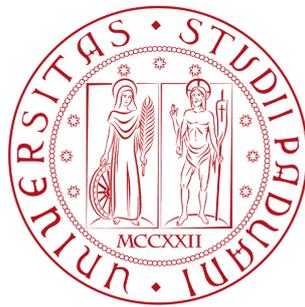


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE
**TEORIA DELLA MISURA E TEORIA DELLA
PROBABILITÀ**

Relatore Prof.ssa Giulia Treu
Dipartimento di Matematica

Laureando: Silvia Brosolo
Matricola n° 1149615

Anno Accademico 2018/2019

Introduzione

L'obiettivo di questa tesi è quello di studiare i fondamenti matematici alla base del calcolo delle probabilità, presentando i principali elementi di teoria della probabilità da un punto di vista rigorosamente matematico, in termini di teoria della misura. Il concetto di misura verrà utilizzato per fornire un modello matematico della nozione intuitiva di probabilità.

Verranno presentati in maniera astratta gli elementi più importanti della teoria della misura: σ -algebre, misure, integrali. In particolare ci si soffermerà sull'esposizione della teoria della misura di Lebesgue.

Nonostante l'idea di integrazione risalga a molti secoli prima, lo studio della teoria dell'integrazione (e della misura), di cui si parla in questa tesi, è proprio dell'inizio del ventesimo secolo. I maggiori contributi furono dati da Emile Borel e Henry Lebesgue tra la fine dell'Ottocento e l'inizio del Novecento. Lebesgue presentò un nuovo concetto di integrale, noto oggi come integrale di Lebesgue, che generalizzava il tradizionale integrale di Riemann ed era un'estensione delle classiche nozioni di lunghezza e di area a insiemi più generali di quelli solitamente associati a curve e superfici.

Lo studio della teoria della probabilità ha origini relativamente recenti, e inizialmente si è sviluppato indipendentemente dall'analisi e dalla teoria della misura: la prima trattazione su questo argomento fu scritta da Gerolamo Cardano, pubblicata postuma nel 1663, riguardante il gioco dei dadi; su questo argomento scrisse anche Galileo Galilei intorno al 1620. Tuttavia, la nascita del calcolo delle probabilità viene solitamente attribuita alla corrispondenza tra Blaise Pascal e Pierre Fermat (avvenuta intorno al 1654) nella quale i due matematici discussero sulla soluzione di un problema nato nel contesto del gioco d'azzardo. Successivamente altri matematici si interessarono a questi argomenti, come Christiaan Huygens e i Bernoulli (in particolare Jakob) nel XVII secolo e Abraham De Moivre, Eulero e Laplace nel XVIII secolo.

Prima dello sviluppo della teoria della misura comunque lo studio della probabilità riguardava problemi specifici e strategie per risolverli: non

esisteva alcuna teoria generale né definizioni fondamentali adeguate su cui basare una tale teoria. L'uso di una misura per descrivere una probabilità fu introdotto da Borel stesso intorno alla fine del XIX secolo.

A. N. Kolmogorov nei primi anni Trenta diede un fondamento assiomatico alla teoria della probabilità mediante l'uso della teoria della misura di Lebesgue.

I cenni storici appena riportati sono stati presi da [3] e [7].

Vi sono diversi approcci filosofici al concetto di probabilità; i principali sono quello classico, quello frequentista e quello soggettivo, che si distinguono da come viene definito il concetto stesso di probabilità [7], ma non è obiettivo di questa tesi approfondirli: la probabilità verrà presentata esclusivamente in termini matematici di teoria della misura, secondo l'impostazione assiomatica formulata da Kolmogorov nel 1933.

I primi due capitoli si concentrano sulla teoria della misura in generale, mentre il terzo presenta gli elementi di teoria della probabilità. In particolare, nel primo capitolo si definiscono i concetti di σ -algebra e misura, e si presenta la misura di Lebesgue come esempio di misura. Nel secondo capitolo invece si espone la teoria degli integrali: si definiscono funzioni misurabili e integrabili, e si definiscono l'integrale in senso generale e l'integrale di Lebesgue. Infine nel terzo capitolo si utilizzano le nozioni introdotte nei primi due per definire gli elementi della teoria della probabilità, come variabili casuali, distribuzioni e densità; inoltre si riportano degli esempi riassuntivi che mostrano come questi elementi siano riconducibili alla teoria della misura.

Capitolo 1

In questo capitolo si definiscono gli elementi fondamentali di teoria della misura: σ -algebra, misura, spazio misurabile; questi sono necessari per poter dare una prima definizione di spazio di probabilità; infine si presenta un esempio di misura importante: la misura di Lebesgue. I risultati riportati in questo capitolo sono noti e si trovano in [4], [5] e [6].

1.1 σ -algebra

Chiamiamo X un generico insieme e $\mathcal{P}(X)$ l'insieme delle parti di X .

Definizione 1. Un sottoinsieme non vuoto \mathcal{A} di $\mathcal{P}(X)$ si dice *algebra* in X se soddisfa le seguenti condizioni:

- (i) $\emptyset, X \in \mathcal{A}$,
- (ii) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$,
- (iii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$.

Definizione 2. Un'algebra \mathcal{A} in X si dice *σ -algebra* se soddisfa l'ulteriore proprietà:

$$A_n, n \in \mathbb{N}, A_n \in \mathcal{A} \implies \bigcup_{n \in \mathbb{N}} A_n.$$

Una σ -algebra dunque è un sottoinsieme dell'insieme delle parti di X che include l'insieme vuoto e l'insieme X stesso ed è chiuso rispetto all'unione numerabile e al complementare. Ne deriva che \mathcal{A} è chiusa anche rispetto all'unione finita e all'intersezione numerabile (e finita); ciò è facilmente dimostrabile utilizzando l'identità di De Morgan e la definizione stessa di σ -algebra.

Definizione 3. Gli elementi di una σ -algebra \mathcal{A} in X si dicono *insiemi misurabili* e la coppia (X, \mathcal{A}) si dice *spazio misurabile*.

L'intersezione di tutte le σ -algebre in X contenenti un sottoinsieme \mathcal{K} di $\mathcal{P}(X)$ è una σ -algebra e si dice *generata* da \mathcal{K} .

Definizione 4. La σ -algebra generata da tutti gli insiemi aperti in X si dice *σ -algebra di Borel* e si denota con $\mathcal{B}(X)$; gli elementi di $\mathcal{B}(X)$ si dicono *insiemi di Borel* o *boreliani*.

In \mathbb{R} , $\mathcal{B}(\mathbb{R})$ è uguale alla σ -algebra generata da \mathcal{I} , ovvero la classe di tutti gli intervalli $[a, b)$ con $a < b$ (o, equivalentemente, (a, b) , o $(a, +\infty)$, o $(-\infty, a]$), $a, b \in \mathbb{R}$; ovvero, è la più piccola σ -algebra in \mathbb{R} che contiene tutti gli intervalli.

Si vedrà più avanti che in probabilità praticamente tutti gli insiemi di interesse sono boreliani; inoltre i boreliani verranno utilizzati per definire le funzioni misurabili.

1.2 Misura

Si vuole definire la nozione di misura, una funzione con delle particolari proprietà, su una σ -algebra; per farlo però è necessario prima chiarire cosa si intende quando si dice che una funzione è additiva, σ -additiva, σ -subadditiva.

Definizione 5. Data \mathcal{A} algebra sull'insieme X , e $\mu : \mathcal{A} \rightarrow [0, \infty]$ funzione tale che $\mu(\emptyset) = 0$, si dice che la funzione μ è:

- (i) *additiva*, se per ogni famiglia finita $A_1, \dots, A_n \in \mathcal{A}$ di insiemi disgiunti si ha che

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mu(A_k);$$

- (ii) *σ -additiva* o *numerabilmente additiva*, se per ogni successione di insiemi disgiunti $(A_n)_n \subset \mathcal{A}$ tale che $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ si ha che

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n);$$

- (iii) *σ -subadditiva* o *numerabilmente subadditiva*, se per ogni successione $(A_n)_n \subset \mathcal{A}$ tale che $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ si ha che

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Si può dimostrare che una funzione σ -additiva μ su \mathcal{A} è anche additiva e σ -subadditiva.

A questo punto è possibile definire una misura e lo spazio di misura ad essa associato.

Definizione 6. Data \mathcal{A} σ -algebra sull'insieme X , si dice *misura* una funzione σ -additiva $\mu : \mathcal{A} \rightarrow [0, \infty]$ tale che $\mu(\emptyset) = 0$.

La terna (X, \mathcal{A}, μ) si chiama *spazio di misura*.

Per quanto detto sopra, una misura, essendo σ -additiva, è anche additiva e σ -subadditiva.

È importante notare che una misura è una funzione monotona, ovvero:

Proposizione 1. Dato uno spazio di misura (X, \mathcal{A}, μ) e due insiemi $A, B \in \mathcal{A}$ tali che $A \subseteq B$, allora $\mu(A) \leq \mu(B)$.

Una misura μ può essere: *finita*, se $\mu(X) < \infty$; σ -*finita*, se esiste una successione $(A_n)_n \subset \mathcal{A}$ tale che $\cup_{n=1}^{\infty} A_n = X$ e $\mu(A_n) < \infty, \forall n \in \mathbb{N}$.

Una misura μ si dice *completa* se soddisfa la seguente proprietà: $A \in \mathcal{A}, B \subset A, \mu(A) = 0 \implies B \in \mathcal{A}$, e dunque $\mu(B) = 0$.

Infine si dice che la misura μ è *concentrata sull'insieme* $A \in \mathcal{A}$ se $\mu(A^c) = 0$; l'insieme A è chiamato *supporto* di μ .

Dato uno spazio di misura (X, \mathcal{A}, μ) , si può considerare per ogni sottoinsieme $A \in \mathcal{A}$ la *restrizione* della misura μ al sottoinsieme A , ovvero la funzione $\mu_A(B) = \mu(A \cap B), \forall B \in \mathcal{A}$. Si può dimostrare che μ_A è una misura su \mathcal{A} .

Si conclude il discorso sulle misure introducendo un'importante distinzione, che sarà utile quando si parlerà di misure di probabilità: dato uno spazio misurabile (X, \mathcal{A}) tale che $\{x\} \in \mathcal{A}, \forall x \in X$, una misura finita o σ -finita μ su (X, \mathcal{A}) è *continua* se $\mu(\{x\}) = 0, \forall x \in X$; è *discreta* se esiste un sottoinsieme numerabile $D \subseteq X$ tale che $\mu(D^c) = 0$, ovvero μ è concentrata su D .

Prima di proseguire, si mostra un primo esempio di misura: la *misura di Dirac in* x, δ_x . Essa è definita nel modo seguente: dato $x \in X$, per ogni $A \in \mathcal{P}(X)$,

$$\delta_x(A) = \begin{cases} 1 & \text{se } x \in A, \\ 0 & \text{se } x \notin A. \end{cases}$$

La misura di Dirac è una misura su $\mathcal{P}(X)$, ed è concentrata sull'insieme $\{x\}$ (e dunque è una misura discreta).

1.3 Spazio di Probabilità

Con le nozioni definite finora è possibile introdurre i concetti fondamentali della teoria della probabilità.

Definizione 7. Dato un insieme Ω e una σ -algebra \mathcal{F} in Ω , una misura P tale che $P(\Omega) = 1$ si dice *misura di probabilità*.

La terna (Ω, \mathcal{F}, P) si dice *spazio di probabilità*.

Si riporta anche la definizione originale di spazio di probabilità data da Kolmogorov nel 1933: (Ω, \mathcal{F}, P) è uno spazio di probabilità se (Ω, \mathcal{F}) è uno spazio misurabile e $P : \mathcal{F} \rightarrow [0, +\infty]$ è una funzione additiva tale che $P(\Omega) = 1$ e $P(B_n) \rightarrow 0$ quando la successione $(B_n)_n$ in \mathcal{F} decresce a \emptyset [7].

In probabilità l'insieme Ω , detto *spazio campionario*, rappresenta l'insieme di tutti i possibili esiti, detti *eventi elementari*, di un esperimento casuale. Il termine *evento* si riferisce al fatto che l'esito dell'esperimento appartenga a un certo sottoinsieme E di Ω , e la misura di E (il numero $P(E)$) si interpreta come la probabilità che l'esito giaccia in E , ovvero la probabilità che si verifichi l'evento. È naturale richiedere che E appartenga a \mathcal{F} , ovvero che sia misurabile, dato che ci interessa la sua misura. Per convenzione, con il termine evento si identifica l'insieme E che lo caratterizza. Si può dunque dare la seguente definizione di eventi:

Definizione 8. Dato uno spazio di probabilità (Ω, \mathcal{F}, P) , gli elementi di \mathcal{F} si dicono *eventi*.

Le proprietà della σ -algebra \mathcal{F} garantiscono che l'unione e l'intersezione, finite o numerabili, e il complementare di eventi siano ancora degli eventi.

1.4 Misura di Lebesgue

Si introduce ora un esempio di misura molto importante, la misura di Lebesgue; in \mathbb{R} , è la misura che assegna agli intervalli la loro lunghezza (e in generale in \mathbb{R}^d assegna ad ogni intervallo d -dimensionale il suo volume). Vedremo nel terzo capitolo la sua importanza in relazione alla teoria della probabilità.

Per costruire la misura di Lebesgue si parte dalla definizione di misura esterna.

Definizione 9. Dato un insieme X , una funzione $\mu^* : \mathcal{P}(X) \rightarrow [0, +\infty]$ si dice *misura esterna* su X se:

- (i) $\mu^*(\emptyset) = 0$;
- (ii) μ^* è monotona, ovvero $A \subset B \implies \mu^*(A) \leq \mu^*(B)$;
- (iii) μ^* è σ -subadditiva.

Una misura può non essere una misura esterna (anzi, lo è se e solo se il suo dominio è $\mathcal{P}(X)$), e in generale una misura esterna non è una misura poiché non è σ -additiva. Vedremo che per ogni misura esterna μ^* su X esiste una σ -algebra \mathcal{M}_{μ^*} su X tale che la restrizione di μ^* a \mathcal{M}_{μ^*} è una misura.

Definizione 10. Si chiama *misura esterna di Lebesgue* su \mathbb{R} la funzione $\lambda^* : \mathcal{P}(X) \rightarrow [0, +\infty]$ definita da

$$\lambda^*(A) = \inf \left\{ \sum_{i=1}^{\infty} (b_i - a_i) : A \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i) \right\}, \quad \forall A \subseteq \mathbb{R},$$

dove (a_i, b_i) sono intervalli aperti limitati.

Proposizione 2. *La misura esterna di Lebesgue su \mathbb{R} è una misura esterna, e assegna ad ogni sottointervallo di \mathbb{R} la sua lunghezza.*

La misura esterna di Lebesgue di un insieme $A \subseteq \mathbb{R}$ è dunque l'estremo inferiore delle lunghezze di tutte le possibili coperture di A , dove una copertura di A è una sequenza di intervalli (a_i, b_i) tali che $A \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i)$. La serie $\sum_{i=1}^{\infty} (a_i, b_i)$ potrebbe divergere per ogni copertura di A , dunque $\lambda^*(A)$ può essere uguale a $+\infty$. L'insieme descritto nella Definizione 10 è limitato inferiormente da 0 dunque l'estremo inferiore esiste sempre.

Si noti che $\lambda^*(\emptyset) = 0$, $\lambda^*({x}) = 0$ per ogni $x \in \mathbb{R}$ e $\lambda^*(\mathbb{Q}) = 0$; in generale, gli insiemi numerabili hanno misura esterna di Lebesgue pari a zero. Anche insiemi con potenza del continuo possono avere misura esterna di Lebesgue pari a zero.

La definizione formale di misura esterna di Lebesgue dunque applica a tutti i sottoinsiemi di \mathbb{R} quella che coincide con l'idea intuitiva di lunghezza di un intervallo; infatti è pari a zero per insiemi costituiti da singoli punti (che sono intervalli di lunghezza 0), o insiemi numerabili (somme di intervalli di lunghezza 0); è proprio uguale alla lunghezza per gli intervalli; infine è monotona, ovvero più grande è l'insieme preso in considerazione, più grande sarà la sua misura esterna (di Lebesgue).

Definizione 11. Data una misura esterna μ^* su X , un insieme $A \in \mathcal{P}(X)$ si dice *additivo* o μ^* -*misurabile* o *misurabile secondo la misura μ^** se

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c), \quad \forall E \in \mathcal{P}(X).$$

Si denota con \mathcal{G} la famiglia di tutti gli insiemi additivi.

Teorema 1 (di Carathéodory). \mathcal{G} è una σ -algebra in X e μ^* è una misura su \mathcal{G} .

Si osservi inoltre che ogni insieme con misura esterna pari a 0 (detto *insieme nullo*) è additivo; dunque la misura μ^* sulla σ -algebra \mathcal{G} è completa.

Un sottoinsieme di \mathbb{R} è *misurabile secondo Lebesgue* se è misurabile secondo la misura esterna di Lebesgue. Chiamiamo \mathcal{M}_{λ^*} la famiglia di tutti gli insiemi misurabili secondo Lebesgue in \mathbb{R} : per il Teorema 1 questa è una σ -algebra e la restrizione di λ^* a \mathcal{M}_{λ^*} è una misura (completa).

Tutti gli insiemi boreliani in \mathbb{R} sono misurabili secondo Lebesgue, infatti \mathcal{M}_{λ^*} è una σ -algebra che contiene tutti gli intervalli (ovvero tutti gli intervalli in \mathbb{R} sono misurabili secondo Lebesgue) e $\mathcal{B}(\mathbb{R})$ è la più piccola σ -algebra che contiene tutti gli intervalli di \mathbb{R} : dunque si ha che $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}_{\lambda^*}$.

Si può dunque definire la misura di Lebesgue su \mathbb{R} :

Definizione 12. La restrizione della misura esterna di Lebesgue su \mathbb{R} alla famiglia \mathcal{M}_{λ^*} di sottoinsiemi di \mathbb{R} misurabili secondo Lebesgue si dice *misura di Lebesgue* e si denota con λ .

Si può ora introdurre un esempio di misura di probabilità: la misura di Lebesgue ristretta all'intervallo $[0, 1]$, $\lambda_{[0,1]}$, è una misura di probabilità sulla σ -algebra $\mathcal{M}_{[0,1]}$ di sottoinsiemi misurabili di $[0, 1]$; dato che tutti i sottointervalli di $[0, 1]$ con la stessa lunghezza hanno uguale misura, la massa di probabilità di $\lambda_{[0,1]}$ è distribuita in modo uniforme sull'intervallo $[0, 1]$.

Capitolo 2

Si è visto cosa si intende quando si parla di insiemi misurabili e misure definite su σ -algebre; in questo capitolo si trattano le funzioni misurabili (rispetto a una certa σ -algebra) e si introduce il concetto generale di integrale di una funzione, e la caratterizzazione di funzioni integrabili, sia in senso generale, sia secondo Lebesgue. La caratterizzazione delle funzioni misurabili permetterà di definire le variabili casuali nel prossimo capitolo. Inoltre si introducono proprietà delle misure (continuità assoluta, singolarità) che saranno utili per sviluppare il discorso sulle variabili casuali e la loro distribuzione. I contenuti di questo capitolo sono riportati in [2], [5], [6] e [8].

2.1 Funzioni Misurabili

Per prima cosa dunque si definiscono le funzioni misurabili.

Definizione 13. Dato uno spazio di misura (X, \mathcal{A}) e un insieme misurabile $A \in \mathcal{A}$, una funzione $f : A \rightarrow [-\infty, +\infty]$ si dice *misurabile rispetto ad \mathcal{A}* se soddisfa una, e dunque tutte, delle seguenti proprietà, tra loro equivalenti:

- (i) $f^{-1}((a, \infty)) \in \mathcal{A}$ per ogni $a \in \mathbb{R}$;
- (ii) $f^{-1}([a, \infty)) \in \mathcal{A}$ per ogni $a \in \mathbb{R}$;
- (iii) $f^{-1}((-\infty, a)) \in \mathcal{A}$ per ogni $a \in \mathbb{R}$;
- (iv) $f^{-1}((-\infty, a]) \in \mathcal{A}$ per ogni $a \in \mathbb{R}$.

Se $X = \mathbb{R}$, una funzione misurabile rispetto alla σ -algebra \mathcal{M}_{λ^*} definita nel capitolo precedente si dice *misurabile secondo Lebesgue*; una funzione misurabile rispetto alla σ -algebra di Borel si dice invece *funzione di Borel*. Naturalmente, essendo $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}_{\lambda^*}$, ogni funzione di Borel è misurabile secondo Lebesgue.

Vediamo alcuni esempi, che illustrano quando alcune particolari funzioni sono misurabili.

- (i) Denotiamo con \mathcal{I}_B la *funzione caratteristica* o *funzione indicatrice* dell'insieme B , ovvero la funzione che associa il valore 1 agli elementi che si trovano in B e 0 agli elementi non in B : allora, dato uno spazio misurabile (X, \mathcal{A}) e $B \in X$, la funzione \mathcal{I}_B è misurabile rispetto ad \mathcal{A} se e solo se B è misurabile, ovvero $B \in \mathcal{A}$.
- (ii) Chiamiamo *semplice* una funzione φ che assume al massimo un numero finito di valori $\alpha_1, \dots, \alpha_n$; essa può essere scritta come combinazione lineare di funzioni indicatrici: $\varphi = \sum_{i=1}^n \alpha_i \mathcal{I}_{A_i}$, dove $A_i = \varphi^{-1}(\{\alpha_i\})$ per ogni $i = 1, \dots, n$. In uno spazio misurabile (X, \mathcal{A}) una funzione semplice $\varphi : X \rightarrow [-\infty, \infty]$ che assume valori $\alpha_1, \dots, \alpha_n$ è misurabile rispetto ad \mathcal{A} se e solo se $A_i = \varphi^{-1}(\{\alpha_i\}) \in \mathcal{A}$ per ogni $i = 1, \dots, n$.
- (iii) Le funzioni costanti sono misurabili secondo Lebesgue.
- (iv) Le funzioni continue sono misurabili secondo Lebesgue.
- (v) Le funzioni monotone sono misurabili secondo Lebesgue.

È interessante notare che il passaggio al limite conserva la misurabilità: se $\{f_n\}$ è una sequenza di funzioni misurabili che assumono valori sulla retta reale estesa definite su un insieme A in uno spazio misurabile (X, \mathcal{A}) , le funzioni $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$, $\liminf_n f_n$, $\lim_n f_n$ (se esiste) sono anch'esse misurabili.

Una funzione f è misurabile rispetto a una σ -algebra se e solo se lo sono la sua parte positiva f^+ e la sua parte negativa f^- (dove $f^+(x) = \max(f(x), 0)$, $f^-(x) = -\min(f(x), 0)$).

Per definire una funzione misurabile si possono usare altri enunciati che sono equivalenti a quelli riportati fin'ora; si può dire che una funzione a valori reali è misurabile rispetto ad \mathcal{A} nello spazio misurabile (X, \mathcal{A}) se la controimmagine di f di ogni sottoinsieme aperto di \mathbb{R} appartiene ad \mathcal{A} , oppure se la controimmagine di f di ogni sottoinsieme chiuso di \mathbb{R} appartiene ad \mathcal{A} , oppure se la controimmagine di f di ogni sottoinsieme di Borel di \mathbb{R} appartiene ad \mathcal{A} .

2.2 Funzioni Integrabili e Integrale

Passiamo ora al concetto di integrale di una funzione, sempre rispetto a una misura, e vediamo quando una funzione è integrabile. Per costruire

l'integrale si parte da funzioni semplici non negative fino ad arrivare alla definizione generale di integrale di funzioni qualsiasi.

Dunque si parte con una misura μ definita su uno spazio misurabile (X, \mathcal{A}) e innanzitutto si definisce l'integrale rispetto a μ di una funzione semplice non negativa misurabile $\varphi : X \rightarrow [0, +\infty]$ definita da $\varphi = \sum_{i=1}^n \alpha_i \mathcal{I}_{A_i}$, dove $\alpha_1, \dots, \alpha_n$ sono numeri reali non negativi e A_1, \dots, A_n sono sottoinsiemi disgiunti di \mathcal{A} :

$$\int \varphi \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i).$$

Si noti che questa somma è pari a un valore reale non negativo o a $+\infty$.

Con questa scrittura si sottintende che l'integrale è calcolato su tutto il dominio di φ (ovvero X); è possibile calcolare l'integrale di φ anche su un sottoinsieme di X (purché questo sia misurabile), e in questo caso si ha che

$$\int_B \varphi \, d\mu = \int \varphi \mathcal{I}_B \, d\mu = \sum_{i=1}^n \alpha_i \mu(A_i \cap B), \quad \text{dove } B \in \mathcal{A}.$$

Il passo successivo è utilizzare la definizione appena data per costruire l'integrale (su $B \in \mathcal{A}$) rispetto a μ di una qualsiasi funzione misurabile (rispetto ad \mathcal{A}) non negativa $f : X \rightarrow [0, +\infty]$:

$$\int_B f \, d\mu = \sup \left\{ \int_B \varphi \, d\mu : 0 \leq \varphi \leq f, \varphi \text{ funzione semplice} \right\}.$$

Il seguente teorema assicura che le φ utilizzate nella definizione esistano sempre:

Teorema 2. *Dato (X, \mathcal{A}) spazio misurabile, se la funzione $f : X \rightarrow [0, +\infty]$ è misurabile, allora esiste una sequenza $\{\varphi_n\}$ di funzioni semplici tali che $0 \leq \varphi_1 \leq \dots \leq f$, $\varphi_n \rightarrow f$ puntualmente e $\varphi_n \rightarrow f$ in modo uniforme su ogni insieme sul quale f è limitata.*

Si può dunque passare alla definizione di integrale per funzioni misurabili qualsiasi: l'integrale di una funzione misurabile $f : X \rightarrow [-\infty, +\infty]$ rispetto alla misura μ esiste se almeno uno dei due integrali $\int f^+ \, d\mu$ e $\int f^- \, d\mu$ è finito. Se l'integrale di f esiste, esso è definito come segue:

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu.$$

Se l'integrale è finito, ovvero se $\int f^+ d\mu$ e $\int f^- d\mu$ sono entrambi finiti, allora f si dice *integrabile* rispetto a μ .

Se $B \in \mathcal{A}$ è un insieme misurabile, la funzione f è *integrabile sull'insieme* B se la funzione $f\mathcal{I}_B$ è integrabile, e in questo caso si ha che l'integrale di f sull'insieme B è definito come $\int_B f d\mu = \int f\mathcal{I}_B d\mu$. La funzione $f\mathcal{I}_B$ è integrabile se lo è la funzione f e l'insieme B è misurabile.

Se $X = \mathbb{R}$ e la misura rispetto a cui si integra è la misura di Lebesgue λ , allora si parla di funzioni integrabili rispetto a Lebesgue e l'integrale di una funzione rispetto a questa misura è l'*integrale di Lebesgue* : $\int f d\lambda$, che spesso viene indicato anche come $\int f(x) dx$.

Si riportano alcune importanti proprietà delle funzioni integrabili:

Proposizione 3. *Dato uno spazio di misura (X, \mathcal{A}, μ) e f, g funzioni integrabili a valori reali su X , e $\alpha \in \mathbb{R}$, allora:*

- (i) αf e $f + g$ sono integrabili;
- (ii) $\int \alpha f d\mu = \alpha \int f d\mu$;
- (iii) $\int (f + g) d\mu = \int f d\mu + \int g d\mu$;
- (iv) se $f(x) \leq g(x)$ per ogni $x \in X$ allora $\int f d\mu \leq \int g d\mu$.

Nella teoria degli integrali vi sono degli importanti risultati che riguardano limiti di successioni di funzioni integrabili; due di questi sono contenuti in due teoremi fondamentali, ovvero il teorema della convergenza monotona e il teorema della convergenza dominata, che si enunciano di seguito. Il primo riguarda funzioni misurabili non negative; il secondo estende, con ipotesi aggiuntive, i risultati del primo a generiche funzioni misurabili.

Teorema 3 (Teorema della Convergenza Monotona). *Dato (X, \mathcal{A}, μ) spazio di misura, se $\{f_n\}$ è una sequenza di funzioni misurabili non negative su X tali che $f_n \leq f$ per ogni n e $f_n \rightarrow f$ puntualmente, allora $\int f_n d\mu \rightarrow \int f d\mu$. In particolare, se $f_1 \leq f_2 \leq \dots \leq f$ allora $\int f_n d\mu \rightarrow \int f d\mu$.*

Teorema 4 (Teorema della Convergenza Dominata di Lebesgue). *Data una sequenza di funzioni misurabili su X $\{f_n\}$ tale che $|f_n| \leq g$ quasi ovunque¹ per ogni n , dove g è una funzione integrabile non negativa, e $f_n \rightarrow f$ quasi ovunque, allora f è integrabile e $\int f_n d\mu \rightarrow \int f d\mu$.*

¹Si dice che una funzione f ha una proprietà *quasi ovunque* se f ha questa proprietà su tutti i punti del suo dominio, eccetto al massimo un insieme nullo di punti, ovvero che ha misura pari a zero. In probabilità si usa invece l'espressione *quasi certamente*.

La convergenza dell'integrale dunque vale per una sequenza di funzioni misurabili che converge quasi ovunque che sia dominata da una funzione integrabile.

È opportuno soffermarsi ancora sull'integrale di Lebesgue per capire meglio come è costruito. Il discorso in realtà vale anche per la costruzione dell'integrale di una funzione rispetto a una misura generica.

L'integrale di una funzione a valori reali qualsiasi si è costruito per approssimazione, partendo da funzioni semplici non negative. L'integrale di una funzione semplice è la somma di un numero finito di valori: questi valori si ottengono moltiplicando il valore assunto dalla funzione in un sottoinsieme del dominio per la misura di quel sottoinsieme. In questo modo il valore totale dell'integrale si è ottenuto operando una divisione della funzione in "pezzi", che poi vengono sommati; la divisione non è avvenuta sul dominio della funzione, ma sull'insieme immagine: i sottoinsiemi dei quali si somma la misura infatti sono le controimmagini dei valori assunti dalla funzione. Nel caso dell'integrale di Lebesgue, questi sottoinsiemi sono intervalli di \mathbb{R} .

Per ottenere l'integrale di Lebesgue di una funzione a valori reali, dunque, si divide l'insieme immagine in piccoli intervalli; si moltiplica la misura (di Lebesgue) di ognuna delle controimmagini di questi intervalli per un valore assunto dalla funzione in quell'intervallo, e si sommano i valori ottenuti. Questa somma è un'approssimazione dell'integrale, che sarà sempre più precisa all'aumentare del numero di intervalli in cui si divide l'insieme immagine.

L'integrale di Lebesgue si differenzia sotto questo aspetto dall'integrale di Riemann, che opera una suddivisione del dominio della funzione (e che non è un argomento trattato in questa tesi) e fu Henri Lebesgue, nel 1902, a presentare questa divisione dell'immagine della funzione invece che del suo dominio, per rendere il concetto di integrale applicabile a funzioni più generali rispetto a quelle integrabili secondo Riemann.

Infine, per come è definito l'integrale è chiaro vedere che la misurabilità di una funzione è necessaria per la sua integrabilità (infatti si è partiti da funzioni semplici definite su insiemi misurabili). Nel caso di funzioni limitate definite su insiemi di misura finita, la misurabilità della funzione è anche condizione sufficiente perché essa sia integrabile:

Proposizione 4. *Data una funzione f limitata e definita su un insieme di misura finita, allora f è integrabile se e solo se f è misurabile.*

2.3 Continuità Assoluta e Singolarità

Si illustrano ora delle importanti proprietà che una misura può avere rispetto ad un'altra. Partiamo con il concetto di continuità assoluta.

Definizione 14. Date due misure μ e ν su uno spazio misurabile (X, \mathcal{A}) , si dice che ν è *assolutamente continua rispetto a μ* e si indica con $\nu \ll \mu$ se per ogni $A \in \mathcal{A}$ tale che $\mu(A) = 0$ allora $\nu(A) = 0$.

Se la misura ν è finita, la continuità assoluta rispetto alla misura μ è equivalente a un'altra condizione che chiaramente è una forma di continuità: $\nu \ll \mu$ se e solo se per ogni $\varepsilon > 0$ esiste un $\delta > 0$ tale che per ogni $A \in \mathcal{A}$ che soddisfa $\mu(A) < \delta$ si ha che $\nu(A) < \varepsilon$, ovvero $\nu(A) \rightarrow 0$ se $\mu(A) \rightarrow 0$.

Un esempio notevole è il seguente. Data una funzione non negativa f integrabile rispetto alla misura μ , si definisce $\beta(B) = \int_B f d\mu$. Si può dimostrare che β è una misura finita, e che inoltre $\beta(B) \rightarrow 0$ se $\mu(B) \rightarrow 0$, ovvero $\beta \ll \mu$. Infatti se $\mu(B) = 0$, allora $f \mathcal{I}_B$ è quasi ovunque nulla e dunque $\beta(B) = 0$.

In realtà, tutte le misure finite assolutamente continue rispetto a una misura σ -finita μ possono essere costruite come integrali; questo è garantito dal teorema di Radon-Nikodym.

Teorema 5 (di Radon-Nikodym). *Date β e μ misure σ -finite su uno spazio misurabile (X, \mathcal{A}) tali che $\beta \ll \mu$, esiste un'unica funzione $f : X \rightarrow [0, +\infty]$ misurabile rispetto a μ definita su X tale che*

$$\beta(B) = \int_B f d\mu$$

per ogni $B \in \mathcal{A}$, ovvero $d\beta = f d\mu$.

La funzione f appena menzionata viene detta *derivata di Radon-Nikodym* di β rispetto a μ , e si scrive $f = \frac{d\beta}{d\mu}$.

Vi è una proprietà che descrive la relazione opposta rispetto alla continuità assoluta tra due misure: la singolarità. Due misure infatti sono tra loro singolari quando una misura è concentrata su un insieme e l'altra è concentrata sul suo complementare.

Definizione 15. Date due misure μ e ν su uno spazio misurabile (X, \mathcal{A}) , si dice che μ è *singolare* rispetto a ν (o viceversa, o che μ e ν sono *mutuamente singolari*), e si indica con $\mu \perp \nu$, se esistono A, B insiemi disgiunti misurabili tali che $A \cup B = X$ e $\mu(A) = \nu(B) = 0$.

Per esempio, la misura di Lebesgue sull'intervallo $(0, 1)$ e la misura di Dirac in $x = \frac{1}{2}$ sono mutuamente singolari: se si pensa a una misura come a una distribuzione di massa su un insieme, la misura di Lebesgue è una distribuzione di massa uniforme su tutto l'insieme, mentre la misura di Dirac in x concentra tutta la massa su un singolo punto, ovvero l'insieme $\{x\}$, che ha misura di Lebesgue pari a 0.

A questo punto si può introdurre la *decomposizione di Lebesgue* di una misura, ovvero la decomposizione della misura in una parte singolare e una assolutamente continua rispetto a una misura di riferimento.

Teorema 6. *Date due misure σ -finite μ e ν su (X, \mathcal{A}) , esistono e sono uniche due misure σ -finite ν_a e ν_s su (X, \mathcal{A}) tali che $\nu_a \ll \mu$, $\nu_s \perp \mu$, e ν può essere scritta come somma $\nu = \nu_a + \nu_s$.*

Mettendo insieme i risultati presentati in questo paragrafo si arriva alla conclusione che una misura σ -finita ν può essere scomposta nella somma di due misure, una assolutamente continua e una singolare rispetto a una misura di riferimento μ , e che la parte assolutamente continua può essere ricavata come integrale di una funzione non negativa misurabile rispetto alla misura μ ; dunque, dato un insieme $A \in \mathcal{A}$, si può scrivere:

$$\nu(A) = \nu_a(A) + \nu_s(A) = \int_A f \, d\mu + \nu_s(A),$$

dove f è la derivata di Radon-Nikodym di ν_a rispetto a μ .

Capitolo 3

In questo capitolo si utilizzano i concetti introdotti finora per parlare in maniera rigorosa di variabili casuali e loro distribuzione e proprietà, riconducendo elementi di teoria della probabilità a nozioni più generali di teoria della misura. Si riportano risultati noti che si possono trovare in [5], [6] e [8].

3.1 Variabili Casuali e Distribuzione

Sia dato uno spazio di probabilità (Ω, \mathcal{F}, P) . Il termine *variabile casuale* si riferisce a una funzione misurabile; ovvero la funzione $X : \Omega \rightarrow \mathbb{R}$ è una variabile casuale se per ogni $a \in \mathbb{R}$ l'insieme $X^{-1}([a, \infty)) = \{\omega \in \Omega : X(\omega) \geq a\}$ appartiene a \mathcal{F} . Se $\Omega \subset \mathbb{R}$ e $\mathcal{F} = \mathcal{B}(\Omega)$, le variabili casuali sono le funzioni boreliane da \mathbb{R} in \mathbb{R} . Nella pratica, il significato di variabile casuale è il seguente: l'insieme Ω rappresenta i possibili esiti di un esperimento casuale, che si possono osservare tramite varie misurazioni; queste assegnano valori numerici agli esiti dell'esperimento. Una variabile casuale dunque rappresenta una misurazione, o una quantità d'interesse, il cui valore dipende dall'esito dell'esperimento.

Una variabile casuale X genera una σ -algebra: questa è definita da $X^{-1}(\mathcal{B}(\mathbb{R})) = \{S \subset \mathcal{F} : S = X^{-1}(B) \text{ per qualche } B \in \mathcal{B}(\mathbb{R})\}$ e si denota con \mathcal{F}_X . \mathcal{F}_X è la più piccola σ -algebra che contiene le immagini inverse di tutti gli insiemi boreliani. \mathcal{F}_X è un sottoinsieme di \mathcal{F} ma può essere molto più piccolo: la grandezza di \mathcal{F}_X dipende dal grado di complessità di X . Infatti, nel caso più semplice in cui X è costante ($X \equiv a$), $X^{-1}(B)$ può essere Ω o \emptyset a seconda che $a \in B$ o no; dunque la σ -algebra generata da X è composta da soli due elementi: $\mathcal{F}_X = \{\emptyset, \Omega\}$. Se X assume due valori $a \neq b$, \mathcal{F}_X conterrà quattro elementi: $\mathcal{F}_X = \{\emptyset, \Omega, X^{-1}(\{a\}), X^{-1}(\{b\})\}$. La grandezza della σ -algebra generata dunque cresce con il livello di complessità della variabile casuale. In pratica, \mathcal{F}_X rappresenta la quantità di

informazioni fornite dalla variabile casuale: i valori delle misurazioni (ovvero della variabile casuale) sono tutto ciò che è possibile osservare, ed è da questi che si deduce la complessità dell'esperimento, ovvero la grandezza di Ω e \mathcal{F}_X .

Dati uno spazio di probabilità (Ω, \mathcal{F}, P) , uno spazio di misura (Ω', \mathcal{F}') e una funzione misurabile $\varphi : \Omega \rightarrow \Omega'$, allora la misura P induce una *misura immagine* P_φ su Ω' definita come $P_\varphi(E) = P(\varphi^{-1}(E))$.

Il risultato appena presentato è importante per definire la distribuzione di probabilità di una variabile casuale X . Infatti se $\varphi = X$, $\Omega' = \mathbb{R}$, $\mathcal{F}' = \mathcal{B}(\mathbb{R})$ la misura indotta P_X è la *distribuzione di probabilità* della variabile casuale X :

$$P_X(B) = P(X^{-1}(B)) \text{ con } B \in \mathcal{B}(\mathbb{R}).$$

Essendo P_X una misura, ed essendo $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$, è chiaro che $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ è anch'esso uno spazio di probabilità.

La *funzione di distribuzione (cumulata)* $F_X : \mathbb{R} \rightarrow \mathbb{R}$ di una variabile casuale X invece è definita nel modo seguente:

$$F_X(x) = P_X((-\infty, x]) = P(\{\omega : X(\omega) \leq x\}).$$

F_X è limitata, non decrescente, continua da destra, e tende a 0 per x che tende a $-\infty$ e a 1 per x che tende a $+\infty$.

Se $\{X_\alpha\}_{\alpha \in A}$ è una famiglia di variabili casuali tali che $P_{X_\alpha} = P_{X_\beta}$ per ogni $\alpha, \beta \in A$, allora le X_α si dicono essere *identicamente distribuite*. Inoltre, data una sequenza finita di variabili casuali X_1, \dots, X_n , allora $X = (X_1, \dots, X_n)$ è una variabile casuale da Ω a \mathbb{R}^n , e la misura $P_X = P_{(X_1, \dots, X_n)}$ su \mathbb{R}^n viene detta *distribuzione congiunta* di X_1, \dots, X_n .

Ricapitoliamo dunque quanto detto finora sugli elementi fondamentali del calcolo delle probabilità. Dato uno spazio di probabilità (Ω, \mathcal{F}, P) , la σ -algebra \mathcal{F} è lo spazio degli eventi (ovvero gli insiemi misurabili); la misura P è una funzione che ad ogni evento di \mathcal{F} assegna un valore reale non negativo, e in particolare assegna all'insieme $\Omega \in \mathcal{F}$ il valore 1; questo valore rappresenta la probabilità che si verifichi l'evento in questione. Una variabile casuale X è una funzione definita su Ω a valori in \mathbb{R} misurabile, ovvero è tale che la controimmagine di insiemi del tipo $[a, +\infty)$, $a \in \mathbb{R}$, appartenga a \mathcal{F} e dunque sia un evento; nella pratica rappresenta la misurazione di un fenomeno osservato. La distribuzione di probabilità di una variabile casuale, P_X , è una misura (e dunque una funzione) che assegna un valore ai sottoinsiemi dell'immagine di X , che sono sottoinsiemi di \mathbb{R} , in particolare ai sottoinsiemi boreliani $B \subset \mathcal{B}(\mathbb{R})$; questo valore corrisponde alla misura di probabilità

della controimmagine dell'insieme, ovvero di $X^{-1}(B)$, che è un sottoinsieme di Ω (e di \mathcal{F}): $P_X(B) = P(X^{-1}(B))$. In termini più informali, si può dire che la distribuzione di probabilità dice quanto peso la variabile casuale dà a un insieme B in termini di misura di probabilità, ovvero quanto è grande la sua controimmagine, o in altre parole, quanto è probabile l'evento a cui corrispondono i valori nell'insieme B .

In probabilità spesso si utilizza una notazione abbreviata per indicare la controimmagine di un insieme: per esempio $X^{-1}([0, +\infty)) = \{\omega \in \Omega : X(\omega) \geq 0\}$ si scrive semplicemente $\{X \geq 0\}$. Inoltre si scrive $P(X \geq 0)$ invece che $P(\{X \geq 0\})$. D'ora in avanti si utilizzerà questa notazione.

Infine, una variabile casuale X si dice *discreta* se l'insieme dei possibili valori (detto *supporto*) è finito o numerabile; se X è discreta, la sua distribuzione P_X è una misura discreta, ovvero è concentrata su un insieme discreto (finito o numerabile): esiste un insieme discreto $C \in \mathbb{R}$ tale che $P_X(C) = 1$; è chiaro vedere che in questo caso P_X è singolare rispetto alla misura di Lebesgue λ . Invece si dice che X è una variabile casuale *continua* se il suo supporto ha la potenza del continuo; in questo caso la distribuzione di X è continua, e dunque P_X è assolutamente continua rispetto a λ .

3.2 Alcuni Esempi

3.2.1

Vediamo ora un esempio nel quale si applicano i concetti visti finora: partendo da una variabile casuale costante, vediamo come è definita la distribuzione di probabilità di una generica variabile casuale discreta.

Partiamo prendendo una variabile casuale costante: $X \equiv a$, $a \in \mathbb{R}$. Come si è già visto, in questo caso le possibili controimmagini di un insieme B sono due, l'insieme Ω se a è in B o l'insieme vuoto se non lo è. Dunque la distribuzione di probabilità di X assumerà solo due valori: 1 nel primo caso, 0 nel secondo. In questo caso quindi P_X è la misura di Dirac concentrata in a :

$$P_X(B) = \delta_a(B) = \begin{cases} 1 & \text{se } a \in B, \\ 0 & \text{se } a \notin B. \end{cases}$$

La funzione di distribuzione cumulata sarà la seguente:

$$F_X(x) = \begin{cases} 0 & \text{per } x < a, \\ 1 & \text{per } x \geq a. \end{cases}$$

Successivamente, vediamo cosa succede se X assume due valori, a e b , e l'insieme su cui X assume valore a ha misura di probabilità p , mentre l'insieme su cui X vale b ha misura di probabilità $1-p$, ovvero $P(X = a) = p$, $P(X = b) = 1 - p$. Questo si può scrivere nel seguente modo:

$$X(\omega) = \begin{cases} a & \text{con probabilità } p, \\ b & \text{con probabilità } 1 - p. \end{cases}$$

La distribuzione di probabilità di X in questo caso è la seguente:

$$P_X(B) = \begin{cases} 1 & \text{se } a, b \in B, \\ p & \text{se } a \in B, b \notin B, \\ 1 - p & \text{se } b \in B, a \notin B, \\ 0 & \text{altrimenti,} \end{cases}$$

e si può scrivere $P_X(B) = p\delta_a(B) + (1 - p)\delta_b(B)$.

Analogamente, la distribuzione di una variabile casuale discreta X che assume valori a_i con probabilità $p_i > 0$, $i = 1, 2, \dots$, $\sum_{i=1}^{\infty} p_i = 1$, ha la forma

$$P_X(B) = \sum_{i=1}^{\infty} p_i \delta_{a_i}(B).$$

In questo caso, p_i indica la misura di probabilità della controimmagine del valore a_i : $p_i = P(X = a_i)$.

Per esempio, nel caso della distribuzione di Poisson si ha che $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$. Nella distribuzione geometrica invece $p_i = (1 - q)q^i$ per qualche $q \in (0, 1)$.

3.2.2

Si illustra ora un altro esempio, stavolta di un comune esperimento casuale nel quale si cerca di riconoscere gli elementi teorici definiti.

Si suppone di lanciare una moneta equilibrata (ovvero tale che la probabilità di ottenere testa sia $\frac{1}{2}$) due volte. Vi sono quattro possibili esiti: si

ottengono due teste (indichiamo questo esito con TT), si ottengono due croci (CC), si ottiene prima una testa e poi una croce (TC), o si ottiene prima una croce e poi una testa (CT). L'insieme degli eventi elementari in questo caso è $\Omega = \{TT, CC, TC, CT\}$. In questo contesto ha senso prendere una σ -algebra \mathcal{F} tale che contenga tutti i sottoinsiemi di Ω ; uno di questi per esempio è $\{TC, CT\}$, ovvero l'evento nel quale si ottiene esattamente una testa nei due lanci. Definiamo la misura di probabilità P su \mathcal{F} tale che valga $\frac{1}{4}$ per ognuno dei quattro eventi elementari, mentre per ogni evento in \mathcal{F} sarà uguale al numero di eventi elementari al suo interno moltiplicato per $\frac{1}{4}$; per esempio, $P(\{TC, CT\}) = \frac{1}{2}$. Notiamo che, come deve essere, $P(\Omega) = 1$. Abbiamo dunque definito uno spazio di probabilità (Ω, \mathcal{F}, P) .

Il numero di teste uscite nei due lanci si può rappresentare con una variabile casuale X , definita come segue:

$$X(\omega) = \begin{cases} 0 & \text{se } \omega = CC, \\ 1 & \text{se } \omega = TC \text{ o } \omega = CT, \\ 2 & \text{se } \omega = TT. \end{cases}$$

La distribuzione di probabilità di X allora è

$$\begin{aligned} P_X &= P(X = 0)\delta_0 + P(X = 1)\delta_1 + P(X = 2)\delta_2 \\ &= P(\{CC\})\delta_0 + P(\{TC, CT\})\delta_1 + P(\{TT\})\delta_2 \\ &= \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2. \end{aligned}$$

Supponiamo invece di lanciare la moneta ripetutamente e di fermarci dopo aver ottenuto la prima testa. Stavolta lo spazio degli eventi elementari contiene infiniti elementi (è comunque numerabile): $\Omega = \{T, CT, CCT, CCCT, \dots\}$. Anche in questo caso è ragionevole assumere $\mathcal{F} = \mathcal{P}(\Omega)$. Definiamo P ponendo $P(\{T\}) = \frac{1}{2}$, $P(\{CT\}) = \frac{1}{4}$, $P(\{CCT\}) = \frac{1}{8}$, e così via. Le probabilità degli eventi di \mathcal{F} saranno determinate per σ -additività dai valori delle probabilità degli eventi elementari. La misura P così definita è una misura di probabilità, infatti la somma della serie geometrica $\sum_{n=1}^{\infty} \frac{1}{2^n}$ è 1.

3.2.3

Ora invece supponiamo di scegliere un numero reale all'interno dell'intervallo $[a, b]$ in modo tale che la probabilità che il numero scelto appartenga a un sottointervallo di $[a, b]$ sia proporzionale alla sua lunghezza. Si può descrivere questa situazione con lo spazio di probabilità $([a, b], \mathcal{B}[a, b], P)$ dove la misura P è data da $P(A) = c\lambda(A)$, dove λ è la misura di Lebesgue e

$c = \frac{1}{\lambda([a,b])} = \frac{1}{b-a}$. Chiaramente $P([a,b]) = 1$. In questo caso si ha una *distribuzione uniforme* su $[a,b]$. Se l'intervallo $[a,b]$ è $[0,1]$, allora P è la misura di Lebesgue ristretta all'intervallo $[0,1]$, della quale si è parlato alla fine del primo capitolo.

In questo caso, ovvero quando $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $P = \lambda_{[0,1]}$, possiamo fare un esempio di variabile casuale: sia $X(\omega) = a\omega + b$, con $a, b \in \mathbb{R}$. L'insieme immagine di X , $X([0,1])$, è l'intervallo $[b, a+b]$, e la distribuzione di probabilità di X è $P_X = \frac{1}{a}\lambda_{[b,a+b]}$, ovvero $P_X(B) = \frac{\lambda(B \cap [b,a+b])}{a}$.

3.3 Integrazione Rispetto a Distribuzioni

Sia data una variabile casuale X definita su uno spazio di probabilità (Ω, \mathcal{F}, P) con distribuzione P_X . Si riporta un teorema che permette di cambiare la misura di un integrale di una funzione di X . Questo è importante quando si applica la teoria degli integrali alle probabilità, ed è utile quando si conosce la forma di P_X .

Teorema 7. *Data una variabile casuale $X : \Omega \rightarrow \mathbb{R}$, si ha che*

$$\int_{\Omega} g(X(\omega)) dP(\omega) = \int_{\mathbb{R}} g(x) dP_X(x).$$

Dunque si è passati dalla misura $P(\omega)$ alla misura $P_X(x)$ dove $x = X(\omega)$.

Si illustra la formula presentata con un semplice esempio. Sia X una variabile casuale costante ($X \equiv a$); allora il membro a sinistra della formula è l'integrale di una funzione costante: $\int_{\Omega} g(a) dP(\omega) = g(a)P(\Omega) = g(a)$; a destra si ha che $P_X = \delta_a$. In questo modo si è definito un metodo per integrare rispetto alla misura di Dirac: $\int_{\mathbb{R}} g(x) d\delta_a(x) = g(a)$.

Se X è una variabile casuale discreta che assume valori a_i con probabilità p_i la sua funzione di distribuzione, come si è già visto, è $P_X = \sum_i p_i \delta_{a_i}$. Si può dimostrare che l'integrale di una funzione rispetto a una combinazione lineare di misure è la combinazione lineare degli integrali della funzione rispetto alle misure, dunque il membro a destra dell'equazione in questo caso diventa $\int g(x) dP_X(x) = \sum_i p_i \int g(x) d\delta_{a_i}(x)$. Dunque per il Teorema 7 si ottiene una nota formula della teoria della probabilità:

$$\int_{\Omega} g(X(\omega)) dP(\omega) = \sum_i g(a_i) p_i.$$

3.4 Misure Assolutamente Continue e Densità

Si è già visto, grazie al teorema di Radon-Nikodym, che se una misura P è assolutamente continua rispetto a una misura λ , allora esiste una funzione integrabile non negativa f tale che $P(A) = \int_A f d\lambda$. Se λ è la misura di Lebesgue, P si dice semplicemente *assolutamente continua*; perché P sia una misura di probabilità chiaramente è necessario imporre la condizione $\int f d\lambda = 1$.

La funzione f si dice *densità di P rispetto alla misura di Lebesgue*, o semplicemente *densità di P* (o di una variabile casuale che ha distribuzione di probabilità P).

Come si è già visto, se una variabile casuale X è continua, allora la sua distribuzione di probabilità P_X è assolutamente continua, e dunque X ha una densità f_X ; dato B sottoinsieme boreliano di \mathbb{R} , si ha che:

$$P_X(B) = P(X \in B) = \int_B f_X d\lambda = \int_B f_X(x) dx.$$

Dunque anche la misura di probabilità P dello spazio su cui è definita X è assolutamente continua.

Il seguente risultato è fondamentale per calcolare integrali relativi a distribuzioni assolutamente continue:

Teorema 8. *Se P_X è una distribuzione definita su \mathbb{R} con densità f_X e $g : \mathbb{R} \rightarrow \mathbb{R}$ è integrabile rispetto a P , allora*

$$\int_{\mathbb{R}} g(x) dP_X(x) = \int_{\mathbb{R}} f_X(x)g(x) dx.$$

La funzione di distribuzione cumulata di una variabile casuale continua si può esprimere tramite la sua densità:

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]) = \int_{-\infty}^x f_X(x) dx.$$

Vediamo ora qualche esempio di densità.

L'esempio più semplice è la densità di una variabile casuale con distribuzione *uniforme*. Se $\Omega \subset \mathcal{B}(\mathbb{R})$ ha misura di Lebesgue finita, allora la densità è definita come segue:

$$f_X(x) = \begin{cases} \frac{1}{\lambda(\Omega)} & \text{se } x \in \Omega, \\ 0 & \text{altrimenti.} \end{cases}$$

Questo caso è già stato introdotto nell'esempio della Sezione 3.2.3; la probabilità di un insieme B si ottiene come $P(X \in B) = \frac{\lambda(B)}{\lambda(\Omega)}$.

Se per esempio Ω è l'intervallo $[a, b]$, la densità di X è $f_X = \frac{1}{b-a}$ su $[a, b]$ e 0 altrove, e $P(X \in B) = \frac{\lambda(B)}{b-a}$.

La densità più famosa è sicuramente quella di una variabile casuale con distribuzione *normale* o *Gaussiana*:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ dove } \mu \in \mathbb{R}, \sigma \in (0, \infty).$$

In questo caso f_X è simmetrica rispetto a $x = \mu$ e il limite per x che tende a $\pm\infty$ è 0.

La densità di una variabile casuale con distribuzione esponenziale è data da:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0, \\ 0 & \text{altrimenti,} \end{cases} \text{ dove } \lambda \in (0, \infty).$$

Infine è importante notare che, come si è visto, una variabile casuale può essere discreta o continua, ma non necessariamente una delle due. Si possono per esempio definire variabili casuali che sono una mistura delle due; in questo caso la distribuzione di probabilità sarà una misura né singolare, né assolutamente continua rispetto a Lebesgue, ma si potrà scomporre come nel Teorema 6.

Un esempio è il seguente: si supponga che una macchina lasci la città A casualmente tra le ore 12 e le 13. Essa viaggia con una velocità di 50 km/h verso la città B , che dista 25 km dalla città A . Si rappresenta la distanza tra la macchina e la città B alle ore 13 con una variabile casuale $X(\omega)$, dove $\omega \in [0, 1]$ rappresenta il tempo di partenza:

$$X(\omega) = \begin{cases} 0 & \text{se } \omega \in [0, \frac{1}{2}], \\ 50\omega - 25 & \text{se } \omega \in (\frac{1}{2}, 1]; \end{cases}$$

infatti chiaramente la distanza è 0 con probabilità $\frac{1}{2}$, ovvero se la macchina parte prima delle 12.30. La variabile casuale in questione ha una distribuzione di probabilità che non è né discreta, né assolutamente continua:

$$P_X = \frac{1}{2}P_1 + \frac{1}{2}P_2, \text{ dove } P_1 = \delta_0, P_2 = \frac{1}{25}\lambda_{[0,25]}.$$

P_X è dunque una combinazione della misura di Dirac in 0, che è singolare rispetto a Lebesgue, e la misura di Lebesgue ristretta all'intervallo $[0, 25]$, chiaramente assolutamente continua.

3.5 Indipendenza

La teoria della probabilità è caratterizzata da un concetto molto importante, quello di indipendenza. Si consideri uno spazio di probabilità (Ω, \mathcal{F}, P) e un evento E tale che $P(E) > 0$. Allora, per $F \in \mathcal{F}$, la funzione $P_E(F) = P(E|F) = P(E \cap F)/P(E)$, è una misura di probabilità su Ω chiamata *probabilità condizionata rispetto a E* ; $P(E|F)$ rappresenta la probabilità che si verifichi l'evento F , dato che si è verificato l'evento E . F è indipendente da E se la probabilità che si verifichi F non cambia se si verifica o meno E , ovvero $P(F|E) = P(F)$. La definizione formale è la seguente.

Definizione 16. Due eventi E, F si dicono *indipendenti* se

$$P(E \cap F) = P(E)P(F).$$

Questa definizione è simmetrica rispetto ad E e F e inoltre non richiede l'assunzione che $P(E) > 0$. Il concetto si può estendere a un numero finito arbitrario di eventi nel modo seguente.

Definizione 17. Gli eventi della collezione $\{E_\alpha\}_{\alpha \in A}$ sono *indipendenti* se, per ogni $n \in \mathbb{N}$ e per ogni $\alpha_1, \dots, \alpha_n \in A$, si ha che

$$P(E_{\alpha_1} \cap \dots \cap E_{\alpha_n}) = \prod_{j=1}^n P(E_{\alpha_j}).$$

Si può dimostrare che se un evento E è indipendente da F , allora lo è anche il suo complementare E^c ; da questo si deriva che se due eventi sono indipendenti, allora tutti gli elementi delle σ -algebre da essi generate sono indipendenti, visto che queste σ -algebre sono semplicemente $\mathcal{F}_E = \{\emptyset, E, E^c, \Omega\}$ e $\mathcal{F}_F = \{\emptyset, F, F^c, \Omega\}$. Si può dunque estendere la definizione di indipendenza anche alle σ -algebre: due σ -algebre \mathcal{F}_1 e \mathcal{F}_2 sono indipendenti se per ogni coppia di insiemi $E_1 \in \mathcal{F}_1$ e $E_2 \in \mathcal{F}_2$ si ha che $P(E_1 \cap E_2) = P(E_1)P(E_2)$.

Si può parlare di indipendenza anche tra variabili casuali:

Definizione 18. Le variabili casuali della collezione $\{X_\alpha\}_{\alpha \in A}$ si dicono *indipendenti* se gli eventi $\{X_\alpha \in B_\alpha\} = X_\alpha^{-1}(B_\alpha)$ sono indipendenti per ogni $B_\alpha \subset \mathcal{B}(\mathbb{R})$.

Si può dimostrare che le variabili casuali di una collezione sono indipendenti se e solo se la distribuzione congiunta di ogni sottoinsieme finito di variabili della collezione è uguale al prodotto delle distribuzioni individuali delle variabili. Si mostra questa condizione nel caso più semplice, ovvero con due variabili casuali X e Y : se X e Y sono indipendenti, allora per ogni coppia di insiemi $B, C \subset \mathcal{B}(\mathbb{R})$, $X^{-1}(B)$ e $X^{-1}(C)$ sono indipendenti, ovvero $P(X^{-1}(B) \cap Y^{-1}(C)) = P(X^{-1}(B))P(X^{-1}(C))$, che, per come si è definita la distribuzione di una variabile casuale, si può scrivere come $P_{(X,Y)}(B, C) = P_X(B)P_Y(C)$.

Infine, un insieme di variabili casuali è indipendente se e solo se le σ -algebre generate da ogni variabile casuale sono indipendenti.

3.6 Valore Atteso e Varianza

Infine, per concludere il discorso sulla teoria della probabilità, vediamo brevemente cosa sono e come si ottengono il valore atteso di una variabile casuale e la sua varianza.

Data una variabile casuale X definita sullo spazio di probabilità (Ω, \mathcal{F}, P) ,

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

si dice *valore atteso* o *speranza matematica* di X . Se X è integrabile, si dice che X ha valore atteso finito o semplicemente che ha valore atteso.

Per il Teorema 7, il valore atteso si può scrivere anche come integrale rispetto alla distribuzione di probabilità: $E(X) = \int_{\mathbb{R}} x \, dP_X(x)$. Nel caso di una variabile casuale discreta che assume valori x_i con $i = 1, \dots, n$, per i risultati visti nella Sezione 3.3, si avrà che

$$E(X) = \sum_i x_i P(X = x_i).$$

Invece se X è assolutamente continua con densità f_X , per il Teorema 8, il suo valore atteso si calcolerà come

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

Il valore atteso del quadrato di una variabile casuale X , $E(X^2)$, si dice *momento secondo* di X ; se X ha momento secondo finito, ha anche valore atteso finito, e in questo caso si definisce la *varianza* della variabile X come $\text{Var}(X) = E[(X - E(X))^2]$. Si dimostra che $\text{Var}(X) = E(X^2) - (E(X))^2$.

3.7 Esempi Riassuntivi

In conclusione, si presentano alcuni esempi nei quali si prendono in considerazione alcune variabili casuali e si illustrano gli elementi trattati in questa tesi; si mostra dunque lo spazio di probabilità su cui sono definite, la loro distribuzione, se sono discrete o assolutamente continue.

3.7.1

Per prima cosa vediamo un semplice esperimento casuale, ovvero il lancio di un dado equilibrato a sei facce, e vediamo qualche esempio di variabile casuale discreta a valori finiti. Vedremo che anche con un esperimento semplice come il lancio di un dado si possono definire diverse variabili casuali.

Innanzitutto, lo spazio degli eventi elementari è $\Omega = \{1, 2, 3, 4, 5, 6\}$. È ragionevole prendere come σ -algebra quella che contiene tutti i possibili sottoinsiemi di Ω : $\mathcal{F} = \mathcal{P}(\Omega)$. Intuitivamente, si definisce la misura di probabilità $P : \mathcal{F} \rightarrow [0, +\infty)$ nel modo seguente: $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$; la misura di probabilità di insiemi con più di un elemento sarà tale da rispettare il requisito di σ -additività di P , dunque per esempio $P(\{1, 2\}) = P(\{1\}) + P(\{2\}) = \frac{1}{3}$. Chiaramente $P(\Omega) = 1$; P è concentrata sull'insieme discreto $\{1, 2, 3, 4, 5, 6\}$, e dunque è singolare rispetto alla misura di Lebesgue. Abbiamo dunque definito uno spazio di probabilità (Ω, \mathcal{F}, P) .

Definiamo una variabile casuale $X : \Omega \rightarrow \mathbb{R}$:

$$X(\omega) = \begin{cases} 1 & \text{se } \omega \text{ è pari,} \\ 0 & \text{se } \omega \text{ è dispari.} \end{cases}$$

In questo caso le informazioni fornite dalla variabile casuale sono poche, infatti l'esperimento è indistinguibile dal lancio di una moneta; questo si vede anche dalla σ -algebra generata da X , che contiene solo quattro elementi: $\mathcal{F}_X = \{\emptyset, \Omega, X^{-1}(1) = \{2, 4, 6\}, X^{-1}(0) = \{1, 3, 5\}\}$.

Si ha che $P(\{2, 4, 6\}) = P(X = 1) = P_X(1) = \frac{1}{2}$ e analogamente $P(\{1, 3, 5\}) = P(X = 0) = P_X(0) = \frac{1}{2}$; dunque, per $B \subset \mathcal{B}(\mathbb{R})$:

$$P_X(B) = \begin{cases} 1 & \text{se } 0, 1 \in B, \\ \frac{1}{2} & \text{se } 0 \in B, 1 \notin B \text{ o } 1 \in B, 0 \notin B, \\ 0 & \text{altrimenti,} \end{cases}$$

ovvero $P_X(B) = \frac{1}{2}\delta_0(B) + \frac{1}{2}\delta_1(B)$.

Naturalmente $P_X(\mathbb{R}) = 1$. X è una variabile casuale discreta, infatti il suo supporto è discreto e di conseguenza la sua distribuzione di probabilità è singolare rispetto alla misura di Lebesgue. Il suo valore atteso è $E(X) = \sum_i x_i P(X = x_i) = 1P(X = 1) + 0P(X = 0) = \frac{1}{2}$.

Ora definiamo un'altra variabile casuale $Y : \Omega \rightarrow \mathbb{R}$ tale che $Y(\omega) = \omega$, ovvero Y vale 1 se lanciando il dado esce 1, 2 se esce 2 e così via. Questa variabile casuale è più complessa di quella precedente e fornisce più informazioni; infatti la σ -algebra generata in questo caso è più grande, e contiene tutti gli eventi elementari di Ω e tutte le loro possibili combinazioni, ovvero $\mathcal{F}_Y = \mathcal{F}$.

Si ha che $P_Y(1) = P_Y(2) = \dots = P_Y(6) = \frac{1}{6}$ e la distribuzione è tale che $P_Y(B) = \frac{1}{6}\delta_1(B) + \frac{1}{6}\delta_2(B) + \dots + \frac{1}{6}\delta_6(B)$ per ogni $B \subset \mathcal{B}(\mathbb{R})$. Anche qui è facile vedere che $P_Y(\mathbb{R}) = 1$, e che P_Y è singolare rispetto a Lebesgue. Si ottiene che $E(Y) = \sum_i y_i P(Y = y_i) = 1P(Y = 1) + \dots + 6P(Y = 6) = 3,5$.

Infine definiamo una variabile casuale $Z : \Omega \rightarrow \mathbb{R}$ che, a differenza di X e Y , assume valori diversi con diverse probabilità:

$$Z(\omega) = \begin{cases} 0 & \text{se } \omega \in \{1, 3, 5\}, \\ 1 & \text{se } \omega = 2 \text{ o } \omega = 4, \\ 2 & \text{se } \omega = 6. \end{cases}$$

Si può dire che Z è più complessa di X ma meno complessa di Y , infatti la σ -algebra generata da Z contiene più elementi rispetto a quella generata da X ma è più piccola di quella generata da Y : $\mathcal{F}_X \subset \mathcal{F}_Z \subset \mathcal{F}_Y = \mathcal{F} = \mathcal{P}(\Omega)$.

In questo caso si ha che $P_Z(0) = P(\{1, 3, 5\}) = \frac{1}{2}$, $P_Z(1) = P(\{2, 4\}) = \frac{1}{3}$, $P_Z(2) = P(\{6\}) = \frac{1}{6}$, e dunque la distribuzione di probabilità è definita come $P_Z(B) = \frac{1}{2}\delta_0 + \frac{1}{3}\delta_1 + \frac{1}{6}\delta_2$. Naturalmente $P_Z(\mathbb{R}) = 1$ e $P_Z \perp \lambda$. Z è discreta e il suo valore atteso è $E(Z) = \sum_i z_i P(Z = z_i) = \frac{2}{3}$.

3.7.2

Vediamo ora un esempio di variabile casuale discreta che assume un numero infinito di valori, ovvero una variabile casuale con distribuzione di Poisson. Essa solitamente viene usata per esprimere le probabilità per il numero di eventi che si verificano successivamente ed indipendentemente in un dato intervallo di tempo. Un caso pratico in cui questa variabile casuale viene applicata è, per esempio, il numero di automobili che passano in un determinato tratto di strada in un prefissato intervallo di tempo.

Solitamente, quando si utilizza una variabile casuale con distribuzione di Poisson è ragionevole assumere che $\Omega = \mathbb{N}$ e $\mathcal{F} = \mathcal{P}(\mathbb{N})$. La misura di probabilità P è definita in modo tale che, per ogni $n \in \mathbb{N}$, $P(\{n\}) = \frac{\lambda^n}{n!} e^{-\lambda}$, dove $\lambda \in (0, \infty)$ [1]. Sapendo che lo sviluppo in serie di e^λ è $e^\lambda = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$, si ha che $P(\Omega) = P(\mathbb{N}) = \sum_{n=0}^{\infty} P(n) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = 1$.

Una variabile casuale $X : \mathbb{N} \rightarrow \mathbb{R}$ con distribuzione di Poisson è definita nel modo seguente:

$$X(n) = n \text{ con probabilità } \frac{\lambda^n}{n!} e^{-\lambda}.$$

X è una variabile casuale discreta. La σ -algebra generata da X è \mathcal{F} , che in questo caso contiene infiniti elementi.

La distribuzione di probabilità di X calcolata in un singolo valore n è $P_X(n) = P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}$; come si è già mostrato nella Sezione 3.2.1, per un qualsiasi insieme $B \subset \mathcal{B}(\mathbb{R})$, la distribuzione di X è la seguente:

$$P_X(B) = \sum_{n=0}^{\infty} P(X = n) \delta_n(B) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \delta_n(B).$$

Come si è visto, $P_X(\mathbb{R}) = P(\mathbb{N}) = 1$. La misura P_X è concentrata sull'insieme \mathbb{N} , dunque è singolare rispetto alla misura di Lebesgue.

La funzione di distribuzione cumulata è la seguente:

$$F_X(n) = P_X((-\infty, n]) = P_X(\{0, 1, \dots, n\}) = \sum_{i=0}^n \frac{\lambda^i}{i!} e^{-\lambda}.$$

Il valore atteso di X è

$$E(X) = \sum_{n=0}^{\infty} n P(X = n) = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda.$$

3.7.3

Come ultimo esempio, si presenta una variabile casuale con distribuzione assolutamente continua: la distribuzione normale standard.

In questo caso, l'insieme Ω è tutta la retta reale; come σ -algebra si prende quella dei boreliani $\mathcal{B}(\mathbb{R})$. Sullo spazio misurabile $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si definisce una misura di probabilità P che soddisfi le seguenti condizioni:

- (i) $P(\omega) = 0$ per ogni $\omega \in \Omega = \mathbb{R}$,
- (ii) $P(B) \geq 0$ per ogni $B \subset \mathcal{B}(\mathbb{R})$,

(iii) $P(\mathbb{R}) = 1$.

Dato lo spazio di probabilità $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ appena descritto, si definisce una variabile casuale $X : \mathbb{R} \rightarrow \mathbb{R}$ che assegna ad ogni valore sulla retta reale un valore reale: $X(\omega) = x \in \mathbb{R}$. La sua distribuzione di probabilità è definita come:

(i) $P_X(x) = P(X^{-1}(x)) = P(\omega) = 0$ per ogni $x \in \mathbb{R}$,

(ii) $P_X(B) = P(X^{-1}(B)) = P(X \in B) = \int_B f_X dx$ per ogni $B \subset \mathcal{B}(\mathbb{R})$,

dove $f_X : \mathbb{R} \rightarrow [0, \infty)$ è la densità della variabile casuale, ovvero la derivata di Radon-Nikodym della sua distribuzione P_X rispetto alla misura di Lebesgue. P_X è dunque assolutamente continua rispetto a λ . Se $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, la variabile casuale ha distribuzione normale standard; considerando densità diverse si ottengono altre variabili casuali continue.

La funzione di distribuzione cumulata è:

$$F_X(x) = P_X((-\infty, x]) = \int_{-\infty}^x f_X dx.$$

Come si è già mostrato, il valore atteso di una variabile casuale continua X è $E(X) = \int_{\mathbb{R}} x f_X dx$, e nel caso della distribuzione normale standard si può dimostrare che questo integrale è uguale a zero.

Bibliografia

- [1] Agresti, Alan, "Foundations of Linear and Generalized Linear Models", John Wiley & Sons Inc, 2015
- [2] Bear, Herbert S., "A Primer of Lebesgue Integration", Academic Press, 1995
- [3] Boyer, Carl B., "Storia della Matematica", Mondadori, 1980
- [4] Cannarsa, Piermarco e D'Aprile, Teresa, "Introduzione alla Teoria della Misura e all'Analisi Funzionale", Springer, 2008
- [5] Capiński, Marek and Kopp, Ekkehard, "Measure, Integral and Probability", Springer, 1999
- [6] Cohn, Donald L., "Measure Theory", Birkhäuser, 2013
- [7] Dall'Aglio, Giorgio, "Calcolo delle Probabilità", Zanichelli, 2003
- [8] Folland, Gerald B., "Real Analysis", J. Wiley, 1999