in seguito:

Lemma A.5.1. Siano Y_1 e Y_2 due variabili aleatorie con densità gaussiane rispettivamente $N(0, 1)$ e $N(\mu, \sigma^2)$. Dette F_1 e F_2 le due funzioni di ripartizione si ha

$$F_1(t) = F_2\left(\frac{t - \mu}{\sigma}\right).$$

Dimostrazione. È sufficiente scrivere le due funzioni di ripartizione e fare un cambio di variabile negli integrali. \square

A.6 Inferenza statistica

Precisiamo che possiamo formulare un modello canonico anche in statistica. Generalmente basterà considerare $\Omega = \mathbb{R}^n$, $\mathcal{A} = B(\mathbb{R}^n)$ e $P^\vartheta = m_\vartheta \otimes \cdots \otimes m_\vartheta$ (prodotto fatto n volte) e come variabili aleatorie le solite proiezioni canoniche.

Esercizio A.6.1. Siano $X_i \sim N(\mu_i, \sigma_i^2)$ con $i = 0, 1$ due variabili aleatorie con densità gaussiana; sia poi Y una variabile bernoulliana di parametro p . Supponiamo che X_0, X_1 e Y siano indipendenti e poniamo $W = X_Y$, ossia

$$W(\omega) = \begin{cases} X_0(\omega) & \text{se } Y(\omega) = 0 \\ X_1(\omega) & \text{se } Y(\omega) = 1 \end{cases}.$$

- (1) Mostrare che W ha densità e calcolare $E(W)$ e $E(W^2)$;
- (2) supponendo $\sigma_0 = \sigma_1 = 1$ e $\mu_0 = -1$ e $\mu_1 = 1$, mostrare che $U_n = \frac{\overline{W}_n + 1}{2}$ stima p in modo corretto e consistente;
- (3) calcolare l'intervallo di fiducia col quantile $\alpha = \frac{1}{10}$.

Soluzione. (1) Intanto per il lemma che chiude il precedente paragrafo

$$\begin{aligned} P(W \leq t) &= P(W \leq t \mid Y = 0)P(Y = 0) + P(W \leq t \mid Y = 1)P(Y = 1) = \\ &= (1-p)F\left(\frac{t-\mu_0}{\sigma_0}\right) + pF\left(\frac{t-\mu_1}{\sigma_1}\right), \end{aligned}$$

dove abbiamo posto $F(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds$. Visto che la funzione di ripartizione è di classe C^∞ , possiamo ottenere la densità derivandola; si ha allora

$$f_W(t) = (1-p) \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(t-\mu_0)^2}{2\sigma_0^2}} + p \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}}.$$

Inoltre si ha $E(W) = (1-p)E(X_0) + pE(X_1) = p\mu_1 + (1-p)\mu_0$ e dunque similmente

$$E(W^2) = (1-p)E(X_0^2) + pE(X_1^2) = p(\mu_1^2 + \sigma_1^2) + (1-p)(\mu_0^2 + \sigma_0^2);$$

possiamo ricavare facilmente la speranza una volta calcolata $E(W)$ in quanto vale che $Var(X_i) = \sigma_i^2 = E(X_i^2) - E(X_i)^2$.

(2) Sia $\vartheta = (p, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2)$. Per quanto già calcolato si ha $E^\vartheta(U) = 2p - 1$ e $E^\vartheta(W^2) = 0$, da cui $Var^\vartheta(W) = 1 - 4p^2 + 4p$. Si ha infine

$$E^\vartheta(W) = E\left(\frac{\overline{W}_n + 1}{2}\right) = \frac{E(W) + 1}{2} = p,$$

e ciò mostra che lo stimatore è corretto. Per mostrare che è consistente basta

$$P^\vartheta(|U_n - p| > \varepsilon) = P^\vartheta(|\overline{W}_n - (2p - 1)| > 2\varepsilon) \leq \frac{Var^\vartheta(W)}{n(2\varepsilon)^2} \leq \frac{2}{4n\varepsilon^2} = \frac{1}{2n\varepsilon^2},$$

e ciò mostra che con $\varepsilon \rightarrow 0$ si ha convergenza in probabilità.

(3) Vogliamo che

$$\alpha = \frac{1}{2n\varepsilon^2},$$

da cui $\varepsilon = \sqrt{\frac{5}{n}}$. Quindi l'intervallo di fiducia è $\left[U_n - \sqrt{\frac{5}{n}}, U_n + \sqrt{\frac{5}{n}} \right]$.

Esercizio A.6.2. Un importatore di hard disk acquista i suoi prodotti da due ditte A e B : si sa che gli hard disk prodotti da A sono ben funzionanti con probabilità $0,95$, mentre quelli prodotti da B con $0,85$. Si ignora però qual è la percentuale ϑ dei pezzi prodotti dalla ditta A ($0 < \vartheta < 1$): per fare questo si verificano n pezzi e si denotano con X_1, \dots, X_n le variabili aleatorie che indicano se l' i -esimo pezzo è ben funzionante.

(1) Esprimere un modello statistico che descrive la situazione;

(2) dire se esiste una stima corretta e consistente per ϑ ;

(3) determinare un intervallo di fiducia con $\alpha = 0,05$.

Soluzione. (1) Vale che X_i è una variabile aleatoria di Bernoulli con parametro

$$\varphi(\vartheta) = \vartheta \cdot 0,95 + (1 - \vartheta) \cdot 0,85,$$

che si può ottenere per disintegrazione rispetto al sistema di alternative dato da gli eventi "l'hard disk proviene da A " e il suo complementare. Il modello statistico è dunque

$$(\{0, 1\}^n, \mathcal{P}(\Omega), P^\vartheta),$$

con P^ϑ probabilità prodotto di n marginali di Bernoulli; si può anche scrivere che

$$P^\vartheta(\{(x_1, \dots, x_n)\}) = \varphi(\vartheta)^{x_1 + \dots + x_n} (1 - \varphi(\vartheta))^{n - (x_1 + \dots + x_n)}.$$

Le variabili X_i sono le proiezioni canoniche.

(2) Denotiamo con \bar{X} la media empirica delle variabili aleatorie. Allora

$$E(\bar{X}) = \varphi(\vartheta) \quad \text{e} \quad \text{Var}(\bar{X}) = \frac{\varphi(\vartheta)(1 - \varphi(\vartheta))}{n},$$

da cui si vede che la media è uno stimatore corretto e consistente per $\varphi(\vartheta)$. Ma noi volevamo stimare ϑ , dunque basta porre

$$U = \frac{\bar{X} - 0,85}{0,1} \quad \text{7}$$

e si ha che U è uno stimatore corretto e consistente per ϑ . (3) Intanto abbiamo scelto $\alpha = 0,05$ e dalle tavole si legge $q_{0,975} = 1,96$. Grazie al teorema di de Moivre-Laplace e alle considerazioni fatte nella parte di teoria si ha

$$p_\vartheta = P^\vartheta \left(\sqrt{n} \frac{|\bar{X} - \varphi(\vartheta)|}{\sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))}} \leq q_{0,975} \right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-q_{0,975}}^{q_{0,975}} e^{-\frac{t^2}{2}} dt = 0,95 \quad \text{8}.$$

⁷abbiamo ricavato questa espressione invertendo l'equazione che dà $\varphi(\vartheta)$.

⁸per il valore dell'integrale si fa una semplice considerazione sulle aree.

Semplici passaggi algebrici mostrano allora che

$$p_{\vartheta} = P \left(\bar{X} - \frac{1,96 \cdot \sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))}}{\sqrt{n}} \leq \varphi(\vartheta) \leq \bar{X} + \frac{1,96 \cdot \sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))}}{\sqrt{n}} \right).$$

A questo punto però ci serve un intervallo che non dipenda da $\varphi(\vartheta)$ e dunque dobbiamo trovare una maggiorazione per $\sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))}$. In effetti

$$\max_{\vartheta \in (0,1)} \sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))} = \max_{\varphi(\vartheta) \in (0,85, 0,95)} \sqrt{\varphi(\vartheta)(1 - \varphi(\vartheta))} \approx 0,35.$$

Dunque possiamo dire che

$$p_{\vartheta} \leq P \left(\bar{X} - \frac{1,96 \cdot 0,35}{\sqrt{n}} \leq \varphi(\vartheta) \leq \bar{X} + \frac{1,96 \cdot 0,35}{\sqrt{n}} \right),$$

e allora, se ψ è l'inversa di φ ⁹, avremo finalmente che

$$p_{\vartheta} \leq P \left(\psi \left(\bar{X} - \frac{0,68}{\sqrt{n}} \right) \leq \vartheta \leq \psi \left(\bar{X} + \frac{0,68}{\sqrt{n}} \right) \right).$$

⁹l'avevamo già ricavata, più precisamente

$$\psi(\eta) = \frac{\eta - 0,85}{0,1}.$$