La Scuola di dottorato di ricerca in Finanza
è stata istituita con decreto rettorale n. 1188
del 14 luglio 2006 ed ha il compito di promuovere,
organizzare e coordinare la gestione di progetti
formativi di livello dottorale nei campi scientifici
della finanza teorica e della finanza sperimentale
a carattere interdisciplinare e internazionale,
facenti parte dell'offerta formativa
di terzo livello dell' Ateneo.

La prima lezione della Scuola
è stata tenuta dal prof. Giampaolo De Ferra,
membro del Consiglio Scientifico,
il giorno 24 gennaio 2007.

ISBN 978-88-8303-221-9

EUT - Edizioni Università di Trieste
p.zzale Europa, 1 – 34127 Trieste
http://eut.units.it

Elementi di teoria delle decisioni statistiche

Silvano Holzer

EUT EDIZIONI UNIVERSITÀ DI TRIESTE

Quelli che s’innamoran di pratica sanza scienza, son come ‘l nocchiere, ch’entra in navilio sanza timone o bussola, che mai ha certezza dove si vada.

Leonardo da Vinci

Indice

Introduzione

L'attività umana è completamente permeata dall'incertezza, essendo l'uomo costantemente chiamato a prendere decisioni pur non conoscendone, in modo preciso, le relative conseguenze che possono essere più o meno gradite. Si pensi, ad esempio, ad una persona che sceglie la località ove trascorrere le ferie; ad un giornalaio che programma la quantità di giornali e riviste da acquistare; ad un medico che prescrive un farmaco per curare una data malattia; ad un giocatore interessato a comperare biglietti di lotterie. Nelle situazioni descritte gli individui si trovano di fronte a diverse opportunità di scelta (ad esempio, la persona può prendere in esame le diverse proposte di un tour operator; il giocatore trova disponibili biglietti di svariate lotterie). Sorge allora il problema di come scegliere tra le varie possibili alternative a disposizione, ricorrendo anche (qualora sia possibile e conveniente) ad informazioni di natura statistica che facciano luce su questioni connesse con il problema decisionale (ad esempio, per scegliere la località di villeggiatura, la persona potrebbe informarsi sulle previsioni del tempo relative alle varie scelte possibili; il giornalaio potrebbe effettuare una serie di osservazioni sul venduto giornaliero prima di decidere; il medico potrebbe far ricorso ai risultati della sperimentazione medica per valutare l'efficacia dei farmaci presi in considerazione).

Sin dall'antichità l'uomo ha cercato di formulare, nella sua "lotta contro l'incertezza", regole di comportamento sensate che gli consentissero di affrontare razionalmente il problema della scelta sopra esemplificato. A tale proposito, conviene ricordare che nel tempo l'individuazione di tali regole è stata notevolmente condizionata da due interpretazioni sulla loro natura: quella *normativa* e quella *descrittiva*. La prima, analizzando teoricamente i comportamenti di una persona ideale perfettamente razionale, tende ad identificarle come norme canoniche astratte; la seconda invece, partendo dall'osservazione "sul campo" dei soggetti, tende ad identificarle con il comportamento delle persone reali e "ragionevolmente" razionali.

Tralasciando le vicissitudini storiche che, alla luce di questa dicotomia, hanno condotto a modificare e/o reinterpretare via via le metodologie proposte per l'analisi delle scelte individuali in condizioni d'incertezza, ci limiteremo a presentare e ad analizzare, nei suoi aspetti basilari, il risultato finale "più popolare" di questo processo nell'ambito dei contesti non competitivi[1].

La sua costruzione deriva dalla "fusione" del *calcolo delle probabilità* (analisi quantitativa dei fenomeni aleatori) con la *teoria dell'utilità* (studio delle rappresentazioni numeriche delle preferenze individuali) e con la *statistica* (applicazioni del metodo induttivo concernenti osservazioni - possibilmente numerose - in qualche senso analoghe). Infatti, limitandoci ad una descrizione succinta, fa riferimento ai seguenti elementi fondamentali:

- l'insieme delle decisioni a disposizione del decisore;
- l'insieme degli stati di natura, dei quali l'unico "vero" è sconosciuto al decisore;
- la funzione di danno $\mathrm{L}(\theta, d)$ rappresentante la disutilità che il decisore subisce se prende la decisione d quando θ è lo stato di natura "vero";
- l'esperimento consistente nell'osservare un ente aleatorio - con distribuzione di probabilità dipendente dallo stato di natura "vero" - che potrebbe aiutare (sperabilmente) il decisore a contenere i danni;
- i "comportamenti" del decisore, cioè funzioni che ad ogni possibile osservazione associano una decisione;
- la funzione di rischio $R_\delta(\theta)$ che fornisce il danno medio che il decisore subisce se adotta il comportamento δ quando θ è lo stato di natura "vero",

e si basa sull'assunzione che l'obiettivo principale, se non l'unico, del decisore sia quello di perseguire un comportamento che comporti un rischio più piccolo

[1] Cioè, con riferimento a situazioni nelle quali le decisioni comportano conseguenze non dipendenti dalla volontà di individui diversi da quello chiamato a scegliere. Per chi fosse interessato anche ai contesti competitivi, consigliamo Aliprantis, C.D. - Chakrabarti, S.K., Games and Decision Making, Oxford University Press, Oxford (2000) che fornisce una ottima introduzione all'argomento e Osborne, M.J. - Rubinstein, A., A course in Game Theory, The MIT Press, Cambridge, Massachusetts (1994) per una trattazione avanzata.

possibile (in qualche senso[2]).

Si ottiene così una trattazione delle scelte individuali in condizione di incertezza che, presentando la metodologia statistica come una entità unica e omogenea, consente di annullare l'impressione, piuttosto diffusa, che la Statistica sia un agglomerato di tecniche più o meno slegate tra loro. Si potrà così comprendere appieno la logica sottostante le soluzioni "standard" che sono state proposte nell'ambito dei problemi statistici classici (e.g. di stima puntuale, di verifica di ipotesi, etc.).

Nella stesura del testo si è posta particolare cura agli aspetti logico-matematici cercando di mettere in piena luce sia le ipotesi (e quindi i limiti) che gli sviluppi formali sottostanti le problematiche statistiche considerate. Sono stati inseriti numerosi esempi per una comprensione migliore degli aspetti, sia teorici che applicativi, inerenti gli argomenti trattati. Inoltre, per rendere sufficientemente completa l'esposizione, sono state incluse delle parti opzionali (di norma scritte in carattere più piccolo) la cui lettura potrà essere omessa (e rimandata eventualmente ad un secondo momento) senza pregiudicare in alcun modo la comprensione della maggior parte del materiale esposto.

Si è ritenuto opportuno inserire anche due ampie appendici che forniscono sia un riferimento per il linguaggio e la simbologia usati, che il materiale - della teoria della misura e del calcolo delle probabilità - indispensabile per una piena comprensione del testo. A tale proposito ampio spazio è stato riservato, a causa della loro "delicatezza", alle nozioni generali di speranza matematica condizionata e di funzione di regressione (fornendone le proprietà principali, come pure alcuni esempi chiarificatori) che, come è ben noto, sono bagaglio indispensabile per chiunque voglia cimentarsi, oltre che nell'analisi statistica, anche nell'analisi quantitativa dei mercati finanziari[3].

Il testo è stato scritto seguendo il filo delle lezioni svolte da parecchi anni nel corso di Statistica Matematica (laurea in "Scienze Statistiche ed Attuariali") e, per quanto riguarda le appendici, nel corso progredito di Calcolo delle Probabilità (laurea in "Statistica e Informatica per l'Azienda, la Finanza e l'Assicurazione") tenuti presso la Facoltà di Economia dell'Università di Trieste. Alcuni degli argomenti trattati sono stati inclusi anche nel ciclo di

[2] Come potrebbe essere, ad esempio, quello di minimizzare il massimo rischio oppure il rischio medio calcolato rispetto a qualche distribuzione di probabilità sugli stati di natura.

[3] Essendo sviluppate in dettaglio le dimostrazioni dei risultati di teoria della misura e di probabilità considerati, queste appendici possono essere usate sia come utile ripasso per chi già ne conosca il contenuto che come una introduzione per chi lo incontrasse per la prima volta.

lezioni di Calcolo delle Probabilità e di Teoria delle Decisioni per gli studenti della Scuola di Dottorato di Ricerca in Finanza afferente alla Facoltà di Economia dell'Università di Trieste.

Per quanto riguarda la bibliografia, sono riportati solamente quei testi che forniscono, a nostro parere, una guida sicura e sufficientemente esaustiva per chiunque volesse approfondire le tematiche inerenti la moderna teoria delle decisioni statistiche. Nel corso della trattazione, sono riportati, a piè di pagina, anche alcuni testi non citati in bibliografia che forniscono risultati o impostazioni utili per alcuni argomenti specifici.

Concludiamo aggiungendo che giudizi, critiche e suggerimenti da parte dei diretti interessati (gli studenti) o da altri eventuali lettori saranno accolti con gratitudine.

Silvano Holzer

Trieste, dicembre 2007

Preliminari

Riportiamo le principali notazioni (come pure qualche precisazione terminologica) relative alla teoria degli insiemi e agli spazi numerici reali che saranno costantemente adoperate nel corso dell'esposizione.

Logica

- $\wedge$ connettivo logico di congiunzione.
- $\vee$ connettivo logico di disgiunzione (inclusiva).
- $\Rightarrow$ connettivo logico di implicazione.
- $\iff$ connettivo logico di biimplicazione.
- $\exists$ quantificatore esistenziale.
- $\forall$ quantificatore universale.
- $\neg p$ negazione della proposizione p.
- $\not\asymp$ negazione della relazione binaria $\asymp$.
- $=$ relazione di uguaglianza.

Insiemi

- $\emptyset$ insieme vuoto.
- $\in$ relazione di appartenenza.
- $A \subseteq B$ (o $B \supseteq A$) significa: A sottoinsieme di B.
- $A \subset B$ (o $B \supset A$) significa: $A \subseteq B$ e $A \neq B$.

- 2^A insieme delle parti dell'insieme A.
- $\cap$ operazione di intersezione.
- $\cup$ operazione di unione.
- $\setminus$ operazione di differenza insiemistica.
- $A^c = B \setminus A$ insieme complementare di A rispetto all'"ambiente" B.
- $\{a \in A : P(a)\}$ insieme degli elementi $a \in A$ che rendono vera la proposizione: "a possiede la proprietà espressa dal predicato $P(\cdot)$".
- $\mathbb{N}$ insieme dei numeri naturali.
  m, n (dotati o no di apici o pedici) numeri naturali non nulli.
- $\{a_1, \dots, a_n\}$ insieme finito costituito dagli elementi $a_1, \dots, a_n$.

  $\{a_1, a_2, \dots\}$ insieme numerabile costituito dai termini della successione $a_1, a_2, \dots$.
  Un insieme è *discreto* se è finito o numerabile; un *singoletto* se ha un solo elemento.
- Data una famiglia indiciata $(A_i)_{i\in I}$ di insiemi, $\bigcap_{i\in I} A_i$ è l'insieme degli elementi a tali che $a \in A_i$ per ogni $i \in I$; $\bigcup_{i\in I} A_i$ quello degli elementi a tali che $a \in A_i$ per qualche $i \in I$.[4]

  In particolare, queste notazioni sono sostituite, rispettivamente, dalle $\bigcap_{h=m}^{n} A_{i_h}$ e $\bigcup_{h=m}^{n} A_{i_h}$, se $I = \{i_m, i_{m+1}, \dots, i_n\}$ e $m < n$; dalle $\bigcap_{n\geq m} A_{i_n}$ e $\bigcup_{n\geq m} A_{i_n}$, se $I = \{i_m, i_{m+1}, i_{m+2}, \dots\}$.
- La famiglia di insiemi $(A_i)_{i\in I}$ è *disgiunta* se è formata da insiemi a due a due disgiunti (cioè tali che $A_i \cap A_j = \emptyset$ se $i \neq j$).

  La successione di insiemi $(A_{i_n})_{n\geq m}$ è *non decrescente* se $A_{i_n} \subseteq A_{i_{n+1}}$ per ogni $n \geq m$; *non crescente* se $A_{i_n} \supseteq A_{i_{n+1}}$ per ogni $n \geq m$.
- $\mathbf{a} = (a_1, \dots, a_n)$ significa: $\mathbf{a}$ è la n-pla ordinata avente come primo termine a_1, come secondo termine a_2, ... , come ultimo termine a_n.

[4] Delle regole di calcolo riguardanti le operazioni insiemistiche, ricordiamo, in particolare, le leggi di De Morgan $\left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$, $\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c$ e le proprietà distributive $A \cap \left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} (A \cap A_i)$, $A \cup \left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} (A \cup A_i)$.

$(a_n)_{n\geq 1}$ successione avente come primo termine a_1, come secondo termine a_2, come terzo termine a_3, ...

- $A_1 \times \cdots \times A_n$ prodotto (cartesiano) degli insiemi $A_1, \ldots, A_n$.

  A^n prodotto di n copie dell'insieme A.

Preordinamenti

- Una relazione binaria $\asymp$ su un insieme A è:

  - un *preordinamento* se è riflessiva ($a \asymp a$ per ogni $a \in A$) e transitiva ($a \asymp a' \wedge a' \asymp a'' \Rightarrow a \asymp a''$ per ogni $a, a', a'' \in A$);

  - un *ordinamento* se è un preordinamento antisimmetrico ($a \asymp a' \wedge a' \asymp a \Rightarrow a = a'$ per ogni $a, a' \in A$);

  - un *ordinamento stretto* se è transitiva e asimmetrica ($a \asymp a' \Rightarrow a' \not\asymp a$ per ogni $a, a' \in A$);

  - un *preordinamento completo* se è un preordinamento che verifica la completezza ($a \asymp a'$ o $a' \asymp a$ per ogni $a, a' \in A$);

  - una *equivalenza* se è riflessiva, transitiva e simmetrica ($a \asymp a' \Rightarrow a' \asymp a$ per ogni $a, a' \in A$).

- Dato un preordinamento $\succeq$ su A, per:

  parte simmetrica di $\succeq$ si intende la relazione binaria $\sim$ così definita: $a \sim a' \iff a \succeq a' \wedge a' \succeq a$;

  parte asimmetrica di $\succeq$ si intende la relazione binaria $\succ$ così definita: $a \succ a' \iff a \succeq a' \wedge a \not\sim a'$.[5]

- Dato un preordinamento $\succeq$ su A, si chiamano, rispettivamente, *intervallo chiuso* (*aperto, inferiormente semiaperto, superiormente semiaperto*) *di estremi* $a', a'' \in A$ gli insiemi:

$$[a', a''] = \{a \in A : a'' \succeq a \succeq a'\} \qquad]a', a''[= \{a \in A : a'' \succ a \succ a'\}$$
$$]a', a''] = \{a \in A : a'' \succeq a \succ a'\} \qquad [a', a''[= \{a \in A : a'' \succ a \succeq a'\}.$$

[5]Ricordiamo che la parte simmetrica è una equivalenza mentre quella asimmetrica è un ordinamento stretto. Notiamo inoltre che sussiste la proprietà seguente:

$$a_1 \sim a_1' \wedge a_2 \sim a_2' \Rightarrow \big(a_1 \succeq a_2 \iff a_1' \succeq a_2'\big) \wedge \big(a_1 \succ a_2 \iff a_1' \succ a_2'\big)$$

che mette in luce la compatibilità dell'equivalenza $\sim$ sia con il preordinamento $\succeq$ che con l'ordinamento stretto $\succ$.

$A' \subseteq A$ è un insieme $\succeq$-*connesso* se $[a, a'] \subseteq A'$ per ogni $a, a' \in A'$.

Dato un insieme non vuoto $A' \subseteq A$, le notazioni $\inf A'$, $\sup A'$ indicano i suoi estremi, rispettivamente, inferiore e superiore (qualora esistenti).

Applicazioni

- $f : A \mapsto B$ significa: f è un'applicazione di dominio A e codominio B (in breve, *di* A *in* B).

  $b = f(a)$ (o $f : a \mapsto b$) significa: b è l'immagine di a tramite f.

  $f\big|_{A'}$ restrizione di f su $A' \subseteq A$

  $f(A')$ insieme-immagine di $A' \subseteq A$ tramite f.

  $f^{-1}(B')$ (o $\{f \in B'\}$) controimmagine di $B' \subseteq B$ tramite f.[6]

  Se $A = A' \times A''$, $a' \in A'$ e $a'' \in A''$, le restrizioni $f\big|_{\{a'\}\times A''}$ e $f\big|_{A'\times\{a''\}}$ si denotano, rispettivamente, con $f(a', \cdot)$ e $f(\cdot, a'')$.

  Se l'applicazione f è biiettiva, f^{-1} è la sua inversa.

- Date le applicazioni f, g di A in B e una relazione binaria $\asymp$ su B, $\{f \asymp g\}$ è l'insieme degli $a \in A$ tali che $f(a) \asymp g(a)$.

- Dare le applicazioni $f : A \mapsto B$ e $g : B \mapsto C$, l'applicazione composta è indicata con $g \circ f$ (o $g(f)$).[7]

- Dato $S \subseteq A$, la *funzione indicatrice* di S è la funzione $I_S : A \mapsto \{0, 1\}$ tale che $I_S(a) = 1$, se $a \in S$, e $I_S(a) = 0$, se $a \notin S$.[8]

Retta reale ampliata

- $\mathbb{R}^*$ *retta reale ampliata* ottenuta aggiungendo all'insieme $\mathbb{R}$ dei numeri reali (*retta reale*) i simboli $-\infty$ e $+\infty$.

[6]Delle proprietà degli insiemi immagine e controimmagine ricordiamo, in particolare, le seguenti: $f(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} f(A_i)$, $f(\bigcap_{i\in I} A_i) \subseteq \bigcap_{i\in I} f(A_i)$, $f^{-1}(\bigcup_{i\in I} B_i) = \bigcup_{i\in I} f^{-1}(B_i)$, $f^{-1}(\bigcap_{i\in I} B_i) = \bigcap_{i\in I} f^{-1}(B_i)$, $f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2)$ e $f^{-1}(B_1^c) = [f^{-1}(B_1)]^c$.

[7]Ricordiamo che $(g \circ f)^{-1}(C') = f^{-1}(g^{-1}(C'))$ per ogni $C' \subseteq C$.

[8]Ovviamente, I_A è la costante unitaria mentre $I_\emptyset$ è la costante nulla; inoltre, qualunque siano $S', S'' \subseteq A$, si ha $I_{S'\cap S''} = I_{S'}\, I_{S''}$, $I_{S'\cup S''} = I_{S'} + I_{S''} - I_{S'}\, I_{S''}$, $I_{S'\cup S''} = I_{S'} + I_{S''}$ se S', S'' sono disgiunti, $I_{S'\setminus S''} = I_{S'} - I_{S'} I_{S''}$, $I_{S^c} = 1 - I_S$; infine, $I_{\bigcup_{n\geq 1} S_n} = \sum_{n=1}^{+\infty} I_{S_n}$ se $(S_n)_{n\geq 1}$ è una successione disgiunta di sottoinsiemi di A.

- $\leq$ ordinamento completo su $\mathbb{R}^*$ ottenuto, per prolungamento dell'usuale ordinamento per grandezza dei numeri reali, richiedendo che ogni numero reale sia maggiore di $-\infty$ e minore di $+\infty$.[9]

- Per quanto riguarda la struttura topologica, gli aperti di $\mathbb{R}^*$ sono unioni arbitrarie di intervalli del tipo $]a, b[$, $[-\infty, b[$ e $]a, +\infty]$ $(a, b \in \mathbb{R})$.

- Per quanto concerne l'aritmetica di $\mathbb{R}^*$, valgono le regole di calcolo:

$$\begin{aligned}
&a + \infty = a + (+\infty) = +\infty + a = +\infty, \text{ se } -\infty < a \leq +\infty;\\
&a - \infty = a + (-\infty) = -\infty + a = -\infty, \text{ se } -\infty \leq a < +\infty;\\
&a \cdot (-\infty) = -\infty \cdot a = -\infty, \text{ se}\, 0 < a \leq +\infty;\\
&a \cdot (+\infty) = +\infty \cdot a = +\infty, \text{ se}\, 0 < a \leq +\infty;\\
&a \cdot (-\infty) = -\infty \cdot a = +\infty, \text{ se } -\infty \leq a < 0;\\
&a \cdot (+\infty) = +\infty \cdot a = -\infty, \text{ se } -\infty \leq a < 0;\\
&0 \cdot (-\infty) = 0 \cdot (+\infty) = -\infty \cdot 0 = +\infty \cdot 0 = 0;\\
&\frac{a}{-\infty} = \frac{a}{+\infty} = 0, \ \text{se}\, a \in \mathbb{R}.
\end{aligned}$$

  Inoltre, sono escluse dalle espressioni aritmetiche lecite quelle che coinvolgono rapporti di infiniti o rapporti con denominatori nulli, oppure somme di infiniti di segno opposto o differenze di infiniti di ugual segno.

- Dati $\mathbf{a}, \mathbf{b} \in (\mathbb{R}^*)^n$, $\mathbf{a} \leq \mathbf{b}$ significa: $a_i \leq b_i$ $(i = 1, \ldots, n)$; $\mathbf{a} \ll \mathbf{b}$ significa: $a_i < b_i$ $(i = 1, \ldots, n)$.

Funzioni a valori nella retta reale ampliata

- Data una funzione $f : A \mapsto \mathbb{R}^*$, le notazioni $\inf f$, $\sup f$ indicano gli estremi, rispettivamente, inferiore e superiore di $f(A)$.

  Dato un preordinamento $\succeq$ su A, la funzione f è *non decrescente* (*crescente*) se $f(a) \geq f(a')$ $(f(a) > f(a'))$ per ogni a, a' tali che $a \succ a'$; *non crescente* (*decrescente*) se $f(a) \leq f(a')$ $(f(a) < f(a'))$ per ogni a, a' tali che $a \succ a'$.

[9] Quindi $\mathbb{R} =]-\infty, +\infty[$ e $\mathbb{R}^* = [-\infty, +\infty]$; inoltre, $[0, +\infty[$ e $[0, +\infty]$ sono l'insieme dei numeri non negativi, rispettivamente, della retta reale e di quella ampliata. Conseguentemente, il valore assoluto $|a|$ di $a \in \mathbb{R}^*$ coincide con a su $[0, +\infty]$ e con l'opposto di a su $[-\infty, 0[$.

- Date le funzioni f, g di A in $\mathbb{R}^*$, la funzione:

  $fg : a \mapsto f(a)g(a)$ è definita su A;

  $f + g : a \mapsto f(a) + g(a)$ è definita sull'insieme degli elementi a tali che $f(a)$, $g(a)$ non sono infiniti di segno opposto;

  $\frac{f}{g} : a \mapsto \frac{f(a)}{g(a)}$ è definita sull'insieme degli elementi a tali che $g(a) \neq 0$ o $f(a)$, $g(a)$ non sono entrambi infiniti.

  $f \leq g$ significa: $f(a) \leq g(a)$ per ogni $a \in A$.

  $f \ll g$ significa: $f(a) < g(a)$ per ogni $a \in A$.

- Data una famiglia $(f_i)_{i \in I}$ di funzioni di A in $\mathbb{R}^*$, le funzioni $\inf\limits_{i \in I} f_i$ e $\sup\limits_{i \in I} f_i$ sono così definite:

  $\inf\limits_{i \in I} f_i : a \mapsto \inf \big\{ f_i(a) : i \in I \big\}$;

  $\sup\limits_{i \in I} f_i : a \mapsto \sup \big\{ f_i(a) : i \in I \big\}$.

- Data una successione $(f_n)_{n \geq 1}$ di funzioni di A in $\mathbb{R}^*$, la successione è *non decrescente* (*crescente*) se $f_n(a) \leq f_{n+1}(a)$ ($f_n(a) < f_{n+1}(a)$) per ogni a e per ogni n; *non crescente* (*decrescente*) se $f_n(a) \geq f_{n+1}(a)$ ($f_n(a) > f_{n+1}(a)$) per ogni a e per ogni n.

  $f_n \to f$ significa: la successione converge (puntualmente) alla funzione f (cioè, $f(a) = \lim\limits_{n \to +\infty} f_n(a)$ per ogni $a \in A$).

  $f_n \uparrow f$ significa: la successione è non decrescente e converge a f.

  $f_n \downarrow f$ significa: la successione è non crescente e converge a f.

  $\sum\limits_{n \geq 1} f_n$ serie di funzioni di termine generale f_n.

Capitolo 1

Preferenze e utilità

Un individuo, il *decisore* (che chiamiamo DM[1]), intende scegliere un elemento nell'ambito di un dato insieme $\mathfrak{D}$ di alternative (**decisioni**) possibili[2]. Assumiamo che la scelta di una decisione d comporti per DM una **conseguenza** che non sempre è in grado di individuare nel suo stato d'informazione, salvo ritenerla elemento di un dato insieme C_d (determinato senza possibilità di equivoci) di conseguenze possibili[3]. Supposto che l'insieme $\mathfrak{C} = \bigcup_{d \in \mathfrak{D}} C_d$

[1]Abbreviazione della locuzione inglese "decision maker". Precisiamo che con il termine "individuo" intendiamo sia una persona che un gruppo di persone, pensate però formanti un'unità (come, ad esempio, una famiglia, un consiglio d'amministrazione, la Corte dei conti, la Banca d'Italia, etc.).

[2]Si pensi, ad esempio, a scelte inerenti il menu da ordinare al ristorante, la terapia medica da prescrivere, il corso di laurea da attivare in una data facoltà, il piano d'investimento d'adottare, etc..

[3]Che possono essere di varia natura; ad esempio, morale (gradimento dei cibi previsti dal menu, etc.), probabilistica (probabilità di guarigione della terapia, etc.), d'immagine (laureati ben preparati con ottime possibilità d'inserimento nel mondo del lavoro, etc.), monetaria (incremento del capitale alla scadenza, etc.).

Ovviamente, C_d dovrebbe fornire, da un punto di vista ideale, un elenco esaustivo delle conseguenze connesse con la decisione d. Purtroppo, a causa delle limitazioni (sia deduttive che predittive) umane, ciò non è possibile (salvo casi specifici) per cui C_d conterrà, in pratica, solamente le conseguenze che DM riterrà più rilevanti per il suo problema decisionale (tra quelle considerate).

Chiaramente, la conseguenza (relativa alla decisione d) sarà certa, se C_d ha un solo elemento, e aleatoria, se ne ha più di uno (non dipendendo soltanto dalla scelta d del decisore, ma anche da altre circostanze che per DM sono ben definite ma le cui realizzazioni sono, nel suo stato d'informazione, sconosciute).

delle conseguenze abbia almeno due elementi[4] e che il decisore sia un *individuo razionale*, DM terrà conto nella sua scelta dei vantaggi e degli svantaggi che la corrispondente conseguenza (certa o aleatoria) può procurargli; verrà, cioè, guidato dalle sue *preferenze* tra le conseguenze di $\mathfrak{C}$.

1.1 Relazioni di preferenza e di indifferenza

Al fine di individuare le proprietà che sembra ragionevole richiedere al sistema di preferenze (sulle conseguenze) di un decisore razionale, immaginiamo di porre a DM, per ogni coppia $c_1, c_2 \in \mathfrak{C}$, la seguente domanda: c'è tra le conseguenze c_1 e c_2 una che per te sia migliore dell'altra, oppure che almeno non le sia peggiore? Evidentemente, potremo ricevere solo una delle seguenti risposte:
- c_i non è peggiore di c_j;
- c_1 non è peggiore di c_2 e c_2 non è peggiore di c_1;
- c_i è migliore di c_j;
- nè "c_1 non è peggiore di c_2", nè "c_2 non è peggiore di c_1".
Si vengono così ad individuare su $\mathfrak{C}$ tre relazioni:
- la relazione di **preferenza**: $c \succeq c' \iff c$ non è peggiore di c';
- la relazione di **indifferenza**: $c \sim c' \iff c \succeq c' \wedge c' \succeq c$;
- la relazione di **preferenza stretta**: $c \succ c' \iff c \succeq c' \wedge c \not\sim c'$.

Quali proprietà queste relazioni dovrebbero ragionevolmente possedere? Chiaramente, ogni conseguenza non è peggiore di se stessa; inoltre, appare naturale richiedere che se la conseguenza c non è peggiore della conseguenza c' e la conseguenza c' non è peggiore della conseguenza c'', allora la conseguenza c non è peggiore della conseguenza c''. Supponiamo quindi che la relazione $\succeq$ verifichi gli assiomi:

A1 Riflessività: $c \succeq c$ per ogni $c \in \mathfrak{C}$.

A2 Transitività: $c \succeq c' \wedge c' \succeq c'' \Rightarrow c \succeq c''$ per ogni $c, c', c'' \in \mathfrak{C}$.

Conseguentemente, la relazione di preferenza diviene un preordinamento e quelle di indifferenza e di preferenza stretta, rispettivamente, un'equivalenza

[4]In caso contrario, non sussiste alcun problema di scelta; DM va incontro ad una conseguenza totalmente indipendente dalle sue decisioni.

e un ordinamento stretto (osservato che la prima è la parte simmetrica e la seconda è quella asimmetrica di $\succeq$)[5].

Per rendere significativo il problema di scelta, supponiamo inoltre che ci siano almeno due conseguenze confrontabili e non indifferenti, cioè che la relazione di preferenza stretta verifichi l'assioma:

A3 NON DEGENERAZIONE: $c^* \succ c_*$ per qualche $c^*, c_* \in \mathfrak{C}$.

Supposto infine, per semplificare il contesto decisionale, che DM non fornisca mai l'ultima delle risposte considerate all'inizio della sezione[6], assumiamo, tramite il prossimo assioma, la completezza della relazione di preferenza.

A4 COMPLETEZZA: $c \succeq c'$ o $c' \succeq c$ per ogni $c, c' \in \mathfrak{C}$.

Richiedere, come è stato fatto, che la relazione di preferenza sia transitiva, se da una parte sembra del tutto ragionevole, dall'altra esclude alcuni modelli introdotti per descrivere le preferenze del decisore in contesti specifici. A titolo d'esempio, ne riportiamo uno riguardante la stima puntuale. Considerato il problema di stimare il parametro incognito reale Z, siano Θ l'insieme delle sue determinazioni possibili e X un ente aleatorio osservabile con distribuzione di probabilità P_θ dipendente da $\theta \in \Theta$. Al fine di valutare la bontà degli stimatori di Z, Edwin Pitman introdusse nel 1937 la relazione di preferenza stretta tra stimatori:

$$\delta \succ_{\mathrm{P}} \delta' \iff \forall \theta \in \Theta \Big(\mathrm{P}_\theta \big(|\delta(X) - \theta| < |\delta'(X) - \theta| \big) > \frac{1}{2} \Big).$$

Lo stimatore δ è dunque migliore dello stimatore δ' se, qualunque sia il valore "vero" del parametro, è maggiore di $\frac{1}{2}$ la probabilità che δ approssimi Z meglio di quanto lo faccia δ'. Proviamo ora che $\succ_{\mathrm{P}}$ non è una relazione transitiva. A tal fine, sia $X = YZ$, con Y distribuito uniformemente nell'intervallo $[-0.9, 1.1]$. Supposto $Z > 0$, consideriamo gli stimatori $\delta(X) = X$, $\delta'(X) = a|X|$ e $\delta''(X) = b|X|$ $(0 < a < 1 < b)$. Riesce allora

$$\begin{aligned}
\mathrm{P}_\theta\big(|\delta(X) - Z| < |\delta'(X) - Z|\big) &= \mathrm{P}_\theta\big(|X - Z| < |\, a|X| - Z\,|\big) = \mathrm{Pr}\big(|Y - 1| < |\, a|Y| - 1\,|\big) \\
&= \mathrm{Pr}\Big(0 < Y < \frac{2}{1+a}\Big) \\
\mathrm{P}_\theta\big(|\delta''(X) - Z| < |\delta(X) - Z|\big) &= \mathrm{P}_\theta\big(|\, b|X| - Z\,| < |\, X - Z|\big) = \mathrm{Pr}\big(|\, b|Y| - 1| < |\, Y - 1|\big) \\
&= \mathrm{Pr}\Big(\frac{2}{1-b} < Y < \frac{2}{1+b}\Big)
\end{aligned}$$

[5] Non possono quindi, nel campo delle preferenze strette, formarsi dei cicli (cioè sequenze del tipo $c \succ c'$, $c' \succ c''$ e $c'' \succ c$).

[6] E quindi, a differenza dell'asino di Buridano, non abbia mai (sapendo sempre quello che vuole) dei dubbi, delle incertezze su come scegliere tra due conseguenze qualsiasi.

$$\begin{aligned}
\mathrm{P}_\theta\big(|\delta'(X)-Z|<|\delta''(X)-Z|\big) &= \mathrm{P}_\theta\big(|\,a|X|-Z\,|<|\,b|X|-Z\,|\big)\\
&= \mathrm{Pr}\big(|\,a|Y|-1\,|<|\,b|Y|-1\,|\big)= \mathrm{Pr}\Big(|Y|>\frac{2}{a+b}\Big)\\
&= \mathrm{Pr}\Big(Y>\frac{2}{a+b}\Big)+\mathrm{Pr}\Big(-\frac{2}{a+b}<Y<0\Big)
\end{aligned}$$

da cui, posto $a=0.9$ e $b=3.2$, otteniamo

$$\begin{aligned}
\mathrm{P}_\theta\big(|\delta(X)-Z|<|\delta'(X)-Z|\big) &= \frac{10}{19}\\
\mathrm{P}_\theta\big(|\delta''(X)-Z|<|\delta(X)-Z|\big) &= \frac{289}{410}\\
\mathrm{P}_\theta\big(|\delta'(X)-Z|<|\delta''(X)-Z|\big) &= \frac{11}{20}
\end{aligned}$$

e quindi il ciclo $\delta \succ_{\mathrm{P}} \delta'$, $\delta' \succ_{\mathrm{P}} \delta''$ e $\delta'' \succ_{\mathrm{P}} \delta$. Ne segue la non transitività di $\succ_{\mathrm{P}}$.

1.2 Utilità ordinale

Essendo un preordinamento completo, la relazione di preferenza $\succeq$ ammette una rappresentazione numerica nella retta reale. Basta infatti considerare la funzione:

$$\phi(c_1,c_2)=\begin{cases}1 & \text{se } c_1\succ c_2\\ 0 & \text{se } c_1\sim c_2\\ -1 & \text{se } c_2\succ c_1\end{cases}$$

di dominio $\mathfrak{C}^2$ e osservare che $c_1 \succeq c_2 \iff \phi(c_1,c_2)\geq 0$ per ogni c_1, $c_2\in\mathfrak{C}$. Sebbene utile in alcuni contesti specifici, questa descrizione quantitativa delle preferenze non è in grado di introdurre alcuna “scala numerica” che a conseguenze più gradite faccia corrispondere valori numerici più elevati, e viceversa; basta infatti osservare che, qualunque sia la conseguenza c' di “riferimento”, l'applicazione indotta $\phi(\cdot,c')$ associa ad ogni conseguenza c migliore di c' sempre valore unitario. D'altronde, soprattutto in vista delle applicazioni, l'esistenza di una scala numerica è di notevole interesse; consente infatti di ridurre l'individuazione delle conseguenze “migliori” per DM ad un problema di massimizzazione numerica (la determinazione dei punti di massimo della funzione $v:\mathfrak{C}\mapsto\mathbb{R}$ che realizza la scala), consentendo così il ricorso - qualora l'insieme $\mathfrak{C}$ e la funzione v siano “fatti bene” - alle usuali tecniche di ottimizzazione. Ciò osservato, con la definizione seguente, identifichiamo le utilità ordinali di DM con le scale numeriche relative alle sue preferenze[7].

Definizione 1.2.1 Una funzione $v : \mathfrak{C} \mapsto \mathbb{R}$ è una **utilità ordinale** (per $\succeq$) se: $c \succeq c' \iff v(c) \geq v(c')$ per ogni $c, c' \in \mathfrak{C}$.

Il risultato seguente rileva che un'utilità ordinale è anche in grado di descrivere le relazioni di indifferenza e di preferenza stretta; inoltre, assicura che è definita a meno di trasformazioni crescenti (e quindi differenze tra valori d'utilità indicano solamente l'ordine di preferenza tra le relative conseguenze e non di quanto più preferita sia una conseguenza rispetto all'altra[8]).

Teorema 1.2.2 *Sia v un'utilità ordinale. Riesce allora:*

(i) $c \sim c' \iff v(c) = v(c')$*;*

(ii) $c \succ c' \iff v(c) > v(c')$

per ogni $c, c' \in \mathfrak{C}$. Inoltre, $\tilde{v} : \mathfrak{C} \mapsto \mathbb{R}$ è un'utilità ordinale per $\succeq$ se e solo se $\tilde{v}$ è una trasformata crescente di v (cioè, se esiste una funzione crescente $f : v(\mathfrak{C}) \mapsto \mathbb{R}$ tale che $\tilde{v} = f \circ v$).

DIMOSTRAZIONE Essendo le dimostrazioni di (i) e di (ii) immediate, ci limitiamo a verificare l'ultima parte della tesi. Notato che la condizione sufficiente è banale, passiamo a quella necessaria. Sia quindi $\tilde{v}$ un'utilità ordinale per $\succeq$. Dalla (ii) segue allora

$$v(c) > v(c') \iff \tilde{v}(c) > \tilde{v}(c') \tag{1.1}$$

[7] Storicamente, il termine "utilità" (centrale per il pensiero economico) ha assunto diversi significati che si possono, grosso modo, far risalire a tre distinti periodi. Nel primo, corrispondente all'economia politica classica (Adam Smith, David Ricardo e Karl Marx), veniva inteso come una *proprietà dei beni*: il loro "valore d'uso", cioè l'attitudine a soddisfare un bisogno biologico o culturale. Nel secondo, relativo all'economia del benessere utilitarista (iniziata da Jeremy Bentham), veniva visto come una *grandezza psichica degli individui*: il loro "benessere" (ritenuto suscettibile di misura e di segno, positivo per il piacere e negativo per la pena). Nell'ultimo, corrispondente alla moderna teoria delle decisioni, viene inteso (anche per noi) come una *rappresentazione numerica delle preferenze individuali* (non necessariamente connesse ad un contesto economico) di carattere: *ordinale* se riguardano alternative certe (Vilfredo Pareto), *cardinale* se riguardano distribuzioni di probabilità sulle alternative certe (Frank P. Ramsey, John von Neumann e Oskar Morgenstern).

[8] Giustificando così l'associazione dell'aggettivo "ordinale" alla nozione d'utilità considerata.

per ogni $c, c' \in \mathfrak{C}$.

Considerato l'insieme-immagine $T = \{v(c) + \tilde{v}(c) : c \in \mathfrak{C}\}$ della funzione $v + \tilde{v}$, associamo ad ogni suo elemento t il sottoinsieme $C_t = \{c \in \mathfrak{C} : v(c) + \tilde{v}(c) = t\}$ e verifichiamo che, qualunque sia $t \in T$, le utilità v, $\tilde{v}$ sono costanti su C_t. Dato $t \in T$ e scelti $c, c' \in C_t$, si ha $v(c)+\tilde{v}(c) = t = v(c')+\tilde{v}(c')$ e quindi

$$v(c) - v(c') = -[\tilde{v}(c) - \tilde{v}(c')]. \tag{1.2}$$

Allora, $v(c) = v(c')$; infatti, se così non fosse, da (1.1) si avrebbe $\tilde{v}(c)-\tilde{v}(c') > 0$, se $v(c)-v(c') > 0$, e $\tilde{v}(c)-\tilde{v}(c') < 0$, se $v(c)-v(c') < 0$, contraddicendo in ogni caso (1.2). Conseguentemente, tramite (1.2), risulta anche $\tilde{v}(c) = \tilde{v}(c')$.

Indicati, per ogni $t \in T$, con $\alpha(t)$ e $\tilde{\alpha}(t)$ i valori assunti su C_t, rispettivamente, da v e $\tilde{v}$, si vengono a individuare due funzioni $\alpha : T \mapsto v(\mathfrak{C})$ e $\tilde{\alpha} : T \mapsto \tilde{v}(\mathfrak{C})$ tali che

$$v(c) = \alpha(v(c) + \tilde{v}(c)) \tag{1.3}$$

$$\tilde{v}(c) = \tilde{\alpha}(v(c) + \tilde{v}(c)) \tag{1.4}$$

per ogni conseguenza c.

Proviamo ora che queste funzioni sono crescenti. A tal fine, siano $t, t' \in T$ con $t < t'$. Esistono allora due conseguenze c, c' tali che $v(c) + \tilde{v}(c) = t < t' = v(c') + \tilde{v}(c')$. Ne segue

$$v(c) - v(c') < -[\tilde{v}(c) - \tilde{v}(c')]$$

da cui otteniamo $v(c') > v(c)$; infatti, in caso contrario, risulterebbe $v(c) - v(c') \geq 0$ da cui discenderebbe $\tilde{v}(c) - \tilde{v}(c') < 0$ e quindi, per (1.1), si avrebbe $v(c) - v(c') < 0$ (Contraddizione!). Dalla $v(c') > v(c)$, tenuto ancora conto di (1.1), si ha anche $\tilde{v}(c') > \tilde{v}(c)$. Ne segue, tramite (1.3) e (1.4),

$$\alpha(t) = \alpha(v(c) + \tilde{v}(c)) = v(c) < v(c') = \alpha(v(c') + \tilde{v}(c')) = \alpha(t')$$
$$\tilde{\alpha}(t) = \tilde{\alpha}(v(c) + \tilde{v}(c)) = \tilde{v}(c) < \tilde{v}(c') = \tilde{\alpha}(v(c') + \tilde{v}(c')) = \tilde{\alpha}(t').$$

Dato che la funzione α è crescente, possiamo considerare la sua inversa $\alpha^{-1} : v(\mathfrak{C}) \mapsto T$. Allora, la funzione composta $\tilde{\alpha} \circ \alpha^{-1}$ è crescente (l'inversa α^{-1} è crescente!) e inoltre, per (1.3) e (1.4),

$$\begin{aligned}\tilde{v}(c) &= \tilde{\alpha}(v(c) + \tilde{v}(c)) = \tilde{\alpha}\big(\alpha^{-1}\big(\alpha(v(c) + \tilde{v}(c))\big)\big) = (\tilde{\alpha} \circ \alpha^{-1})(v(c)) \\ &= [(\tilde{\alpha} \circ \alpha^{-1}) \circ v](c)\end{aligned}$$

per ogni conseguenza c. Dunque, $\tilde{v}$ è una trasformata crescente di v. ■

Non tutte le relazioni di preferenza ammettono un'utilità ordinale. Proviamo infatti che non è possibile descrivere l'usuale ordinamento lessicografico:

$$(a,b) \succeq (a',b') \iff a > a' \vee (a = a' \wedge b \geq b')$$

su $\mathbb{R}^2$ mediante un'utilità ordinale. A tal fine, supponiamo (per assurdo) che v sia un'utilità ordinale per $\succeq$. Considerato allora un numero reale a qualsiasi, risulta $(a,1) \succ (a,0)$ e quindi, per il Teorema 1.2.2(ii), $v((a,1)) > v((a,0))$. Esiste dunque un numero razionale q_a tale che $v((a,1)) > q_a > v((a,0))$. Si viene così a creare una funzione iniettiva di $\mathbb{R}$ nei numeri razionali; infatti, da $a > a'$ segue $(a,0) \succ (a',1)$ e quindi, per il Teorema 1.2.2(ii), $q_a > v((a,0)) > v((a',1)) > q_{a'}$. Conseguentemente, l'insieme numerabile dei numeri razionali include un insieme avente la cardinalità del continuo (Contraddizione!).

Di particolare interesse è la caratterizzazione seguente che identifica le preferenze che ammettono un'utilità ordinale con quelle che consentono di individuare un insieme discreto di classi di indifferenza la cui unione si "comporti" (rispetto alla preferenza) come il sottoinsieme dei numeri razionali rispetto all'usuale ordinamento per grandezza dei numeri reali. Ricordiamo, in proposito, che $C \subseteq \mathfrak{C}$ è **$\succeq$- denso** (nel senso di Birkhoff) se, per ogni $c_1, c_2 \notin C$ tali che $c_2 \succ c_1$, esiste una conseguenza $c \in C$ peggiore di c_2 e migliore di c_1 (cioè, appartenente all'intervallo aperto $]c_1, c_2[$). Precisiamo inoltre che $[c']$ denota la classe di indifferenza $\{c \in \mathfrak{C} : c \sim c'\}$ della conseguenza c'.

Teorema 1.2.3 *Esiste un'utilità ordinale per $\succeq$ se e solo se esiste un insieme discreto C di conseguenze tale che l'unione $\bigcup_{c \in C}[c]$ è un insieme $\succeq$-denso.*[9]

Dimostrazione Selezionato in ogni classe di indifferenza un elemento, denotiamo con C_0 l'insieme così costituito e, per ogni conseguenza c, con $\sigma(c)$ quell'unico elemento di C_0 tale che $\sigma(c) \sim c$.

Supponiamo intanto che v sia un'utilità ordinale per $\succeq$. Allora, la restrizione $v|_{C_0}$ è un'applicazione iniettiva; infatti, dati $c, c' \in C_0$ tali che $v(c) = v(c')$, per il Teorema 1.2.2(i), si ha $c \sim c'$, cioè $[c] = [c']$ e quindi $c = c'$ (per definizione di C_0).

Sia ora $\mathcal{I}$ la famiglia (numerabile) degli intervalli reali aperti con estremi razionali che contengono l'immagine, tramite v, di qualche elemento di C_0. Si ha quindi

$$\mathcal{I} = \big\{]q,q'[\subset \mathbb{R} : q, q' \text{ razionali} \wedge]q,q'[\cap v(C_0) \neq \emptyset \big\}.$$

Selezionato, per ogni intervallo $I \in \mathcal{I}$, un elemento c_I in C_0 tale che $v(c_I) \in I$, consideriamo l'insieme discreto $C_{\mathcal{I}} = \{c_I : I \in \mathcal{I}\}$ e poniamo:

$$C_1 = \{c \in C_0 \setminus C_{\mathcal{I}} : \exists c' \in C_0 (c \succ c' \wedge]c', c[\cap C_{\mathcal{I}} = \emptyset)\}.$$

Una conseguenza c di C_0 che non è in $C_{\mathcal{I}}$ appartiene dunque a C_1 solo se è possibile trovare una conseguenza $\gamma(c) \in C_0$ peggiore di c e tale che non sia possibile individuare una conseguenza di $C_{\mathcal{I}}$ che sia peggiore di c e migliore di $\gamma(c)$.

[9] Ammettono pertanto un'utilità ordinale le relazioni di preferenza definite su insiemi di conseguenze discreti oppure aventi al più un numerabile di classi di indifferenza.

Per provare che C_1 è un insieme discreto basta verificare (ricordando l'iniettività di $v|_{C_0}$) che la famiglia di intervalli reali non vuoti:

$$\mathcal{J} = \{I(c) =]v(\gamma(c)), v(c)[: c \in C_1\}$$

è disgiunta. A tal fine, supponiamo (per assurdo) che esistano $c, c' \in C_1$ tali che $I(c) \cap I(c') \neq \emptyset$. Possiamo allora assumere (senza perdere in generalità) $v(\gamma(c)) < v(\gamma(c')) < v(c)$. Considerato quindi un intervallo $I \in \mathcal{I}$ tale che $v(\gamma(c')) \in I \subset I(c)$, riesce $v(c_I) \in I \subset I(c)$; ne segue, $v(c) > v(c_I) > v(\gamma(c))$ da cui, per il Teorema 1.2.2(ii), $c \succ c_I \succ \gamma(c)$ e quindi una contraddizione, osservato che $c_I \in C_{\mathcal{I}}$.

Poichè C_1 è discreto, è pure discreto l'insieme $C = C_1 \cup C_{\mathcal{I}}$. Per provare che l'insieme $\widehat{C} = \bigcup_{c \in C}[c]$ è $\succeq$-denso, siano $c_1, c_2 \notin \widehat{C}$ con $c_1 \succ c_2$. Per definizione di C_0, esistono $\widetilde{c}_1, \widetilde{c}_2 \in C_0$ tali che $c_1 \sim \widetilde{c}_1$ e $c_2 \sim \widetilde{c}_2$. Ne segue, $\widetilde{c}_1 \succ \widetilde{c}_2$ e $\widetilde{c}_1 \notin C = C_1 \cup C_{\mathcal{I}}$ (per definizione di $\widehat{C}$). Allora, $\widetilde{c}_1 \in C_0 \setminus C_{\mathcal{I}}$ e $\widetilde{c}_1 \notin C_1$. Per $\widetilde{c}_1 \succ \widetilde{c}_2$, esiste quindi $c_I \in C_{\mathcal{I}} \subseteq C$ tale che $\widetilde{c}_1 \succ c_I \succ \widetilde{c}_2$. Ne segue, $c_1 \succ c_I \succ c_2$ con $c_I \in \widehat{C}$.

Passando alla condizione sufficiente, supponiamo che C sia un insieme discreto tale che l'insieme $\bigcup_{c \in C}[c]$ risulti $\succeq$-denso. Procediamo per casi considerando la preferenza prima antisimmetrica e poi arbitraria.

• Caso 1: $\succeq$ antisimmetrica. Allora, ogni conseguenza è indifferente solo a se stessa e quindi $C = \bigcup_{c \in C}[c]$ è un insieme $\succeq$-denso. Proviamo, innanzitutto, che l'insieme:

$$S = \{\, (c_1, c_2) \in \mathfrak{C}^2 : c_2 \succ c_1 \wedge]c_1, c_2[= \emptyset\}$$

è discreto[10]. Osservato che, per la $\succeq$-densità di C, una almeno delle due componenti di una coppia di S deve appartenere a C, consideriamo l'applicazione $f : S \mapsto C$ così definita:

$$f(c_1, c_2) = \begin{cases} c_1 & \text{se } c_1 \in C \\ c_2 & \text{se } c_1 \notin C. \end{cases}$$

La finitezza o numerabilità di S sarà evidentemente provata se verifichiamo che la controimmagine (tramite f) di un qualsiasi elemento di C ha al più due elementi. Siano pertanto $(c_1, c_2), (\widetilde{c}_1, \widetilde{c}_2), (\widehat{c}_1, \widehat{c}_2) \in S$ tali che $f(\widetilde{c}_1, \widetilde{c}_2) = f(c_1, c_2) = f(\widehat{c}_1, \widehat{c}_2)$ e $(\widetilde{c}_1, \widetilde{c}_2), (\widehat{c}_1, \widehat{c}_2) \neq (c_1, c_2)$. Poichè nelle due situazioni possibili $c_1 \in C$, $c_1 \notin C$ i procedimenti dimostrativi sono analoghi, ci limitiamo a riportare quello relativo alla prima. Assumiamo pertanto $c_1 \in C$. Allora, $f(c_1, c_2) = c_1$. Se $\widetilde{c}_1 \in C$, otteniamo $\widetilde{c}_1 = f(\widetilde{c}_1, \widetilde{c}_2) = c_1$ e quindi, tenuto conto della $]c_1, c_2[=]\widetilde{c}_1, \widetilde{c}_2[= \emptyset$ e della completezza della preferenza, $c_2 = \widetilde{c}_2$, cioè $(c_1, c_2) = (\widetilde{c}_1, \widetilde{c}_2)$. Se invece $\widetilde{c}_1 \notin C$, si ha $\widetilde{c}_2 = f(\widetilde{c}_1, \widetilde{c}_2) = c_1$. Ora, nel caso $\widehat{c}_1 \in C$, riesce $\widehat{c}_1 = f(\widehat{c}_1, \widehat{c}_2) = c_1$ e quindi $\widehat{c}_2 = c_2$ $(]c_1, c_2[=]\widehat{c}_1, \widehat{c}_2[= \emptyset!)$, cioè $(c_1, c_2) = (\widehat{c}_1, \widehat{c}_2)$; nel caso $\widehat{c}_1 \notin C$, risulta infine $\widehat{c}_2 = f(\widehat{c}_1, \widehat{c}_2) = c_1 = \widetilde{c}_2$ e quindi $\widehat{c}_1 = \widetilde{c}_1$ $(]\widetilde{c}_1, \widetilde{c}_2[=]\widehat{c}_1, \widehat{c}_2[= \emptyset!)$, cioè $(\widetilde{c}_1, \widetilde{c}_2) = (\widehat{c}_1, \widehat{c}_2)$. Dunque, in ogni caso, almeno due delle tre coppie considerate coincidono.

Poichè S è un insieme discreto, lo è banalmente anche l'insieme:

$$C' = \{c \in \mathfrak{C} : \exists \bar{c}\,\big((c, \bar{c}) \in S \vee (\bar{c}, c) \in S\big)\}.$$

[10] Ad esempio, con riferimento all'usuale ordinamento per grandezza, $S = \emptyset$, se $\mathfrak{C} = \mathbb{R}$, e $S = \{(n, n+1) : n \in \mathbb{N}\}$, se $\mathfrak{C} = \mathbb{N}$.

Scelta allora una numerazione $\widetilde{c}_1, \ldots, \widetilde{c}_n, \ldots$ dell'insieme discreto $\widetilde{C} = C \cup C'$, consideriamo la funzione di dominio $\mathfrak{C}$:

$$v(c) = \sum_{n \geq 1} \frac{1}{2^n} I_{]\widetilde{c}_n \rightarrow [}(c),$$

avendo posto $]\widetilde{c}_n, \rightarrow [= \{c : c \succ \tilde{c}_n\}$ per ogni n.

Proviamo ora che v è un'utilità ordinale per $\succeq$. Date due conseguenze c_1, c_2 con $c_2 \succeq c_1$, l'implicazione $I_{]\widetilde{c}_n, \rightarrow [}(c_1) = 1 \Rightarrow I_{]\widetilde{c}_n, \rightarrow [}(c_2) = 1$ sussiste per ogni n; ne segue, per definizione di v, che $v(c_2) \geq v(c_1)$. Viceversa, tenuto conto della completezza della preferenza, basta provare che, date due conseguenze c_1, c_2 con $c_2 \succ c_1$, si ha $v(c_2) > v(c_1)$. Procediamo per casi supponendo l'esistenza o no di m tale che $c_2 \succ \widetilde{c}_m \succeq c_1$.

• Caso 1.1: $c_2 \succ \widetilde{c}_m \succeq c_1$ per qualche m. Allora, $I_{]\widetilde{c}_m, \rightarrow [}(c_1) = 0 < 1 = I_{]\widetilde{c}_m, \rightarrow [}(c_2)$. Inoltre, considerata una generica conseguenza $\widetilde{c}_n$, si ha $I_{]\widetilde{c}_n, \rightarrow [}(c_1) = 0 = I_{]\widetilde{c}_n, \rightarrow [}(c_2)$, se $\widetilde{c}_n \succeq c_2$; $I_{]\widetilde{c}_n, \rightarrow [}(c_1) = 0 < 1 = I_{]\widetilde{c}_n, \rightarrow [}(c_2)$, se $c_2 \succ \widetilde{c}_n \succeq c_1$; $I_{]\widetilde{c}_n, \rightarrow [}(c_1) = 1 = I_{]\widetilde{c}_n, \rightarrow [}(c_2)$, se $c_1 \succ \widetilde{c}_n$. Riesce pertanto $v(c_2) = \sum_{n \geq 1} 2^{-n} I_{]\widetilde{c}_n, \rightarrow [}(c_2) > \sum_{n \geq 1} 2^{-n} I_{]\widetilde{c}_n, \rightarrow [}(c_1) = v(c_1)$.

• Caso 1.2: Non esiste m tale che $c_2 \succ \widetilde{c}_m \succeq c_1$. Allora, $c_1 \notin \widetilde{C}$. Ne segue $]c_1, c_2[\neq \emptyset$ (in caso contrario si avrebbe $(c_1, c_2) \in S$ da cui seguirebbe la contraddizione $c_1 \in C'$) e quindi esiste una conseguenza $\widehat{c}$ tale che $c_2 \succ \widehat{c} \succ c_1$. Allora, $\widehat{c} \notin \widetilde{C}$. Per la $\succeq$-densità di C esiste m tale che $c_2 \succ \widehat{c} \succ \widetilde{c}_m \succ c_1$ e quindi, per il Caso 1.1, $v(c_2) > v(c_1)$.

• Caso 2: $\succeq$ arbitraria. Chiaramente, la restrizione $\succeq_0$ della preferenza $\succeq$ su C_0 è antisimmetrica. Inoltre, posto $\widehat{C} = \bigcup_{c \in C} [c]$, l'insieme discreto $C_0' = C_0 \cap \widehat{C} = \{\sigma(c) : c \in C\}$ è $\succeq_0$-denso; infatti, dati $c_1, c_2 \in C_0 \setminus C_0'$ con $c_1 \succ_0 c_2$, si ha $c_1, c_2 \notin \widehat{C}$ da cui, per la $\succeq$-densità di $\widehat{C}$, esiste $c \in \widehat{C}$ tale che $c_1 \succ c \succ c_2$ e quindi $\sigma(c) \in C_0'$ e $c_1 \succ_0 \sigma(c) \succ_0 c_2$. Esiste allora, per il Caso 1, un'utilità ordinale $v_0 : C_0 \mapsto \mathbb{R}$ per $\succeq_0$. Risulta quindi $c_1 \succeq_0 c_2 \iff v_0(c_1) \geq v_0(c_2)$ per ogni $c_1, c_2 \in C_0$. Possiamo, a questo punto, ottenere un'utilità ordinale considerando la funzione $v : c \mapsto v_0(\sigma(c))$. Si ha infatti

$$c_1 \succeq c_2 \iff \sigma(c_1) \succeq_0 \sigma(c_2) \iff v_0(\sigma(c_1)) \geq v_0(\sigma(c_2)) \iff v(c_1) \geq v(c_2)$$

qualunque siano le conseguenze c_1 e c_2. ■

1.3 Utilità cardinale

Se le decisioni comportano conseguenze certe, DM sarà guidato nella sua scelta solamente dalla relazione di preferenza $\succeq$; conseguentemente, nel caso particolare che egli esprima le sue preferenze mediante un'utilità ordinale v, il problema decisionale si riduce ad un problema di ottimizzazione: individuare punti di massimo della funzione v. Purtroppo, nelle situazioni reali, le scelte in "condizioni di certezza" sono, per così dire, un caso limite; infatti, gli individui prendono decisioni che usualmente comportano conseguenze aleatorie[11]. Per affrontare il problema decisionale nella sua generalità, DM non può quindi limitarsi a confrontare solamente conseguenze certe ma deve anche elicere le sue preferenze nell'ambiente molto più ampio delle conseguenze aleatorie. Analogamente a quanto fatto per la relazione $\succeq$, viene allora naturale chiedersi quali proprietà questa nuova relazione di preferenza dovrebbe ragionevolmente possedere. Tra le varie risposte che sono via via apparse, andiamo ora a considerare quella che, senza ombra di dubbio, è a tutt'oggi la più "popolare".

Iniziamo col considerare una decisione d che comporti una conseguenza avente un numero finito di specificazioni possibili. Posto $C_d = \{c_1, \dots, c_n\}$, introduciamo la partizione dell'evento certo costituita dagli eventi:

$$E_i : \text{"}d \text{ comporta la conseguenza } c_i\text{"} \quad (i = 1, \dots, n).$$

Supposto che DM sia in grado di esprimere la sua opinione sul verificarsi di tali eventi tramite una distribuzione di probabilità P_d, la scelta di d comporta per DM l'accesso ad una particolare lotteria: subire "a sorte" la conseguenza c_1 con probabilità $\mathrm{P}_d(E_1)$, ... , la conseguenza c_n con probabilità $\mathrm{P}_d(E_n)$. Ovviamente, nel caso delle conseguenze certe ($n = 1$), la distribuzione si concentra tutta su c_1 e quindi la lotteria degenera fornendo la conseguenza c_1 con certezza.

Abbandonando il caso finito e volendo ripercorrere, in qualche modo, le considerazioni fatte, la situazione si complica notevolmente poichè bisogna considerare distribuzioni di probabilità su insiemi infiniti. Per proseguire, scegliamo d'adottare l'usuale impostazione del calcolo delle probabilità basata sulle misure di probabilità. Data una decisione d che comporti una conseguenza avente un'infinità di specificazioni possibili, fissiamo dunque una σ-algebra $\mathcal{C}_d$ su C_d includente i singoletti che descriva, dal punto di vista interpretativo, gli insiemi $C \subseteq C_d$ per i quali l'evento:

$$E_C : \text{"}d \text{ comporta una conseguenza appartenente a } C\text{"}$$

presenti qualche interesse per DM. Supposto, analogamente al caso finito, che DM sia in grado di esprimere la sua opinione sul verificarsi degli eventi E_C in

[11] Si pensi, ad esempio, all'acquisto di un'automobile. Sceglierne una comporta come conseguenza, tra l'altro, il prezzo d'acquisto (che è certo) e gli importi futuri relativi alla sua manutenzione, chiaramente non noti (al momento dell'acquisto) dipendendo da fattori del tutto casuali (come il verificarsi di un guasto, di un incidente, etc.).

termini probabilistici, veniamo a individuare una probabilità P_d su $\mathcal{C}_d$. Conseguentenente, la scelta della decisione d comporta per DM l'accesso ad una particolare lotteria: ottenere "a sorte" l'evento "la conseguenza appartiene a C" con probabilità $\mathrm{P}_d(C)$ (e quindi, in particolare, subire la conseguenza $c \in C_d$ con probabilità $\mathrm{P}_d(\{c\})$).

La capacità di formulare valutazioni probabilistiche per gli eventi E_i (caso finito) ed E_C (caso infinito) consente dunque a DM di interpretare la conseguenza (certa o aleatoria) relativa alla decisione d come una lotteria governata da una opportuna probabilità P_d definita sugli "eventi" di una σ-algebra $\mathcal{C}_d$ su C_d[12].

Poichè tali probabilità si possono sempre estendere agli elementi di una σ-algebra "di riferimento" $\mathcal{C}$ su $\mathfrak{C}$ che contenga i singoletti e includa la famiglia $\bigcup_{d\in\mathfrak{D}} \mathcal{C}_d$[13], DM sarà naturalmente condotto, per descrivere le sue preferenze tra le conseguenze (certe o aleatorie), a confrontare tra di loro le probabilità (intese su $\mathcal{C}$) che regolano le corrispondenti lotterie.

Passando ad una formalizzazione di quanto esposto, assumiamo che DM, per risolvere il problema decisionale, consideri una relazione di preferenza $\succeq^*$ su un opportuno insieme $\pounds$ di probabilità definite su $\mathcal{C}$ (dette **lotterie**) che, oltre a contenere le probabilità P_d, includa anche le restrizioni su $\mathcal{C}$ delle misure di Dirac su $2^{\mathfrak{C}}$ (che chiameremo **lotterie degeneri**) e sia **chiuso per misture**[14] (includendo così anche le **lotterie semplici**, cioè le probabilità concentrate su insiemi finiti di conseguenze[15]). Conveniamo inoltre di denotare con ℓ (dotata o no di apici o pedici) una lotteria generica di $\pounds$ e, per ogni conseguenza c, con $\mathit{1}_c$ la restrizione su $\mathcal{C}$ della misura di Dirac δ_c su $2^{\mathfrak{C}}$. Per quanto riguarda le proprietà della preferenza $\succeq^*$, assumiamo intanto che sia un preordinamento completo compatibile con la preferenza $\succeq$.

A1* Riflessività: $\ell \succeq^* \ell$.

A2* Transitività: $\ell \succeq^* \ell' \wedge \ell' \succeq^* \ell'' \Rightarrow \ell \succeq^* \ell''$.

A3* Completezza: $\ell \succeq^* \ell'$ o $\ell' \succeq^* \ell$.

A4* Coerenza: $\mathit{1}_{c''} \succeq^* \mathit{1}_{c'} \iff c'' \succeq c'$ per ogni $c', c'' \in \mathfrak{C}$.

[12] Che, nel caso finito, coincide con l'insieme delle parti di C_d.

[13] Ponendo $\mathrm{P}_d(C) = \mathrm{P}_d(C \cap C_d)$ per ogni $C \in \mathcal{C}$.

[14] Nel senso che, qualunque siano $\ell_1, \ldots, \ell_n \in \pounds$ e $\vartheta_1, \ldots, \vartheta_n \in [0,1]$ tali che $\sum_{i=1}^n \vartheta_i = 1$, la probabilità $\sum_{i=1}^n \vartheta_i \ell_i$ (detta **mistura** delle lotterie $\ell_1, \ldots, \ell_n$ con **pesi** $\vartheta_1, \ldots, \vartheta_n$) appartiene a $\pounds$.

[15] Ricordiamo che una probabilità P su $\mathcal{C}$ si dice **concentrata su** $C \in \mathcal{C}$ se $\mathrm{P}(C) = 1$.

Naturalmente, come nel caso della preferenza $\succeq$, accanto alla relazione $\succeq^*$ considereremo anche le relative relazioni di indifferenza:

$$\ell \sim^* \ell' \iff \ell \succeq^* \ell' \wedge \ell' \succeq^* \ell$$

e di preferenza stretta:

$$\ell \succ^* \ell' \iff \ell \succeq^* \ell' \wedge \ell \not\sim^* \ell'$$

che sono, per A1*, A2* e A4*, rispettivamente, un'equivalenza coerente con $\sim$ e un ordinamento stretto coerente con $\succ$.

Passando a considerare ulteriori proprietà che la preferenza $\succeq^*$ dovrebbe avere, viene naturale ritenere che una lotteria ℓ' non sia peggiore (migliore) della lotteria ℓ, se ogni suo possibile esito non è peggiore (migliore) di ℓ.

A5* DOMINANZA: Siano $C' \in \mathcal{C}$ e ℓ' tali che $\ell'(C') = 1$. Riesce allora:
- $\ell' \succeq^* \ell$, se $\mathit{1}_{c'} \succeq^* \ell$ per ogni $c' \in C'$;
- $\ell \succeq^* \ell'$, se $\ell \succeq^* \mathit{1}_{c'}$ per ogni $c' \in C'$.

Al fine di individuare un'altra proprietà "naturale", conviene considerare un tipo abbastanza comune di scommessa sui cavalli: la *scommessa doppia*, consistente nel fare una puntata p su un dato cavallo nella prima corsa specificando che, in caso di vittoria, la vincita verrà utilizzata per puntare su un certo cavallo nella seconda corsa al fine di acquisire, in caso di vittoria, l'importo w (al netto della puntata p). Indicate con π_1 la probabilità che lo scommettitore associa all'evento "il primo cavallo scelto perde la corsa" e con π_2 quella che associa all'evento "il secondo cavallo scelto perde la corsa", i possibili risultati della scommessa doppia sono:
- perdere, con probabilità π_1, la puntata p;
- guadagnare, con probabilità $1 - \pi_1$, una scommessa per la seconda corsa avente come possibili risultati:
  - perdere la puntata p con probabilità π_2;
  - ricevere l'importo w con probabilità $1 - \pi_2$.

Andiamo ora a rappresentare la scommessa doppia in termini di lotterie concentrate sulle conseguenze $c_1 = -p$ e $c_2 = w$. Indicata con la coppia ordinata (l_1, l_2) la lotteria che associa probabilità l_1 alla conseguenza c_1 e probabilità l_2 alla conseguenza c_2, la scommessa relativa alla seconda corsa può essere rappresentata dalla lotteria $\ell_2 = (\pi_2, 1 - \pi_2)$. Considerata allora la lotteria degenere $\mathit{1}_{c_1} = (1, 0)$ che rappresenta l'evento "perdere l'importo p con certezza", la scommessa doppia può essere realizzata ricorrendo alla procedura:

"Usare la prima corsa come meccanismo di sorteggio per ottenere la lotteria $\mathit{1}_{c_1}$ con probabilità π_1 e la lotteria ℓ_2 con probabilità $1-\pi_1$"

che consente di interpretarla come una particolare **lotteria composta** (cioè un meccanismo di sorteggio - di nota legge probabilistica - i cui esiti sono delle lotterie).

Andiamo infine a valutare la probabilità di vittoria della scommessa doppia, cioè la probabilità π di vincere sia nella prima che nella seconda corsa. Supposto che lo scommettitore ritenga i risultati delle due corse tra loro indipendenti, si ha $\pi = (1-\pi_1)(1-\pi_2) = 1-[\pi_1+(1-\pi_1)\pi_2]$. Pertanto, le probabilità di perdere l'importo p e di ricevere l'importo w nella scommessa doppia coincidono con quelle relative alla lotteria mistura $\ell = \pi_1 \mathit{1}_{c_1} + (1-\pi_1)\ell_2$ delle due lotterie $\mathit{1}_{c_1}$, ℓ_2 con pesi π_1 e $1-\pi_1$.

Ritenendo, come peraltro naturale, che lo scommettitore valuti le scommesse sui cavalli basandosi unicamente sulle relative probabilità di vittoria, possiamo concludere che, per lui, la lotteria composta e la lotteria mistura sopra considerate rappresentano, malgrado la loro differente natura (aleatoria e certa, rispettivamente), il medesimo oggetto: la scommessa doppia.

Prendendo spunto da questo esempio, assumiamo quindi che, date due qualsiasi lotterie ℓ', ℓ'' e un arbitrario numero reale $\vartheta \in [0,1]$, la lotteria mistura $\ell'\vartheta\ell'' = \vartheta\ell' + (1-\vartheta)\ell''$ sia per DM del tutto equivalente alla lotteria composta retta da un meccanismo di sorteggio che fornisca la lotteria ℓ' con probabilità ϑ e la lotteria ℓ'' con probabilità $1-\vartheta$[16].

[16]Per realizzarlo (nel caso $\mathfrak{C} \subseteq \mathbb{R}$) si potrà ricorre ai metodi della simulazione osservando che, dato un numero pseudocasuale X distribuito uniformemente in $[0,1]$, la trasformata $Y = f(X)$ mediante la funzione:

$$f(t) = \begin{cases} -\infty & \text{se } t = 0 \\ 1-\lambda & \text{se } 0 < t \leq 1-\lambda \\ 1 & \text{se } 1-\lambda < t \leq 1 \end{cases}$$

è tale che $\Pr(Y=0) = 1-\lambda$ e $\Pr(Y=1) = \lambda$. Più in generale, ricordiamo che, considerata una v.a. X distribuita uniformemente in $[0,1]$ e data una funzione di ripartizione F, la v.a. $Y = f(X)$, trasformata di X mediante la funzione:

$$f(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\} = \sup\{x \in \mathbb{R} : F(x) < t\},$$

ammette F come funzione di ripartizione (per una dimostrazione si veda la Sezione B.7 del testo di Schervish).

L'identificazione delle lotterie mistura con quelle composte suggerisce di richiedere alla relazione di preferenza $\succeq^*$ la seguente ulteriore proprietà.

A6* INDIPENDENZA: $\ell'' \succeq^* \ell' \iff \ell''\vartheta\ell \succeq^* \ell'\vartheta\ell \qquad (\vartheta \in]0,1[)$.

Infatti, poichè in entrambe le lotterie composte $\ell''\vartheta\ell$ e $\ell'\vartheta\ell$, DM ottiene la lotteria ℓ con probabilità $1-\vartheta$, viene naturale ritenere che questo "termine costante" non influenzi in alcun modo la sua scelta. Conseguentemente, la preferenza di DM tra le due lotterie composte dipenderà unicamente dal "termine variabile" nelle due misture, cioè dalla sua preferenza tra le lotterie ℓ' e ℓ''.

Una ulteriore (e ultima) proprietà "naturale" che assumiamo per $\succeq^*$ è la seguente: se la lotteria ℓ'' è migliore della lotteria ℓ, è sempre possibile trovare, per quanto possa essere indesiderabile per DM la lotteria ℓ', una lotteria composta - che fornisca ℓ'' con probabilità $\alpha > 0$ e ℓ' con probabilità $1-\alpha > 0$ - che "scavalca" nella preferenza la lotteria ℓ; analogamente, se ℓ è migliore di ℓ', è sempre possibile trovare, per quanto possa essere desiderabile per DM la lotteria ℓ'', una lotteria composta - che fornisca ℓ'' con probabilità $\beta > 0$ e ℓ' con probabilità $1-\beta > 0$ - che viene "scavalcata" nella preferenza dalla lotteria ℓ.[17]

A7* PROPRIETÀ ARCHIMEDEA: Sia $\ell'' \succ^* \ell'$. Allora, per ogni ℓ tale che:
- $\ell'' \succ^* \ell \succeq^* \ell'$, esiste $\alpha \in]0,1[$ tale che $\ell'' \succ^* \ell''\alpha\,\ell' \succ^* \ell$;
- $\ell'' \succeq^* \ell \succ^* \ell'$, esiste $\beta \in]0,1[$ tale che $\ell \succ^* \ell''\beta\,\ell' \succ^* \ell'$.

Il prossimo teorema (chiave di volta della teoria delle scelte individuali in condizioni di incertezza) mette in evidenza che le proprietà sin qui ammesse per le preferenze di DM sulle lotterie permettono d'individuare - qualora $\mathcal{C}$ includa gli insiemi $\succeq$-connessi di $\mathfrak{C}$ e $\pounds$ le **lotterie a supporto limitato**[18]- un'utilità ordinale (per $\succeq$) che consente, nell'ambito delle lotterie a supporto limitato, una rappresentazione numerica di $\succeq^*$ in termini di speranze matematiche[19].

[17]Si potrebbe pensare che questa proprietà venga violata nei casi in cui tra i possibili esiti della lotteria ℓ' ci sia la morte del decisore (che, salvo casi estremi, può ragionevolmente ritenersi la conseguenza peggiore che gli possa capitare). Ma se così fosse, nessuno si assumerebbe, ad esempio, il rischio di essere investito e di morire attraversando una strada trafficata per andare a bere un caffè ritenuto migliore di quello offerto nel bar sotto casa.

[18]Cioè, probabilità su $\mathcal{C}$ che sono concentrate su intervalli chiusi di $\mathfrak{C}$.

[19]La relativa dimostrazione (alquanto complessa) viene riportata, in corpo minore, alla fine della sezione in quanto, essendo essenzialmente tecnica, non interviene significativamente nello sviluppo del contesto teorico che stiamo delineando.

Teorema 1.3.1 *Sussistono le seguenti proposizioni:*

(i) Esista un'utilità ordinale u (per $\succeq$) integrabile rispetto ad ogni lotteria. Allora, la relazione $\succeq^*$ su £ così definita:

$$\ell \succeq^* \ell' \iff \mathrm{E}_\ell(u) = \int_{\mathfrak{C}} u \, d\ell \geq \int_{\mathfrak{C}} u \, d\ell' = \mathrm{E}_{\ell'}(u)$$

è un preordinamento che verifica la proprietà archimedea e quelle di completezza, coerenza, dominanza e indipendenza;

(ii) La σ-algebra $\mathcal{C}$ includa gli insiemi $\succeq$-connessi di $\mathfrak{C}$ e la famiglia £ le lotterie a supporto limitato. Inoltre, $\succeq^*$ sia un preordinamento su £ che verifica la proprietà archimedea e quelle di completezza, coerenza, dominanza e indipendenza. Esiste allora un'utilità ordinale u (per $\succeq$) tale che $\ell \succeq^* \ell' \iff \mathrm{E}_\ell(u) \geq \mathrm{E}_{\ell'}(u)$ per ogni lotteria ℓ, ℓ' a supporto limitato.

L'assunzione che le lotterie mistura siano del tutto equivalenti alle corrispondenti lotterie composte ha consentito di giustificare, in qualche modo, le proprietà di indipendenza e archimedea (e quindi, in ultima analisi, il ricorso alle speranze matematiche di un'opportuna utilità ordinale per esprimere le preferenze sulle lotterie). A causa di questo ruolo chiave nell'impostazione concettuale delle scelte individuali in condizioni di incertezza, particolare attenzione è stata posta, dagli psicologi e dagli studiosi del comportamento umano, per individuare il reale valore descrittivo sia di questa identificazione che della "conseguente" proprietà di indipendenza. Per ottenerlo, il metodo comunemente adottato è stato quello di organizzare degli esperimenti, proponendo a delle popolazioni campione (scelte in qualche modo) opportuni problemi di decisione in condizioni di incertezza, al fine di stabilire se le risposte ottenute siano o no in accordo con l'identificazione fatta. Pur essendo tale metodologia soggetta a numerose critiche[20], i risultati ottenuti hanno generato una notevole mole di ricerche che hanno condotto, anche recentemente, a nuove impostazioni del problema delle scelte in condizioni di incertezza[21] .

Dei vari esperimenti effettuati, riportiamo una versione di uno dei più celebri (dovuto a Maurice Allais) che riguarda la proprietà di indipendenza. Date le lotterie:

- ℓ_1: assegna un premio di 27.500 euro con probabilità 0,33, un premio di 24.000 euro con probabilità 0,66 e un premio nullo con probabilità 0,01;

[20]Dovute alla scelta della popolazione (usualmente matricole universitarie) da cui estrarre il campione; dalla mancanza di incentivi che inducano i soggetti ad impegnarsi seriamente nel rispondere alle domande poste; alla reale portata delle risposte ottenute nell'illuminare il comportamento degli individui in situazioni reali; etc..

[21]Si veda, ad esempio, Schmidt, U., Axiomatic Utility Theory under Risk: Non-Archimedean Representations and Applications to Insurance Economics, *Lectures Notes in Economics and Mathematical Systems*, Springer (1998).

- ℓ_2: assegna un premio di 24.000 euro con certezza;

- ℓ_3: assegna un premio di 27.500 euro con probabilità 0,33 e un premio nullo con probabilità $0,67$;

- ℓ_4: assegna un premio di 24.000 euro con probabilità 0,34 e un premio nullo con probabilità 0,66

si considerino le opportunità di scelta:

- Opportunità 1: Il soggetto sceglie tra le lotterie ℓ_1 e ℓ_2;

- Opportunità 2: Il soggetto sceglie tra le lotterie ℓ_3 e ℓ_4.

La risposta più frequente data dai soggetti sottoposti all'esperimento:

"Scelta di ℓ_2 nella prima opportunità e di ℓ_3 nella seconda"

contraddice purtroppo la proprietà di indipendenza (**paradosso di Allais**). Infatti, considerate le lotterie:

- ℓ_5: assegna un premio di 27.500 euro con probabilità $\frac{33}{34}$ e un premio nullo con probabilità $\frac{1}{34}$;

- ℓ_6: assegna un premio nullo con certezza,

riesce $\ell_1 = 0,34\,\ell_5 + 0,66\,\ell_2$, $\ell_2 = 0,34\,\ell_2 + 0,66\,\ell_2$, $\ell_3 = 0,34\,\ell_5 + 0,66\,\ell_6$ e $\ell_4 = 0,34\,\ell_2 + 0,66\,\ell_6$. Ne segue, per la proprietà di indipendenza,

$$\ell_2 \succ^* \ell_1 \iff \ell_2 \succ^* \ell_5 \iff \ell_4 \succ^* \ell_3.$$

Si potrebbe allora pensare che questo "scollamento" tra le teoria e la pratica metta in discussione solamente la validità descrittiva della proprietà di indipendenza e non quella relativa all'identificazione fatta tra lotterie mistura e lotterie composte. Purtroppo non è così come ha messo in evidenza l'esperimento seguente. Date le lotterie composte:

- L_1: assegna ℓ_5 con probabilità 0,34 e un premio di 24.000 euro con probabilità 0,66;

- L_3: assegna ℓ_5 con probabilità 0,34 e un premio nullo con probabilità $0,66$

si considerino le opportunità di scelta:

- Opportunità I: Il soggetto sceglie tra le lotterie L_1 e ℓ_2;

- Opportunità II: Il soggetto sceglie tra le lotterie L_3 e ℓ_4.

In questo caso, la frequenza delle risposte coerenti con la teoria ("L_1 e L_3" oppure "ℓ_2 e ℓ_4") è significativamente maggiore che nel caso del paradosso di Allais, pur essendo tali lotterie composte "coincidenti" con le lotterie considerate nel paradosso. Quindi, da un punto di vista descrittivo, proporre delle scommesse nella forma di lotterie composte altera la scelta degli individui rispetto al caso in cui siano proposte in termini di lotterie semplici.

La "debolezza" applicativa del Teorema 1.3.1, rilevata da questi esperimenti, ha suggerito una sua interpretazione (che adottiamo senz'altro) di tipo normativo: esso fornisce un modello di comportamento che può essere utilmente utilizzato sia per controllare la razionalità delle scelte degli individui che per individuare la decisione "migliore", in situazioni (più o meno complesse) di scelte in condizioni di incertezza.

Prendendo spunto dal Teorema 1.3.1, la prossima definizione fissa l'attenzione su quelle particolari utilità ordinali che consentono a DM di descrivere le sue preferenze sulle lotterie in termini di confronti tra speranze matematiche (anche non necessariamente finite).

Definizione 1.3.2 Un'utilità ordinale (per $\succeq$) u è una **utilità cardinale** (per $\succeq^*$) se: $\ell \succeq^* \ell' \iff \mathrm{E}_\ell(u) \geq \mathrm{E}_{\ell'}(u)$ per ogni lotteria ℓ e ℓ'. Inoltre, l'**utilità attesa** della lotteria ℓ è la speranza matematica $\mathrm{E}_\ell(u) = \int_{\mathfrak{C}} u \, d\ell$.[22]

Il risultato seguente rileva che un'utilità cardinale è in grado di descrivere anche le relazioni di indifferenza e di preferenza stretta associate a $\succeq^*$; inoltre, assicura che è definita a meno di trasformazioni affini positive.

Teorema 1.3.3 *Sia u un'utilità cardinale. Riesce allora:*

(i) $\ell \sim^* \ell' \iff \mathrm{E}_\ell(u) = \mathrm{E}_{\ell'}(u)$;

(ii) $\ell \succ^* \ell' \iff \mathrm{E}_\ell(u) > \mathrm{E}_{\ell'}(u)$.

Inoltre, $\tilde{u} : \mathfrak{C} \mapsto \mathbb{R}$ è un'utilità cardinale per $\succeq^*$ se e solo se esistono due numeri reali, $a > 0$ e b, tali che $\tilde{v} = au + b$.

Dimostrazione Essendo le dimostrazioni di (i) e di (ii) immediate, ci limitiamo a verificare l'ultima parte della tesi. Osservato che la condizione sufficiente è banale, passiamo a quella necessaria assumendo che $\tilde{u}$ sia un'utilità cardinale per $\succeq^*$. Allora, per la proprietà di non degenerazione A3 e (ii), $u(c^*) > u(c_*)$ e $\tilde{u}(c^*) > \tilde{u}(c_*)$. Possiamo quindi considerare le trasformate affini positive, rispettivamente, di u e $\tilde{u}$:

$$u_1 = \frac{1}{u(c^*) - u(c_*)}\,[u - u(c_*)], \qquad u_2 = \frac{1}{\tilde{u}(c^*) - \tilde{u}(c_*)}\,[\tilde{u} - \tilde{u}(c_*)]$$

[22] Conseguentemente, per l'Osservazione A.3.9, l'utilità attesa di una lotteria degenere 1_c è il valore $u(c)$ che l'utilità cardinale assume su c e, per il Teorema A.3.4(ii), il valore atteso di una lotteria mistura $\sum_{i=1}^n \vartheta_i \ell_i$ è la mistura $\sum_{i=1}^n \vartheta_i \mathrm{E}_{\ell_i}(u)$ dei valori attesi delle lotterie che la compongono (ogniqualvolta tale mistura non contempli espressioni aritmetiche illecite).

ottenendo così due utilità cardinali (per $\succeq^*$) tali che $u_i(c_*) = 0$ e $u_i(c^*) = 1$ $(i = 1, 2)$.

A questo punto basta verificare che riesce $u_1 = u_2$; infatti, dall'uguaglianza

$$\frac{1}{u(c^*) - u(c_*)}\,[u - u(c_*)] = \frac{1}{\tilde{u}(c^*) - \tilde{u}(c_*)}\,[\tilde{u} - \tilde{u}(c_*)]$$

risulta

$$\tilde{u} = \frac{\tilde{u}(c^*) - \tilde{u}(c_*)}{u(c^*) - u(c_*)}\,u + \left[\tilde{u}(c_*) - \frac{\tilde{u}(c^*) - \tilde{u}(c_*)}{u(c^*) - u(c_*)}\,u(c_*)\right]$$

con $\frac{\tilde{u}(c^*) - \tilde{u}(c_*)}{u(c^*) - u(c_*)} > 0$ e quindi $\tilde{u}$ è una trasformata affine positiva di u.

Per provare che le utilità u_1, u_2 coincidono, osserviamo intanto che, date le conseguenze c, c' e c'' con $c \in [c', c'']$, risulta $\mathit{1}_c \sim^* \mathit{1}_{c''}\vartheta\mathit{1}_{c'}$ per qualche $\vartheta \in [0, 1]$. Infatti, dalla $c'' \succeq c \succeq c'$ otteniamo $u(c'') \geq u(c) \geq u(c')$ e quindi esiste $\vartheta \in [0, 1]$ tale che $u(c) = \vartheta u(c'') + (1 - \vartheta)u(c')$. Ne segue

$$\begin{aligned}\mathrm{E}_{\mathit{1}_c}(u) &= u(c) = \vartheta u(c'') + (1 - \vartheta)u(c') = \vartheta\mathrm{E}_{\mathit{1}_{c''}}(u) + (1 - \vartheta)\mathrm{E}_{\mathit{1}_{c'}}(u)\\ &= \mathrm{E}_{\vartheta\mathit{1}_{c''} + (1-\vartheta)\mathit{1}_{c'}}(u)\end{aligned}$$

da cui, tramite (i), $\mathit{1}_c \sim^* \vartheta\mathit{1}_{c''} + (1 - \vartheta)\mathit{1}_{c'} = \mathit{1}_{c''}\vartheta\mathit{1}_{c'}$.

Sia ora $i \in \{1, 2\}$. Per valutare $u_i(c)$ procediamo esaminando le possibili "posizioni" di c rispetto c_* e c^* ($\succeq$ è completa!). Sia intanto $c \in [c_*, c^*]$. Per quanto provato, esiste $\vartheta \in [0, 1]$ tale che $\mathit{1}_c \sim^* \mathit{1}_{c^*}\vartheta\mathit{1}_{c_*}$. Da (i) segue allora

$$u_i(c) = \mathrm{E}_{\vartheta\mathit{1}_{c^*} + (1-\vartheta)\mathit{1}_{c_*}}(u_i) = \vartheta u_i(c^*) + (1 - \vartheta)u_i(c_*) = \vartheta,$$

ricordato che $u_i(c_*) = 0$. Sia ora $c \succ c^*$. Poichè $c^* \in [c_*, c]$, esiste $\vartheta \in]0, 1[$ tale che $\mathit{1}_{c^*} \sim^* \mathit{1}_c\vartheta\mathit{1}_{c_*}$. Da (i) si ha allora

$$1 = u_i(c^*) = \mathrm{E}_{\vartheta\mathit{1}_c + (1-\vartheta)\mathit{1}_{c_*}}(u_i) = \vartheta u_i(c) + (1 - \vartheta)u_i(c_*) = \vartheta u_i(c)$$

da cui risulta $u_i(c) = \frac{1}{\vartheta}$. Sia infine $c_* \succ c$. Poichè $c_* \in [c, c^*]$, esiste $\vartheta \in]0, 1[$ tale che $\mathit{1}_{c_*} \sim^* \mathit{1}_{c^*}\vartheta\mathit{1}_c$. Da (i) riesce allora

$$0 = u_i(c_*) = \mathrm{E}_{\vartheta\mathit{1}_{c^*} + (1-\vartheta)\mathit{1}_c}(u_i) = \vartheta u_i(c^*) + (1 - \vartheta)u_i(c) = \vartheta + (1 - \vartheta)u_i(c)$$

da cui segue $u_i(c) = \frac{\vartheta}{\vartheta - 1}$. In ogni caso si ha dunque $u_1(c) = u_2(c)$. ■

Osservazione 1.3.4 Per il teorema appena provato, le utilità cardinali "conservano i rapporti", nel senso che, se u, $\tilde{u}$ sono utilità cardinali per $\succeq^*$, risulta

$$\frac{\tilde{u}(c'') - \tilde{u}(c)}{\tilde{u}(c') - \tilde{u}(c)} = \frac{u(c'') - u(c)}{u(c') - u(c)}$$

qualunque siano le conseguenze c, c' e c'' tali che $c' \not\sim c$. Tali utilità hanno dunque la medesima *natura cardinale* (da cui la loro denominazione) degli usuali sistemi di misurazione come, ad esempio, quelli della temperatura (gradi Celsius, Farhenheit, Kelvin, etc.), della distanza (Km, anni-luce, nodi, etc.), del peso (Kg, libbre, once, etc.) e così via. Natura che, per il Teorema 1.2.2, non è certamente inerente alle utilità ordinali (basti pensare, ad esempio, all'utilità ordinale ottenuta dall'utilità cardinale u tramite la trasformazione crescente $t \mapsto t^3$). ♦

Dimostrazione del Teorema 1.3.1 La proposizione (i) si ottiene facilmente ricorrendo alle proprietà elementari dell'integrale di Lebesgue. Per verificare la proposizione (ii), iniziamo col provare un lemma che fornisce alcune utili conseguenze delle proprietà di indipendenza e di completezza. Le prime due assicurano che la proprietà di indipendenza sussiste anche per le relazioni di preferenza stretta e di indifferenza; la terza mette in evidenza che la mistura di due lotterie è una lotteria sempre compresa tra di esse; la quarta invece mostra che se la lotteria ℓ_2 è migliore della lotteria ℓ_1, allora la loro mistura $\ell_2\vartheta\ell_1$ aumenta la sua preferibilità all'aumentare del peso $\vartheta \in [0, 1]$.

Lemma 1.3.5 *Risulta:*

(i) $\ell_2 \succ^* \ell_1 \iff \ell_2\vartheta\ell \succ^* \ell_1\vartheta\ell$;

(ii) $\ell_2 \sim^* \ell_1 \iff \ell_2\vartheta\ell \sim^* \ell_1\vartheta\ell$

per ogni $\vartheta \in]0, 1[$. *Inoltre:*

(iii) Se $\ell_2 \succ^* \ell_1$, *allora* $\ell_2 \succ^* \ell_2\vartheta\ell_1 \succ^* \ell_1$ *per ogni* $\vartheta \in]0, 1[$;

(iv) Se $\ell_2 \succ^* \ell_1$ *e* $1 \geq \vartheta'' > \vartheta' \geq 0$, *allora* $\ell_2\vartheta''\ell_1 \succ^* \ell_2\vartheta'\ell_1$;

(v) Se $\ell'_i \succeq^* \ell_i$, $\vartheta_i \geq 0\, (i = 1, \ldots, n)$ *e* $\sum_{i=1}^n \vartheta_i = 1$, *allora* $\sum_{i=1}^n \vartheta_i\ell'_i \succeq^* \sum_{i=1}^n \vartheta_i\ell_i$.

Dimostrazione (i) Sia intanto $\ell_2 \succ^* \ell_1$, cioè $\ell_2 \succeq^* \ell_1$ e $\ell_2 \not\sim^* \ell_1$. Assumiamo (per assurdo) $\ell_2\vartheta\ell \not\succ^* \ell_1\vartheta\ell$ con $\vartheta \in]0, 1[$. Allora, per la proprietà di completezza, $\ell_1\vartheta\ell \succeq^* \ell_2\vartheta\ell$ da cui, per la proprietà di indipendenza, otteniamo $\ell_1 \succeq^* \ell_2$ e quindi $\ell_2 \sim^* \ell_1$ (Contraddizione!). In modo analogo si prova l'implicazione opposta.

(ii) Per la proprietà di indipendenza, si ha

$$\ell_2 \sim^* \ell_1 \iff \ell_2 \succeq^* \ell_1 \wedge \ell_1 \succeq^* \ell_2 \iff \ell_2\vartheta\ell \succeq^* \ell_1\vartheta\ell \wedge \ell_1\vartheta\ell \succeq^* \ell_2\vartheta\ell \iff \ell_2\vartheta\ell \sim^* \ell_1\vartheta\ell.$$

(iii) Siano $\ell_2 \succ^* \ell_1$ e $\vartheta \in]0, 1[$. Allora, $1 - \vartheta \in]0, 1[$ da cui, posto $\ell = \ell_2$ in (i), segue $\ell_2 = \ell_2(1 - \vartheta)\ell_2 \succ^* \ell_1(1 - \vartheta)\ell_2 = \ell_2\vartheta\ell_1$ e quindi, posto $\ell = \ell_1$ in (i), $\ell_2 \succ^* \ell_2\vartheta\ell_1 \succ^* \ell_1\vartheta\ell_1 = \ell_1$.

(iv) Supposto $\ell_2 \succ^* \ell_1$ e $1 \geq \vartheta'' > \vartheta' \geq 0$, da (iii) otteniamo

$$\ell_2 \succ^* \ell_2 \vartheta' \ell_1. \tag{1.5}$$

Posto $\alpha = \frac{1-\vartheta''}{1-\vartheta'}$, si ha $0 \leq \alpha < 1$ e inoltre

$$\begin{aligned}
\alpha(\ell_2\vartheta'\ell_1) + (1-\alpha)\ell_2 &= \frac{1-\vartheta''}{1-\vartheta'}\big[\vartheta'\ell_2 + (1-\vartheta')\ell_1\big] + \Big(1 - \frac{1-\vartheta''}{1-\vartheta'}\Big)\ell_2 \\
&= \frac{1-\vartheta''}{1-\vartheta'}\,\vartheta'\ell_2 + (1-\vartheta'')\ell_1 + \frac{\vartheta''-\vartheta'}{1-\vartheta'}\ell_2 \\
&= \vartheta''\ell_2 + (1-\vartheta'')\ell_1 = \ell_2\vartheta''\ell_1.
\end{aligned}$$

Se $\alpha = 0$, allora $\ell_2\vartheta''\ell_1 = \ell_2$ e quindi, per (1.5), $\ell_2\vartheta''\ell_1 \succ^* \ell_2\vartheta'\ell_1$. Se invece $\alpha > 0$, da (i) e (1.5), otteniamo $\ell_2\vartheta''\ell_1 = \alpha(\ell_2\vartheta'\ell_1) + (1-\alpha)\ell_2 \succ^* \alpha(\ell_2\vartheta'\ell_1) + (1-\alpha)\ell_2\vartheta'\ell_1 = \ell_2\vartheta'\ell_1$.

(v) Assumiamo $\ell_i' \succeq^* \ell_i$, $\vartheta_i \geq 0$ $(i = 1, \ldots, n)$, $\sum_{i=1}^n \vartheta_i = 1$ e procediamo per induzione su n. Osservato che la tesi è banale per $n = 1$, supponiamo che sussista per $n = m$ (ipotesi induttiva) e proviamola per $n = m+1$. Poichè nel caso $\vartheta_{m+1} = 1$ la tesi è ancora banale, sia $\vartheta_{m+1} < 1$. Ne segue $\vartheta = \sum_{i=1}^m \vartheta_i = 1 - \vartheta_{m+1} > 0$ e quindi possiamo considerare il suo reciproco $\frac{1}{\vartheta}$. Riesce allora

$$\sum_{i=1}^{m+1} \vartheta_i \ell_i' = \vartheta \sum_{i=1}^{m} \frac{\vartheta_i}{\vartheta}\, \ell_i' + \vartheta_{m+1}\ell_{m+1}'$$

da cui, tenuto conto dell'ipotesi induttiva $\sum_{i=1}^m \frac{\vartheta_i}{\vartheta}\, \ell_i' \succeq^* \sum_{i=1}^m \frac{\vartheta_i}{\vartheta}\, \ell_i$ e della proprietà di indipendenza A6*, otteniamo

$$\sum_{i=1}^{m+1} \vartheta_i \ell_i' \succeq^* \vartheta \sum_{i=1}^{m} \frac{\vartheta_i}{\vartheta}\, \ell_i + \vartheta_{m+1}\ell_{m+1}'$$

e quindi, ancora per A6*,

$$\sum_{i=1}^{m+1} \vartheta_i \ell_i' \succeq^* \vartheta \sum_{i=1}^{m} \frac{\vartheta_i}{\vartheta}\, \ell_i + \vartheta_{m+1}\ell_{m+1} = \sum_{i=1}^{m} \vartheta_i \ell_i + \vartheta_{m+1}\ell_{m+1} = \sum_{i=1}^{m+1} \vartheta_i \ell_i.$$

La dimostrazione è così conclusa. ■

Siamo ora in grado di provare che la restrizione della preferenza $\succeq^*$ su un intervallo di lotterie è descrivibile in termini di un'utilità ordinale (definita su tale intervallo) che risulta lineare sulle misture.

Lemma 1.3.6 *Sia $\ell_2 \succ^* \ell_1$. Considerato allora l'intervallo chiuso $[\ell_1, \ell_2]$ di £, esiste un'applicazione suriettiva $U_{\ell_1,\ell_2} : [\ell_1, \ell_2] \mapsto [0, 1]$ tale che:*

(i) $\ell \sim^* \ell_2\vartheta\ell_1 \iff \vartheta = U_{\ell_1,\ell_2}(\ell)$;

(ii) $\ell'' \succ^* \ell' \iff U_{\ell_1,\ell_2}(\ell'') > U_{\ell_1,\ell_2}(\ell')$;

(iii) $\ell'' \succeq^* \ell' \iff U_{\ell_1,\ell_2}(\ell'') \geq U_{\ell_1,\ell_2}(\ell')$;

(iv) $U_{\ell_1,\ell_2}(\ell'\vartheta\ell'') = \vartheta U_{\ell_1,\ell_2}(\ell') + (1-\vartheta)U_{\ell_1,\ell_2}(\ell'')$

per ogni $\ell, \ell', \ell'' \in [\ell_1, \ell_2]$ *e* $\vartheta \in [0,1]$.

Dimostrazione Al fine di individuare l'applicazione $\tilde{U} = U_{\ell_1,\ell_2}$, scegliamo una lotteria qualsiasi $\ell \in [\ell_1, \ell_2]$ e poniamo:

$$A_\ell = \{\vartheta' \in [0,1] : \ell \succeq^* \ell_2\vartheta'\ell_1\}, \qquad B_\ell = \{\vartheta'' \in [0,1] : \ell_2\vartheta''\ell_1 \succeq^* \ell\}.$$

Allora, $0 \in A_\ell$, $1 \in B_\ell$ e, per la completezza di $\succeq^*$, $A_\ell \cup B_\ell = [0,1]$. Inoltre, si ha $\sup A_\ell \leq \inf B_\ell$; infatti, se così non fosse, esisterebbero $\vartheta' \in A_\ell$ e $\vartheta'' \in B_\ell$ tali che $\inf B_\ell < \vartheta'' < \vartheta' < \sup A_\ell$ e quindi, per il Lemma 1.3.5(iv), si avrebbe $\ell \succeq^* \ell_2\vartheta'\ell_1 \succ^* \ell_2\vartheta''\ell_1 \succeq^* \ell$ (Contraddizione!). Riesce pertanto $\sup A_\ell = \inf B_\ell$ per cui possiamo porre:

$$\tilde{U}(\ell) = \sup A_\ell = \inf B_\ell.$$

Si ha quindi $\tilde{U}(\ell_1) = 0$ e $\tilde{U}(\ell_2) = 1$ (infatti $0 \in B_{\ell_1}$ e $1 \in A_{\ell_2}$).

(i) Posto $\vartheta = \tilde{U}(\ell)$, assumiamo (per assurdo) $\ell \not\sim^* \ell_2\vartheta\ell_1$, cioè $\ell \succ^* \ell_2\vartheta\ell_1$ o $\ell_2\vartheta\ell_1 \succ^* \ell$ ($\succeq^*$ è completa!). Nel primo caso, per la proprietà archimedea (ponendo $\ell'' = \ell_2$ e $\ell' = \ell_2\vartheta\ell_1$ in A7*), esiste $\beta \in]0,1[$ tale che

$$\ell \succ^* \ell_2\beta(\ell_2\vartheta\ell_1) = \beta\ell_2 + (1-\beta)\big[\vartheta\ell_2 + (1-\vartheta)\ell_1\big] = \ell_2\big[\beta + (1-\beta)\vartheta\big]\ell_1.$$

Allora $\beta + (1-\beta)\vartheta \in A_\ell$ da cui otteniamo $\beta + (1-\beta)\vartheta \leq \sup A_\ell = \vartheta$; ne segue $\vartheta = 1$ e quindi $\ell_2 \succeq^* \ell \succ^* \ell_2\vartheta\ell_1 = \ell_2$ (Contraddizione!). Passando a considerare il secondo caso, si giunge in modo analogo ad una contradizione.

(ii) Dati $\ell', \ell'' \in [\ell_1, \ell_2]$, sia intanto $\ell'' \succ^* \ell'$, cioè $\ell'' \not\sim^* \ell'$ e $\ell'' \succeq^* \ell'$. Allora, per (i), $\tilde{U}(\ell') \neq \tilde{U}(\ell'')$ e inoltre $A_{\ell'} \subseteq A_{\ell''}$. Ne segue $\tilde{U}(\ell') = \sup A_{\ell'} \leq \sup A_{\ell''} = \tilde{U}(\ell'')$ e quindi $\tilde{U}(\ell') < \tilde{U}(\ell'')$. In modo analogo si prova l'implicazione opposta.

(iii) Conseguenza immediata di (i) e (ii).

(iv) Siano $\ell', \ell'' \in [\ell_1, \ell_2]$, $\vartheta \in [0,1]$ e $\ell = \ell'\vartheta\ell''$. Posto $\vartheta' = \tilde{U}(\ell')$ e $\vartheta'' = \tilde{U}(\ell'')$, da (i) otteniamo $\ell' \sim^* \ell_2\vartheta'\ell_1$ e $\ell'' \sim^* \ell_2\vartheta''\ell_1$. Ne segue, tramite il Lemma 1.3.5(ii),

$$\ell \sim^* [\ell_2\vartheta'\ell_1]\,\vartheta\,[\ell_2\,\vartheta''\ell_1] = \vartheta\big[\vartheta'\ell_2 + (1-\vartheta')\ell_1\big] + (1-\vartheta)\big[\vartheta''\ell_2 + (1-\vartheta'')\ell_1\big]$$

e quindi

$$\ell \sim^* \ell_2\big[\vartheta\vartheta' + (1-\vartheta)\vartheta''\big]\ell_1. \tag{1.6}$$

Proviamo ora che $\ell_2 \succeq^* \ell$. Dalla $\ell_2 \succeq^* \ell''$, per la proprietà di indipendenza A6*, si ha $\ell_2 = \ell_2(1-\vartheta)\ell_2 \succeq^* \ell''(1-\vartheta)\ell_2 = \ell_2\vartheta\ell''$. Inoltre, sempre per A6*, da $\ell_2 \succeq^* \ell'$ otteniamo $\ell_2\vartheta\ell'' \succeq^* \ell'\vartheta\ell'' = \ell$. Riesce dunque $\ell_2 \succeq^* \ell$. Poichè in modo analogo si prova $\ell \succeq^* \ell_1$, si ha $\ell \in [\ell_1, \ell_2]$ e quindi, tramite (i) e (1.6), $\tilde{U}(\ell) = \vartheta\tilde{U}(\ell') + (1-\vartheta)\tilde{U}(\ell'')$.

Osservato infine che, per $\tilde{U}(\ell_1) = 0$, $\tilde{U}(\ell_2) = 1$ e (iv), $\tilde{U}(\ell_2\vartheta\ell_1) = \vartheta$ per ogni $\vartheta \in [0,1]$, possiamo concludere che la funzione $\tilde{U}$ è suriettiva. ■

Tramite la famiglia $\{U_{\ell_1,\ell_2} : \ell_2 \succ^* \ell_1\}$ delle utilità ordinali associate alle "preferenze intervallari" $\succeq^*\big|_{[\ell_1,\ell_2]}$[23], possiamo ora individuare un'utilità ordinale U (per $\succeq^*$) lineare sulle misture. Data un'arbitraria lotteria ℓ, l'idea guida per la costruzione di $U(\ell)$ è considerare la "posizione" di ℓ rispetto alle due lotterie degeneri $\ell_* = 1_{c_*}$ e $\ell^* = 1_{c^*}$. Poniamo infatti:

$$U(\ell) = \begin{cases} -\dfrac{U_{\ell,\ell^*}(\ell_*)}{1 - U_{\ell,\ell^*}(\ell_*)} & \text{se } \ell_* \succ^* \ell \\ U_{\ell_*,\ell^*}(\ell) & \text{se } \ell \in [\ell_*, \ell^*] \\ \dfrac{1}{U_{\ell_*,\ell}(\ell^*)} & \text{se } \ell \succ^* \ell^*. \end{cases}$$

Riesce allora $U(\ell_*) = 0$ e $U(\ell^*) = 1$. Inoltre, sussiste l'implicazione:

$$\ell_2 \succeq^* \ell^* \succ^* \ell_* \succeq^* \ell_1 \Rightarrow \forall \ell \in [\ell_1, \ell_2] \Big(U_{\ell_1,\ell_2}(\ell) = \big[U_{\ell_1,\ell_2}(\ell^*) - U_{\ell_1,\ell_2}(\ell_*)\big]U(\ell) + U_{\ell_1,\ell_2}(\ell_*)\Big). \tag{1.7}$$

Infatti, supposto $\ell_2 \succeq^* \ell^* \succ^* \ell_* \succeq^* \ell_1$ e data $\ell \in [\ell_1, \ell_2]$, sia intanto $\ell \in [\ell_*, \ell^*]$. Allora, per il Lemma 1.3.6(i), $\ell \sim^* \ell^*\ U_{\ell_*,\ell^*}(\ell)\ \ell_* = \ell^*\ U(\ell)\ \ell_*$ da cui, tramite il Lemma 1.3.6(iv), si ha $U_{\ell_1,\ell_2}(\ell) = U(\ell)\,U_{\ell_1,\ell_2}(\ell^*) + (1 - U(\ell))\,U_{\ell_1,\ell_2}(\ell_*)$ e quindi

$$U_{\ell_1,\ell_2}(\ell) = \big[U_{\ell_1,\ell_2}(\ell^*) - U_{\ell_1,\ell_2}(\ell_*)\big]U(\ell) + U_{\ell_1,\ell_2}(\ell_*). \tag{1.8}$$

Sia ora $\ell_* \succ^* \ell$. Dunque $\ell_* \in [\ell, \ell^*]$ e $U(\ell) = -\frac{U_{\ell,\ell^*}(\ell_*)}{1-U_{\ell,\ell^*}(\ell_*)}$. Allora, per il Lemma 1.3.6(i),

$$\ell_* \sim^* \ell^*\ U_{\ell,\ell^*}(\ell_*)\ \ell = -\frac{U(\ell)}{1-U(\ell)}\ \ell^* + \frac{1}{1-U(\ell)}\ \ell$$

da cui, per il Lemma 1.3.6(ii),(iv), otteniamo

$$U_{\ell_1,\ell_2}(\ell_*) \quad = \quad -\frac{U(\ell)}{1-U(\ell)}\,U_{\ell_1,\ell_2}(\ell^*) + \frac{1}{1-U(\ell)}\,U_{\ell_1,\ell_2}(\ell)$$

e quindi, riordinando opportunamente i termini, (1.8). Supposto infine $\ell \succ^* \ell^*$, si ha $\ell^* \in [\ell_*, \ell]$ e $U(\ell) = \frac{1}{U_{\ell_*,\ell}(\ell^*)}$. Allora,

$$\ell^* \sim^* \ell\ U_{\ell_*,\ell}(\ell^*)\ \ell_* = \frac{1}{U(\ell)}\,\ell + \Big(1 - \frac{1}{U(\ell)}\Big)\,\ell_*$$

da cui, sempre per il Lemma 1.3.6(ii),(iv), risulta

$$U_{\ell_1,\ell_2}(\ell^*) = \frac{1}{U(\ell)}\,U_{\ell_1,\ell_2}(\ell) + \Big(1 - \frac{1}{U(\ell)}\Big)U_{\ell_1,\ell_2}(\ell_*).$$

Conseguentemente, riordinando opportunamente i termini, si ha ancora (1.8), completando così la sua verifica.

La prima proposizione del prossimo lemma mette in evidenza che l'applicazione U è effettivamente un'utilità ordinale (per $\succeq^*$); la seconda che è lineare sulle misture mentre la terza che muta intervalli (di lotterie) in intervalli (di numeri reali).

[23]Certamente non vuota per le proprietà di non degenerazione e di coerenza.

Lemma 1.3.7 *Risulta:*

(i) $\ell'' \succeq^* \ell' \iff U(\ell'') \geq U(\ell')$;

(ii) $U(\ell'\vartheta\ell'') = \vartheta U(\ell') + (1-\vartheta)U(\ell'')$ *per ogni* $\vartheta \in [0,1]$;

(iii) $U([\ell', \ell'']) = [U(\ell'), U(\ell'')]$.

Dimostrazione (i) + (ii) Seguono facilmente da (1.7) e dal Lemma 1.3.6, una volta posto $\ell_1 = \min\{\ell_*, \ell^*, \ell', \ell''\}$ e $\ell_2 = \max\{\ell_*, \ell^*, \ell', \ell''\}$.

(iii) Conseguenza immediata di (ii), una volta osservato che, per il Lemma 1.3.5(iii), $\ell''\vartheta\ell' \in [\ell', \ell'']$ per ogni $\vartheta \in [0,1]$. ■

Una conseguenza immediata della proprietà di coerenza e della proposizione (i) del lemma precedente è che l'applicazione $u : c \mapsto U(\mathit{1}_c)$ è un'utilità ordinale per $\succeq$. Allora, qualunque sia il numero reale t, la controimmagine $\{u \leq t\}$ è un insieme $\succeq$-connesso e quindi appartiene alla σ-algebra $\mathcal{C}$ (che, ricordiamolo, contiene tutti gli insiemi $\succeq$-connessi di $\mathfrak{C}$). Possiamo pertanto affermare, per il criterio standard di misurabilità A.2.2, che la funzione u è $\mathcal{C}$-Borel misurabile. Se ora proviamo che, qualunque sia la lotteria ℓ a supporto limitato, il valore $U(\ell)$ è proprio l'integrale di Lebesgue di u rispetto a ℓ, avremmo dimostrato la proposizione (ii) del Teorema 1.3.1, in quanto (sempre per la proposizione (i) del lemma precedente) riesce $\ell'' \succeq^* \ell' \iff \mathrm{E}_{\ell''}(u) \geq \mathrm{E}_{\ell'}(u)$ per ogni ℓ, ℓ' a supporto limitato.

Lemma 1.3.8 *Se ℓ è una lotteria a supporto limitato, allora $U(\ell) = \mathrm{E}_\ell(u)$.*

Dimostrazione Data una lotteria ℓ a supporto limitato, esistono $c', c'' \in \mathfrak{C}$ tali che $\ell([c', c'']) = 1$. Allora, le funzioni u e $\widetilde{u} = I_{[c',c'']}u$ sono uguali ℓ-quasi certamente. Ne segue $\mathrm{E}_\ell(u) = \mathrm{E}_\ell(\widetilde{u})$ e quindi basta provare l'uguaglianza $U(\ell) = \int_{\mathfrak{C}} \widetilde{u}\, d\ell$. A tal fine, sia $\epsilon > 0$. Scelti $h+1$ numeri reali $t_0, t_1, \dots, t_h$ in modo tale che risulti:

$$t_0 = u(c') \leq t_1 \leq \dots \leq t_{h-1} \leq u(c'') = t_h, \qquad t_i - t_{i-1} < \epsilon \;\; (i = 1, \dots, h),$$

consideriamo le controimmagini $C_i = u^{-1}([t_{i-1}, t_i[)$ $(i = 1, \dots, h-1)$, $C_h = u^{-1}([t_{h-1}, t_h])$ che sono tutte in $\mathcal{C}$ (u è $\mathcal{C}$-Borel misurabile!).

Siano ora $C'_1, \dots, C'_n$ le controimmagini C_i aventi probabilità positiva, cioè tali che $\ell(C_i) > 0$. Allora, preso $j \in \{1, \dots, n\}$, si ha

$$\sup u(C'_j) - \inf u(C'_j) < \epsilon \tag{1.9}$$

e, per l'inclusione $u(C'_j) \subseteq [u(c'), u(c'')]$ e il Lemma 1.3.7(iii),

$$\inf u(C'_j), \sup u(C'_j) \in [u(c'), u(c'')] = [U(\mathit{1}_{c'}), U(\mathit{1}_{c''})] = U([\mathit{1}_{c'}, \mathit{1}_{c''}]).$$

Esistono dunque due lotterie $\ell'_j, \ell''_j \in [\mathit{1}_{c'}, \mathit{1}_{c''}]$ tali che $U(\ell'_j) = \inf u(C'_j)$ e $U(\ell''_j) = \sup u(C'_j)$. Inoltre, considerata la probabilità condizionata:

$$Q_j = \ell(\cdot \mid C'_j) = \frac{\ell(\cdot \cap C'_j)}{\ell(C'_j)}$$

si ha $Q_j(C'_j) = 1 = Q_j([c', c''])$ e quindi $Q_j \in \pounds$. Inoltre, osservato che, per ogni $c \in C'_j$, si ha $U(\ell'_j) = \inf u(C'_j) \leq u(c) = U(1_c) \leq \sup u(C'_j) = U(\ell''_j)$ e quindi $\ell''_j \succeq^* 1_c \succeq^* \ell'_j$ (Lemma 1.3.7(i)), dalla proprietà di dominanza otteniamo $\ell''_j \succeq^* Q_j \succeq^* \ell'_j$.

Dall'arbitrarietà della scelta di j, tramite il Lemma 1.3.5(v), si ha quindi

$$\sum_{j=1}^{n} \ell(C'_j)\ell''_j \succeq^* \sum_{j=1}^{n} \ell(C'_j)Q_j \succeq^* \sum_{j=1}^{n} \ell(C'_j)\ell'_j,$$

osservato che $\sum_{j=1}^{n} \ell(C'_j) = 1$. Poichè

$$\begin{aligned}\ell(C) = \ell(C \cap [c', c'']) &= \ell\big(C \cap \bigcup_{i=1}^{h} C_i\big) = \sum_{i=1}^{h} \ell\big(C \cap C_i\big)\\ &= \sum_{j=1}^{n} \ell(C \cap C'_j) = \sum_{j=1}^{n} \ell(C \mid C'_j)\,\ell(C'_j) = \sum_{j=1}^{n} Q_j(C)\,\ell(C'_j)\end{aligned}$$

per ogni $C \in \mathcal{C}$, riesce allora $\sum_{j=1}^{n} \ell(C'_j)\ell''_j \succeq^* \ell \succeq^* \sum_{j=1}^{n} \ell(C'_j)\ell'_j$ e quindi, per il Lemma 1.3.7(i),(ii),

$$\begin{aligned}\int_{\mathfrak{C}} \sum_{j=1}^{n} \inf u(C'_j)\, I_{C'_j}\, d\ell &= \sum_{j=1}^{n} \ell(C'_j) \inf u(C'_j) = \sum_{j=1}^{n} \ell(C'_j)U(\ell'_j)\\ &= U\big(\sum_{j=1}^{n} \ell(C'_j)\ell'_j\big) \leq U(\ell) \leq U\big(\sum_{j=1}^{n} \ell(C'_j)\ell''_j\big)\\ &= \sum_{j=1}^{n} \ell(C'_j)U(\ell''_j) = \sum_{j=1}^{n} \ell(C'_j) \sup u(C'_j) = \int_{\mathfrak{C}} \sum_{j=1}^{n} \sup u(C'_j)\, I_{C'_j}\, d\ell.\end{aligned}$$

Tenuto conto della disuguaglianza $\sum_{j=1}^{n} \inf u(C'_j)\, I_{C'_j} \leq \widetilde{u} \leq \sum_{j=1}^{n} \sup u(C'_j)\, I_{C'_j}$, si ha anche

$$\int_{\mathfrak{C}} \sum_{j=1}^{n} \inf u(C'_j)\, I_{C'_j}\, d\ell \leq \int_{\mathfrak{C}} \widetilde{u}\, d\ell \leq \int_{\mathfrak{C}} \sum_{j=1}^{n} \sup u(C'_j)\, I_{C'_j}\, d\ell$$

e quindi, per (1.9),

$$\begin{aligned}\Big|U(\ell) - \int_{\mathfrak{C}} \widetilde{u}\, d\ell\Big| &\leq \int_{\mathfrak{C}} \sum_{j=1}^{n} \sup u(C'_j)\, I_{C'_j}\, d\ell - \int_{\mathfrak{C}} \sum_{j=1}^{n} \inf u(C'_j)\, I_{C'_j}\, d\ell\\ &= \int_{\mathfrak{C}} \sum_{j=1}^{n} \big[\sup u(C'_j) - \inf u(C'_j)\big]\, I_{C'_j}\, d\ell\\ &\leq \epsilon \int_{\mathfrak{C}} \sum_{j=1}^{n} I_{C'_j}\, d\ell = \epsilon \sum_{j=1}^{n} \int_{\mathfrak{C}} I_{C'_j}\, d\ell = \epsilon \sum_{j=1}^{n} \ell(C'_j) = \epsilon.\end{aligned}$$

Passando infine al limite per $\epsilon \to 0^+$, otteniamo $U(\ell) = \int_{\mathfrak{C}} \widetilde{u}\, d\ell$. ∎

1.4 Avversione al rischio

In quest'ultima sezione consideriamo il caso particolare, ma di estrema importanza per la teoria delle scelte del consumatore, di conseguenze espresse in termini monetari[24]. Essendo ragionevole ritenere che DM desideri aumentare, qualunque sia il livello di ricchezza raggiunto, il suo capitale (cioè, sia un *individuo non saziato*), assumiamo che l'insieme $\mathfrak{C}$ delle conseguenze (degli **importi**) sia un intervallo (limitato o no) superiormente aperto della retta reale e che la preferenza $\succeq$ sia l'usuale ordinamento per grandezza. La considerazione di un'utilità ordinale è quindi, in questo contesto, superflua. Non lo è invece quella di un'utilità cardinale poichè consente di introdurre, come vedremo, una nozione chiave per la teoria del consumatore: quella di avversione al rischio. In questa sezione, $\mathcal{C}$ sarà la traccia della σ-algebra di Borel su $\mathfrak{C}$ e $\pounds$ coinciderà con la famiglia delle lotterie a supporto limitato; inoltre, u indicherà un'utilità cardinale per $\succeq^*$ (che chiameremo semplicemente "utilità")[25]. Considerata infine un'arbitraria lotteria ℓ, denoteremo con $\mathrm{M}(\ell)$ l'**importo medio** $\int_{\mathfrak{C}} x\,\ell(dx)$ ad essa relativo (certamente finito in quanto, essendo ℓ a supporto limitato, l'identità di $\mathfrak{C}$ è limitata ℓ-quasi certamente).

Nella teoria della scelta del consumatore è usuale intendere l'importo medio di una lotteria come **prezzo equo** della medesima. Poichè questa interpretazione può sembrare strana a coloro che non hanno familiarità con la teoria economica, riportiamo le considerazioni che la giustificano. Iniziamo con l'introdurre la nozione di sistema di prezzi equo per una famiglia $\Im$ di importi aleatori limitati definiti su un arbitrario insieme Ω di casi elementari e a valori in $\mathfrak{C}$. Indicato con $\mathrm{p}(X)$ il prezzo (d'acquisto e di vendita) dell'importo aleatorio $X \in \Im$, il profitto del consumatore relativo all'acquisto di X è $G(X) = X - \mathrm{p}(X)$ mentre quello relativo alla sua vendita è $\mathrm{p}(X) - X = -G(X)$. Conseguentemente, il profitto totale inerente l'acquisto degli importi $X_1, \dots, X_m \in \Im$ e la vendita degli importi $Y_1, \dots, Y_n \in \Im$ è $\sum_{i=1}^m G(X_i) - \sum_{j=1}^n G(Y_j)$. Assunto, come appare del tutto naturale, che il consumatore desideri avere un profitto totale non negativo in ogni operazione che comporti acquisti e/o vendite di importi di $\Im$, chiamiamo il sistema di prezzi p (inteso come funzione di dominio $\Im$) *equo* se consente al consumatore di *evitare la perdita certa*, cioè se:

$$\sup\Big[\sum_{i=1}^{m} G(X_i) - \sum_{j=1}^{n} G(Y_j)\Big] \geq 0$$

[24]Chi non fosse interessato a questa problematica può omettere la lettura della sezione senza pregiudicare in alcun modo la comprensione degli argomenti che seguono.

[25]Quindi, $\succeq^*$ è un preordinamento completo che verifica la proprietà archimedea e quelle di coerenza, dominanza e di indipendenza (Teorema 1.3.1(i)) e u è una funzione crescente su $\mathfrak{C}$ (Teorema 1.3.3(ii)).

qualunque siano gli importi acquistati $X_1, \dots, X_m \in \Im$ e venduti $Y_1, \dots, Y_n \in \Im$.

Come si può costruire un sistema di prezzi equo? Basta considerare una probabilità P definita su una σ-algebra $\mathcal{A}$ su Ω che renda misurabili tutti gli elementi di $\Im$ e scegliere come prezzo di un qualsiasi importo aleatorio $X \in \Im$ la sua speranza matematica $\mathrm{E_P}(X) = \int_\Omega X \, d\mathrm{P}$. Riesce infatti

$$\begin{aligned}\sum_{i=1}^m G(X_i) - \sum_{j=1}^n G(Y_j) &= \sum_{i=1}^m \big[X_i - \mathrm{E_P}(X_i)\big] - \sum_{j=1}^n \big[Y_j - \mathrm{E_P}(Y_j)\big] \\ &= \sum_{i=1}^m X_i - \sum_{j=1}^n Y_j - \Big[\sum_{i=1}^m \mathrm{E_P}(X_i) - \sum_{j=1}^n \mathrm{E_P}(Y_j)\Big] \\ &= \sum_{i=1}^m X_i - \sum_{j=1}^n Y_j - \mathrm{E_P}\big(\sum_{i=1}^m X_i - \sum_{j=1}^n Y_j\big)\end{aligned}$$

da cui, per l'internalità della speranza matematica, otteniamo

$$\sup\Big[\sum_{i=1}^m G(X_i) - \sum_{j=1}^n G(Y_j)\Big] = \sup\Big[\sum_{i=1}^m X_i - \sum_{j=1}^n Y_j\Big] - \mathrm{E_P}\big(\sum_{i=1}^m X_i - \sum_{j=1}^n Y_j\big) \geq 0.$$

Ricordato che, per (B.2), $\mathrm{E_P}(X) = \int_\mathbb{R} x \, \mathrm{P}_X(dx) = \mathsf{M}(\mathrm{P}_X)$ (ove $\mathrm{P}_X \in \pounds$ è la legge dell'importo aleatorio X), la scelta della speranza matematica come prezzo equo di X comporta che tale prezzo sia da intendersi relativo alla lotteria P_X piuttosto che all'importo aleatorio X (in quanto, così scegliendo, importi aleatori equidistribuiti hanno il medesimo prezzo equo!). Cogliendo questa osservazione, viene allora naturale interpretare (e così faremo) l'importo medio $\mathrm{M}(\ell)$ di una lotteria ℓ come suo prezzo equo (e anche come prezzo equo di un qualsiasi importo aleatorio limitato avente ℓ come legge).

Interpretato l'importo medio di una lotteria come prezzo equo della medesima, supponiamo che DM debba decidere se acquistare o no una data lotteria ℓ. Per farlo, DM dovrà mettere a confronto la possibilità di ricevere certamente il prezzo equo $\mathrm{M}(\ell)$ con quella di partecipare alla lotteria ℓ assumendosi il relativo rischio; dovrà, in altri termini, vedere quale dei seguenti tre casi $\ell \succ^* \mathit{1}_{\mathrm{M}(\ell)}$, $\mathit{1}_{\mathrm{M}(\ell)} \succ^* \ell$, $\ell \sim^* \mathit{1}_{\mathrm{M}(\ell)}$ si verifica. Se sussiste il primo, allora per DM è preferibile possedere l'importo aleatorio ℓ piuttosto che l'importo certo $\mathrm{M}(\ell)$ e quindi acquisterà la lotteria; se invece si verifica il secondo, allora per DM è preferibile possedere l'importo certo $\mathrm{M}(\ell)$ piuttosto che l'importo aleatorio ℓ e quindi non acquisterà la lotteria; se infine risulta valido l'ultimo, allora per DM è del tutto indifferente possedere l'importo aleatorio ℓ oppure l'importo certo $\mathrm{M}(\ell)$ e quindi sarà libero di acquistare o no la lotteria.

Ovviamente, in generale, il comportamento di DM non sarà "uniforme";

egli acquisterà certe lotterie mentre ne rifiuterà altre[26]. D'altra parte, è usuale nella teoria della scelta del consumatore assumere (almeno in prima istanza) un comportamento "uniforme" da parte dei consumatori. Viene così legittimata la seguente definizione fondamentale.

Definizione 1.4.1 Il decisore è:
- **avverso al rischio** se $\mathit{1}_{\mathrm{M}(\ell)} \succeq^* \ell$ per ogni lotteria ℓ;
- **propenso al rischio** se $\ell \succeq^* \mathit{1}_{\mathrm{M}(\ell)}$ per ogni lotteria ℓ;
- **neutrale al rischio** se $\mathit{1}_{\mathrm{M}(\ell)} \sim^* \ell$ per ogni lotteria ℓ.[27]

Il teorema seguente fornisce una caratterizzazione analitica dell'avversione, propensione e neutralità rispetto al rischio in termini di concavità, convessità e affinità della funzione di utilità.

Teorema 1.4.2 *Il decisore è:*

(i) avverso al rischio se e solo se u è una funzione concava;

(ii) propenso al rischio se e solo se u è una funzione convessa;

(iii) neutrale al rischio se e solo se u è una funzione affine.

Dimostrazione Poichè la proposizione (iii) segue immediatamente dalle prime due e le dimostrazioni di queste ultime sono simili, ci limitiamo a provare la proposizione (i). Assumiamo intanto DM avverso al rischio. Dati gli importi c', c'' e $\vartheta \in [0,1]$, sia $c = \vartheta c' + (1-\vartheta)c''$. Considerata allora la lotteria $\ell = \mathit{1}_{c'}\, \vartheta\, \mathit{1}_{c''}$, riesce $\mathrm{M}(\ell) = c$ da cui, per l'avversione al rischio, otteniamo $\mathit{1}_c \succeq^* \ell$ e quindi $u(\vartheta c' + (1-\vartheta)c'') = u(c) = \mathrm{E}_{\mathit{1}_c}(u) \geq \mathrm{E}_\ell(u) = \vartheta \mathrm{E}_{\mathit{1}_{c'}}(u) + (1-\vartheta)\mathrm{E}_{\mathit{1}_{c''}}(u) = \vartheta u(c') + (1-\vartheta)u(c'')$.

Sia ora u una funzione concava. Considerata un'arbitraria lotteria ℓ, per il teorema di Jensen B.2.5 (prendendo come v.a. X l'identità e ponendo $g = -u$), si ha $\mathrm{E}_\ell(u) \leq u(\mathrm{M}(\ell)) = \mathrm{E}_{\mathit{1}_{\mathrm{M}(\ell)}}(u)$ da cui otteniamo $\mathit{1}_{\mathrm{M}(\ell)} \succeq^* \ell$. ■

[26]È bene tenere presente che tale comportamento (indotto dalla preferenza $\succeq^*$) va sempre riferito ad un *fissato livello di ricchezza* del decisore e ad un *dato contesto*; può capitare infatti che DM, se messo in condizioni di ricchezza differenti o in contesti differenti (come, ad esempio, il campo degli affari o quello dei giochi d'azzardo), dall'avversione per certe lotterie passi alla propensione per esse (e viceversa).

[27]Una chiara interpretazione psicologica delle nozioni introdotte può essere ottenuta facendo riferimento alle lotterie che fanno vincere o perdere un medesimo importo con uguale probabilità. Ad esempio, l'avversione al rischio può essere interpretata (in questo contesto) come la prevalenza nel decisore del timore di conseguire una perdita rispetto alla speranza di procurarsi, con pari probabilità, una vincita di uguale importo.

Osservazione 1.4.3 Supposto che DM non sia neutrale al rischio, l'acquisizione, al livello di ricchezza z, di un importo certo $x > 0$ comporta per DM un incremento di utilità $u(z+x) - u(z)$ dipendente (oltre che dal guadagno x) dalla dotazione di ricchezza z; precisamente, decrescente (crescente) o al più costante al crescere della ricchezza z se DM è avverso (propenso) al rischio. Verifichiamolo nel caso di avversione al rischio. Per la caratterizzazione precedente, l'utilità u risulta concava. Conseguentemente, qualunque siano gli importi z', z'' tali che $z' < z''$, dal Lemma B.2.4(i) otteniamo

- $\frac{u(z'+x)-u(z')}{(z'+x)-z'} \geq \frac{u(z'')-u(z')}{z''-z'} \geq \frac{u(z''+x)-u(z'')}{(z''+x)-z''}$ se $z' + x \leq z''$;
- $\frac{u(z'+x)-u(z')}{(z'+x)-z'} \geq \frac{u(z'+x)-u(z'')}{(z'+x)-z''} \geq \frac{u(z''+x)-u(z'')}{(z''+x)-z''}$ se $z'' < z' + x$

e quindi $u(z'+x) - u(z') \geq u(z''+x) - u(z'')$. ♦

Per ottenere ulteriori caratterizzazioni significative del comportamento di DM rispetto al rischio, introduciamo l'equivalente certo di una lotteria che può essere interpretato come il prezzo "soggettivo" che DM attribuisce (nel suo sistema di preferenze) alla lotteria.

Definizione 1.4.4 L'**equivalente certo** della lotteria ℓ è quell'unico importo (se esistente) $\mathrm{C}(\ell)$ tale che $\mathit{1}_{\mathrm{C}(\ell)} \sim^* \ell$, cioè tale che $\mathrm{C}(\ell) = u^{-1}(\mathrm{E}_\ell(u))$[28].

Forniamo ora una caratterizzazione analitica dell'esistenza dell'equivalente certo per ogni lotteria in termini di continuità della funzione di utilità.

Teorema 1.4.5 *Esiste l'equivalente certo di ogni lotteria se e solo se u è una funzione continua.*

Dimostrazione Esista intanto l'equivalente certo di ogni lotteria. Supposto (per assurdo) che $\tilde{c}$ sia un punto di discontinuità di u, dalla crescenza di u otteniamo $u(\tilde{c}) < \alpha = \lim_{c \to \tilde{c}^+} u(c) = \inf_{c > \tilde{c}} u(c)$ o $u(\tilde{c}) > \beta = \lim_{c \to \tilde{c}^-} u(c) = \sup_{c < \tilde{c}} u(c)$. Nel primo caso, scelto $c' > \tilde{c}$, esiste $\vartheta \in]0,1[$ tale che $u(\tilde{c}) < \vartheta u(\tilde{c}) + (1-\vartheta)u(c') < \alpha$. Considerata allora la lotteria $\ell = \mathit{1}_{\tilde{c}}\, \vartheta\, \mathit{1}_{c'}$ e posto $c = \mathrm{C}(\ell)$, otteniamo $u(c) = \mathrm{E}_\ell(u) = \vartheta u(\tilde{c}) + (1-\vartheta)u(c')$ e quindi $u(\tilde{c}) < u(c) < \alpha$ (Contraddizione!). Passando a considerare il secondo caso, si giunge in modo analogo ad una contraddizione.

Sia ora u una funzione continua. Data una lotteria ℓ, esistono due importi c', c'' tali che $\ell([c', c'']) = 1$. Allora $c' \leq c''$ e quindi, per la crescenza e

[28] In forza della crescenza di u e del Teorema 1.3.3(i).

continuità di u, si ha $u([c', c'']) = [u(c'), u(c'')]$, tenuto presente che le funzioni continue trasformano intervalli in intervalli. Ne segue, tramite il teorema della media integrale A.3.6(iv),

$$u(c') = u(c')\,\ell([c', c'']) \leq \int_{[c',c'']} u\,d\ell \leq u(c'')\,\ell([c', c'']) = u(c'')$$

e quindi esiste $c \in [c', c'']$ tale che $u(c) = \int_{[c',c'']} u\,d\ell = \int_{\mathfrak{C}} u\,d\ell = \mathrm{E}_\ell(u)$. Allora, $c = u^{-1}(\mathrm{E}_\ell(u)) = \mathrm{C}(\ell)$. ■

Supposto che tutte le lotterie ammettano equivalente certo, otteniamo $\ell \succeq^* \ell' \iff \mathit{1}_{\mathrm{C}(\ell)} \succeq^* \mathit{1}_{\mathrm{C}(\ell')} \iff \mathrm{C}(\ell) \geq \mathrm{C}(\ell')$ per ogni lotteria ℓ e ℓ'. Conseguentemente, l'equivalente certo - inteso come una funzione di $\mathcal{L}$ in $\mathfrak{C}$ - è un'utilità ordinale per la preferenza $\succeq^*$. Inoltre, come messo in evidenza dal prossimo teorema (di facile verifica), il comportamento di DM rispetto al rischio può essere descritto confrontando l'equivalente certo (prezzo "soggettivo" della lotteria) con l'importo medio (prezzo equo della lotteria).

Teorema 1.4.6 *Se esiste l'equivalente certo di ogni lotteria, allora il decisore è:*

(i) avverso al rischio se e solo se $\mathrm{C}(\ell) \leq \mathrm{M}(\ell)$ *per ogni lotteria* ℓ*;*

(ii) propenso al rischio se e solo se $\mathrm{C}(\ell) \geq \mathrm{M}(\ell)$ *per ogni lotteria* ℓ*;*

(iii) neutrale al rischio se e solo se $\mathrm{C}(\ell) = \mathrm{M}(\ell)$ *per ogni lotteria* ℓ*.*

Caratterizzato il comportamento di DM rispetto al rischio in termini di concavità e/o convessità dell'utilità e di equivalente certo, procediamo introducendo due quantità numeriche di estrema importanza. Onde evitare complicazioni di natura matematica, *considereremo (fino alla fine della sezione) solamente utilità* u *aventi derivata prima* $u' \gg 0$ *e derivata seconda* u'' *continua.*

Essendo, in queste ipotesi, u una funzione continua, per il Teorema 1.4.5, ogni lotteria ℓ ammette equivalente certo; possiamo quindi introdurre il numero $\mathrm{R}(\ell) = \mathrm{M}(\ell) - \mathrm{C}(\ell)$, detto **costo del rischio** (o, con linguaggio assicurativo, **premio di rischio**), che rappresenta:
- il massimo importo che DM è disposto a defalcare dal prezzo equo per evitare il rischio connesso con la lotteria, nel caso che sia avverso al rischio;
- in valore assoluto il massimo importo che DM è disposto ad aggiungere al

prezzo equo per accedere alla lotteria, nel caso che sia propenso al rischio. Proviamolo, a titolo d'esempio, nel caso dell'avversione al rischio. Data una lotteria ℓ e posto $c = \mathrm{M}(\ell)$, si ha $\mathit{1}_c \succ^* \ell$. Allora, $u(c) > \mathrm{E}_\ell(u)$ e quindi, qualunque sia l'importo $c' > 0$, DM preferirà l'importo $c - c'$ alla lotteria, se $u(c-c') > \mathrm{E}_\ell(u)$, e preferirà la lotteria all'importo $c-c'$, se $\mathrm{E}_\ell(u) > u(c-c')$. Conseguentemente, il massimo importo che DM sarà disposto a defalcare da c (per evitare il rischio connesso alla lotteria) sarà l'importo $\hat{c}$ che realizza l'uguaglianza $u(c-\hat{c}) = \mathrm{E}_\ell(u)$. Ne segue $c-\hat{c} = u^{-1}(\mathrm{E}_\ell(u))$ da cui otteniamo $\hat{c} = c - \mathrm{C}(\ell) = \mathrm{R}(\ell)$.

Inoltre, essendo $u' > 0$, ha anche senso considerare la funzione continua:

$$r_u = -\,\frac{u''}{u'}$$

di dominio $\mathfrak{C}$, detta **coefficiente di avversione al rischio** (**assoluta**) del decisore, che misura la concavità relativa della funzione u. Osservato che $\frac{d}{dt} \ln u'(t) = -r_u(t)$, è facile ottenere, mediante il teorema di Torricelli e tramite due integrazioni successive, la seguente relazione tra u e r_u:

$$u(c) = u'(c_0) \int_{c_0}^{c} \exp\Big[-\int_{c_0}^{x} r_u(t)dt\Big]dx + u(c_0), \tag{1.10}$$

ove c_0 è un elemento arbitrario di $\mathfrak{C}$.

Il teorema seguente assicura che il coefficiente di avversione al rischio è in grado sia di caratterizzare la preferenza $\succeq^*$ che di descrivere il comportamento di DM rispetto al rischio.

Teorema 1.4.7 *Il decisore è:*

(i) avverso al rischio se e solo se $r_u \geq 0$*;*

(ii) propenso al rischio se e solo se $r_u \leq 0$*;*

(iii) neutrale al rischio se e solo se $r_u = 0$*.*

Inoltre, r_u *è invariante per trasformazioni affini positive di* u*. Infine, se* $\tilde{u} : \mathfrak{C} \mapsto \mathbb{R}$ *è una funzione con derivata prima positiva, derivata seconda continua e tale che* $r_{\tilde{u}} = r_u$*, allora* $\tilde{u}$ *è un'utilità cardinale per* $\succeq^*$*.*

Dimostrazione La caratterizzazione del comportamento di DM rispetto al rischio discende immediatamente dal Teorema 1.4.2 e dalla caratterizzazione del secondo ordine della convessità. L'invarianza rispetto alle trasformazioni affini positive è banale. Proviamo infine la terza parte della tesi. Da (1.10) otteniamo

$$\begin{aligned} u(c) &= a \int_{c_0}^{c} \exp\Big[-\int_{c_0}^{x} r_u(t)dt\Big] dx + b \\ \tilde{u}(c) &= \tilde{a} \int_{c_0}^{c} \exp\Big[-\int_{c_0}^{x} r_u(t)dt\Big] dx + \tilde{b} \end{aligned}$$

con a, $\tilde{a} > 0$. Riesce allora $\tilde{u} = \frac{\tilde{a}}{a}\, u + \big[\tilde{b} - \frac{\tilde{a}b}{a}\big]$ e quindi, per il Teorema 1.3.3, $\tilde{u}$ è un'utilità per $\succeq^*$. ■

Il comportamento di DM rispetto al rischio è stato sin qui analizzato facendo riferimento (come puntualizzato nella nota 26 di p. 27) ad un *dato* livello di ricchezza (seppur non precisato) del decisore. L'introduzione del coefficiente di avversione al rischio consente di caratterizzare non solo le preferenze e l'attitudine verso il rischio ma anche una opportuna formulazione "dinamica" dell'avversione al rischio che tenga conto del divenire della ricchezza posseduta. Per introdurla, consideriamo un livello di ricchezza z e supponiamo che DM (ritenuto avverso al rischio) sia interessato alla lotteria $\ell = \sum_{i=1}^{n} \vartheta_i \mathit{1}_{c_i}$ che assegna il premio c_i con probabilità $\vartheta_i > 0\,(i = 1, \ldots, n)$. Allora, una volta che DM decidesse di partecipare alla lotteria, la sua dotazione di ricchezza diverrebbe aleatoria e sarebbe descritta dalla lotteria $\ell_z = \sum_{i=1}^{n} \vartheta_i \mathit{1}_{c_i+z}$ che assegna il premio $c_i + z$ con probabilità $\vartheta_i\,(i = 1, \ldots, n)$, cioè dalla lotteria che si ottiene dalla ℓ aumentandone i premi di un importo pari alla ricchezza posseduta e lasciando invariate le probabilità. Osservato che tale lotteria ha la medesima varianza della lotteria iniziale[29], possiamo concludere che le lotterie ℓ_{z_1} e ℓ_{z_2} (corrispondenti ai livelli di ricchezza z_1 e z_2) sono *ugualmente rischiose.* Conseguentemente, l'evolversi del "grado" di avversione al rischio di DM, al variare della sua ricchezza, potrebbe essere descritto confrontando i costi del rischio connessi con le due lotterie. Infatti, qualora riuscisse, ad esempio, $\mathrm{R}(\ell_{z_1}) > \mathrm{R}(\ell_{z_2})$, DM risulterebbe, da un punto di vista intuitivo, più avverso al rischio possedendo il patrimonio z_1 che il patrimonio z_2, in quanto sarebbe disposto, per evitare la situazione rischiosa descritta dalla lotteria ℓ, a defalcare dal prezzo equo

[29]Ricordiamo che per *varianza* di una lotteria ℓ si intende la varianza di una qualsiasi v.a. avente ℓ come legge, cioè l'integrale $\mathrm{Var}(\ell) = \int_{\mathfrak{C}} [c - \mathrm{M}(\ell)]^2\, \ell(dc)$, certamente finito essendo l'applicazione "elevamento al quadrato" ℓ-quasi certamente limitata.

un importo che, nello stato di ricchezza z_1, è maggiore di quello relativo allo stato z_2.

Passando ad una formalizzazione di queste considerazioni, per ogni conseguenza z e per ogni $C \in \mathcal{C}$, poniamo $C + z = \{c + z : c \in C\}$ e consideriamo la σ-algebra "traslata" $\mathcal{C}_z = \{C + z : C \in \mathcal{C}\}$ di $\mathcal{C}$ su $\mathfrak{C} + z$. Infine, data una qualsiasi lotteria ℓ, denotiamo con $\ell \oplus z$ la lotteria su $\mathcal{C}_z$ tale che $(\ell \oplus z)(C + z) = \ell(C)$ per ogni $C \in \mathcal{C}$, cioè la lotteria che aumenta tutti i "premi" di ℓ dell'importo z lasciandone inalterate le "probabilità". Allora, per le considerazioni precedenti, la variazione (all'evolversi della dotazione di ricchezza) del grado di avversione di DM ad accedere alla situazione rischiosa rappresentata dalla lotteria ℓ può essere ragionevolmente espressa precisando il comportamento del costo del rischio $\mathrm{R}(\ell \oplus z)$ all'aumentare della ricchezza z. Diamo pertanto la definizione seguente.

Definizione 1.4.8 Il decisore (supposto avverso al rischio) è:
- **avverso al rischio in modo non crescente** se il costo del rischio $\mathrm{R}(\ell \oplus z)$ è una funzione non crescente in z, qualunque sia la lotteria ℓ;
- **avverso al rischio in modo non decrescente** se il costo del rischio $\mathrm{R}(\ell \oplus z)$ è una funzione non decrescente in z, qualunque sia la lotteria ℓ;
- **avverso al rischio in modo costante** se il costo del rischio $\mathrm{R}(\ell \oplus z)$ è una funzione costante in z, qualunque sia la lotteria ℓ.[30]

Il teorema successivo (al quale premettiamo un lemma) mette in evidenza che tramite il coefficiente di avversione al rischio siamo in grado di caratterizzare queste nozioni "dinamiche" di avversione al rischio.

Lemma 1.4.9 *Siano f_1, f_2 funzioni di dominio $\mathfrak{C}$ con derivata prima positiva e derivata seconda continua. Allora, la funzione composta $\phi = f_1 \circ f_2^{-1} : f_2(\mathfrak{C}) \to \mathbb{R}$ è crescente e continua. Inoltre, sono equivalenti le seguenti proposizioni:*

(i) $\mathrm{E}_\ell(f_1) \leq \phi(\mathrm{E}_\ell(f_2))$ *per ogni lotteria ℓ;*

(ii) ϕ *è una funzione concava;*

(iii) $r_{f_1} \geq r_{f_2}$.

[30] Ricordiamo che, nella teoria della scelta del consumatore, viene normalmente assunto che gli individui, oltre a essere avversi al rischio, siano anche avversi al rischio in modo non crescente o al più costante.

Dimostrazione Per quanto riguarda la crescenza e la continuità della funzione composta basta osservare che la funzione f_2^{-1} è crescente e continua (in quanto inversa di una funzione crescente definita su un intervallo (limitato o no) della retta reale). Passando alla seconda parte della tesi, procediamo provando le equivalenze $(i) \iff (ii)$ e $(ii) \iff (iii)$.

$(i) \iff (ii)$. Assumiamo intanto (i). Per verificare la concavità, siano $\vartheta \in [0,1]$ e $a', a'' \in f_2(\mathfrak{C})$ con $a' < a''$. Esistono allora due conseguenze c', c'' tali che $f_2(c') = a'$ e $f_2(c'') = a''$. Considerata la lotteria $\ell = 1_{c'}\,\vartheta\, 1_{c''}$, si ha

$$\vartheta f_1(c') + (1-\vartheta) f_1(c'') = \mathrm{E}_\ell(f_1) \le \phi(\mathrm{E}_\ell(f_2)) = \phi\left(\vartheta f_2(c') + (1-\vartheta) f_2(c'')\right)$$

e quindi

$$\begin{aligned}
\vartheta\phi(a') + (1-\vartheta)\phi(a'') &= \vartheta\phi(f_2(c')) + (1-\vartheta)\phi(f_2(c'')) \\
&= \vartheta f_1(c') + (1-\vartheta) f_1(c'') \\
&\le \phi(\vartheta f_2(c') + (1-\vartheta) f_2(c'')) = \phi(\vartheta a' + (1-\vartheta)a'').
\end{aligned}$$

Assumiamo ora (ii). Data un'arbitraria lotteria ℓ, dal teorema di Jensen (ponendo $X = f_2$ e $g = \phi$) risulta $\mathrm{E}_\ell(f_1) = \mathrm{E}_\ell\big[(f_1 \circ f_2^{-1}) \circ f_2\big] = \mathrm{E}_\ell[\phi(f_2)\,] \le \phi[\mathrm{E}_\ell(f_2)]$.

$(ii) \iff (iii)$. Basta osservare che riesce

$$\begin{aligned}
\phi'(f_2(c)) &= \frac{f_1'(c)}{f_2'(c)} > 0 \\
\phi''(f_2(c)) &= \frac{f_1''(c) f_2'(c) - f_1'(c) f_2''(c)}{f_2'(c)^2} = \frac{f_1'(c)}{f_2'(c)}\left[\frac{f_1''(c)}{f_1'(c)} - \frac{f_2''(c)}{f_2'(c)}\right] \\
&= \frac{f_1'(c)}{f_2'(c)}\left[r_{f_2}(c) - r_{f_1}(c)\right]
\end{aligned}$$

per ogni conseguenza c. ∎

Teorema 1.4.10 *L'utilità u ammetta derivata terza continua e sia inoltre $u'' \ll 0$. Allora il decisore è avverso al rischio in modo:*

(i) non crescente se e solo se r_u è una funzione non crescente;

(ii) non decrescente se e solo se r_u è una funzione non decrescente;

(iii) costante se e solo se r_u è una funzione costante.

Inoltre, in caso di avversione al rischio in modo costante, u è una trasformata affine positiva della funzione $\tilde{u} : c \mapsto -\exp(-r_u c)$.

Dimostrazione Poichè la proposizione (iii) segue immediatamente dalle prime due e le dimostrazioni di queste ultime sono analoghe, ci limitiamo a provare la proposizione (i).

Supposto intanto DM avverso al rischio in modo non crescente, consideriamo la funzione biunivoca $\tau_z : \mathfrak{C} \mapsto \mathfrak{C} + z$ tale che $\tau_z(c) = c + z$. Allora, data una lotteria ℓ, si ha $(\ell \oplus z)(C + z) = \ell(C) = \ell(\tau_z^{-1}(C)) = \ell \circ \tau_z^{-1}(C)$ per ogni $C \in \mathcal{C}$. Conseguentemente, $\ell \oplus z$ è la probabilità immagine di ℓ mediante τ e quindi, per il teorema della misura immagine A.4.1,

$$\mathrm{E}_{\ell \oplus z}(f) = \int_{\mathfrak{C}+z} f(c)\,(\ell \oplus z)(dc) = \int_{\mathfrak{C}} f(c+z)\,\ell(dc)$$

per ogni funzione $f : \mathfrak{C} + z \mapsto \mathbb{R}$ che sia $(\ell \oplus z)$-integrabile. Riesce pertanto

$$\begin{aligned}\mathrm{R}(\ell \oplus z) = \mathrm{M}(\ell \oplus z) - \mathrm{C}(\ell \oplus z) &= \int_{\mathfrak{C}} (c+z)\,\ell(dc) - u^{-1}\Big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\Big) \\ &= z + \mathrm{M}(L) - u^{-1}\Big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\Big)\end{aligned}$$

da cui, per il teorema di derivazione sotto il segno d'integrale A.3.13, risulta

$$\begin{aligned}\frac{d}{dz}\mathrm{R}(\ell \oplus z) &= 1 - \frac{\frac{d}{dz}\int_{\mathfrak{C}} u(c+z)\,\ell(dc)}{u'\Big(u^{-1}\big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\big)\Big)} \\ &= 1 - \frac{\int_{\mathfrak{C}} u'(c+z)\,\ell(dc)}{u'\Big(u^{-1}\big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\big)\Big)}\end{aligned}$$

e quindi, tenuto conto della disuguaglianza $\frac{d}{dz}\mathrm{R}(\ell \oplus z) \leq 0$ (la funzione $\mathrm{R}(\ell \oplus \cdot)$ è non crescente!),

$$1 - \frac{\int_{\mathfrak{C}} u'(c+z)\,\ell(dc)}{u'\Big(u^{-1}\big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\big)\Big)} \leq 0,$$

cioè $u'\big(u^{-1}(\int_{\mathfrak{C}} u(c+z)\,\ell(dc))\big) \leq \int_{\mathfrak{C}} u'(c+z)\,\ell(dc)$. Ne segue, per la decrescenza di u' (u'' è negativa!),

$$u^{-1}\Big(\int_{\mathfrak{C}} u(c+z)\,\ell(dc)\Big) \geq (u')^{-1}\Big(\int_{\mathfrak{C}} u'(c+z)\,\ell(dc)\Big). \tag{1.11}$$

Consideriamo ora le funzioni di dominio $\mathfrak{C}$:

$$f_1 : c \mapsto -u'(c+z), \qquad f_2 : c \mapsto u(c+z).$$

Entrambe hanno derivata prima positiva e derivata seconda continua; inoltre, come facilmente si verifica[31], da (1.11) risulta $f_1^{-1}\bigl(\int_{\mathfrak{C}} f_1\, d\ell\bigr) \leq f_2^{-1}\bigl(\int_{\mathfrak{C}} f_2\, d\ell\bigr)$, cioè $\mathrm{E}_\ell(f_1) \leq (f_1 \circ f_2^{-1})(\mathrm{E}_\ell(f_2))$. Conseguentemente, dall'equivalenza delle proposizioni (i), (iii) del Lemma 1.4.9 otteniamo

$$-\frac{u'''}{u''} = -\frac{f_1''}{f_1'} = r_{f_1} \geq r_{f_2} = -\frac{f_2''}{f_2'} = -\frac{u''}{u'}.$$

Ne segue

$$\frac{-(u'')^2 + u'u'''}{u'u''} \leq 0,$$

cioè $-(u'')^2 + u'u''' \geq 0$ (essendo $u'u'' < 0$). Riesce allora

$$(r_u)' = \frac{(u'')^2 - u'u'''}{(u')^2} = -\frac{-(u'')^2 + u'u'''}{(u')^2} \leq 0$$

e quindi il coefficiente di avversione al rischio è non crescente.

Sia ora r_u una funzione non crescente. Ripercorrendo a ritroso la dimostrazione precedente si ottiene che DM è avverso al rischio in modo non crescente. La dimostrazione di (i) è così conclusa.

Passando all'ultima parte della tesi, supponiamo che il coefficiente di avversione al rischio sia costante, cioè $r_u = \alpha$ per qualche numero reale $\alpha > 0$ ($u'' < 0$!). Allora, da (1.10) otteniamo, con facili passaggi, che esistono due numeri reali, $a > 0$ e b, tali che $u(c) = a\,[-\exp(-kc)] + b$ per ogni importo c. Conseguentemente, u è una trasformata affine positiva della funzione $\tilde{u}$.$\blacksquare$

Nell'esempio seguente presentiamo alcune utilità cardinali (usualmente adoperate nelle applicazioni economiche) individuandone i relativi coefficienti di avversione al rischio.

[31]Osservando che, qualunque sia l'applicazione invertibile $f : \mathfrak{C} \mapsto \mathbb{R}$, si ha $f(\cdot + z) = f \circ \tau_z$ e quindi $f^{-1}(\cdot + z) = (f \circ \tau_z)^{-1} = \tau_z^{-1} \circ f^{-1} = f^{-1} - z$; inoltre, che risulta $(-f)^{-1}(t) = f^{-1}(-t)$ per ogni numero reale t.

Esempio 1.4.11 (Festival delle utilità della moneta) (i) UTILITÀ LOGARITMICA. Si pone $\mathfrak{C} =]0, +\infty[$ e $u : c \mapsto \ln c$.[32] Poichè $r_u(c) = \frac{1}{c}$, tale utilità descrive un'avversione al rischio in modo non crescente.

(ii) UTILITÀ ESPONENZIALE. È un'utilità del tipo $u : c \mapsto -\exp(-\alpha c)$ $(\alpha > 0)$. Essendo il relativo coefficiente di avversione al rischio la costante di valore α, tale utilità descrive un'avversione al rischio in modo costante.[33]

(iii) UTILITÀ QUADRATICA. È un'utilità del tipo $u : c \mapsto ac - \frac{b}{2}c^2$ $(a,\, b > 0)$ di dominio $\mathfrak{C} =]-\infty, \frac{a}{b}]$ (intervallo di crescenza di u). Poichè $r_u(c) = \frac{b}{a-bc}$, tale utilità descrive un'avversione al rischio in modo non decrescente[34]. Inoltre, l'utilità attesa di una lotteria dipende solamente dall'importo medio e dalla varianza della lotteria; riesce infatti $\mathrm{E}_\ell(u) = a\mathrm{M}(\ell) - \frac{b}{2}\left[\mathrm{M}(\ell)^2 + \mathrm{Var}(\ell)\right]$.

(iv) UTILITÀ POTENZA (in senso esteso). È un'utilità del tipo

$$u : c \mapsto \frac{1}{b-1}(a+bc)^{\frac{b-1}{b}} \quad (a \neq 0,\, b \neq 1,\, b > 0)$$

di dominio $]-\frac{a}{b}, +\infty[$ (intervallo di crescenza di u). Poichè $r_u(c) = (a+bc)^{-1}$, tale utilità descrive un'avversione al rischio in modo non crescente. ▲

Considerati infine due differenti decisori (detti "primo" e "secondo") e indicate con $\succeq_1^*$ e $\succeq_2^*$, rispettivamente, le loro relazioni di preferenza sulle lotterie di $\mathcal{L}$, formalizziamo, con la prossima definizione, l'idea che il primo decisore è avverso al rischio almeno quanto il secondo se quest'ultimo è disposto a correre un dato rischio ogniqualvolta lo è anche il primo.

[32] Osservato che ad uguali incrementi di ricchezza non sempre corrisponde un uguale incremento del grado di soddisfazione dell'individuo (si pensi, ad esempio, ad un eventuale aumento di stipendio di un lavoratore dipendente: un incremento da 6.000 a 12.000 euro netti all'anno gli consente un miglioramento del tenore di vita nettamente superiore a quello che otterrebbe con un incremento da 30.000 a 36.000 euro), Daniel Bernoulli propose nel 1730 (basandosi su un celeberrimo esempio, noto come *Paradosso di San Pietroburgo*) di valutare l'importo connesso ad una scommessa e/o operazione aleatoria non in termini di speranza matematica dei relativi possibili risultati, bensì in termini del valore atteso (che chiamò *speranza morale*) dei loro logaritmi naturali. L'utilità logaritmica può quindi ritenersi il primo esempio di utilità cardinale apparso nell'evoluzione storica della teoria delle decisioni.

[33] Che giustifica, assieme alle sue proprietà analitiche (facilità nei calcoli), il suo largo uso nelle applicazioni.

[34] Che è uno degli inconvenienti più gravi nell'uso dell'utilità quadratica per descrivere le scelte del consumatore. D'altra parte, per le ipotesi analitiche assunte per le funzioni d'utilità, la possiamo utilmente impiegare per approssimare localmente nell'origine una qualsiasi utilità $\tilde{u}$ tale che $\tilde{u}(0) = 0$ e $\tilde{u}''(0) < 0$; infatti, per il teorema di Taylor, il polinomio di secondo grado $\tilde{u}'(0)\, c + \frac{1}{2}\tilde{u}''(0)\, c^2$ approssima la funzione $\tilde{u}$ in un intorno dell'origine a meno d'infinitesimi di ordine superiore al secondo rispetto all'identità.

Definizione 1.4.12 Il primo decisore è **avverso al rischio almeno quanto** il secondo se: $\ell \succeq_1^* \mathit{1}_c \Rightarrow \ell \succeq_2^* \mathit{1}_c$ per ogni lotteria ℓ e ogni conseguenza c[35].

Supposto che i decisori adottino lo **schema dell'utilità attesa** (cioè, esprimano le loro preferenze sulle lotterie in termini di utilità attesa), indichiamo con u_1,u_2 le rispettive utilità e assumiamo che abbiano derivata prima positiva e seconda continua. Il risultato seguente fornisce alcune caratterizzazioni, della nozione appena introdotta, in termini del costo del rischio e del coefficiente di avversione al rischio; inoltre, mette in evidenza che se il primo decisore è avverso al rischio almeno quanto il secondo, allora l'utilità del primo è una trasformata continua, crescente e concava di quella del secondo. Precisiamo che R_i, r_i sono, rispettivamente, il costo del rischio e il coefficiente di avversione al rischio dell'i-simo decisore ($i = 1, 2$).

Teorema 1.4.13 *Le seguenti proposizioni sono equivalenti:*

(i) *Il primo decisore è avverso al rischio almeno quanto il secondo;*

(ii) $\mathrm{C}_1(\ell) \leq \mathrm{C}_2(\ell)$ *per ogni lotteria* ℓ*;*

(iii) $\mathrm{R}_1(\ell) \geq \mathrm{R}_2(\ell)$ *per ogni lotteria* ℓ*;*

(iv) $u_1 = \phi \circ u_2$, *con* $\phi = u_1 \circ u_2^{-1}$ *funzione continua, crescente e concava;*

(v) $r_1 \geq r_2$.

Dimostrazione Le proposizioni (ii), (iii) sono banalmente equivalenti. Inoltre, notato che ϕ è una funzione continua e crescente (Lemma 1.4.9), le equivalenze $(ii) \iff (iv)$, $(iv) \iff (v)$ sono equivalenti rispettivamente alle "$(ii) \iff \phi$ concava", "$(v) \iff \phi$ concava″. Conseguentemente, la prima si ottiene dall'equivalenza delle proposizioni (i), (ii) del Lemma 1.4.9, una volta osservato che, per la crescenza di u_1, si ha

$$\mathrm{C}_1(\ell) \leq \mathrm{C}_2(\ell) \iff u_1^{-1}\big(\mathrm{E}_\ell(u_1)\big) \leq u_2^{-1}\big(\mathrm{E}_\ell(u_2)\big) \iff \mathrm{E}_\ell(u_1) \leq \phi\big(\mathrm{E}_\ell(u_2)\big)$$

qualunque sia la lotteria ℓ; la seconda invece dall'equivalenza delle proposizioni (ii), (iii) sempre del Lemma 1.4.9.

[35]Oppure (la preferenza $\succeq^*$ è completa !), se: $\mathit{1}_c \succ_2^* \ell \Rightarrow \mathit{1}_c \succ_1^* \ell$ per ogni lotteria ℓ e ogni conseguenza c. Conseguentemente, il primo decisore è avverso al rischio almeno quanto il secondo se non è disposto a correre un dato rischio ogniqualvolta non lo è neanche il secondo.

Verifichiamo infine l'equivalenza $(i) \iff (ii)$. Assumiamo intanto (i). Esista (per assurdo) una lotteria ℓ tale che $c' = \mathrm{C}_1(\ell) > \mathrm{C}_2(\ell) = c''$. Scelto allora $c \in]c'', c'[$ si ha $\mathit{1}_{c'} \succ_1^* \mathit{1}_c$ e $\mathit{1}_c \succ_2^* \mathit{1}_{c''}$ da cui otteniamo $\ell \succ_1^* \mathit{1}_c$ e $\mathit{1}_c \succ_2^* \ell$. Poichè il primo decisore è avverso al rischio almeno quanto il secondo, dalla $\ell \succ_1^* \mathit{1}_c$, risulta $\ell \succeq_2^* \mathit{1}_c$ e quindi $\ell \succ_2^* \ell$ (Contraddizione!). Assumiamo ora (ii). Esistano (per assurdo) una lotteria ℓ e una conseguenza c tali che $\ell \succeq_1^* \mathit{1}_c$ e $\ell \not\succeq_2^* \mathit{1}_c$. Allora, posto $c' = \mathrm{C}_1(\ell)$, $c'' = \mathrm{C}_2(\ell)$ e tenuto conto della completezza di $\succeq_2^*$, si ha $\mathit{1}_{c'} \succeq_1^* \mathit{1}_c$ e $\mathit{1}_c \succ_2^* \mathit{1}_{c''}$; ne segue $u_1(c') \geq u_1(c)$, $u_2(c) > u_2(c'')$ e quindi $\mathrm{C}_1(\ell) = c' \geq c > c'' = \mathrm{C}_2(\ell)$ (Contraddizione!). ■

Concludiamo il capitolo illustrando come le nozioni ed i risultati ottenuti in questa sezione possano essere usati per analizzare uno dei mercati più importanti nei quali l'incertezza gioca un ruolo sostanziale: il mercato assicurativo. Nel prossimo esempio - considerato un consumatore che intende assicurarsi contro un danno aleatorio e una compagnia di assicurazione che offre un contratto assicurativo a copertura del danno - perveniamo (sotto opportune ipotesi semplificatrici) alla seguente fondamentale proprietà: il contratto assicurativo non è svantaggioso per entrambi i soggetti, anche se prevede un caricamento del premio equo (a favore della compagnia), qualora il caricamento risulti contenuto (a favore del consumatore) entro una soglia opportuna.

Esempio 1.4.14 (Domanda di assicurazione) Consideriamo due soggetti economici ("assicurato" e "Assicuratore") che intendono concludere un contratto assicurativo che preveda la copertura di un danno aleatorio $X \in]0, \beta]$ incombente sull'assicurato a fronte del pagamento all'Assicuratore di un premio $p > 0$. In una situazione di questo genere, è del tutto naturale ritenere che entrambi i soggetti concordino sulla legge (probabilistica) del danno aleatorio, che siano avversi al rischio e che l'assicurato sia più avverso al rischio dell'Assicuratore. Conseguentemente, supposto che entrambi adottino lo schema dell'utilità attesa e usino utilità aventi derivata prima positiva e seconda continua, per le utilità u_a dell'assicurato e u_A dell'assicuratore assumiamo $u''_a < 0$, $u''_A < 0$ e $r_{u_a} \geq r_{u_A}$, $r_{u_a} \neq r_{u_A}$; inoltre, tenuto presente che le utilità cardinali sono definite a meno di trasformazioni affini positive, possiamo supporre (senza perdere in generalità) $u_a(0) = u_A(0) = 0$ e $u'_a(0) = u'_A(0) = 1$. Indicata infine con P la probabilità sui boreliani della retta reale che rappresenta la comune legge del danno aleatorio X, il contratto sarà non svantaggioso per:

- l'assicurato se $u_a(-p) \geq \int_{\mathbb{R}} u_a(-x)\,\mathrm{P}(dx)$. Infatti, indicata con P' la legge della v.a. $-X$, l'assicurato riterrà pagare il premio una circostanza non peggiore di quella che preveda la perdita aleatoria $-X$ se (nel suo sistema di preferenze) la lotteria certa $\mathit{1}_{-p}$ fosse non peggiore della lotteria P', cioè se $u_a(-p) = \mathrm{E}_{\mathit{1}_{-p}}(u_a) \geq \mathrm{E}_{\mathrm{P}'}(u_a) = \int_{\mathbb{R}} u_a(-x)\,\mathrm{P}(dx)$ (ove l'ultima uguaglianza si ottiene dalla proposizione (ii) del teorema fondamentale del calcolo

delle probabilità B.1.2 considerando la funzione $\tau : x \mapsto -x$);
• l'Assicuratore se $\int_{\mathbb{R}} u_A(p-x)\,\mathrm{P}(dx) \geq 0$. Infatti, indicata con P'' la legge della v.a. $p-X$, l'Assicuratore riterrà subire la perdita aleatoria $p-X$ una circostanza non peggiore di quella che preveda di non stipulare il contratto se (nel suo sistema di preferenze) la lotteria P'' fosse non peggiore della lotteria certa $\mathit{1}_0$, cioè se $\int_{\mathbb{R}} u_A(p-x)\,\mathrm{P}(dx) = \mathrm{E}_{\mathrm{P}''}(u_A) \geq u_A(0) = 0$ (ove la prima uguaglianza si ottiene sempre dalla proposizione (ii) del teorema fondamentale del calcolo delle probabilità considerando la funzione $\tau : x \mapsto p-x$).

Il contratto sarà dunque non svantaggioso per entrambi i soggetti se il premio p è soluzione del sistema:

$$\begin{cases} u_a(-p) \geq \displaystyle\int_{\mathbb{R}} u_a(-x)\,\mathrm{P}(dx) \\ \\ \displaystyle\int_{\mathbb{R}} u_A(p-x)\,\mathrm{P}(dx) \geq 0 \end{cases}.$$

Supposto che il massimo danno possibile β sia "piccolo", possiamo sostituire (con buona approssimazione) le utilità u_a, u_A con le loro approssimazioni quadratiche nell'origine:

$$\tilde{u}_a(x) = x + \frac{u_a''(0)}{2}\,x^2, \qquad \tilde{u}_A(x) = x + \frac{u_A''(0)}{2}\,x^2.$$

Posto $\alpha_a = -\frac{1}{u_a''(0)}$, $\alpha_A = -\frac{1}{u_A''(0)}$ e tenuto conto dell'espressione del coefficiente di avversione al rischio nel caso dell'utilità quadratica, il sistema precedente si trasforma nel sistema di disequazioni di secondo grado:

$$\begin{cases} p^2 + 2\alpha_a\, p \leq 2\alpha_a\,\mathrm{E}(X) + \mathrm{E}(X^2) \\ p^2 - 2\big[\alpha_A + \mathrm{E}(X)\big]p + \big[\,2\alpha_A\mathrm{E}(X) + \mathrm{E}(X^2)\,\big] \geq 0 \end{cases},$$

ove $\alpha_a < \alpha_A$ e $\mathrm{E}(X)$, $\mathrm{E}(X^2)$ sono i primi due momenti dell'importo aleatorio X. Risolvendo le due disequazioni, è facile constatare che il massimo valore di p che verifica la prima è dato dall'importo:

$$\begin{aligned} p^* &= -\alpha_a + \sqrt{\alpha_a^2 + 2\alpha_a\mathrm{E}(X) + \mathrm{E}(X^2)} \\ &= -\alpha_a + \sqrt{\big[\alpha_a^2 + 2\alpha_a\mathrm{E}(X) + \mathrm{E}(X)^2\big] + [\mathrm{E}(X^2) - \mathrm{E}(X)^2]} \\ &= -\alpha_a + \sqrt{\big[\alpha_a + \mathrm{E}(X)\,\big]^2 + \mathrm{Var}(X)} \end{aligned}$$

e il minimo valore di p che verifica la seconda è dato dall'importo:

$$\begin{aligned} p_* &= \alpha_A + \mathrm{E}(X) - \sqrt{\big[\alpha_A + \mathrm{E}(X)\big]^2 - \big[2\alpha_A\mathrm{E}(X) + \mathrm{E}(X^2)\big]} \\ &= \alpha_A + \mathrm{E}(X) - \sqrt{\alpha_A^2 + \mathrm{E}(X)^2 - \mathrm{E}(X^2)} = \alpha_A + \mathrm{E}(X) - \sqrt{\alpha_A^2 - \mathrm{Var}(X)}. \end{aligned}$$

Ricordata, a questo punto, l'approssimazione $\sqrt{a^2 + x} \approx a + \frac{1}{2a}\,x\,(a > 0;\, x$ "piccolo"), supponiamo che il danno aleatorio sia poco disperso. Possiamo allora approssimare la

radice:
- $\sqrt{\big[\alpha_a + \mathrm{E}(X)\big]^2 + \mathrm{Var}(X)}$ con l'importo $\alpha_a + \mathrm{E}(X) + \frac{\mathrm{Var}(X)}{2\,[\alpha_a + \mathrm{E}(X)\,]}$;
- $\sqrt{\alpha_A^2 - \mathrm{Var}(X)}$ con l'importo $\alpha_A + \frac{-\mathrm{Var}(X)}{2\alpha_A}$

e quindi ottenere le seguenti approssimazioni, rispettivamente, di p^* e p_*:

$$\widetilde{p}^* = \mathrm{E}(X) + \frac{\mathrm{Var}(X)}{2\big[\alpha_a + \mathrm{E}(X)\big]}, \qquad \widetilde{p}_* = \mathrm{E}(X) + \frac{\mathrm{Var}(X)}{2\alpha_A}.$$

Supposto infine $\alpha_a + \mathrm{E}(X) < \alpha_A$ (ricordiamo che la disuguaglianza $\alpha_a < \alpha_A$ già sussiste), possiamo concludere (con le cautele dovute alle approssimazioni effettuate) che si possono trovare contratti che prevedano un caricamento del premio equo $\mathrm{E}(X)$ non svantaggiosi per entrambi i soggetti, qualora il caricamento risulti contenuto, sia cioè compreso tra le soglie $\frac{\mathrm{Var}(X)}{2\alpha_A}$ e $\frac{\mathrm{Var}(X)}{2\,[\alpha_a + \mathrm{E}(X)\,]}$. ▲

Capitolo 2

Decisioni statistiche

Adottando lo schema dell'utilità attesa, assumiamo che DM esprima le proprie preferenze sulle lotterie della famiglia $\pounds$ in termini di utilità attesa, tramite un'utilità cardinale u, e che $\pounds$ presenti la massima estensione possibile per tale rappresentazione numerica (cioè, coincida con la famiglia delle lotterie ℓ per le quali esista (finito o no) il valore atteso $\mathrm{E}_\ell(u)$)[1].

Supponiamo inoltre che la scelta di una qualsiasi decisione d comporti, per DM, una conseguenza dipendente (oltre che da d) da un ulteriore elemento, lo **stato di natura** Z, di specificazione non nota a DM e da lui ritenuta dovuta solo al "caso"[2]. Indicati con Θ lo **spazio parametrico** (cioè l'insieme delle determinazioni possibili di Z) e, qualunque siano $d \in \mathfrak{D}$ e $\theta \in \Theta$, con $\gamma(\theta, d)$ la corrispondente conseguenza, possiamo introdurre la **funzione di danno** $\mathrm{L} : (\theta, d) \mapsto -u(\gamma(\theta, d))$ che consente a DM di valutare la "disutilità" $\mathrm{L}_d(\cdot) = \mathrm{L}(\cdot, d)$ della scelta della decisione d[3].

[1] Conseguentemente, per il Teorema 1.3.1(i), la relazione di preferenza $\succeq^*$, considerata solamente sulla famiglia $\pounds'$ delle lotterie aventi utilità attesa finita, è un preordinamento che verifica la proprietà archimedea e quelle di completezza, coerenza, dominanza e indipendenza.

[2] Lo stato di natura rappresenta quindi un qualsiasi "parametro", rilevante per le scelte del decisore, che per DM riesca ben definito e - nello stato di informazione relativo al momento di effettuare la scelta - di determinazione sconosciuta e non influenzata nè dal suo comportamento, nè da quello di altri individui che intendessero favorirlo o danneggiarlo.

Osserviamo anche che l'assunzione di un'unica fonte d'incertezza per le conseguenze non è, in effetti, una condizione restrittiva. Infatti, qualora l'incertezza della conseguenza connessa alla decisione d fosse dovuta ad uno stato di natura Z_d (dipendente da d), basterebbe, per ricondursi ad un'unica fonte d'incertezza, considerare come stato di natura di riferimento la famiglia $Z = (Z_d)_{d \in D}$.

[3] I problemi decisionali che compaiono nelle applicazioni (come, ad esempio, l'allocazione

Assumiamo altresì che a questo contesto decisionale ne sia affiancato uno inferenziale, consentendo così a DM di acquisire ulteriori informazioni sullo stato di natura ricorrendo ad un opportuno esperimento statistico. Precisamente, considerati un fenomeno (per DM) aleatorio X e una σ-algebra $\aleph$ sul relativo **spazio campionario** $\mathfrak{X}$ (cioè, sull'insieme delle realizzazioni possibili di X)[4], supponiamo che DM sia in grado di osservare la realizzazione di X e la ritenga generata in accordo con una legge probabilistica P_Z su $\aleph$ dipendente dalla determinazione "vera" di Z. Poichè questa determinazione gli è sconosciuta, DM sarà portato a considerare la legge P_Z in corrispondenza ad ogni possibile specificazione dello stato di natura, pervenendo così ad una famiglia $(\mathrm{P}_\theta)_{\theta\in\Theta}$ di probabilità su $\aleph$, parametrizzata sullo spazio parametrico. Per quanto riguarda le proprietà da richiedere alle leggi di tale famiglia, ipotizziamo, per semplicità, che ogni suo elemento P_θ ammetta una funzione di densità f_θ rispetto ad una prefissata misura (σ-finita) di riferimento μ su $\aleph$[5]. Precisiamo che chiameremo **probabilità di campionamento** le probabilità P_θ e **densità di campionamento** le densità f_θ; inoltre, che denoteremo con la lettera (dotata o no di apici o pedici):
- θ gli **stati**, cioè gli elementi dello spazio parametrico Θ;
- x i **campioni**, cioè gli elementi dello spazio campionario $\mathfrak{X}$.

Tenendo presente che le utilità cardinali adoperate nelle applicazioni statistiche sono usualmente superiormente limitate (e quindi sostituibili, per il Teorema 1.3.3, con utilità cardinali a valori non positivi), assumiamo infine $\mathrm{L} \geq 0$. Si ha allora, per il Teorema A.3.5(i), che $\pounds$ coincide con la famiglia di tutte le probabilità sulla σ-algebra di riferimento $\mathcal{C}$.

In particolare, a scopo esemplificativo, faremo quasi sempre riferimento a funzioni di danno "standard"; precisamente, nell'ambito dei problemi di:
- *stima puntuale per parametri reali* ($\Theta \subseteq \mathfrak{D} \subseteq \mathbb{R}$), alla funzione di **danno**

---

di un nuovo aereoporto) sono generalmente talmente complessi che una loro formalizzazione (nel senso qui inteso) non risulta sempre agevole. In anni recenti sono state sviluppate varie tecniche di *analisi delle decisioni* che forniscono al decisore un supporto per individuare, in specifici problemi, tutte le componenti rilevanti dello stato di natura Z (e quindi la struttura dello spazio parametrico Θ), l'insieme delle decisioni $\mathfrak{D}$ e, in particolar modo, la forma funzionale della funzione di danno L.

[4] Da un punto di vista interpretativo, $\aleph$ raccoglie tutti i sottoinsiemi A di $\mathfrak{X}$ per i quali l'evento $\{X \in \mathsf{A}\}$ è di interesse per DM.

[5] Pertanto, la probabilità $\mathrm{Pr}_\theta(X \in \mathsf{A})$ che il fenomeno aleatorio X assuma un valore appartenente all'insieme $\mathsf{A} \in \aleph$, in corrispondenza alla determinazione θ dello stato di natura, è data dall'integrale $\mathrm{P}_\theta(\mathsf{A}) = \int_{\mathsf{A}} f_\theta \, d\mu$.

quadratico: $\mathrm{L}(\theta, d) = (\theta - d)^2$, a quella di **danno lineare**:

$$\mathrm{L}(\theta, d) = \begin{cases} k(\theta - d) & \text{se } \theta > d \\ (1-k)(d-\theta) & \text{se } \theta \leq d \end{cases} \qquad (0 \leq k \leq 1)$$

oppure a quella di **danno assoluto**: $\mathrm{L}(\theta, d) = |\theta - d|$;[6]
- *stima intervallare* ($\Theta \subseteq \mathbb{R}$, $\mathfrak{D}$ famiglia di intervalli limitati di Θ), alla funzione di **danno lineare**: $\mathrm{L}(\theta, d) = k \lg(d) + 1 - I_d(\theta)$ $(k > 0)$, ove $\lg(d)$ è la lunghezza dell'intervallo d;
- *verifica d'ipotesi* ($\{\Theta_0, \Theta_1\}$ partizione di Θ formata da insiemi non vuoti e $\mathfrak{D} = \{d_0,\, d_1\}$), a funzioni di danno del tipo:

$$\mathrm{L}_i(\theta) = \mathrm{L}(\theta, d_i) = \begin{cases} 0 & \text{se } \theta \in \Theta_i \\ k_i(\theta) & \text{se } \theta \in \Theta_j \ (j \neq i) \end{cases} \tag{2.1}$$

con k_i funzione a valori positivi definita su Θ_j $(j \neq i;\ i = 1,\, 2)$[7].

2.1 Regole di decisione

Dopo aver descritto il contesto decisionale-induttivo nel quale DM è chiamato a fare le sue scelte, precisiamo ulteriormente il modello di comportamento del decisore assumendo che DM scelga la decisione da adottare *unicamente* sulla base dei possibili risultati sperimentali. Conseguentemente, non essendo a lui

[6] Il danno quadratico, introdotto da Adrien M. Legendre (1805) e Karl F. Gauss (1810) nell'ambito della valutazione degli errori di misura, è senza alcun dubbio la funzione di danno più "popolare" a causa sia della sua stretta connessione con il classico metodo dei minimi quadrati che della sua forma che consente sviluppi formali relativamente semplici. Il danno assoluto è stato invece introdotto da Pierre S. de Laplace nel 1773 in connessione con il problema della stima delle orbite dei pianeti e delle comete.

Ricordiamo che una delle principali critiche rivolte all'uso di questi due danni (oltre a quella di "pesare" allo stesso modo gli errori per difetto e quelli per eccesso) è che, in quello quadratico, vengono penalizzate pesantemente le "grandi" deviazioni dal valore "vero" dello stato di natura, mentre, in quello assoluto, le "piccole".

[7] Poichè la decisione d_i determina un danno nullo se e solo se $\theta \in \Theta_i$ $(i = 1, 2)$, possiamo interpretare d_0 come l'accettazione dell'ipotesi nulla "$\mathrm{H}_0 : Z \in \Theta_0$" e d_1 come l'accettazione dell'ipotesi alternativa "$\mathrm{H}_1 : Z \in \Theta_1$". In questo ordine di idee, k_0 può intendersi come il danno relativo ad un errore di I tipo (accettare l'ipotesi alternativa mentre è vera quella nulla) e k_1 come il danno relativo ad un errore di II tipo (accettare l'ipotesi nulla mentre è vera quella alternativa).

noto a priori quale campione verrà osservato, DM sarà portato a spostare la sua attenzione dalla scelta di una decisione a quella di un "comportamento" che precisi la decisione da adottare in corrispondenza ad ogni possibile campione. È quindi di fondamentale importanza per DM (al fine di risolvere in qualche modo il problema decisionale) riuscire ad elicere una relazione di preferenza che gli consenta di confrontare tra loro i vari "comportamenti".

Interpretati, dal punto di vista matematico, tali "comportamenti" come applicazioni di dominio lo spazio campionario e a valori nell'insieme delle decisioni, supponiamo che DM assuma (in via ipotetica) che θ sia lo stato "vero". Allora, per individuare le conseguenze a cui andrebbe incontro perseguendo il "comportamento" $\delta : \mathfrak{X} \mapsto \mathfrak{D}$, DM considererà la conseguenza aleatoria $\gamma(\theta, \delta) : x \mapsto \gamma(\theta, \delta(x))$. Supposto che tale conseguenza aleatoria sia $(\aleph, \mathcal{C})$-misurabile, rimane individuata una particolare lotteria; precisamente la legge $\ell_\theta^{(\delta)}$ di $\gamma(\theta, \delta)$:

$$\ell_\theta^{(\delta)}(C) = \mathrm{P}_\theta(\gamma(\theta, \delta) \in C)$$

per ogni $C \in \mathcal{C}$. Poichè a DM interessa, in ultima analisi, la conseguenza che subirà a causa del suo "comportamento" piuttosto che il particolare meccanismo aleatorio che la genera, viene naturale ritenere che consideri del tutto equivalenti due "comportamenti" δ_1 e δ_2 che abbiano la medesima legge (cioè tali che $\ell_\theta^{(\delta_1)} = \ell_\theta^{(\delta_2)}$). Ne viene che DM sarà indotto ad esprimere le sue preferenze sui "comportamenti" mediante la relazione di preferenza $\succeq^*$ già considerata per le lotterie. Indicata allora con $\succeq_\theta$ la relazione di preferenza sui "comportamenti" relativa allo stato θ, si avrà

$$\begin{aligned} \delta_1 \succeq_\theta \delta_2 \iff \mathrm{E}_{\ell_\theta^{(\delta_1)}}(u) \geq \mathrm{E}_{\ell_\theta^{(\delta_2)}}(u) &\iff \int_{\mathfrak{C}} u \, d\ell_\theta^{(\delta_1)} \geq \int_{\mathfrak{C}} u \, d\ell_\theta^{(\delta_2)} \\ &\iff \int_{\mathfrak{X}} u(\gamma(\theta, \delta_1)) \, d\mathrm{P}_\theta \geq \int_{\mathfrak{X}} u(\gamma(\theta, \delta_2)) \, d\mathrm{P}_\theta, \end{aligned}$$

ove l'ultima equivalenza sussiste in forza della proposizione (i) del teorema fondamentale del calcolo delle probabilità B.1.2. Risulterà pertanto

$$\delta_1 \succeq_\theta \delta_2 \iff \int_{\mathfrak{X}} \mathrm{L}(\theta, \delta_1(x)) \, \mathrm{P}_\theta(dx) \leq \int_{\mathfrak{X}} \mathrm{L}(\theta, \delta_2(x)) \, \mathrm{P}_\theta(dx).$$

Ora, essendo in realtà ignoto a DM lo stato "vero", egli sarà portato a considerare, per ogni stato, la relativa relazione di preferenza e quindi, in definitiva, ad associare, per quanto visto, ad ogni "comportamento" δ il numero aleatorio $R_\delta(Z) = \int_{\mathfrak{X}} \mathrm{L}(Z, \delta(x)) \, \mathrm{P}_Z(dx)$.

Le considerazioni svolte suggeriscono e giustificano la definizione seguente che introduce le nozioni chiave di regola di decisione e di funzione di rischio (dovute ad Abraham Wald che le considerò in un lavoro del 1939[8]).

Definizione 2.1.1 Una **regola di decisione** è una qualsiasi applicazione $\delta : \mathfrak{X} \mapsto \mathfrak{D}$ che assicuri, per ogni stato θ, l'esistenza e la finitezza dell'integrale $\mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)\big) = \int_{\mathfrak{X}} \mathrm{L}(\theta,\delta(x))\, \mathrm{P}_\theta(dx)$[9]. Inoltre, la **funzione di rischio** della regola di decisione δ è la funzione $R_\delta : \theta \mapsto \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)\big)$[10]. Infine, Δ denota l'insieme delle regole di decisione e δ (dotata o no di apici o pedici) un suo elemento generico.

Ovviamente, tra le regole di decisione ci sono anche le applicazioni costanti di $\mathfrak{X}$ in $\mathfrak{D}$. Osservato che la funzione di rischio della costante di valore d è la funzione L_d (detta **funzione di danno relativa a** d), le applicazioni costanti possono essere intese come quelle particolari regole di decisione atte a rappresentare i "comportamenti" di un decisore che, non intendendo o non potendo usufruire dell'informazione campionaria, effettui la sua scelta solamente riferendosi alle funzioni di danno relative alle decisioni. Ne viene che l'analisi delle scelte individuali in condizioni d'incertezza senza acquisione di dati (obiettivo specifico della *teoria delle decisioni*) può essere condotta, nello schema decisionale-inferenziale qui considerato, supponendo che lo spazio campionario abbia un solo elemento (e quindi identificando le regole di decisione con le decisioni e le funzioni di rischio con quelle di danno relativo a decisioni).

Prima di mostrare che nelle usuali modellizzazioni statistiche si possono trovare funzioni di rischio sia continue che discontinue che costanti, determiniamo le espressioni che la funzione di rischio assume nei problemi di stima

[8]Cogliamo l'occasione per ricordare che, nel medesimo lavoro, Wald introdusse anche le nozioni di regola di decisione ammissibile, bayesiana e minimax che, come vedremo, formano l'impalcatura concettuale della teoria delle decisioni statistiche. Osserviamo inoltre che tale teoria costituisce il riferimento di fondo di una parte rilevante della statistica (sia teorica che applicata).

[9]Quindi, nel caso particolare che $\mathfrak{X}$ sia discreto, ogni applicazione di $\mathfrak{X}$ in $\mathfrak{D}$ è una funzione di decisione (in quanto, in questo caso, $\aleph$ viene identificata con l'insieme delle parti).

[10]Che associa, ad ogni stato θ, il danno medio (finito) a cui va incontro DM adottando il "comportamento" δ, qualora θ sia lo stato "vero". Osserviamo che limitarsi a considerare solamente "comportamenti" che implicano un danno medio finito (in ogni stato), se da una parte esclude la possibilità di sviluppare la teoria delle decisioni statistiche in tutta generalità, dall'altra consente comunque di ottenere i principali risultati che vengono adoperati nelle sue usuali applicazioni a contesti reali.

puntuale con danno quadratico, in quelli di stima intervallare con danno lineare e in quelli di verifica d'ipotesi con danno (2.1).

Osservazione 2.1.2 (i) Con riferimento alla stima puntuale per parametri reali, consideriamo il danno quadratico e chiamiamo **errore quadratico medio** di uno **stimatore** (regola di decisione) δ la sua funzione di rischio che, prendendo spunto dalla corrispondente frase inglese "Mean Squared Error", denotiamo con il simbolo MSE_δ. Riesce pertanto

$$\mathrm{MSE}_\delta(\theta) = \mathrm{E}_\theta\big((\theta-\delta)^2\big) = \int_{\mathfrak{X}} [\theta - \delta(x))^2 \, \mathrm{P}_\theta(dx)$$

per ogni stato θ. Tenuto conto che, per definizione, la funzione $\theta - \delta(\cdot)$ è a quadrato P_θ-integrabile, possiamo concludere che è pure P_θ-integrabile (Teorema A.3.5(v)). Conseguentemente, $\theta - \mathrm{E}_\theta(\delta) = \mathrm{E}_\theta(\theta-\delta) \in \mathbb{R}$ e quindi la speranza matematica $\mathrm{E}_\theta(\delta)$ è finita. Riesce allora

$$\begin{aligned}
\mathrm{MSE}_\delta(\theta) &= \mathrm{E}_\theta\big(\,[(\delta - \mathrm{E}_\theta(\delta)) + (\mathrm{E}_\theta(\delta)-\theta)]^2\big) \\
&= \mathrm{E}_\theta\big((\delta - \mathrm{E}_\theta(\delta))^2\big) + \mathrm{E}_\theta\big(\,(\mathrm{E}_\theta(\delta)-\theta)^2\big) \\
&\qquad\qquad + 2\mathrm{E}_\theta\big(\,(\delta - \mathrm{E}_\theta(\delta))(\mathrm{E}_\theta(\delta)-\theta)\,\big) \\
&= \mathrm{Var}_\theta(\delta) + (\mathrm{E}_\theta(\delta)-\theta)^2 + 2\big(\,\mathrm{E}_\theta(\delta) - \mathrm{E}_\theta(\delta))(\mathrm{E}_\theta(\delta)-\theta)\,\big)
\end{aligned}$$

e quindi

$$\mathrm{MSE}_\delta(\theta) = \mathrm{Var}_\theta(\delta) + \mathrm{Bias}_\theta(\delta)^2, \tag{2.2}$$

avendo posto $\mathrm{Bias}_\delta(\theta) = \mathrm{E}_\theta(\delta) - \theta$. Nello stato θ, l'errore quadratico medio incorpora quindi due componenti: una - $\mathrm{Var}_\theta(\delta)$ - che misura la variabilità (precisione) dello stimatore e l'altra - $\mathrm{Bias}_\theta(\delta)$ - che ne misura la **distorsione** (accuratezza).

(ii) Con riferimento alla stima intervallare, chiamiamo **stimatore intervallare** ogni regola di decisione δ tale che la lunghezza aleatoria $\mathrm{lg}(\delta)$ sia una funzione $\aleph$-Borel misurabile e l'insieme di copertura $\{\theta \in \delta\}$ di θ appartenga a $\aleph$ per ogni stato θ. Considerato ora il danno lineare, per la relativa funzione di rischio, otteniamo $R_\delta(\theta) = \mathrm{E}_\theta(k\,\mathrm{lg}(\delta) + 1 - I_\delta(\theta)) = k\mathrm{E}_\theta(\mathrm{lg}(\delta)) + 1 - \mathrm{E}_\theta(I_\delta(\theta))$ e quindi

$$R_\delta(\theta) = k\mathrm{E}_\theta(\mathrm{lg}(\delta)) + 1 - \mathrm{P}_\theta(\theta \in \delta). \tag{2.3}$$

Nello stato θ, la funzione di rischio incorpora dunque due componenti: una - $\mathrm{E}_\theta(\lg(\delta))$ - che contempla la lunghezza media dell'intervallo aleatorio $\delta(X)$ e l'altra - $\mathrm{P}_\theta(\theta \in \delta)$ - che considera la **probabilità di copertura di** θ, cioè la probabilità che l'intervallo aleatorio contenga θ.

(iii) Con riferimento alla verifica d'ipotesi, chiamiamo **test** ogni regola di decisione δ tale che la **regione di accettazione** $\mathfrak{X}_0^{(\delta)} = \{\delta = d_0\}$ di δ e la **regione di rifiuto** $\mathfrak{X}_1^{(\delta)} = \{\delta = d_1\}$ di δ appartengano a $\aleph$. Inoltre, per **funzione di potenza** del test δ intendiamo la funzione $\beta_\delta : \theta \mapsto \mathrm{P}_\theta(\mathfrak{X}_1^{(\delta)})$ che associa, ad ogni stato, la probabilità relativa a quello stato che il test rifiuti l'ipotesi nulla. Considerato ora il danno (2.1), per la relativa funzione di rischio, otteniamo

$$\begin{aligned}
R_\delta(\theta) &= \int_{\mathfrak{X}_0^{(\delta)}} \mathrm{L}(\theta,\delta)\, d\mathrm{P}_\theta + \int_{\mathfrak{X}_1^{(\delta)}} \mathrm{L}(\theta,\delta)\, d\mathrm{P}_\theta \\
&= \int_{\mathfrak{X}_0^{(\delta)}} \mathrm{L}_0(\theta)\, d\mathrm{P}_\theta + \int_{\mathfrak{X}_1^{(\delta)}} \mathrm{L}_1(\theta)\, d\mathrm{P}_\theta = \mathrm{L}_0(\theta)\, \mathrm{P}_\theta(\mathfrak{X}_0^{(\delta)}) + \mathrm{L}_1(\theta)\, \mathrm{P}_\theta(\mathfrak{X}_1^{(\delta)}) \\
&= \mathrm{L}_0(\theta)\, [1 - \mathrm{P}_\theta(\mathfrak{X}_1^{(\delta)})] + \mathrm{L}_1(\theta)\, \mathrm{P}_\theta(\mathfrak{X}_1^{(\delta)}) = \mathrm{L}_0(\theta) + \beta_\delta(\theta)[\mathrm{L}_1(\theta) - \mathrm{L}_0(\theta)]
\end{aligned}$$

e quindi

$$R_\delta(\theta) = \begin{cases} k_1(\theta)\, \beta_\delta(\theta) & \text{se } \theta \in \Theta_0 \\ k_0(\theta)(1 - \beta_\delta(\theta)) & \text{se } \theta \in \Theta_1 \end{cases}. \tag{2.4}$$

Nello stato θ, la funzione di rischio dipende quindi solamente dalla probabilità $\beta_\delta(\theta)$ che il test rifiuti l'ipotesi nulla. ♦

Esempio 2.1.3 (i) FUNZIONE DI RISCHIO COSTANTE Posto $\Theta = \mathbb{R}$ e $\mathfrak{D}$ l'insieme degli intervalli limitati, sia l'osservabile una v.a. distribuita secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ. Inoltre, per ogni numero reale $c \geq 0$, consideriamo lo stimatore intervallare $\delta_c : x \mapsto [x - c\sigma, x + c\sigma]$. Allora, $\mathrm{E}_\theta(\lg(\delta)) = 2c\sigma$. Osservato che $I_{\{\theta\in\delta(x)\}}(x) = I_{[-c,c]}\big(\frac{x-\theta}{\sigma}\big)$ per ogni x, θ e tenuto conto del Corollario A.4.2, si ha anche

$$\begin{aligned}
\mathrm{P}_\theta(\theta \in \delta) &= \mathrm{E}_\theta\Big(I_{[-c,c]}\Big(\frac{\cdot - \theta}{\sigma}\Big)\Big) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} I_{[-c,c]}\Big(\frac{x-\theta}{\sigma}\Big) \exp\Big[-\frac{(x-\theta)^2}{2\sigma^2}\Big] dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\, \sigma \int_{-\infty}^{+\infty} I_{[-c,c]}(x) \exp\Big[-\frac{x^2}{2}\Big] dx = \frac{1}{\sqrt{2\pi}} \int_{-c}^{c} \exp\Big[-\frac{x^2}{2}\Big] dx \\
&= \frac{1}{\sqrt{2\pi}} \Big[\int_{-c}^{0} \exp\Big[-\frac{x^2}{2}\Big] dx + \int_0^c \exp\Big[-\frac{x^2}{2}\Big] dx\Big] = \frac{2}{\sqrt{2\pi}} \int_0^c \exp\Big[-\frac{x^2}{2}\Big] dx.
\end{aligned}$$

Adottando il danno lineare riesce allora $R_{\delta_c}(\theta) = 2kc\sigma + 1 - \frac{2}{\sqrt{2\pi}} \int_0^c \exp\left[-\frac{x^2}{2}\right] dx$ e quindi la costanza della funzione di rischio.

(ii) Funzione di rischio continua Posto $\Theta = \mathfrak{D} = \mathbb{R}$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ. Allora, per (B.10) dell'Esempio B.2.18(ii), le densità di campionamento assumono la seguente espressione:

$$f_\theta(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}{2\sigma^2}\right] \exp\left[-\frac{n(\bar{\mathbf{x}} - \theta)^2}{2\sigma^2}\right] \tag{2.5}$$

per ogni $\mathbf{x} \in \mathfrak{X} = \mathbb{R}^n$.[11]

Al fine di verificare la continuità dell'errore quadratico medio di un qualsiasi stimatore δ, sia $\tilde{\theta}$ uno stato generico. Allora, qualunque sia lo stato θ, si ha

$$\left|\mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\tilde{\theta} - \delta)^2\big)\right| = \left|\mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2 - (\tilde{\theta} - \delta)^2\big)\right| \leq \mathrm{E}_{\tilde{\theta}}(|\,(\theta - \delta)^2 - (\tilde{\theta} - \delta)^2\,|\,)$$

e quindi (ponendo $d\mathbf{x} = dx_1 \cdots dx_n$)

$$\begin{aligned}
\left|\mathrm{MSE}_\delta(\theta) - \mathrm{MSE}_\delta(\tilde{\theta})\right| &= \left|\,\big[\mathrm{E}_\theta\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big)\big] + \big[\mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\tilde{\theta} - \delta)^2\big)\big]\,\right| \\
&\leq \left|\mathrm{E}_\theta\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big)\right| + \left|\mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\tilde{\theta} - \delta)^2\big)\right| \\
&\leq \left|\mathrm{E}_\theta\big((\theta - \delta)^2\big) - \mathrm{E}_{\tilde{\theta}}\big((\theta - \delta)^2\big)\right| + \mathrm{E}_{\tilde{\theta}}(|\,(\theta - \delta)^2 - (\tilde{\theta} - \delta)^2\,|\,) \\
&= \left|\int_{\mathbb{R}^n} (\theta - \delta(\mathbf{x}))^2\,(f_\theta(\mathbf{x}) - f_{\tilde{\theta}}(\mathbf{x}))\,d\mathbf{x}\right| \\
&\qquad + \int_{\mathbb{R}^n} |\,(\theta - \delta(\mathbf{x}))^2 - (\tilde{\theta} - \delta(\mathbf{x}))^2\,|\,f_{\tilde{\theta}}(\mathbf{x})\,d\mathbf{x} \\
&\leq \int_{\mathbb{R}^n} (\theta - \delta(\mathbf{x}))^2\,|f_\theta(\mathbf{x}) - f_{\tilde{\theta}}(\mathbf{x})|\,d\mathbf{x} \\
&\qquad + \int_{\mathbb{R}^n} |\,(\theta - \delta(\mathbf{x}))^2 - (\tilde{\theta} - \delta(\mathbf{x}))^2\,|\,f_{\tilde{\theta}}(\mathbf{x})\,d\mathbf{x}.
\end{aligned}$$

Posto allora:

$$\begin{aligned}
I(\theta) &= \int_{\mathbb{R}^n} (\theta - \delta(\mathbf{x}))^2\,|f_\theta(\mathbf{x}) - f_{\tilde{\theta}}(\mathbf{x})|\,d\mathbf{x} \\
J(\theta) &= \int_{\mathbb{R}^n} |\,(\theta - \delta(\mathbf{x}))^2 - (\tilde{\theta} - \delta(\mathbf{x}))^2\,|\,f_{\tilde{\theta}}(\mathbf{x})\,d\mathbf{x}
\end{aligned}$$

otteniamo

$$\left|\mathrm{MSE}_\delta(\theta) - \mathrm{MSE}_\delta(\tilde{\theta})\right| \leq I(\theta) + J(\theta). \tag{2.6}$$

Inoltre, dalla

$$\begin{aligned}
(\theta - \delta)^2 = [(\theta - \tilde{\theta}) + (\tilde{\theta} - \delta)]^2 &= (\theta - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2 + 2(\theta - \tilde{\theta})(\tilde{\theta} - \delta) \\
&\leq (\theta - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2 + [(\theta - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2]
\end{aligned}$$

[11] Avendo posto, come usuale nella letteratura statistica, $\bar{\mathbf{x}} = \frac{1}{n}(x_1 + \cdots + x_n)$.

si ha

$$(\theta - \delta)^2 \leq 2[(\theta - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2]. \tag{2.7}$$

Sia ora $\epsilon > 0$. Considerata una successione $(\theta_k)_{k\geq 1}$ nell'intervallo aperto $]\tilde{\theta} - \epsilon, \tilde{\theta} + \epsilon[$ convergente a $\tilde{\theta}$, da (2.7) otteniamo

$$\begin{aligned} |\,(\theta_k - \delta)^2 - (\tilde{\theta} - \delta)^2| &\leq (\theta_k - \delta)^2 + (\tilde{\theta} - \delta)^2 \leq 2\big[(\theta_k - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2\big] + (\tilde{\theta} - \delta)^2 \\ &= 3(\tilde{\theta} - \delta)^2 + 2(\theta_k - \tilde{\theta})^2 \leq 3(\tilde{\theta} - \delta)^2 + 2\epsilon^2 \end{aligned}$$

e quindi le funzioni della successione $(|\,(\theta_k - \delta)^2 - (\tilde{\theta} - \delta)^2|\, f_{\tilde{\theta}})_{k\geq 1}$ sono tutte maggiorate dalla funzione integrabile $(3(\tilde{\theta} - \delta)^2 + 2\epsilon^2)\, f_{\tilde{\theta}}$. Per il teorema della convergenza dominata A.3.11 si ha allora

$$\lim_{k\to+\infty} J(\theta_k) = \int_{\mathbb{R}^n} \lim_{k\to+\infty} |\,(\theta_k - \delta(\mathbf{x}))^2 - (\tilde{\theta} - \delta(\mathbf{x}))^2\,|\, f_{\tilde{\theta}}(\mathbf{x})\, d\mathbf{x} = 0. \tag{2.8}$$

Per quanto riguarda invece la funzione che compare nell'integrale $I(\theta)$, riesce, sempre per (2.7), $(\theta_k - \delta)^2\,|\, f_{\theta_k} - f_{\tilde{\theta}}\,| \leq 2[(\theta_k - \tilde{\theta})^2 + (\tilde{\theta} - \delta)^2]\,[f_{\theta_k} + f_{\tilde{\theta}}]$ e quindi

$$(\theta_k - \delta)^2\,|\, f_{\theta_k} - f_{\tilde{\theta}}\,| \leq 2[(\tilde{\theta} - \delta)^2 + \epsilon^2]\,(f_{\theta_k} + f_{\tilde{\theta}}). \tag{2.9}$$

Ora, qualunque sia il campione $\mathbf{x}$, da (2.5) otteniamo

$$\begin{aligned} f_{\theta_k}(\mathbf{x}) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2 + n\bar{\mathbf{x}}^2}{2\sigma^2}\Big] \exp\Big[-\frac{n\theta_k^2}{2\sigma^2}\Big] \exp\Big[-\frac{n\bar{\mathbf{x}}\theta_k}{\sigma^2}\Big] \\ &\leq (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2 + n\bar{\mathbf{x}}^2}{2\sigma^2}\Big]\, K\, \exp\Big[-\frac{n\bar{\mathbf{x}}\alpha(\bar{\mathbf{x}})}{\sigma^2}\Big], \end{aligned}$$

avendo posto:

$$K = \sup_{\theta\,\in]\tilde{\theta}-\epsilon, \tilde{\theta}+\epsilon[} \exp\Big[-\frac{n\theta^2}{2\sigma^2}\Big], \qquad \alpha(\bar{\mathbf{x}}) = \begin{cases} \tilde{\theta} - \epsilon & \text{se } \bar{\mathbf{x}} \geq 0 \\ \tilde{\theta} + \epsilon & \text{se } \bar{\mathbf{x}} < 0 \end{cases}.$$

Osservato che la funzione

$$g : \mathbf{x} \mapsto (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2 + n\bar{\mathbf{x}}^2}{2\sigma^2}\Big]\, K\, \exp\Big[-\frac{n\bar{\mathbf{x}}\alpha(\bar{\mathbf{x}})}{\sigma^2}\Big]$$

è integrabile, lo è pure la funzione $2[(\tilde{\theta} - \delta)^2 + \epsilon^2]\,(g + f_{\tilde{\theta}})$. Le funzioni della successione $((\theta_k - \delta)^2\,|\, f_{\theta_k} - f_{\tilde{\theta}}\,|)_{k\geq 1}$ sono quindi, per (2.9), tutte maggiorate da una medesima funzione integrabile. Dal teorema della convergenza dominata si ha allora

$$\begin{aligned} \lim_{k\to+\infty} I(\theta_k) &= \int_{\mathbb{R}^n} \lim_{k\to+\infty} (\theta_k - \delta(\mathbf{x}))^2\,|\, f_{\theta_k}(\mathbf{x}) - f_{\tilde{\theta}}(\mathbf{x})\,|\, d\mathbf{x} \\ &= \int_{\mathbb{R}^n} (\tilde{\theta} - \delta(\mathbf{x}))^2 \lim_{k\to+\infty} |\, f_{\theta_k}(\mathbf{x}) - f_{\tilde{\theta}}(\mathbf{x})\,|\, d\mathbf{x} = 0, \end{aligned}$$

tenuto conto di (2.5). Pertanto, tramite (2.6) e (2.8), risulta $\lim_{k\to+\infty} \mathrm{MSE}_\delta(\theta_k) = \mathrm{MSE}_\delta(\tilde{\theta})$. Conseguentemente, data l'arbitrarietà della successione $(\theta_k)_{k\geq 1}$, l'errore quadratico medio è continuo in $\tilde{\theta}$.

(iii) Funzione di rischio discontinua Supposto che lo stato di natura sia un parametro di posizione, poniamo $\Theta = \mathfrak{D} = \mathbb{R}$ e scegliamo la funzione di danno:

$$\mathrm{L}(\theta, d) = \begin{cases} |\theta - d| & \text{se } |\theta - d| \leq 1 \\ 1 & \text{se } |\theta - d| > 1 \end{cases}.$$

Inoltre, consideriamo come osservabile una v.a. avente probabilità di campionamento P_θ tale che

$$\mathrm{P}_\theta(\theta - 1) = \mathrm{P}_\theta(\theta + 1) = \frac{1}{2}$$

per ogni stato θ. Allora, la funzione di rischio relativa alla regola di decisione:

$$\delta(x) = \begin{cases} x + 1 & \text{se } x < 0 \\ x - 1 & \text{se } x \geq 0 \end{cases}$$

è una funzione discontinua. Infatti, dalle $R_\delta(\theta) = \frac{1}{2}\left[\mathrm{L}(\theta, \delta(\theta - 1)) + \mathrm{L}(\theta, \delta(\theta + 1))\right]$ e

$$\delta(\theta - 1) = \begin{cases} \theta & \text{se } \theta < 1 \\ \theta - 2 & \text{se } \theta \geq 1 \end{cases}, \qquad \delta(\theta + 1) = \begin{cases} \theta + 2 & \text{se } \theta < -1 \\ \theta & \text{se } \theta \geq -1 \end{cases}$$

otteniamo

$$R_\delta(\theta) = \begin{cases} \frac{1}{2}\left[\mathrm{L}(\theta, \theta) + \mathrm{L}(\theta, \theta + 2)\right] & \text{se } \theta < -1 \\ \frac{1}{2}\left[\mathrm{L}(\theta, \theta) + \mathrm{L}(\theta, \theta)\right] & \text{se } -1 \leq \theta < 1 \\ \frac{1}{2}\left[\mathrm{L}(\theta, \theta - 2) + \mathrm{L}(\theta, \theta)\right] & \text{se } \theta \geq 1 \end{cases} = \begin{cases} 0 & \text{se } -1 \leq \theta < 1 \\ \frac{1}{2} & \text{altrimenti} \end{cases}$$

e quindi la discontinuità della funzione di rischio. ▲

La possibilità di effettuare osservazioni per acquisire ulteriori informazioni sullo stato di natura, induce DM a sostituire ex ante le decisioni con le regole di decisione e quindi a risolvere il problema decisionale individuandone una che ritenga in qualche senso appetibile. Ora, essendo DM un individuo razionale, appare evidente che date δ e δ', se riesce $R_\delta \leq R_{\delta'}$, allora δ sarà per DM non peggiore di δ'; se invece si ha $R_\delta < R_{\delta'}$[12], allora δ sarà per DM migliore di δ'; se infine risulta $R_\delta = R_{\delta'}$, allora le due regole saranno per DM interscambiabili[13]. Queste considerazioni suggeriscono la seguente

[12]Cioè, $R_\delta \leq R_{\delta'}$ e $R_\delta(\tilde{\theta}) < R_{\delta'}(\tilde{\theta})$ per qualche stato $\tilde{\theta}$.

[13]Pur essendo ovvio, vale la pena rilevare che, data la natura delle funzioni di rischio, queste preferenze si basano sul confronto dei danni medi ottenuti considerando *tutti* i possibili campioni e non sul confronto dei danni medi relativi *a qualche* campione particolare (come potrebbe essere, ad esempio, quello osservato).

definizione che introduce nell'insieme Δ due preordinamenti (il secondo dei quali è un ordinamento stretto) e una equivalenza.

Definizione 2.1.4 Date due regole di decisione δ e δ':
- δ **domina** δ' (in simboli $\delta \succeq \delta'$) se $R_\delta \le R_{\delta'}$;
- δ **domina strettamente** δ' (in simboli $\delta \succ \delta'$) se $R_\delta < R_{\delta'}$;
- δ è **equivalente** a δ' (in simboli $\delta \sim \delta'$) se $R_\delta = R_{\delta'}$.[14]

Diamo ora alcuni esempi di regole di decisione dominate.

Esempio 2.1.5 Posto $\Theta = \mathfrak{D} = \mathbb{R}$, consideriamo come osservabile una v.a. distribuita secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ.

(i) Scelto, per ogni numero reale c, lo stimatore lineare $\delta_c : x \mapsto cx$, otteniamo $\mathrm{E}_\theta(\delta_c) = c\,\theta$ e quindi $\mathrm{Var}_\theta(\delta_c) = c^2\sigma^2$ e $\mathrm{Bias}_\theta(\delta_c) = \mathrm{E}_\theta(\delta_c) - \theta = (c-1)\,\theta$. Tramite (2.2) risulta allora $\mathrm{MSE}_\delta(\theta) = c^2\sigma^2 + (c-1)^2\theta^2$. Conseguentemente, $\delta_1 \succ \delta_c$ ogniqualvolta $|c| > 1$ oppure $c = -1$. Infatti, nel primo caso, $\mathrm{MSE}_{\delta_1}(\theta) = \sigma^2 < c^2\sigma^2 + (c-1)^2\theta^2 = \mathrm{MSE}_{\delta_c}(\theta)$ per ogni stato θ; nel secondo $\mathrm{MSE}_{\delta_1}(0) = \sigma^2 = (-1)^2\sigma^2 = \mathrm{MSE}_{\delta_{-1}}(0)$ e $\mathrm{MSE}_{\delta_1}(\theta) = \sigma^2 < \sigma^2 + (-2)^2\theta^2 = \mathrm{MSE}_{\delta_{-1}}(\theta)$ per ogni stato $\theta \neq 0$. Per quanto concerne invece gli altri valori di c diversi dall'unità, δ_1 e δ_c non sono confrontabili. Infatti, dalla $c^2\sigma^2 + (c-1)^2\theta^2 > \sigma^2$ discende $\theta^2 > \frac{\sigma^2(c+1)}{1-c}$ e quindi, per ogni $c \in]-1, 1[$, otteniamo $\mathrm{MSE}_{\delta_1}(0) = \sigma^2 > c^2\sigma^2 = \mathrm{MSE}_{\delta_c}(0)$ e, ad esempio, $\mathrm{MSE}_{\delta_1}(\theta) < \mathrm{MSE}_{\delta_c}(\theta)$ per $\theta > \sigma\sqrt{\frac{c+1}{1-c}}$.

(ii) Considerata la funzione di danno "0-1":

$$\mathrm{L}(\theta, d) = \begin{cases} 0 & \text{se } \theta \le d \\ 1 & \text{se } \theta > d \end{cases},$$

verifichiamo che ogni stimatore δ è dominato strettamente dallo stimatore "incrementato" $\delta' = \delta + 1$. Osservato che

$$R_\delta(\theta) = \int_{\{\delta < \theta\}} f_\theta(x)\, dx = \mathrm{P}_\theta(\delta < \theta)$$

per ogni δ, otteniamo $R_{\delta'}(\theta) = \mathrm{P}_\theta(\delta' < \theta) = \mathrm{P}_\theta(\delta < \theta - 1) \le \mathrm{P}_\theta(\delta < \theta - 1) + \mathrm{P}_\theta(\theta - 1 \le \delta < \theta) = \mathrm{P}_\theta(\delta < \theta) = R_\delta(\theta)$. Basta allora provare che, per qualche stato $\tilde\theta$, si ha

$$\mathrm{P}_{\tilde\theta}(\tilde\theta - 1 \le \delta < \tilde\theta) = \int_{\{\tilde\theta - 1 \le \delta < \tilde\theta\}} f_\theta(x)\, dx > 0.$$

A tal fine, tenuto conto che $f_\theta \gg 0$ per ogni stato θ e del Teorema A.3.6(vi), proviamo che $\lambda(\tilde\theta - 1 \le \delta < \tilde\theta) > 0$. Considerata una numerazione $q_1, q_2, \ldots$ dei numeri razionali, dalla densità dei razionali otteniamo $\mathbb{R} = \bigcup_{n \ge 1}\{q_n - 1 \le \delta < q_n\}$. Allora

$$+\infty = \lambda\Big(\bigcup_{n\ge1}\{q_n - 1 \le \delta < q_n\}\Big) \le \sum_{n\ge1} \lambda(q_n - 1 \le \delta < q_n)$$

[14] Osserviamo che $\succ$ e $\sim$ sono, rispettivamente, la parte asimmetrica e quella simmetrica della relazione di dominanza.

e quindi esistono numeri razionali che rendono il rispettivo addendo positivo.

(iii) Con riferimento all'Esempio 2.1.3(i), per il teorema di Torricelli, si ha

$$\begin{aligned}\frac{dR_{\delta_c}(\theta)}{dc} &= \frac{d}{dc}\Big[2kc\sigma - \frac{2}{\sqrt{2\pi}}\int_0^c \exp\Big[-\frac{x^2}{2}\Big]dx\Big] \\ &= 2k\sigma - \frac{2}{\sqrt{2\pi}}\frac{d}{dc}\int_0^c \exp\Big[-\frac{x^2}{2}\Big]dx = 2k\sigma - \frac{2}{\sqrt{2\pi}}\exp\Big[-\frac{c^2}{2}\Big]\end{aligned}$$

e quindi $\frac{dR_{\delta_c}(\theta)}{dc} = 0 \iff c^2 = -2\ln\big(k\sqrt{2\pi\sigma^2}\big)$. Conseguentemente, se $k > \frac{1}{\sqrt{2\pi\sigma^2}}$, ogni stimatore intervallare δ_c è dominato strettamente dagli stimatori intervallari $\delta_{c'}$ ($c' < c$); se invece $k \leq \frac{1}{\sqrt{2\pi\sigma^2}}$, allora $c^* = \sqrt{-2\ln\big(k\sqrt{2\pi\sigma^2}\big)}$ è un punto di minimo e quindi $\delta_{c'} \succ \delta_c$, se $c < c' < c^*$ oppure $c^* < c' < c$. ▲

Concludiamo la sezione con un celebre teorema (provato indipendentemente da Radhakrishna C. Rao nel 1945 e da David Blackwell nel 1947) che introduce un metodo per migliorare uno stimatore puntuale, basato sull'uso delle statistiche sufficienti. Ricordato che per *statistica* si intende una qualsiasi funzione misurabile dei campioni, fissiamo una σ-algebra $\mathfrak{T}$ su un insieme $\mathbb{T} \neq \emptyset$ contenente i singoletti e consideriamo un'applicazione $\Upsilon : \mathfrak{X} \mapsto \mathbb{T}$ che sia $(\aleph, \mathfrak{T})$-misurabile. Supponiamo inoltre che, con riferimento ad ogni probabilità di campionamento P_θ, esista la legge condizionale $(\mathrm{Q}_\theta^{(\Upsilon)}(\mathsf{A}|\cdot))_{\mathsf{A}\in\aleph}$ della funzione identica di $\mathfrak{X}$ rispetto Υ. Allora, Υ è una **statistica sufficiente** (**per** Z) se tale legge condizionale non dipende dallo stato θ, nel senso che esiste una famiglia di probabilità $(\mathrm{Q}^{(\Upsilon)}(\cdot|t))_{t\in\mathbb{T}}$ su $\aleph$ che è una versione della legge condizionale $(\mathrm{Q}_\theta^{(\Upsilon)}(\mathsf{A}|\cdot))_{\mathsf{A}\in\aleph}$ per ogni stato θ. Conseguentemente, data una funzione $f : \mathfrak{X} \mapsto \mathbb{R}$ che sia $\mathrm{Q}^{(\Upsilon)}(\cdot|t)$-sommabile per ogni $t \in \mathbb{T}$ e P_θ-integrabile, dal Teorema B.2.12 otteniamo che l'integrale $\int_{\mathfrak{X}} f(x)\,\mathrm{Q}^{(\Upsilon)}(dx|\cdot)$ è una versione della funzione di regressione $\mathrm{E}_\theta(f|\Upsilon = \cdot)$ di f su Υ (riferita alla probabilità P_θ).

Teorema 2.1.6 (di Rao-Blackwell) *Siano $\mathfrak{D}$ un intervallo aperto (limitato o no) della retta reale e $\mathrm{L}(\theta, \cdot)$ una funzione convessa per ogni stato θ. Sia inoltre $\Upsilon : \mathfrak{X} \mapsto \mathbb{T}$ una statistica sufficiente. Allora, dato uno stimatore δ con speranza matematica $\mathrm{E}_\theta(\delta)$ finita per ogni stato θ e $\mathrm{Q}^{(\Upsilon)}(\cdot|t)$-integrabile per ogni $t \in \mathbb{T}$, la funzione $\delta^{(\Upsilon)}$ così definita:*

$$\delta^{(\Upsilon)}(x) = \int_{\mathfrak{X}} \delta(x')\,\mathrm{Q}^{(\Upsilon)}(dx'|\Upsilon(x))$$

per ogni campione x, è uno stimatore che domina δ.

DIMOSTRAZIONE Che $\delta^{(\Upsilon)}$ sia una funzione a valori nell'intervallo $\mathfrak{D}$ segue dal Teorema B.1.5(vi). Inoltre, dato uno stato θ, per il Teorema B.2.9(i) e il teorema fondamentale del calcolo delle probabilità, si ha

$$R_\delta(\theta) = \int_{\mathbb{T}} \mathrm{E}_\theta(\mathrm{L}(\theta,\delta)|\Upsilon = t)\,(\mathrm{P}_\theta)_\Upsilon(dt) = \int_{\mathfrak{X}} \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)|\Upsilon = \Upsilon(x)\big)\,\mathrm{P}_\theta(dx).$$

Ora, per il teorema di Jensen B.2.5 (ponendo $g = \mathrm{L}(\theta,\cdot)$ e $X = \delta$), $0 \le \mathrm{L}\big(\theta,\mathrm{E}_\theta(\delta)\big) \le \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)\big) = R_\delta(\theta) < +\infty$ e $\mathrm{L}\big(\theta,\mathrm{E}_\theta(\delta|\Upsilon)\big) \le \mathrm{E}_\theta(\mathrm{L}(\theta,\delta)|\Upsilon)$ (P_θ-q.c.). Osservato che le funzioni $\mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)|\Upsilon = \Upsilon(\cdot)\big)$, $\mathrm{E}_\theta\big(\delta|\Upsilon = \Upsilon(\cdot)\big)$ sono, rispettivamente, versioni di $\mathrm{E}_\theta(\mathrm{L}(\theta,\delta)|\Upsilon)$ e $\mathrm{E}_\theta(\delta|\Upsilon)$, otteniamo $\mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta)|\Upsilon = \Upsilon(\cdot)\big) \ge \mathrm{L}(\theta,\mathrm{E}_\theta(\delta|\Upsilon = \Upsilon(\cdot))$ (P_θ-q.c.) e quindi

$$\begin{aligned} R_\delta(\theta) &\ge \int_{\mathfrak{X}} \mathrm{L}\big(\theta,\mathrm{E}_\theta(\delta|\Upsilon = \Upsilon(x))\big)\,\mathrm{P}_\theta(dx) = \int_{\mathbb{T}} \mathrm{L}\big(\theta,\mathrm{E}_\theta(\delta|\Upsilon = t)\big)\,(\mathrm{P}_\theta)_\Upsilon(dt) \\ &= \int_{\mathbb{T}} \mathrm{L}\Big(\theta, \int_{\mathfrak{X}} \delta(x')\,\mathrm{Q}^{(\Upsilon)}(dx'|t)\Big)\,(\mathrm{P}_\theta)_\Upsilon(dt) \\ &= \int_{\mathfrak{X}} \mathrm{L}\Big(\theta, \int_{\mathfrak{X}} \delta(x')\,\mathrm{Q}^{(\Upsilon)}(dx'|\Upsilon(x))\Big)\,\mathrm{P}_\theta(dx) = \int_{\mathfrak{X}} \mathrm{L}(\theta,\delta^{(\Upsilon)}(x))\,\mathrm{P}_\theta(dx) \\ &= R_{\delta^{(\Upsilon)}}(\theta). \end{aligned}$$

La dimostrazione è così conclusa. ■

Il teorema appena provato non assicura che il passaggio allo stimatore $\delta^{(\Upsilon)}$ comporti (in generale) un effettivo miglioramento; potrebbe infatti avvenire che i due stimatori $\delta^{(\Upsilon)}$ e δ siano equivalenti. Nell'esempio seguente consideriamo una statistica sufficiente e uno stimatore per il quale tale passaggio consente di ottenere uno stimatore che lo domina strettamente.

Esempio 2.1.7 Posto $\Theta =]0,+\infty[$, sia l'osservabile costituito dalle v.a. X_1, X_2 indipendenti e distribuite secondo la distribuzione di Poisson *Poi*(θ) in corrispondenza ad ogni stato θ. Allora, per le probabilità di campionamento, si ha

$$\mathrm{P}_\theta(\mathbf{x}) = e^{-2\theta}\,\frac{\theta^{2\,\overline{\mathbf{x}}}}{x_1!\,x_2!}$$

per ogni $\mathbf{x} \in \mathfrak{X} = \mathbb{N}^2$. Inoltre, per il Teorema B.1.16, la v.a. somma $X_1 + X_2$ è distribuita secondo la distribuzione di Poisson *Poi*(2θ). Riesce infatti

$$\begin{aligned} (\mathrm{P}_\theta)_{X_1+X_2}(t) &= \sum_{k\ge 0} (\mathrm{P}_\theta)_{X_2}(t-k)(\mathrm{P}_\theta)_{X_1}(k) = \sum_{k=0}^{t} e^{-\theta}\,\frac{\theta^{t-k}}{(t-k)!}\,e^{-\theta}\,\frac{\theta^k}{k!} \\ &= e^{-2\theta}\frac{1}{t!}\sum_{k=0}^{t}\frac{t!}{k!\,(t-k)!}\,\theta^k\,\theta^{t-k} = e^{-2\theta}\frac{1}{t!}\sum_{k=0}^{t}\binom{t}{k}\theta^k\,\theta^{t-k} = e^{-2\theta}\,\frac{(2\theta)^t}{t!} \end{aligned}$$

per ogni $t \in \mathbb{N}$. Ne segue, per la distribuzione della media campionaria $\overline{\mathbf{X}} : \mathbf{x} \mapsto \overline{\mathbf{x}}$,

$$\mathrm{P}_\theta(\overline{\mathbf{X}} = t) = e^{-2\theta}\,\frac{(2\theta)^{2t}}{(2t)!} \qquad \Big(t \in \mathbb{T} = \Big\{\frac{k}{2};\, k \in \mathbb{N}\Big\}\Big)$$

e quindi

$$\mathrm{P}_\theta((x_1, 2t - x_1)\,|\overline{\mathbf{X}} = t) = \frac{e^{-2\theta}\frac{\theta^{2t}}{x_1!\,(2t-x_1)!}}{e^{-2\theta}\frac{(2\theta)^{2t}}{(2t)!}} = \frac{(2t)!}{x_1!\,(2t-x_1)!}\Big(\frac{1}{2}\Big)^{2t} = \binom{2t}{x_1}\Big(\frac{1}{2}\Big)^{2t}.$$

Osservato che

$$\mathrm{P}_\theta(\mathbf{x}|\overline{\mathbf{X}} = t) = \begin{cases} 0 & \text{se } x_1 + x_2 \neq 2t \\ \frac{\mathrm{P}_\theta((x_1, 2t-x_1))}{\mathrm{P}_\theta(\overline{\mathbf{X}}=t)} & \text{se } x_1 + x_2 = 2t \end{cases},$$

possiamo identificare $(\mathrm{Q}^{(\overline{\mathbf{X}})}(\cdot|t))_{t\in\mathbb{T}}$ con la famiglia di probabilità:

$$\mathrm{Q}^{(\overline{\mathbf{X}})}(\mathbf{x}|t) = \begin{cases} 0 & \text{se } x_1 + x_2 \neq 2t \\ \binom{2t}{x_1}\big(\frac{1}{2}\big)^{x_1}\big(\frac{1}{2}\big)^{2t-x_1} & \text{se } x_1 + x_2 = 2t \end{cases} \qquad (t \in \mathbb{T})$$

costruita tramite le distribuzioni binomiali del tipo $Bin(2t, \frac{1}{2})$. Conseguentemente, la media campionaria è una statistica sufficiente.

Volendo ora stimare la probabilità aleatoria $\Pr(X_1 = 0) = e^{-Z}$, poniamo $\mathfrak{D} = [0, 1]$ e adottiamo il danno quadratico. Considerato, in particolare, lo stimatore:

$$\delta(\mathbf{x}) = \frac{1}{2}\left(I_{\{0\}\times\mathbb{N}}(\mathbf{x}) + I_{\mathbb{N}\times\{0\}}(\mathbf{x})\right)$$

che associa, ad ogni campione $\mathbf{x}$, il numero delle componenti nulle rapportato alla totalità delle componenti, otteniamo $\delta' = \delta^{(\overline{\mathbf{X}})} = 4^{-\overline{\mathbf{X}}}$. Si ha infatti

$$\begin{aligned}
\delta'(\overline{\mathbf{x}}) &= \frac{1}{2}\int_{\mathbb{N}^2}\left(I_{\{0\}\times\mathbb{N}}(\mathbf{x}') + I_{\mathbb{N}\times\{0\}}(\mathbf{x}')\right)\mathrm{Q}^{(\overline{\mathbf{X}})}(d\mathbf{x}'|\overline{\mathbf{X}}(\mathbf{x})) \\
&= \frac{1}{2}\Big[\mathrm{Q}^{(\overline{\mathbf{X}})}(\{0\}\times\mathbb{N}|\overline{\mathbf{X}}(\mathbf{x})) + \mathrm{Q}^{(\overline{\mathbf{X}})}(\mathbb{N}\times\{0\}|\overline{\mathbf{X}}(\mathbf{x}))\Big] \\
&= \frac{1}{2}\Big[\sum_{m\geq1}\mathrm{Q}^{(\overline{\mathbf{X}})}((0,m)\,|\overline{\mathbf{X}}(\mathbf{x})) + \sum_{m\geq1}\mathrm{Q}^{(\overline{\mathbf{X}})}((m,0)\,|\overline{\mathbf{X}}(\mathbf{x}))\Big] \\
&= \frac{1}{2}\Big[\mathrm{Q}^{(\overline{\mathbf{X}})}((0,2\overline{\mathbf{X}}(\mathbf{x}))\,|\overline{\mathbf{X}}(\mathbf{x})) + \mathrm{Q}^{(\overline{\mathbf{X}})}((2\overline{\mathbf{X}}(\mathbf{x}),0)\,|\overline{\mathbf{X}}(\mathbf{x}))\Big] \\
&= \frac{1}{2}\Big[\Big(\frac{1}{2}\Big)^{2\overline{\mathbf{X}}(\mathbf{x})} + \Big(\frac{1}{2}\Big)^{2\overline{\mathbf{X}}(\mathbf{x})}\Big] = \Big(\frac{1}{4}\Big)^{\overline{\mathbf{X}}(\mathbf{x})}.
\end{aligned}$$

Passando ai relativi errori quadratici medi, proviamo che $\mathrm{MSE}_{\delta'}(\theta) < \mathrm{MSE}_{\delta}(\theta)$ per ogni stato θ. Notato che, per il Teorema B.2.9(i) e il solito teorema fondamentale,

$$\begin{aligned}
\mathrm{E}_\theta(\delta) &= \int_{\mathbb{T}}\mathrm{E}_\theta(\delta|\overline{\mathbf{X}} = t)\,(\mathrm{P}_\theta)_{\overline{\mathbf{X}}}(dt) = \int_{\mathbb{T}}\Big(\int_{\mathfrak{X}}\delta(\mathbf{x}')\,\mathrm{Q}^{(\overline{\mathbf{X}})}(d\mathbf{x}'|t)\Big)\,(\mathrm{P}_\theta)_{\overline{\mathbf{X}}}(dt) \\
&= \int_{\mathfrak{X}}\Big(\int_{\mathfrak{X}}\delta(\mathbf{x}')\,\mathrm{Q}^{(\overline{\mathbf{X}})}(d\mathbf{x}'|\overline{\mathbf{X}}(\mathbf{x}))\Big)\,\mathrm{P}_\theta(d\mathbf{x}) = \int_{\mathfrak{X}}\delta'(\mathbf{x})\,\mathrm{P}_\theta(d\mathbf{x}) = \mathrm{E}_\theta(\delta'),
\end{aligned}$$

basta verificare, per (2.2), che $\mathrm{Var}_\theta(\delta') < \mathrm{Var}_\theta(\delta)$. Ora, dato un qualsiasi numero reale $a > 0$, si ha

$$\begin{aligned}\mathrm{E}_\theta(a^{\overline{\mathbf{X}}}) &= \sum_{k\geq 0} a^{\frac{k}{2}}\, e^{-2\theta}\, \frac{(2\theta)^k}{k!} = e^{-2\theta} \sum_{k\geq 0} \frac{(2\theta)^k}{k!}\, a^{\frac{k}{2}} = e^{-2\theta} \sum_{k\geq 0} \frac{1}{k!} \Big(\frac{d^{(k)}}{dy}\, e^{2\theta y}\Big)(0)\ (a^{\frac{1}{2}})^k \\ &= e^{-2\theta}\, e^{2\theta\sqrt{a}} = e^{2\theta(\sqrt{a}-1)}\end{aligned}$$

e quindi

$$\begin{aligned}\mathrm{Var}_\theta(\delta') &= \mathrm{E}_\theta((\delta')^2) - \mathrm{E}_\theta(\delta')^2 = \mathrm{E}_\theta\Big(\big(\frac{1}{16}\big)^{\overline{\mathbf{X}}}\Big) - \mathrm{E}_\theta\Big(\big(\frac{1}{4}\big)^{\overline{\mathbf{X}}}\Big)^2 = e^{2\theta(\frac{1}{4}-1)} - \big(e^{2\theta(\frac{1}{2}-1)}\big)^2 \\ &= e^{-2\theta}\,(e^{\frac{\theta}{2}} - 1) = e^{-2\theta} \sum_{n\geq 1} \frac{1}{n!} \Big(\frac{\theta}{2}\Big)^n,\end{aligned}$$

cioè

$$\mathrm{Var}_\theta(\delta') = \frac{\theta e^{-2\theta}}{2} + \frac{e^{-2\theta}}{4} \sum_{n\geq 2} \frac{1}{n!}\, \frac{\theta^n}{2^{n-2}}. \tag{2.10}$$

Passando infine alla varianza di δ, per il Teorema B.1.12, risulta

$$\begin{aligned}\mathrm{E}_\theta(I_{\{0\}\times\mathbb{N}}) &= \mathrm{P}_\theta(\{0\}\times\mathbb{N}) = (\mathrm{P}_\theta)_{X_1}(0) = e^{-\theta} \\ \mathrm{E}_\theta(I_{\mathbb{N}\times\{0\}}) &= \mathrm{P}_\theta(\mathbb{N}\times\{0\}) = (\mathrm{P}_\theta)_{X_2}(0) = e^{-\theta}\end{aligned}$$

da cui otteniamo

$$\mathrm{E}_\theta(I_{\{0\}\times\mathbb{N}}\, I_{\mathbb{N}\times\{0\}}) = \mathrm{E}_\theta(I_{\{(0,0)\}}) = \mathrm{P}_\theta((0,0)) = e^{-2\theta} = \mathrm{E}_\theta(I_{\mathbb{N}\times\{0\}})\, \mathrm{E}_\theta(I_{\mathbb{N}\times\{0\}}),$$

cioè $\mathrm{Cov}_\theta(I_{\{0\}\times\mathbb{N}}, I_{\mathbb{N}\times\{0\}}) = 0$. Ne segue

$$\begin{aligned}\mathrm{Var}_\theta(\delta) &= \frac{1}{4}\Big[\mathrm{Var}_\theta(I_{\{0\}\times\mathbb{N}}) + \mathrm{Var}_\theta(I_{\mathbb{N}\times\{0\}})\Big] = \frac{1}{4}\left[e^{-\theta}(1-e^{-\theta}) + e^{-\theta}(1-e^{-\theta})\right] \\ &= \frac{1}{2}\, e^{-\theta}(1-e^{-\theta}) = \frac{\theta e^{-2\theta}}{2} + \frac{e^{-2\theta}}{2}\,(e^\theta - \theta - 1) \\ &> \frac{\theta e^{-2\theta}}{2} + \frac{e^{-2\theta}}{4}\,(e^\theta - \theta - 1) = \frac{\theta e^{-2\theta}}{2} + \frac{e^{-2\theta}}{4} \sum_{n\geq 2} \frac{1}{n!}\theta^n \\ &\geq \frac{\theta e^{-2\theta}}{2} + \frac{e^{-2\theta}}{4} \sum_{n\geq 2} \frac{1}{n!}\, \frac{\theta^n}{2^{n-2}}\end{aligned}$$

e quindi, per (2.10), $\mathrm{Var}_\theta(\delta') < \mathrm{Var}_\theta(\delta)$. ▲

2.2 Classi complete

Poichè riteniamo DM un individuo razionale, ci aspettiamo che escluda (dai suoi "comportamenti") quelle regole di decisione che constata essere strettamente dominate[15], pervenendo così ad un insieme “più piccolo” di possibili “comportamenti” (e quindi semplificando, sperabilmente, il problema decisionale)[16]. Siamo allora indotti a dare la seguente basilare definizione (dovuta ancora a Wald).

Definizione 2.2.1 Una **classe completa** (di regole di decisione) è ogni sottoinsieme $\Delta' \subseteq \Delta$ tale che esista, per ogni $\delta \notin \Delta'$, una regola di decisione $\delta' \in \Delta'$ che domini strettamente δ. Inoltre, una **classe completa propria** è ogni classe completa diversa da Δ.

Come l'esempio successivo rileva, di classi complete proprie ce ne possono essere una, più di una, oppure nessuna.

Esempio 2.2.2 Poichè consideriamo solo situazioni che contemplano un numero finito di stati e di decisioni, rappresentiamo la funzione di danno mediante una tabella nella quale $\mathrm{L}(\theta_i, d_j)$ è l'elemento corrispondente alla linea i ed alla colonna j. Supponiamo inoltre che lo spazio campionario sia costituito da un solo campione.

(i) Siano $\mathfrak{D} = \{d_1, d_2\}$, $\Theta = \{\theta_1, \theta_2\}$ e la funzione di danno tale che:

	d_1	d_2
θ_1	1001	1000
θ_2	0	1000

Allora, $\mathfrak{D}$ è l'unica classe completa.

(ii) Siano $\mathfrak{D} = \{d_1, d_2, d_3\}$, Θ come in (i) e la funzione di danno tale che:

	d_1	d_2	d_3
θ_1	2	2	1
θ_2	1	3	4

Allora, $\{d_1, d_3\}$ è l'unica classe completa propria.

(iii) Siano $\mathfrak{D} = \{d_0, d_1, d_2, \dots\}$ un insieme numerabile, Θ come in (i) e la funzione di danno tale che:

	d_0	d_1	$\dots$	d_n	$\dots$
θ_1	3	$2+\frac{1}{2}$	$\dots$	$2+\frac{1}{2^n}$	$\dots$
θ_2	0	$2+\frac{1}{3}$	$\dots$	$2+\frac{1}{3^n}$	$\dots$

[15]Questo dal punto di vista teorico, ove si tiene conto solamente delle funzioni di rischio. Da un punto di vista pratico invece può talvolta succedere che DM ritenga più opportuno scegliere una regola di decisione particolarmente semplice e facile da utlizzare ad una più complicata che, pur dominandola strettamente, comporti solamente una trascurabile diminuzione dei danni medi.

[16]E quindi, nell'Esempio 2.1.5(iii), conserverà degli stimatori δ_c, nel caso $k > \frac{1}{\sqrt{2\pi\sigma^2}}$, solo lo stimatore puntuale δ_0 (cioè la funzione identica), mentre, nel caso opposto, solo lo stimatore intervallare, δ_{c^*}.

Allora, le classi complete proprie sono i sottoinsiemi del tipo $\mathfrak{D}' = \{d_0, d_{k_1}, d_{k_2}, \dots\}$ con $(k_n)_{n\geq 1}$ successione crescente. ▲

Per quanto detto, DM sarà portato a scegliere una classe completa propria (qualora esistente) e quindi, nel caso ce ne fosse più di una, ad individuarne una tramite un opportuno criterio di scelta. Criterio che, nel caso ne esista una minima (nel senso dell'inclusione), appare ovvio: selezionare quella minima poichè consente la massima semplificazione del problema decisionale (dal punto di vista delle funzioni di rischio). Al fine di individuarla conviene dare la definizione seguente.

Definizione 2.2.3 Una **regola (di decisione) ammissibile** è una qualsiasi regola di decisione che sia un elemento massimale per la relazione di dominanza, cioè tale che non esista alcuna regola di decisione che la domini strettamente. Inoltre, la **classe ammissibile** è l'insieme Δ^+ delle regole ammissibili.

Mostriamo ora che la classe ammissibile può essere vuota, contenere un solo elemento oppure coincidere con Δ e che l'ammissibilità, se da una parte è senza dubbio una proprietà desiderabile per ogni regola di decisione, dall'altra non ne assicura, in generale, la "ragionevolezza".

Esempio 2.2.4 (i) Con riferimento all'Esempio 2.2.2 si ha: $\mathfrak{D}^+ = D$ in (i); $\mathfrak{D}^+ = \{d_1, d_3\}$ in (ii); $\mathfrak{D}^+ = \{d_0\}$ in (iii). Inoltre, nell'Esempio 2.1.5(ii), $\Delta^+ = \emptyset$.

(ii) Posto $\Theta = \mathfrak{D} = [0,1]$, sia l'osservabile costituito dalle v.a. $X_1, \dots, X_n$ indipendenti e distribuite secondo la distribuzione di Bernoulli $Ber(\theta)$ in corrispondenza ad ogni stato θ. Conseguentemente, per le probabilità di campionamento, si ha

$$\mathrm{P}_\theta(\mathbf{x}) = \theta^{n\bar{\mathbf{x}}}\,(1-\theta)^{n(1-\bar{\mathbf{x}})}$$

per ogni $\mathbf{x} \in \mathfrak{X} = \{0,1\}^n$. Per quanto riguarda la funzione di danno, supponiamo $\mathrm{L}(\theta, d) > 0$, se $\theta \neq d$, e $\mathrm{L}(\theta, d) = 0$, se $\theta = d$.

Completato lo schema decisionale-inferenziale, proviamo che sono regole ammissibili tutte le costanti con valori in $]0,1[$. Indicata con $\delta_{\tilde{\theta}}$ la costante di valore $\tilde{\theta}$, basta evidentemente verificare che $\delta \succeq \delta_{\tilde{\theta}} \Rightarrow \delta = \delta_{\tilde{\theta}}$. Supposto quindi $R_\delta \leq R_{\delta_{\tilde{\theta}}} = \mathrm{L}_{\tilde{\theta}}$, otteniamo

$$0 = \mathrm{L}(\tilde{\theta}, \tilde{\theta}) = \mathrm{L}_{\tilde{\theta}}(\tilde{\theta}) \geq R_\delta(\tilde{\theta}) = \sum_{\mathbf{x}\in\mathfrak{X}} \mathrm{L}(\tilde{\theta}, \delta(\mathbf{x}))\,\tilde{\theta}^{n\bar{\mathbf{x}}}\,(1-\tilde{\theta})^{n(1-\bar{\mathbf{x}})} \geq 0$$

e dunque

$$\sum_{\mathbf{x}\in\mathfrak{X}} \mathrm{L}(\tilde{\theta}, \delta(\mathbf{x}))\,\tilde{\theta}^{n\bar{\mathbf{x}}}\,(1-\tilde{\theta})^{n(1-\bar{\mathbf{x}})} = 0.$$

Qualunque sia il campione $\mathbf{x}$, si ha quindi $\mathrm{L}(\tilde{\theta}, \delta(\mathbf{x})) = 0$ ($\tilde{\theta} \neq 0, 1!$). Poichè $\mathrm{L}(\tilde{\theta}, d) > 0$ per ogni $d \neq \tilde{\theta}$, possiamo concludere che $\delta(\mathbf{x}) = \tilde{\theta}$ per ogni campione $\mathbf{x}$.

Conseguentemente, la scelta di stimare pari a $\tilde{\theta} \neq 0, 1$ lo stato di natura *indipendentemente* dal risultato sperimentale (perseguendo così un "comportamento" da ritenersi senza alcun dubbio non ragionevole volendo accedere all'esperimento) non può essere in alcun modo migliorata, in termini di dominanza, da un "comportamento" più razionale basato sull'informazione campionaria. ▲

Il prossimo risultato evidenzia il legame intercorrente tra le nozioni di ammissibilità (di una regola di decisione) e di completezza (di un insieme di regole di decisione).

Teorema 2.2.5 *La classe ammissibile è l'intersezione di tutte le classi complete.*

Dimostrazione Indicata con Δ_1 l'intersezione di tutte le classi complete, verifichiamo intanto che riesce $\Delta^+ \subseteq \Delta_1$. Per assurdo, sia $\delta \in \Delta^+ \backslash \Delta_1$. Esiste allora una classe completa Δ' tale che $\delta \notin \Delta'$. Si può quindi trovare $\delta' \in \Delta'$ tale che $\delta' \succ \delta$. Ne segue $\delta \notin \Delta^+$ (Contraddizione!). Per provare l'inclusione opposta, sia (per assurdo) $\delta_1 \in \Delta_1 \setminus \Delta^+$. Esiste quindi δ'' tale che $\delta'' \succ \delta_1$. Considerato allora l'insieme $\Delta' = \Delta \setminus \{\delta_1\}$, riesce $\delta'' \in \Delta'$; inoltre, presa una qualunque regola di decisione $\delta \notin \Delta'$, si ha $\delta = \delta_1$ e quindi $\delta'' \succ \delta$. Conseguentemente, Δ' è una classe completa tale che $\delta_1 \notin \Delta'$ (Contraddizione!).■

Nello schema decisionale considerato nell'Esempio 2.2.2(iii), la classe ammissibile $\{d_0\}$ non è una classe completa. D'altra parte, qualora Δ^+ fosse completa, sarebbe, per il teorema precedente, la più piccola classe completa (nel senso dell'inclusione). Il risultato successivo asssicura che, nel caso di completezza, la classe ammissibile è l'unica classe completa **minimale** (cioè, tale che ogni suo sottoinsieme proprio non è una classe completa).

Teorema 2.2.6 *Se Δ^+ è una classe completa, allora è l'unica classe completa minimale.*

Dimostrazione Sia Δ^+ una classe completa. Essendo ovvia la sua minimalità, proviamone l'unicità. Supponiamo quindi che Δ' sia una classe completa minimale. Poichè $\Delta^+ \subseteq \Delta'$ (Teorema 2.2.5), esista (per assurdo) $\delta' \in \Delta' \backslash \Delta^+$. Allora, $\delta'' \succ \delta'$ per qualche regola di decisione δ''.

Proviamo ora che δ' è dominata strettamente da qualche elemento di $\Delta'' = \Delta' \setminus \{\delta'\}$. Tenuto conto che $\delta'' \succ \delta'$, basta considerare il caso $\delta'' \notin \Delta'$.

Poichè Δ' è una classe completa, esiste $\delta_1 \in \Delta'$ tale che $\delta_1 \succ \delta''$. Allora, $\delta_1 \succ \delta'$ e quindi $\delta_1 \neq \delta'$. Dunque, δ' è dominata strettamente da $\delta_1 \in \Delta''$.

Al fine di verificare che Δ'' è una classe completa, sia $\delta \notin \Delta''$. Ne segue $\delta \notin \Delta'$ oppure $\delta = \delta'$; quindi, per quanto provato, possiamo limitarci al caso $\delta \notin \Delta'$. Poichè Δ' è una classe completa, esiste $\delta_2 \in \Delta'$ tale che $\delta_2 \succ \delta$. Ora, se $\delta_2 \neq \delta'$, si ha $\delta_2 \in \Delta''$; se invece $\delta_2 = \delta'$, per quanto provato, δ_2 - e quindi δ - è dominata strettamente da un elemento di Δ''. Pertanto, Δ'' è una classe completa, contraddicendo così la minimalità di Δ' ($\Delta'' \subset \Delta'$!). ■

Il primo e il terzo dei teoremi seguenti riguardano la completezza della classe ammissibile (e quindi l'esistenza della classe completa minimale). Il primo rileva che, nel caso di un numero finito di regole di decisione, la classe ammissibile è sempre completa; il terzo invece fornisce una condizione "geometrica" per la completezza, nel caso di un numero finito di stati[17].

Teorema 2.2.7 *Se Δ è finito, allora Δ^+ è una classe completa.*

DIMOSTRAZIONE Supposto Δ finito, sia (per assurdo) Δ^+ non completa. Esiste allora $\delta_1 \notin \Delta^+$ non dominata strettamente da regole ammissibili. Poichè δ_1 non è ammissibile, esiste δ_2 tale che $\delta_2 \succ \delta_1$ (e quindi $\delta_2 \neq \delta_1$). Osservato che $\delta_2 \notin \Delta^+$, esiste δ_3 tale che $\delta_3 \succ \delta_2$ (e quindi $\delta_3 \neq \delta_2$). Ne segue $\delta_3 \succ \delta_1$ (e quindi $\delta_3 \neq \delta_1$). Allora $\delta_3 \notin \Delta^+$ e quindi esiste δ_4 tale che etc.. Così procedendo otteniamo una successione $(\delta_n)_{n\geq 1}$ di regole di decisione tutte tra loro distinte, contraddicendo l'ipotesi di finitezza. ■

Passando al caso in cui gli stati sono finiti, introduciamo la nozione di insieme di rischio che fornisce la base "geometrica" di riferimento per ottenere sia una caratterizzazione dell'ammissibilità che una condizione sufficiente per la completezza della classe ammissibile.

Definizione 2.2.8 Supposto $\Theta = \{\theta_1, \ldots, \theta_m\}$, l'**insieme di rischio** è l'insieme $\mathrm{R} = \{(R_\delta(\theta_1), \ldots, R_\delta(\theta_m)) : \delta \in \Delta\} \subseteq [0, +\infty[^m$ dei **vettori di rischio** $\mathbf{y}^{(\delta)} = (R_\delta(\theta_1), \ldots, R_\delta(\theta_m))$[18].

Denotata, per ogni sottoinsieme non vuoto $B \subseteq \mathbb{R}^m$, con $\overline{B}$ la chiusura topologica di B e considerata la **chiusura inferiore**:

$$\partial_L B = \{\mathbf{y} \in \overline{B} : \neg\, \exists \mathbf{b} \in B\, (\mathbf{b} < \mathbf{y})\}$$

[17] Dispensando così DM dal ricercare un criterio di selezione che individui una classe completa che sia, ai suoi fini, "ragionevolmente piccola".

[18] Certamente non vuoto in quanto $(\mathrm{L}_d(\theta_1), \ldots, \mathrm{L}_d(\theta_m)) \in \mathrm{R}$ per ogni decisione d.

di B[19], otteniamo il risultato seguente che fornisce una caratterizzazione "geometrica" delle regole ammissibili.

Teorema 2.2.9 *Sia Θ finito. Risulta allora:*

(i) $\delta \in \Delta^+ \iff \mathbf{y}^{(\delta)} \in \partial_L \mathrm{R}$*;*

(ii) $\Delta^+ \neq \emptyset \iff \partial_L \mathrm{R} \cap \mathrm{R} \neq \emptyset$*.*

Dimostrazione Poichè (ii) discende banalmente da (i), proviamo la proposizione (i). Sia intanto $\delta \in \Delta^+$. Supposto (per assurdo) $\mathbf{y}^{(\delta)} \notin \partial_L \mathrm{R}$, esiste $\mathbf{y} \in \mathrm{R}$ tale che $\mathbf{y} < \mathbf{y}^{(\delta)}$. Allora, $\mathbf{y} = \mathbf{y}^{(\delta')}$ per qualche regola di decisione δ'. Ne segue $\mathbf{y}^{(\delta')} < \mathbf{y}^{(\delta)}$, cioè $\delta' \succ \delta$ (Contraddizione!). Sia ora $\mathbf{y}^{(\delta)} \in \partial_L \mathrm{R}$. Supposto (per assurdo) $\delta \notin \Delta^+$, esiste una regola di decisione δ' tale che $\delta' \succ \delta$. Allora $\mathbf{y}^{(\delta')} < \mathbf{y}^{(\delta)}$ e quindi $\mathbf{y}^{(\delta)} \notin \partial_L \mathrm{R}$ (Contraddizione!). ■

Il lemma successivo consentirà di provare (nel caso di un numero finito di stati) che le regole ammissibili formano una classe completa ogniqualvolta l'insieme di rischio include la sua chiusura inferiore.

Lemma 2.2.10 *La chiusura inferiore di un insieme non vuoto inferiormente limitato di m-ple reali non è vuota.*

Dimostrazione Sia $B \subset \mathbb{R}^m$ un insieme non vuoto inferiormente limitato. Esiste allora un numero reale α tale che $b_i \geq \alpha$ $(i = 1, \ldots, m)$ per ogni $\mathbf{b} \in B$. Considerata la funzione $f : \mathbf{y} \mapsto y_1 + \cdots + y_m$, l'insieme-immagine $f(B)$ è banalmente un sottoinsieme non vuoto inferiormente limitato della retta reale. Conseguentemente, $\beta = \inf f(B)$ è un numero reale.

Sia ora $(\mathbf{b}^{(n)})_{n \geq 1}$ una successione in B tale che $f(\mathbf{b}^{(n)}) \to \beta$ come $n \to +\infty$. Allora, la successione è superiormente limitata. Infatti, supposto (per assurdo) che per ogni n esistano due numeri naturali h_n, k_n tali che $b_{k_n}^{(h_n)} > n$, otteniamo $\lim\limits_{n \to +\infty} b_{k_n}^{(h_n)} = +\infty$. Allora, dalla

$$f(\mathbf{b}^{(h_n)}) = \sum_{i=1}^{k_n - 1} b_i^{(h_n)} + b_{k_n}^{(h_n)} + \sum_{i=k_n+1}^{m} b_i^{(h_n)} \geq (m-1)\alpha + b_{k_n}^{(h_n)}$$

[19]Riesce quindi $\mathbf{y}' \in \partial_L B \iff \overline{B} \cap \{\mathbf{y} : \mathbf{y} \leq \mathbf{y}'\} = \{\mathbf{y}'\}$ per ogni m-pla $\mathbf{y}'$. Ciò osservato, diamo alcuni esempi di chiusure inferiori nel caso $m = 2$. Posto $B =]-\infty, 0]^2$ si ha $\partial_L B = \emptyset$; $B =]0,1[^2$ riesce $\partial_L \overline{B} = \{\mathbf{0}\} \subset \{(a,0) : 0 \leq a \leq 1\} \cup \{(0,b) : 0 \leq b \leq 1\} = \partial_L B$; $B = \{(3,0), (\frac{5}{2}, \frac{7}{3}), \ldots, (2 + \frac{1}{2^n}, 2 + \frac{1}{3^n}), \ldots\}$, risulta $\partial_L B = \partial_L \overline{B} = \{(3,0), (2,2)\}$.

si ha $\beta = \lim_{n\to+\infty} f(\mathbf{b}^{(h_n)}) \geq \lim_{n\to+\infty}[(m-1)\alpha + b_{k_n}^{(h_n)}] = +\infty$ (Contraddizione!).

Poichè la successione è anche inferiormente limitata (essendo in B), esiste una sottosuccessione $(\mathbf{b}^{(i_n)})_{n\geq 1}$ convergente a qualche m-pla $\mathbf{y}'$. Chiaramente $\mathbf{y}' \in \overline{B}$; inoltre $f(\mathbf{y}') = \lim_{n\to+\infty} f(\mathbf{b}^{(i_n)}) = \lim_{n\to+\infty} f(\mathbf{b}^{(n)}) = \beta$ (f è una funzione continua!). Ne segue $\mathbf{y}' \in \partial_L B$. Infatti, esista (per assurdo) $\mathbf{b} \in B$ tale che $\mathbf{b} < \mathbf{y}'$. Allora, $b_i \leq y_i'$ $(i = 1, \ldots, m)$ e $b_j < y_j'$ per qualche j. Riesce pertanto $\inf f(B) = \beta = f(\mathbf{y}') = \sum_{i=1}^m y_i' > \sum_{i=1}^m b_i = f(\mathbf{b}) \in f(B)$ (Contraddizione!). Dunque $\partial_L B \neq \emptyset$. ■

Teorema 2.2.11 *Supposto Θ finito, sia $\partial_L \mathrm{R} \subseteq \mathrm{R}$. Allora, Δ^+ è una classe completa.*

Dimostrazione Supposto $\Theta = \{\theta_1, \ldots, \theta_m\}$ e $\partial_L \mathrm{R} \subseteq \mathrm{R}$, sia $\delta' \notin \Delta^+$. Esiste quindi una regola di decisione δ tale che $\delta \succ \delta'$. Se δ è una regola ammissibile, allora δ' risulta dominata strettamente da un elemento di Δ^+. Assumiamo pertanto $\delta \notin \Delta^+$. Allora, per il Teorema 2.2.9(i), $\mathbf{y}^{(\delta)} \notin \partial_L \mathrm{R}$. Considerato l'insieme non vuoto:

$$B = \{\mathbf{y} \in \mathrm{R} : \mathbf{y} \leq \mathbf{y}^{(\delta)}\} \subseteq [0, +\infty[^m,$$

esiste, per il lemma precedente, $\mathbf{y} \in \partial_L B$. Allora $\mathbf{y} \in \overline{B} \subseteq \{\mathbf{y} : \mathbf{y} \leq \mathbf{y}^{(\delta)}\} \cap \overline{\mathrm{R}}$. Al fine di provare che $\mathbf{y} \in \partial_L \mathrm{R}$, assumiamo (per assurdo) che esista $\mathbf{b} \in \mathrm{R}$ tale che $\mathbf{b} < \mathbf{y}$. Allora, $\mathbf{b} < \mathbf{y}^{(\delta)}$ da cui otteniamo $\mathbf{b} \in B$ e quindi $\mathbf{y} \notin \partial_L B$ (Contraddizione!). Pertanto $\mathbf{y} \in \partial_L \mathrm{R}$. Per l'ipotesi $\partial_L \mathrm{R} \subseteq \mathrm{R}$, esiste infine una regola di decisione δ'' tale che $\mathbf{y}^{(\delta'')} = \mathbf{y} \in \partial_L \mathrm{R}$. Conseguentemente, per il Teorema 2.2.9(i), $\delta'' \in \Delta^+$; inoltre, essendo $\mathbf{y}^{(\delta'')} = \mathbf{y} \leq \mathbf{y}^{(\delta)}$ e $\mathbf{y}^{(\delta)} < \mathbf{y}^{(\delta')}$ ($\delta \succ \delta'$!), si ha $\delta'' \succ \delta'$. ■

2.3 Classi essenzialmente complete

Una semplificazione del problema decisionale può avvenire anche in assenza di una classe completa propria. Infatti, considerato il problema decisionale (senza acquisizione di dati): $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mathfrak{D} = \{d_1, d_2, d_3\}$ e funzione di danno data dalla tabella:

	d_1	d_2	d_3
θ_1	1	0	0
θ_2	2	4	4
θ_3	2	3	3

,

possiamo ridurre, pur essendo $\mathfrak{D}$ l'unica classe completa, l'insieme delle regole di decisione procedendo per esclusione di decisioni equivalenti; precisamente, eliminando la decisione d_3 e sostituendo $\mathfrak{D}$ con l'insieme $\{d_1, d_2\}$, oppure eliminando la decisione d_2 e rimpiazzando $\mathfrak{D}$ con l'insieme $\{d_1, d_3\}$. Queste considerazioni suggeriscono la seguente definizione (introdotta da Erich L. Lehmann nel 1947).

Definizione 2.3.1 Una **classe essenzialmente completa** (di regole di decisione) è ogni sottoinsieme $\Delta' \subseteq \Delta$ tale che esista, per ogni $\delta \notin \Delta'$, una regola di decisione $\delta' \in \Delta'$ che domini δ.

Ovviamente, ogni classe completa è essenzialmente completa. Inoltre, sussiste il risultato seguente (di verifica immediata) che fornisce la relazione intercorrente tra le regole ammissibili e le classi essenzialmente complete.

Teorema 2.3.2 *Sia Δ' una classe essenzialmente completa. Se una regola ammissibile δ non appartiene a Δ', allora esiste una regola ammissibile $\delta' \in \Delta'$ tale che $\delta' \sim \delta$.*

L'introduzione delle classi essenzialmente complete consente di formulare uno degli obiettivi centrali della teoria delle decisioni statistiche: l'individuazione di classi complete o almeno essenzialmente complete. In questa direzione sono inquadrabili sia teoremi di carattere generale (come il teorema della classe completa che verrà considerato nella sezione quinta) che teoremi riguardanti specifici problemi statistico-decisionali. Nell'ambito dei problemi di stima puntuale, abbiamo già incontrato uno di questi teoremi: il teorema di Rao-Blackwell 2.1.6 che consente, qualora si adotti il danno quadratico o quello assoluto, di costruire classi essenzialmente complete ricorrendo a statistiche sufficienti. Per quanto riguarda invece i problemi di verifica d'ipotesi, un ruolo chiave viene svolto dal prossimo lemma (provato da Jerzy Neyman e Egon S. Pearson[20] nel 1933) che fornisce una classe essenzialmente completa nel caso particolare di ipotesi semplici: $H_0 : Z = \theta_0$ e $H_1 : Z = \theta_1$. Osserviamo che, in questo contesto, $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$ e, per (2.4), $\delta \succeq \delta'$ se e solo se $\beta_\delta(\theta_0) \leq \beta_{\delta'}(\theta_0)$ e $\beta_\delta(\theta_1) \geq \beta_{\delta'}(\theta_1)$. Possiamo inoltre assumere, senza perdere in generalità, che le due densità di campionamento f_{θ_0} e f_{θ_1} siano a valori finiti.

[20] Figlio dello statistico, matematico e biologo Karl Pearson, fondatore della biometria e autore di risultati fondamentali sul metodo dei momenti e sulla distribuzione del test Chi-quadrato.

Lemma 2.3.3 (di Neyman-Pearson) *Per ogni $t \in [0, +\infty]$, sia $\delta^{(t)}$ il test avente come regione di rifiuto l'insieme:*

$$\mathfrak{X}_1^{(t)} = \begin{cases} \{f_{\theta_1} > t f_{\theta_0}\} & \textit{se } t \neq +\infty \\ \{f_{\theta_0} = 0\} & \textit{se } t = +\infty \end{cases}.$$

Riesce allora:

(i) $\beta_{\delta^{(t)}}(\theta_0) \geq \beta_\delta(\theta_0) \Rightarrow \beta_{\delta^{(t)}}(\theta_1) \geq \beta_\delta(\theta_1)$ *per ogni test* δ *e* $t \in [0, +\infty]$.

Inoltre, supposto $\mathrm{P}_{\theta_0}(f_{\theta_1} = t\, f_{\theta_0}) = 0$ *per ogni numero reale* $t \geq 0$*, sussistono le proposizioni:*

(ii) Per ogni $\vartheta \in [0, 1]$ *esiste* $t^* \in [0, +\infty]$ *tale che* $\beta_{\delta^{(t^*)}}(\theta_0) = \vartheta$*;*

(iii) La classe dei **test di Neyman-Pearson** $\Delta^{(NP)} = \{\delta^{(t)} : 0 \leq k \leq +\infty\}$ *è una classe essenzialmente completa costituita da test ammissibili.*

Dimostrazione Posto, per semplicità, $f_i = f_{\theta_i}$, $\mathrm{P}_i = \mathrm{P}_{\theta_i}$ $(i = 0, 1)$ e indicata con β_t la funzione potenza del test $\delta^{(t)}$, proviamo che risulta

$$\beta_t(\theta_1) - \beta_\delta(\theta_1) \geq t\,[\beta_t(\theta_0) - \beta_\delta(\theta_0)] \qquad (t \neq +\infty) \tag{2.11}$$

per ogni test δ. A tal fine, posto $\varphi = I_{\mathfrak{X}_1^{(\delta)}}$ e $\varphi_t = I_{\mathfrak{X}_1^{(t)}}$, consideriamo la funzione $g = (\varphi_t - \varphi)(f_1 - t\, f_0)$. Allora,

$$g(x) = \begin{cases} (1 - \varphi(x))(f_1(x) - t\, f_0(x)) \geq 0 & \text{se } f_1(x) > t\, f_0(x) \\ -\varphi(x)(f_1(x) - t\, f_0(x)) \geq 0 & \text{se } f_1(x) < t\, f_0(x) \\ 0 & \text{se } f_1(x) = t\, f_0(x) \end{cases}$$

da cui otteniamo

$$\begin{aligned} 0 &\leq \int_{\mathfrak{X}} g\, d\mu = \int_{\mathfrak{X}} \varphi_t f_1\, d\mu - \int_{\mathfrak{X}} \varphi f_1\, d\mu - t \int_{\mathfrak{X}} \varphi_t f_0\, d\mu + t \int_{\mathfrak{X}} \varphi f_0\, d\mu \\ &= \int_{\mathfrak{X}_1^{(t)}} f_1\, d\mu - \int_{\mathfrak{X}_1^{(\delta)}} f_1\, d\mu - t\Big[\int_{\mathfrak{X}_1^{(t)}} f_0\, d\mu - \int_{\mathfrak{X}_1^{(\delta)}} f_0\, d\mu\Big] \\ &= \beta_t(\theta_1) - \beta_\delta(\theta_1) - t\,[\beta_t(\theta_0) - \beta_\delta(\theta_0)] \end{aligned}$$

e quindi (2.11).

(i) Supposto $\beta_t(\theta_0) \geq \beta_\delta(\theta_0)$, sia $t \neq +\infty$. Allora, per (2.11), $\beta_t(\theta_1) - \beta_\delta(\theta_1) \geq 0$, cioè $\beta_t(\theta_1) \geq \beta_\delta(\theta_1)$. Sia ora $t = +\infty$. Ne segue $\beta_\delta(\theta_0) \leq \beta_t(\theta_0) = \int_{\{f_0=0\}} f_0\, d\mu = 0$ e quindi, posto $\widetilde{\mathfrak{X}} = \{f_0 > 0\}$,

$$0 = \beta_\delta(\theta_0) = \int_{\mathfrak{X}_1^{(\delta)}} f_0\, d\mu = \int_{\mathfrak{X}_1^{(\delta)} \cap \widetilde{\mathfrak{X}}} f_0\, d\mu = \int_{\mathfrak{X}} \varphi I_{\widetilde{\mathfrak{X}}} f_0\, d\mu.$$

Allora, per il Teorema A.3.6(v), $\varphi I_{\widetilde{\mathfrak{X}}} f_0 = 0$ (μ-q.o.), cioè $\varphi I_{\widetilde{\mathfrak{X}}} = 0$ (μ-q.o.) e quindi $\varphi I_{\widetilde{\mathfrak{X}}} f_1 = 0$ (μ-q.o.). Risulta dunque $\int_{\widetilde{\mathfrak{X}}} \varphi f_1\, d\mu = 0$ da cui otteniamo

$$\begin{aligned}
\beta_t(\theta_1) - \beta_\delta(\theta_1) &= \int_{\mathfrak{X}_1^{(t)}} f_1\, d\mu - \int_{\mathfrak{X}_1^{(\delta)}} f_1\, d\mu = \int_{\mathfrak{X}} (\varphi_t - \varphi) f_1\, d\mu \\
&= \int_{\widetilde{\mathfrak{X}}} (\varphi_t - \varphi) f_1\, d\mu + \int_{\widetilde{\mathfrak{X}}^c} (\varphi_t - \varphi) f_1\, d\mu \\
&= -\int_{\widetilde{\mathfrak{X}}} \varphi f_1\, d\mu + \int_{\widetilde{\mathfrak{X}}^c} (1 - \varphi) f_1\, d\mu = \int_{\widetilde{\mathfrak{X}}^c} (1 - \varphi) f_1\, d\mu \geq 0,
\end{aligned}$$

cioè $\beta_t(\theta_1) \geq \beta_\delta(\theta_1)$.

Assumiamo ora $\mathrm{P}_0(f_1 = t\, f_0) = 0$ per ogni numero reale $t \geq 0$.

(ii) Se $\vartheta = 0$, allora $\beta_{+\infty}(\theta_0) = \vartheta$ e quindi $t^* = +\infty$. Sia pertanto $0 < \vartheta \leq 1$. Considerata la v.a. Y su $\mathfrak{X}$ così definita:

$$Y(x) = \begin{cases} \frac{f_1(x)}{f_0(x)} & \text{se } x \in \widetilde{\mathfrak{X}} \\ 0 & \text{se } x \notin \widetilde{\mathfrak{X}}, \end{cases}$$

proviamo che la funzione di ripartizione $F_0(t) = \mathrm{P}_0(Y \leq t)$ è una funzione continua. A tal fine, per il Teorema B.1.1(vi), basta verificare che, dato $t \geq 0$, si ha $\mathrm{P}_0(Y = t) = 0$. Dalla $\mathrm{P}_0(\widetilde{\mathfrak{X}}^c) = \mathrm{P}_0(f_0 = 0) = 0$ otteniamo

$$\begin{aligned}
\mathrm{P}_0(Y = t) &= \mathrm{P}_0\big(\{Y = t\} \cap \widetilde{\mathfrak{X}}\big) + \mathrm{P}_0\big(\{Y = t\} \cap \widetilde{\mathfrak{X}}^c\big) = \mathrm{P}_0\big(\{Y = t\} \cap \widetilde{\mathfrak{X}}\big) \\
&= \mathrm{P}_0\Big(\Big\{\frac{f_1}{f_0} = t\Big\} \cap \widetilde{\mathfrak{X}}\Big) = \mathrm{P}_0\big(\{f_1 = t\, f_0\} \cap \widetilde{\mathfrak{X}}\big) \leq \mathrm{P}_0(f_1 = t\, f_0) = 0.
\end{aligned}$$

Poichè $0 \leq 1 - \vartheta < 1$, $\lim_{t \to +\infty} F_0(t) = 1$ e F_0 è una funzione continua, esiste $t^* \geq 0$ tale che $F_0(t^*) = 1 - \vartheta$. Riesce allora

$$\begin{aligned}
\beta_{t^*}(\theta_0) &= \mathrm{P}_0(f_1 > t^* f_0) = \mathrm{P}_0\big(\{f_1 > t^* f_0\} \cap \widetilde{\mathfrak{X}}\big) = \mathrm{P}_0\Big(\Big\{\frac{f_1}{f_0} > t^*\Big\} \cap \widetilde{\mathfrak{X}}\Big) \\
&= \mathrm{P}_0\big(\{Y > t^*\} \cap \widetilde{\mathfrak{X}}\big) = \mathrm{P}_0(Y > t^*) = 1 - \mathrm{P}_0(Y \leq t^*) = F_0(t^*) = \vartheta.
\end{aligned}$$

(iii) Dato un test δ, esiste (ponendo $\vartheta = \beta_\delta(\theta_0)$ in (ii)) $t^* \in [0, +\infty]$ tale che $\beta_{t^*}(\theta_0) = \beta_\delta(\theta_0)$. Da (i) otteniamo allora $\beta_{t^*}(\theta_1) \geq \beta_\delta(\theta_1)$ e quindi $\delta^{(t^*)} \succeq \delta$. Rimane da verificare che i test di Neyman-Pearson sono ammissibili. A tal fine, basta constatare che, dati i test $\delta^{(t')}$ e $\delta^{(t'')}$, non può essere $\delta^{(t')} \succ \delta^{(t'')}$. Supposto (per assurdo) $\delta^{(t')} \succ \delta^{(t'')}$, otteniamo, in particolare, $\delta^{(t')} \succeq \delta^{(t'')}$ e quindi $\beta_{t'}(\theta_0) \leq \beta_{t''}(\theta_0)$, $\beta_{t'}(\theta_1) \geq \beta_{t''}(\theta_1)$. Allora, per (i), $\beta_{t'}(\theta_1) \leq \beta_{t''}(\theta_1)$, cioè $\beta_{t'}(\theta_1) = \beta_{t''}(\theta_1)$. Riesce dunque $\beta_{t'}(\theta_0) < \beta_{t''}(\theta_0)$. Scelto ϑ tale che $\beta_{t'}(\theta_0) < \vartheta < \beta_{t''}(\theta_0)$, esiste, per (ii), un test $\delta^{(\tilde{t})}$ con $\beta_{\tilde{t}}(\theta_0) = \vartheta$ e quindi $\beta_{t'}(\theta_0) < \beta_{\tilde{t}}(\theta_0) < \beta_{t''}(\theta_0)$. Allora, $\tilde{t} \neq t''$ e $\tilde{t}, t'' \neq +\infty$, ricordato che $\beta_{+\infty}(\theta_0) = 0$. Infine, per (i), si ha anche $\beta_{\tilde{t}}(\theta_1) = \beta_{t'}(\theta_1)$. Ponendo allora, in (2.11), $\delta = \delta^{(t')}$ e $t = \tilde{t}$, se $\tilde{t} > 0$, e $t = t''$, se $t'' > 0$, otteniamo $0 = \beta_t(\theta_1) - \beta_{t'}(\theta_1) \geq t\,[\beta_t(\theta_0) - \beta_{t'}(\theta_0)] > 0$ (Contraddizione!). ■

Osservazione 2.3.4 Qualora non sussista l'ipotesi "$\mathrm{P}_{\theta_0}(f_{\theta_1} = t\, f_{\theta_0}) = 0$ per ogni numero reale $t \geq 0$", può avvenire che la classe $\Delta^{(NP)}$ non sia essenzialmente completa. Supponiamo infatti che lo spazio campionario sia costituito dagli elementi $\{x_1, x_2\}$ e che le probabilità di campionamento siano fornite dalla tabella seguente, nella quale riportiamo anche i valori del rapporto $\frac{\mathrm{P}_{\theta_1}}{\mathrm{P}_{\theta_0}}$.

	x_1	x_2
P_{θ_0}	$\frac{1}{4}$	$\frac{3}{4}$
P_{θ_1}	$\frac{1}{8}$	$\frac{7}{8}$
$\frac{\mathrm{P}_{\theta_1}}{\mathrm{P}_{\theta_0}}$	$\frac{1}{2}$	$\frac{7}{6}$

In questa situazione, per il test δ avente regione di rifiuto $\mathfrak{X}_1^{(\delta)} = \{x_1\}$ risulta $\beta_\delta(\theta_0) = \frac{1}{4}$ e $\beta_\delta(\theta_1) = \frac{1}{8}$. Inoltre, si ha

$$\mathfrak{X}_1^{(t)} = \begin{cases} \mathfrak{X} & \text{se } 0 \leq t < \frac{1}{2} \\ \{x_2\} & \text{se } \frac{1}{2} \leq t < \frac{7}{6} \\ \emptyset & \text{se } t \geq \frac{7}{6} \end{cases}$$

da cui otteniamo che la funzione potenza $\beta_{\delta^{(t)}}$ assume i valori riportati nella tabella:

	$0 \leq t < \frac{1}{2}$	$\frac{1}{2} \leq t < \frac{7}{6}$	$t \geq \frac{7}{6}$
θ_0	1	$\frac{3}{4}$	0
θ_1	1	$\frac{7}{8}$	0

Conseguentemente, nessun test Neyman-Pearson domina δ e quindi $\Delta^{(NP)}$ non è una classe essenzialmente completa. ♦

Nell'esempio seguente, con riferimento al campionamento normale con media incognita, determiniamo la struttura dei test Neyman-Pearson e verifichiamo che costituiscono una classe essenzialmente completa.

Esempio 2.3.5 Considerati due numeri reali θ_0, θ_1 con $\theta_0 < \theta_1$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta_i, \sigma^2)$ in corrispondenza allo stato θ_i ($i = 0, 1$). Allora, tramite (2.5), per le densità di campionamento si ha

$$f_{\theta_i}(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2}{2\sigma^2}\Big] \exp\Big[-\frac{n(\bar{\mathbf{x}} - \theta_i)^2}{2\sigma^2}\Big] \quad (i = 0, 1)$$

per ogni $\mathbf{x} \in \mathfrak{X} = \mathbb{R}^n$. Ne segue $\mathfrak{X}_1^{(+\infty)} = \emptyset$ e

$$\frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} = \exp\Big[-\frac{n(\bar{\mathbf{x}} - \theta_1)^2}{2\sigma^2} + \frac{n(\bar{\mathbf{x}} - \theta_0)^2}{2\sigma^2}\Big] = \exp\big[\frac{n(\theta_0^2 - \theta_1^2)}{2\sigma^2} + \frac{n(\theta_1 - \theta_0)\,\bar{\mathbf{x}}}{\sigma^2}\big].$$

Per ogni $t \in [0, +\infty[$, si ha allora

$$\mathfrak{X}_1^{(t)} = \Big\{\bar{\mathbf{x}} : \frac{n(\theta_0^2 - \theta_1^2)}{2\sigma^2} + \frac{n(\theta_1 - \theta_0)\,\bar{\mathbf{x}}}{\sigma^2} > \ln t\Big\} = \Big\{\bar{\mathbf{x}} : \bar{\mathbf{x}} > \frac{\sigma^2}{n(\theta_1 - \theta_0)} \ln t + \frac{\theta_0 + \theta_1}{2}\Big\}.$$

Conseguentemente, osservato che il secondo membro della disuguaglianza che definisce $\mathfrak{X}_1^{(t)}$ è una funzione continua di t che diverge a $-\infty$, come $t \to 0^+$, e diverge a $+\infty$, come $t \to +\infty$, possiamo concludere che $\Delta^{(NP)}$ coincide con la famiglia dei test aventi come regione di rifiuto, oltre all'insieme vuoto, gli insiemi del tipo $\{\bar{\mathbf{x}} : \bar{\mathbf{x}} > \alpha\}$, cioè i semispazi aperti positivi individuati dagli iperpiani di $\mathbb{R}^n$ di equazione $x_1 + \cdots + x_n = n\alpha$.

Notato infine che la misura di Lebesgue n-dimensionale di un iperpiano di $\mathbb{R}^n$ è nulla[21], otteniamo $\mathrm{P}_{\theta_0}(f_{\theta_1} = t\, f_{\theta_0}) = \int_{\{f_{\theta_1} = t\, f_{\theta_0}\}} f_0\, d\lambda^{(n)} = 0$ e quindi, per la proposizione (iii) del lemma di Neyman-Pearson, $\Delta^{(NP)}$ è una classe essenzialmente completa. ▲

2.4 Indici di preferibilità

Nei problemi di decisione statistica connessi a situazioni reali, l'individuazione delle regole ammissibili non sempre si presenta agevole; inoltre, la loro mancanza oppure la loro molteplicità (nel caso formino una classe completa) pone a DM - che comunque deve scegliere un "comportamento" - il problema di individuarne uno che sia, in qualche modo, "il migliore possibile". La metodologia che storicamente è stata proposta per la sua identificazione si basa

[21]Posto $B = \{\mathbf{x} : x_1 + \cdots + x_{n+1} = a\}$, risulta $B((x_1, \ldots, x_n)) = \{x_{n+1} : x_1 + \cdots + x_{n+1} = a\} = \{a - (x_1 + \cdots + x_n)\}$ e quindi (per definizione della misura prodotto) $\lambda^{(n+1)}(B) = \int_{\mathbb{R}^n} \lambda\big(B((x_1, \ldots, x_n))\big)\, dx_1 \cdots dx_n = 0$.

sull'idea di adottare, come *criterio di scelta* di regole di decisione "buone", la *minimizzazione di un opportuno funzionale* definito sull'insieme delle funzioni di rischio e a valori nella retta reale ampliata. In questo ordine di idee, un ruolo cruciale svolgono le nozioni di indice di preferibilità e di regola di decisione ottima che ora introduciamo.

Definizione 2.4.1 Un funzionale $\imath : \{R_\delta : \delta \in \Delta\} \mapsto \mathbb{R}^*$ è un **indice di preferibilità (per la dominanza)** se: $\delta \succeq \delta' \Rightarrow \imath(R_\delta) \leq \imath(R_{\delta'})$ per ogni $\delta, \delta' \in \Delta$. Inoltre, una **regola (di decisione) $\imath$-ottima** è ogni regola di decisione δ^* tale che $\imath(R_{\delta^*}) = \inf_\delta \imath(R_\delta) < +\infty$[22].

L'individuazione di un indice di preferibilità adeguato per un dato problema di decisione statistica non è di solito semplice e dipende comunque sia dagli obiettivi di DM che dal suo atteggiamento verso l'incertezza. Nell'esempio seguente riportiamo alcuni indici di preferibilità "classici". Poichè alcuni richiedono di considerare distribuzioni di probabilità sullo spazio parametrico, è necessario, prima di procedere, fare alcune puntualizzazioni di natura "tecnica".

Le misure (in particolare probabilità) che considereremo sullo spazio parametrico sono tutte definite su una prefissata σ-algebra di riferimento $\mathcal{T}$ su Θ[23]*. Inoltre, assumeremo che, per ogni decisione d, la funzione L_d sia $\mathcal{T}$-Borel misurabile e prenderemo in esame solamente regole di decisione δ che rendano $\mathrm{L}(\cdot, \delta)$ una funzione $\mathcal{T} \otimes \aleph$-Borel misurabile*[24].

Precisiamo inoltre che, per semplicità, useremo nel corso dell'esposizione la notazione più snella $\imath(\delta)$ al posto della $\imath(R_\delta)$.

Esempio 2.4.2 (Festival degli indici di preferibilità) Gli indici individuati dai criteri seguenti sono (come facilmente si verifica) esempi di indici di preferibilità[25].

(i) CRITERIO DEL MINIMAX Si pone $\imath(\delta) = \sup_\theta R_\delta(\theta)$ per ogni δ. Il criterio sostituisce quindi l'incertezza sui danni medi con il danno medio "massimo" che si può ottenere; conseguentemente, rappresenta un atteggiamento estremamente pessimistico del decisore.

---

[22] Cioè, tale che nella relativa funzione di rischio R_{δ^*} l'indice di preferibilità assume valore sia finito che minimo.

[23] Che raccoglie, da un punto di vista interpretativo, tutti i sottoinsiemi Θ' di Θ per i quali l'evento $\{Z \in \Theta'\}$ è di interesse per DM.

[24] Assicurando in questo modo, per il teorema di Tonelli A.5.4, la $\mathcal{T}$-Borel misurabilità delle relative funzioni di rischio.

[25] A differenza, ad esempio, di quello relativo al *criterio media-varianza* che si basa sulla scelta di una probabilità P su $\mathcal{T}$, di un numero reale $\alpha > 0$ e sulla minimizzazione dell'indice $\imath(\delta) = \mathrm{E}_\mathrm{P}(R_\delta) + \lambda \mathrm{Var}_\mathrm{P}(R_\delta)$. Infatti, considerato il problema decisionale (senza acquisizione di dati): $\Theta = \{\theta_1, \theta_2\}$, $\mathfrak{D} = \{d_1, d_2\}$ e funzione di danno data dalla tabella:

Le relative regole ottime assicurano a DM il "meno peggio", cioè di incorrere al più nel "minimo" dei "massimi" danni medi[26].

Applicando il criterio alle situazioni considerate nell'Esempio 2.2.2, otteniamo che le decisioni ottime sono d_2 per (i); d_1 per (ii); nessuna per (iii).

(ii) Criterio del pessimismo-ottimismo (Leonid Hurwicz, 1951) Si sceglie $\vartheta \in [0,1]$ e si pone $\imath(\delta) = \vartheta \sup_\theta R_\delta(\theta) + (1-\vartheta) \inf_\theta R_\delta(\theta)$ per ogni δ. Il criterio sostituisce dunque l'incertezza sui danni medi con la mistura (di peso ϑ) dei danni medi "massimo" e "minimo" che si possono ottenere. Conseguentemente, il parametro ϑ può essere inteso come una "misura" del grado di pessimismo del decisore (a $\vartheta = 1$ corrisponde il massimo pessimismo; a $\vartheta = 0$ il massimo ottimismo).

Con riferimento all'Esempio 2.2.2, le relative decisioni ottime sono, per (ii), d_1, se $\vartheta > 0$, e d_1, d_3, se $\vartheta = 0$; per (iii), d_0, se $\vartheta \le \frac{2}{3}$, e nessuna, se $\vartheta > \frac{2}{3}$.

(iii) Criterio del rimpianto (Leonard J. Savage, 1951) Per ogni regola di decisione δ, chiamiamo **funzione di rimpianto (di** δ**)** la funzione $R_\delta^{(r)} : \theta \mapsto R_\delta(\theta) - \inf_\delta R_\delta(\theta)$ che - fornendo, per ogni stato θ, l'incremento di danno medio rispetto al danno medio "ineliminabile" $\inf_\delta R_\delta(\theta)$ - può essere intesa come una "misura" del rimpianto di DM nell'aver seguito il "comportamento" δ. Il criterio si ottiene ora applicando il criterio minimax alle funzioni di rimpianto, ponendo cioè $\imath(\delta) = \sup_\theta R_\delta^{(r)}(\theta)$ per ogni δ.

Con riferimento all'Esempio 2.2.2, le funzioni di rimpianto relative alle situazioni (i), (ii) e (iii) sono date, rispettivamente, dalle tabelle:

	d_1	d_2
θ_1	1	0
θ_2	0	1000

	d_1	d_2	d_3
θ_1	1	1	0
θ_2	0	2	3

---

	d_1	d_2
θ_1	4,5	5
θ_2	0,5	5

otteniamo $d_1 \succ d_2$. D'altra parte, posto $\mathrm{P}(\theta_1) = \frac{1}{2}$, si ha $\imath(d_2) = \mathrm{E_P}(\mathrm{L}_{d_2}) + \mathrm{Var_P}(\mathrm{L}_{d_2}) = 5 < \frac{13}{2} = \frac{5}{2} + 4 = \mathrm{E_P}(\mathrm{L}_{d_1}) + \mathrm{Var_P}(\mathrm{L}_{d_1}) = \imath(d_1)$.

[26]Una giustificazione del criterio può ottenersi ricorrendo alla teoria dei giochi. Consideriamo infatti il gioco a somma nulla, avente la Natura come primo giocatore e DM come secondo, nel quale le strategie della Natura sono gli stati, quelle del decisore le regole di decisione e la funzione di vincita (della Natura) la funzione $v(\theta, \delta) = R_\delta(\theta)$. Allora le regole ottime (secondo il criterio minimax) coincidono con le strategie prudenti di DM che sono anche sue strategie di equilibrio (se il gioco ammette equilibri di Nash). Conseguentemente, il decisore sarà portato ad adottare il criterio del minimax ogniqualvolta ritenga la Natura un avversario intenzionato ad arrecargli il massimo danno (medio) possibile.

Il criterio del minimax, introdotto nel 1933 da Jerzy Neyman e Egon S. Pearson nell'ambito della teoria dei test, svolge un ruolo fondamentale nell'assestamento concettuale della statistica matematica operato da Wald negli anni dal 1939 al 1950. La connessione con i giochi a somma nulla sopra richiamata è stata rilevata da Wald nel 1945, immediatamente dopo l'uscita del celebre trattato dedicato alla teoria dei giochi di von Neumann, J. - Morgenstern, O., Theory of Games and Economic Behavior, Princenton University Press, Princenton (1944).

	d_0	d_1	$\ldots$	d_n	$\ldots$
θ_1	1	$\frac{1}{2}$	$\ldots$	$\frac{1}{2^n}$	$\ldots$.
θ_2	0	$2+\frac{1}{3}$	$\ldots$	$2+\frac{1}{3^n}$	$\ldots$

Conseguentemente, le relative decisioni ottime sono d_1 per (i)[27]; d_1 per (ii); d_0 per (iii).

Al fine di introdurre indici di preferibilità che, a differenza dei precedenti, tengano conto di tutti i possibili valori delle funzioni di rischio e non solamente di quelli estremi, consideriamo una misura di probabilità P sulla σ-algebra $\mathcal{T}$.

(iv) Criterio del valor medio Si pone $\imath(\delta) = \mathrm{E}_{\mathrm{P}}(R_\delta) = \int_\Theta R_\delta \, d\mathrm{P}$ per ogni δ. Il criterio, già proposto nel primo trattato di calcolo delle probabilità (Christiaan Huygens, *Ratiociniis in aleae ludo* (1657)), sostituisce quindi l'incertezza sui danni medi con il loro valore atteso. Conseguentemente, le relative regole ottime assicurano a DM di incorrere al più nella "minima" media dei danni medi.

Applicando il criterio alle situazioni considerate nell'Esempio 2.2.2 con $\mathrm{P}(\theta_1) = \pi$, otteniamo che, nella situazione (ii), le decisioni ottime sono d_1, se $\pi < \frac{3}{4}$, e d_1, d_3, se $\pi = \frac{3}{4}$, e d_3, se $\pi > \frac{3}{4}$. Per quanto riguarda invece la situazione (iii), esistono decisioni ottime (e precisamente d_0) solamente nel caso $\pi \leq \frac{2}{3}$.

(v) Criterio della soglia critica Si sceglie una "soglia" $\lambda > 0$ e si pone $\imath(\delta) = \mathrm{P}(R_\delta > \lambda)$ per ogni δ. Il criterio sostituisce quindi l'incertezza sui danni medi con la probabilità che essi superino la soglia considerata. Conseguentemente, le relative regole ottime consentono a DM di garantirsi la "minima" probabilità di superare tale soglia[28].

Applicando il criterio all'Esempio 2.2.2, con $\lambda = \frac{21}{10}$ e probabilità positiva in ogni stato, otteniamo che le decisioni ottime sono d_1 per (i); d_1 per (ii); d_n ($n \geq 4$) per (iii).

(vi) Criterio di Hodges-Lehmann (1952) Si sceglie $\vartheta \in [0,1]$ e si pone $\imath(\delta) = \vartheta \mathrm{E}_{\mathrm{P}}(R_\delta) + (1-\vartheta)\sup_\theta R_\delta(\theta)$ per ogni δ. Il criterio sostituisce quindi l'incertezza sui danni medi con la mistura (di peso ϑ) della media dei danni medi e del "massimo" danno medio[29].

Applicando il criterio all'Esempio 2.2.2(ii) con $\mathrm{P}(\theta_1) = \frac{7}{8}$, otteniamo che le decisioni ottime sono d_1, se $\vartheta < \frac{4}{5}$, e d_1, d_3, se $\vartheta = \frac{4}{5}$, e d_3, se $\vartheta > \frac{4}{5}$. ▲

[27] Ottenendo così una soluzione più "naturale" di quella (pari a d_2) fornita dal criterio minimax. Infatti, d_1 sembra più appetibile di d_2 in quanto, a fronte della perdita certa pari a 1000 (relativa a d_2), assicura una perdita nulla in corrispondenza a θ_2, contro un incremento di perdita unitario in corrispondenza a θ_1.

[28] Interpretate le funzioni di rischio come danni monetari medi e la soglia come il livello che non deve essere superato per evitare il "fallimento", le regole ottime sono quelle a cui corrisponde la "massima" probabilità di non incorrere nel fallimento.

[29] Una giustificazione intuitiva del criterio può ottenersi osservando che, in generale, le opinioni che DM ha sullo stato di natura non sono in grado di individuare una precisa misura di probabilità su $\mathcal{T}$, ma solamente alcune sue proprietà (come, ad esempio, alcuni suoi momenti). Conseguentemente, il parametro ϑ può essere inteso come una "misura" del grado di fiducia che DM (sulla base del suo stato d'informazione) ha sulla capacità della probabilità P di descrivere l'incertezza connessa allo stato di natura. Se la fiducia è massima (e questo avviene quando lo stato d'informazione individua P), DM adotta il criterio del valor medio ($\vartheta = 1$), mentre, man mano che tale fiducia decresce, DM è sempre più propenso a scegliere il "meno peggio", cioè il criterio del minimax ($\vartheta = 0$).

Considerato un indice di preferibilità $\imath$, non esiste (in generale) alcun legame tra le regole di decisione $\imath$-ottime e quelle ammissibili. Infatti, con riferimento al problema decisionale (senza acquisizione di dati): $\Theta = \{\theta_1, \theta_2, \theta_3\}$, $\mathfrak{D} = \{d_1, d_2, d_3\}$ e funzione di danno data dalla tabella:

	d_1	d_2	d_3
θ_1	1	1	0
θ_2	2	5	2
θ_3	3	3	4

,

si ha $\mathfrak{D}^+ = \{d_1.d_3\}$. Inoltre, usando il criterio del valor medio con $\mathrm{P}(\theta_1) = \frac{1}{4}$ e $P(\theta_3) = \frac{3}{4}$, otteniamo che le decisioni ottime sono d_1 e d_2. Conseguentemente, d_3, pur essendo ammissibile, non è ottima e d_2, pur essendo ottima, non è ammissibile.

I risultati seguenti forniscono delle condizioni sufficienti affinchè il procedimento di minimizzazione dell'indice di preferibilità possa essere visto come un particolare criterio di selezione di regole ammissibili.

Teorema 2.4.3 *Sia $\imath$ un indice di preferibilità tale che tutte le regole $\imath$-ottime hanno medesima funzione di rischio. Allora, ogni regola $\imath$-ottima è ammissibile. In particolare, se δ^* è l'unica regola $\imath$-ottima, allora δ^* è ammissibile.*

Dimostrazione Assumiamo (per assurdo) che la regola $\imath$-ottima δ^* non sia ammissibile. Esiste allora una regola di decisione δ tale che $\delta \succ \delta^*$. Ne segue $R_\delta \neq R_{\delta^*}$ e $\imath(\delta) \leq \imath(\delta^*)$. Pertanto, $\imath(\delta) = \imath(\delta^*)$ e quindi δ è una regola $\imath$-ottima avente funzione di rischio diversa da quella di δ^* (Contraddizione!).$\blacksquare$

Al fine di ottenere due criteri di ammissibilità (il primo dei quali è di verifica immediata) anche nel caso di regole ottime aventi funzioni di rischio diverse, introduciamo due particolari tipi di indici di preferibilità.

Definizione 2.4.4 Un indice di preferibilità $\imath$ è **strettamente monotono** se: $\delta \succ \delta' \Rightarrow \imath(\delta) < \imath(\delta')$ per ogni $\delta, \delta' \in \Delta$. Inoltre, considerata una misura ν su $\mathcal{T}$, l'indice di preferibilità $\imath$ è **ν-strettamente monotono** se: $\delta \succ \delta' \wedge \nu(\{R_\delta < R_{\delta'}\}) > 0 \Rightarrow \imath(\delta) < \imath(\delta')$ per ogni $\delta, \delta' \in \Delta$.

Teorema 2.4.5 *Sia $\imath$ un indice di preferibilità strettamente monotono. Allora ogni regola $\imath$-ottima è ammissibile.*

Teorema 2.4.6 *Siano ν una misura su $\mathcal{T}$ e $\imath$ un indice di preferibilità ν-strettamente monotono. Allora, ogni regola $\imath$-ottima è ammissibile ogniqualvolta sussista una delle due condizioni:*

(a) $\nu(\{\theta\}) > 0$ *per ogni stato* θ*;*

(b) $\Theta \subseteq \mathbb{R}^m$ *e* $\nu(W \cap \Theta) > 0$ *per ogni aperto* W *di* $\mathbb{R}^m$ *tale che* $W \cap \Theta \neq \emptyset$*; inoltre, la funzione di rischio* R_δ *è continua per ogni* δ[30]*.*

Dimostrazione Per il teorema precedente basta provare la stretta monotonia dell'indice $\imath$. Sia quindi $\delta \succ \delta'$. Allora, $R_\delta \leq R_{\delta'}$ e $R_\delta(\tilde{\theta}) < R_{\delta'}(\tilde{\theta})$ per qualche stato $\tilde{\theta}$.

Assumiamo intanto (a). Da $\tilde{\theta} \in \{R_\delta < R_{\delta'}\}$ otteniamo $0 < \nu(\{\tilde{\theta}\}) \leq \nu(\{R_\delta < R_{\delta'}\})$ e quindi, per la ν-stretta monotonia, $\imath(\delta) < \imath(\delta')$.

Supponiamo ora che sussista (b). Per il teorema della permanenza del segno (la funzione $R_\delta - R_{\delta'}$ è continua!), da $(R_\delta - R_{\delta'})(\tilde{\theta}) < 0$ otteniamo l'esistenza di un intorno aperto W di $\tilde{\theta}$ tale che $(R_\delta - R_{\delta'})(\theta) < 0$ per ogni $\theta \in W$. Allora, $\emptyset \neq W \cap \Theta \subseteq \{R_\delta < R_{\delta'}\}$ da cui, per l'ipotesi relativa agli aperti, si ha $0 < \nu(W \cap \Theta) \leq \nu(\{R_\delta < R_{\delta'}\})$. Ne segue, per la ν-stretta monotonia, $\imath(\delta) < \imath(\delta')$. ■

Concludiamo la sezione provando che gli indici di preferibilità a valori finiti - ottenuti considerando una probabilità P su $\mathcal{T}$ e operando una trasformazione crescente della speranza matematica - sono P-strettamente monotoni. Conseguentemente, nelle situazioni considerate nel teorema precedente, la probabilità P può essere intesa come *mero strumento formale per selezionare regole ammissibili* (mediante il criterio del valor medio) piuttosto che *mezzo atto a rappresentare l'incertezza del decisore* sullo stato di natura[31].

[30] La condizione concernente gli aperti è banalmente verificata se Θ è un insieme aperto o la chiusura di un insieme aperto e ν è la misura di Lebesgue m-dimensionale sui boreliani di Θ (come spesso avviene nelle applicazioni).

[31] All'interpretazione della probabilità come strumento (privo talvolta di reale significato per lo stato di natura) utile per individuare regole ammissibili fa riferimento l'impostazione **classico-frequentista** della statistica, dovuta a Ronald A. Fisher che la sviluppò nei primi decenni del '900. Basandosi su una concezione della probabilità intesa come "limite" (in qualche senso) di frequenze osservabili, tale impostazione non è in grado di strutturare tutti i problemi statistico-decisionali in termini totalmente probabilistici, in quanto non ogni tipo di incertezza è "ripetibile" e quindi probabilizzabile (nel senso qui inteso).

All'interpretazione della probabilità come mezzo atto a rappresentare l'incertezza che DM ha (in base al suo stato d'informazione) nei riguardi di un dato fenomeno fa invece

Teorema 2.4.7 *Data una misura ν su $\mathcal{T}$, sia $\varphi : \mathbb{R} \mapsto \mathbb{R}$ una funzione crescente tale che le funzioni composte $\varphi(R_\delta)$ risultino tutte ν-integrabili. Allora, l'indice di preferibilità così definito:*

$$\imath(\delta) = \int_\Theta \varphi(R_\delta(\theta))\, \nu(d\theta) \in \mathbb{R}$$

per ogni δ, è ν-strettamente monotono[32].

Dimostrazione Sia $\delta \succ \delta'$ e $\nu(\{R_\delta < R_{\delta'}\}) > 0$. Allora, $R_\delta \leq R_{\delta'}$ e quindi

$$\begin{aligned}\imath(\delta') - \imath(\delta) &= \int_\Theta \varphi(R_{\delta'})\, d\nu - \int_\Theta \varphi(R_\delta)\, d\nu \\ &= \int_\Theta [\varphi(R_{\delta'}) - \varphi(R_\delta)]\, d\nu = \int_{\{R_\delta < R_{\delta'}\}} [\varphi(R_{\delta'}) - \varphi(R_\delta)]\, d\nu > 0,\end{aligned}$$

tenendo presente la linearità dell'integrale e il Teorema A.3.6(vi). ■

2.5 Regole di decisione bayesiane

Tra i vari criteri considerati nell'Esempio 2.4.2, di particolare importanza è quello del valor medio in quanto è l'unico criterio che può ritenersi coerente con l'impostazione bayesiana della statistica. Infatti, supposto che DM, seguendo lo schema di Bayes (Sezione B.2.4), esprima le sue opinioni sullo stato di natura tramite una densità iniziale π (rispetto alla misura di riferimento σ-finita ν su $\mathcal{T}$) e consideri sulla σ-algebra prodotto la probabilità $P^{(\mathrm{sb})}$ così definita:

$$P^{(\mathrm{sb})}(\hat{A}) = \int_{\hat{A}} f_\theta(x)\pi(\theta)\ \nu \times \mu(d\theta \times dx)$$

riferimento l'impostazione **neo-bayesiana** (brevemente **bayesiana**) della statistica, essenzialmente dovuta a Bruno de Finetti che la sviluppò nella prima metà del '900. Tale impostazione, basandosi su una concezione della probabilità intesa come "grado di fiducia soggettivo" sul verificarsi degli eventi, non richiede che l'incertezza abbia una natura particolare e quindi è in grado di fornire uno sviluppo totalmente probabilistico dei problemi statistico-decisionali. Non va peraltro taciuto che la richiesta di tradurre l'incertezza sullo stato di natura in una ben precisata distribuzione di probabilità non è sempre facilmente attuabile nell'usuale pratica statistica (specialmente nel caso multidimensionale ove lo stato di natura potrebbe essere un vettore aleatorio con un numero elevato di componenti).

[32] Conseguentemente, nel caso particolare che ν sia una probabilità, si ottiene un indice ν-strettamente monotono associando ad ogni funzione di rischio R_δ il suo momento n-simo $\mathrm{E}_\nu(R_\delta^n)$.

per ogni $\hat{A} \in \mathcal{T} \otimes \aleph$, scegliamo un'arbitraria regola di decisione δ. Rimane allora individuata una particolare lotteria: precisamente la legge $\ell^{(\delta)}$ di $\gamma(\cdot, \delta)$:

$$\ell^{(\delta)}(C) = P^{(\text{sb})}(\gamma(\cdot, \delta) \in C)$$

per ogni $C \in \mathcal{C}$. Per quanto riguarda la sua utilità attesa, tramite il teorema fondamentale del calcolo delle probabilità e quello di Tonelli, si ha

$$\begin{aligned}
\mathrm{E}_{\ell^{(\delta)}}(u) &= \int_{\mathfrak{C}} u \, d\ell^{(\delta)} = \int_{\Theta \times \mathfrak{X}} u(\gamma(\theta, \delta(x))) \; P^{(\text{sb})}(d\theta \times dx) \\
&= \int_{\Theta} \Big[\int_{\mathfrak{X}} u(\gamma(\theta, \delta(x))) f_\theta(x) \, \mu(dx) \Big] \, \pi(\theta) \, \nu(d\theta) \\
&= \int_{\Theta} \Big[\int_{\mathfrak{X}} -\mathrm{L}(\theta, \delta(x)) \, \mathrm{P}_\theta(dx) \Big] \pi(\theta) \, \nu(d\theta) = - \int_{\Theta} R_\delta(\theta) \pi(\theta) \, \nu(d\theta).
\end{aligned}$$

Supposto infine, come già fatto all'inizio del capitolo per giustificare l'introduzione della funzione di rischio, che DM consideri equivalenti due regole di decisione che hanno la medesima legge, DM sarà indotto ad esprimere la relazione di preferenza $\succeq_\pi$ su tali regole tramite la relazione di preferenza $\succeq^*$ definita sulle lotterie. Si avrà quindi

$$\delta_1 \succeq_\pi \delta_2 \iff \mathrm{E}_{\ell^{(\delta_1)}}(u) \geq \mathrm{E}_{\ell^{(\delta_2)}}(u) \iff \int_\Theta R_{\delta_1} \pi \, d\nu \leq \int_\Theta R_{\delta_2} \pi \, d\nu.$$

Pertanto DM, per confrontare regole di decisione, ricorrerà al confronto dei valori medi delle corrispondenti funzioni di rischio (calcolati con la probabilità generata dalla densità iniziale). Possiamo dunque concludere che, nella impostazione bayesiana, DM è "forzato" ad usare il criterio del valor medio per scegliere una regola di decisione "buona".

Le considerazioni svolte giustificano la prossima definizione che introduce la nozione centrale di regola di decisione bayesiana.

Definizione 2.5.1 Data una probabilità P su $\mathcal{T}$:
- il **rischio di Bayes di** P è la quantità $\rho(\mathrm{P}) = \inf_\delta \mathrm{E}_{\mathrm{P}}(R_\delta) \geq 0$[33];

[33] Riferendoci all'impostazione bayesiana, dal teorema di Tonelli otteniamo

$$\mathrm{E}_{\mathrm{P}}(R_\delta) = \int_\Theta \Big[\int_{\mathfrak{X}} \mathrm{L}(\theta, \delta) f_\theta \, d\mu \Big] \pi(\theta) \, \nu(d\theta) = \int_{\Theta \times \mathfrak{X}} \mathrm{L}(\theta, \delta(x)) f_\theta(x) \pi(\theta) \, \nu \times \mu(d\theta \times dx)$$

e quindi il rischio di Bayes può interpretarsi come danno medio "ineliminabile" connesso con l'esperimento statistico e le opinioni pre-sperimentali di DM sullo stato di natura.

- una **regola (di decisione) bayesiana per** P è ogni regola di decisiona δ^* tale che $\rho(\mathrm{P}) = \mathrm{E}_{\mathrm{P}}(R_{\delta^*}) < +\infty$[34].

Inoltre, una **regola (di decisione) bayesiana** è una qualsiasi regola di decisione che è bayesiana per qualche probabilità P su $\mathcal{T}$.

Mostriamo ora che le regole bayesiane per una data probabilità P possono anche non esistere e che i test di Neyman-Pearson sono esempi di regole bayesiane nel problema della verifica d'ipotesi semplici.

Esempio 2.5.2 (i) Considerato il problema decisionale (senza acquisizione di dati): $\mathfrak{D} = \Theta = \mathbb{N}$ e funzione di danno:

$$\mathrm{L}(\cdot, 0) = 1, \qquad \mathrm{L}(\theta, n) = \begin{cases} 0 & \text{se } \theta \leq n \\ 2 & \text{se } \theta > n \end{cases} \quad (n \geq 1),$$

sia P una probabilità su $\mathbb{N}$ tale che $\pi_k = \mathrm{P}(k) > 0$ per ogni $k \in \mathbb{N}$. Riesce allora

$$\mathrm{E}_{\mathrm{P}}(\mathrm{L}_k) = \begin{cases} 1 & \text{se } k = 0 \\ 2\sum_{n>k} \pi_n > 0 & \text{se } k \geq 1 \end{cases}$$

e quindi $\lim\limits_{k \to +\infty} \mathrm{E}_{\mathrm{P}}(\mathrm{L}_k) = 0$. Conseguentemente, non esistono regole bayesiane per P.

(ii) Con riferimento al problema della verifica d'ipotesi semplici, consideriamo una distribuzione di probabilità P su Θ e poniamo $\pi = \mathrm{P}(\theta_0)$. Dato un test δ, da (2.4) (ponendo $K_i = k_i(\theta_j)$ $(j \neq i; i = 0, 1)$) risulta allora

$$\begin{aligned} \mathrm{E}_{\mathrm{P}}(R_\delta) &= K_1\beta_\delta(\theta_0)\pi + K_0\big[1 - \beta_\delta(\theta_1)\big](1-\pi) = K_0(1-\pi) + K_1\pi\beta_\delta(\theta_0) - K_0(1-\pi)\beta_\delta(\theta_1) \\ &= K_0(1-\pi) + K_1\pi \int_{\mathfrak{X}_1^{(\delta)}} f_{\theta_0}\, d\mu - K_0(1-\pi)\int_{\mathfrak{X}_1^{(\delta)}} f_{\theta_1}\, d\mu \\ &= K_0(1-\pi) + \int_{\mathfrak{X}_1^{(\delta)}} \big[K_1\pi f_{\theta_0} - K_0(1-\pi) f_{\theta_1}\big]\, d\mu \\ &= K_0(1-\pi) + \int_{\mathfrak{X}} \big[K_1\pi f_{\theta_0} - K_0(1-\pi) f_{\theta_1}\big]\, I_{\mathfrak{X}_1^{(\delta)}}\, d\mu. \end{aligned}$$

Per ottenere un test bayesiano per P, basta dunque considerare (per la monotonia dell'integrale) un test che minimizzi la funzione $\big[K_1\pi f_{\theta_0} - K_0(1-\pi) f_{\theta_1}\big]\, I_{\mathfrak{X}_1^{(\delta)}}$. Ora, posto $g = K_1\pi f_{\theta_0} - K_0(1-\pi) f_{\theta_1}$, si ha

$$\big[K_1\pi f_{\theta_0} - K_0(1-\pi) f_{\theta_1}\big]\, I_{\mathfrak{X}_1^{(\delta)}} = (g^+ - g^-) I_{\mathfrak{X}_1^{(\delta)}} = g^+ I_{\mathfrak{X}_1^{(\delta)}} - g^- I_{\mathfrak{X}_1^{(\delta)}}$$

e quindi la funzione in oggetto sarà minimizzata se $g^+ I_{\mathfrak{X}_1^{(\delta)}} = 0$ e $g^- I_{\mathfrak{X}_1^{(\delta)}} = g^-$. Pertanto, ogni test δ^* tale che $\{g < 0\} \subseteq \mathfrak{X}_1^{(\delta^*)}$ e $\{g > 0\} \subseteq \mathfrak{X}_0^{(\delta^*)}$ è un test bayesiano per P.

[34] Cioè, ogni regola di decisione con rischio medio finito che sia ottima rispetto al criterio del valor medio.

Ciò osservato, supponiamo intanto $\pi = 1$. Allora, $g = K_1 f_{\theta_0}$ e quindi $\{f_{\theta_0} = 0\} \cap \mathfrak{X}_1^{(\delta^*)} \neq \emptyset$ o $\mathfrak{X}_1^{(\delta^*)} = \emptyset$. Conseguentemente, il test di Neyman-Pearson $\delta^{(+\infty)}$ è un test bayesiano per la probabilità che concentra l'intera massa in θ_0. Sia ora $\pi < 1$. Allora,

$$g = K_0(1-\pi)\Big[\frac{K_1\pi}{K_0(1-\pi)}\, f_{\theta_0} - f_{\theta_1}\Big]$$

e quindi

$$\Big\{f_1 > \frac{K_1\pi}{K_0(1-\pi)} f_0\Big\} \subseteq \mathfrak{X}_1^{(\delta^*)}, \qquad \Big\{f_1 < \frac{K_1\pi}{K_0(1-\pi)} f_0\Big\} \subseteq \mathfrak{X}_0^{(\delta^*)}.$$

Conseguentemente, considerato un numero reale $t \geq 0$, il test di Neyman-Pearson $\delta^{(t)}$ è un test bayesiano per la probabilità che concentra massa $\frac{K_0 t}{K_0 t + K_1}$ in θ_0. ▲

Il teorema seguente fornisce una condizione "geometrica" per l'esistenza, nel caso di spazi parametrici finiti, di regole bayesiane per ogni probabilità che concentra massa positiva su ogni stato.

Teorema 2.5.3 *Supposto Θ finito, assumiamo che l'insieme di rischio includa la sua chiusura inferiore. Allora, per ogni probabilità* P *tale che* $\mathrm{P}(\theta) > 0$ *per ogni stato θ, esistono regole bayesiane per* P.

Dimostrazione Posto $\Theta = \{\theta_1, \ldots, \theta_m\}$, sia $\pi_i = \mathrm{P}(\theta_i) > 0$ $(i = 1, \ldots, m)$. Allora, per ogni δ, si ha $\mathrm{E}_{\mathrm{P}}(R_\delta) = \pi_1 y_1^{(\delta)} + \cdots + \pi_m y_m^{(\delta)}$. Conseguentemente, considerata la funzione $f : \mathbf{y} \mapsto \pi_1 y_1 + \cdots + \pi_m y_m$, bisogna provare che l'insieme-immagine $f(R) \subseteq [0, +\infty[$ ammette minimo. A tal fine, poniamo $\beta = \inf f(R)$.

Sia ora $(\mathbf{y}^{(n)})_{n\geq 1}$ una successione in R tale che $f(\mathbf{y}^{(n)}) \to \beta$ come $n \to +\infty$. Allora, la successione è superiormente limitata. Infatti, supposto, per assurdo, che per ogni n esistano due numeri naturali h_n, k_n tali che $y_{k_n}^{(h_n)} > n$, otteniamo $\lim\limits_{n\to+\infty} y_{k_n}^{(h_n)} = +\infty$. Allora, dalla

$$f(\mathbf{y}^{(h_n)}) = \sum_{i=1}^{k_n-1} \pi_i y_i^{(h_n)} + \pi_{k_n} y_{k_n}^{(h_n)} + \sum_{i=k_n+1}^{m} \pi_i y_i^{(h_n)} \geq \pi_{k_n} y_{k_n}^{(h_n)}$$

si ha $\beta = \lim\limits_{n\to+\infty} f(\mathbf{y}^{(h_n)}) \geq \lim\limits_{n\to+\infty} \pi_{k_n} y_{k_n}^{(h_n)} = +\infty$ (Contraddizione!).

Poichè la successione è anche inferiormente limitata (essendo in R), esiste una sottosuccessione $(\mathbf{y}^{(i_n)})_{n\geq 1}$ convergente a qualche m-pla $\mathbf{y}'$. Chiaramente $\mathbf{y}' \in \overline{R}$; inoltre, $f(\mathbf{y}') = \lim\limits_{n\to+\infty} f(\mathbf{y}^{(i_n)}) = \lim\limits_{n\to+\infty} f(\mathbf{y}^{(n)}) = \beta$ (f è una funzione continua!). Ne segue $\mathbf{y}' \in \partial_L R$. Infatti, esista (per assurdo) $\mathbf{y} \in R$ tale che $\mathbf{y} < \mathbf{y}'$. Allora $y_i \leq y'_i \, (i = 1, \ldots, m)$ e $y_j < y'_j$ per qualche j. Riesce pertanto $\inf f(R) = \beta = f(\mathbf{y}') = \sum_{i=1}^m \pi_i y'_i > \sum_{i=1}^m \pi_i y_i = f(\mathbf{y}) \in f(R)$ (Contraddizione!).

Ora, per l'ipotesi $\partial_L R \subset R$, esiste una regola di decisione δ' tale che $\mathbf{y}^{(\delta')} = \mathbf{y}'$. Allora, $f(\mathbf{y}^{(\delta')}) = \beta$ e quindi δ' è una regola bayesiana per P. ■

Come già osservato, il criterio del valor medio può essere utilmente utilizzato per individuare, nelle situazioni considerate nel Teorema 2.4.6, regole ammissibili. Viene allora naturale chiedersi se questa metodologia sia (per l'ammissibilità) anche esaustiva, nel senso che per ogni regola ammissibile δ è possibile specificare una probabilità P su $\mathcal{T}$ rispetto alla quale δ è bayesiana. Per mostrare che (in generale) la risposta è negativa, consideriamo il problema decisionale (senza acquisizione di dati): $\Theta = \{\theta_1, \theta_2\}$, $\mathfrak{D} = \{d_1, d_2, d_3\}$ e funzione di danno data dalla tabella:

	d_1	d_2	d_3
θ_1	1	0	$\frac{1}{2}$
θ_2	0	1	$\frac{3}{4}$

.

Allora, la decisione d_3, pur essendo ammissibile, non è (come facilmente si verifica) bayesiana.

Proviamo ora che, per spazi parametrici finiti, la convessità dell'insieme di rischio è una condizione sufficiente affinchè ogni regola ammissibile sia bayesiana.

Teorema 2.5.4 *Supposto Θ finito, sia l'insieme di rischio convesso. Allora, ogni regola ammissibile è bayesiana.*

DIMOSTRAZIONE Posto $\Theta = \{\theta_1, \ldots, \theta_m\}$, sia δ' una regola ammissibile. Allora, per il Teorema 2.2.9(i), $\mathbf{y}^{(\delta')} = (y'_1, \ldots, y'_m) \in \partial_L$R. Pertanto l'insieme $\{\mathbf{y} : \mathbf{y} \ll \mathbf{y}^{(\delta')}\}$ risulta disgiunto da R. Notato che tali insiemi sono convessi, per il teorema dell'iperpiano separatore[35], esistono una m-pla $\mathbf{a} \neq \mathbf{0}$ e un

[35]Per *teorema dell'iperpiano separatore* intendiamo il seguente risultato: Siano $B', B'' \neq \emptyset$ due sottoinsiemi convessi e disgiunti di $\Re^m$. Esistono allora una m-pla $\mathbf{a} \neq \mathbf{0}$ e un numero reale α tali che $a_1 y_1 + \cdots + a_m y_m \leq \alpha$, per ogni $\mathbf{y} \in B'$, e $a_1 y_1 + \cdots + a_m y_m \geq \alpha$, per ogni $\mathbf{y} \in B''$ (per una dimostrazione si veda Bazaraa, M.S.-Sherali, H.D.-Shetty, C.M., Nonlinear Programming. Theory and Algorithms (Second Edition), John Wiley, New York (1979), teorema 2.4.8).

numero reale α tali che

$$a_1 y_1^{(\delta)} + \cdots + a_m y_m^{(\delta)} \geq \alpha, \quad \text{per ogni } \delta \in \Delta \tag{2.12}$$

$$a_1 y_1 + \cdots + a_m y_m \leq \alpha, \quad \text{per ogni } \mathbf{y} \ll \mathbf{y}^{(\delta')}. \tag{2.13}$$

Per provare che

$$a_1 y_1' + \cdots + a_m y_m' = \alpha \tag{2.14}$$

consideriamo un numero reale $\epsilon > 0$. Allora $(y_1' - \epsilon, \ldots, y_m' - \epsilon) \ll \mathbf{y}^{(\delta')}$ da cui, tramite (2.13), otteniamo $a_1 y_1' + \cdots + a_m y_m' - (a_1 + \cdots + a_m)\epsilon \leq \alpha$. Ne segue, passando al limite per $\epsilon \to 0^+$, $a_1 y_1' + \cdots + a_m y_m' \leq \alpha$ e quindi, per (2.12), $a_1 y_1' + \cdots + a_m y_m' = \alpha$.

Per provare che riesce $\mathbf{a} \geq \mathbf{0}$, sia (per assurdo) $a_j < 0$ per qualche j. Dato n, si ha $(y_1' - 1, \ldots, y_{j-1}' - 1, y_j' - n, y_{j+1}' - 1, \ldots, y_m' - 1) \ll \mathbf{y}^{(\delta')}$ da cui, per (2.13), risulta

$$\sum_{i \neq j} a_i (y_i' - 1) + a_j (y_j' - n) \leq \alpha.$$

Tenuto conto della $a_j < 0$, riesce allora

$$\alpha \geq \lim_{n \to +\infty} \Big[\sum_{i \neq j} a_i (y_i' - 1) + a_j (y_j' - n)\Big] = +\infty$$

ottenendo così una contraddizione.

Sia ora P la distribuzione di probabilità sugli stati tale che:

$$\pi_i = \mathrm{P}(\theta_i) = \frac{a_i}{a_1 + \cdots + a_m} \quad (i = 1, \ldots, m).$$

Considerata una regola δ, da (2.12), (2.14) otteniamo

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}}(R_\delta) &= p_1 y_1^{(\delta)} + \cdots + p_m y_m^{(\delta)} = \frac{a_1 y_1^{(\delta)} + \cdots + a_m y_m^{(\delta)}}{a_1 + \cdots + a_m} \\
&\geq \frac{\alpha}{a_1 + \cdots + a_m} = \frac{a_1 y_1' + \cdots + a_m y_m'}{a_1 + \cdots + a_m} \\
&= p_1 y_1' + \cdots + p_m y_m' = \mathrm{E}_{\mathrm{P}}(R_{\delta'}).
\end{aligned}$$

Conseguentemente, data l'arbitrarietà di δ, la regola δ' è bayesiana per P. ■

Conseguenza immediata del teorema appena provato e del Teorema 2.2.11 è il seguente importante risultato che delinea una situazione nella quale la **classe bayesiana** (cioè l'insieme delle regole bayesiane) è una classe completa. In tale contesto DM (per risolvere il problema decisionale) può dunque limitare la ricerca di un "comportamento" che, per i suoi scopi, sia il "migliore possibile" nell'ambito delle regole bayesiane.

Teorema 2.5.5 (della classe completa) *Supposto Θ finito, sia l'insieme di rischio un insieme convesso che includa la sua chiusura inferiore. Allora la classe bayesiana è una classe completa che include la classe completa minimale delle regole ammissibili.*

2.6 Regole bayesiane formali

Poichè la classe bayesiana svolge, come abbiamo constatato, un ruolo centrale nei problemi statistico-decisionali, viene naturale cercare di individuarne gli elementi tramite un procedimento che risulti, in qualche modo, costruttivo. A tal fine, consideriamo una probabilità iniziale P avente π come funzione di densità rispetto a una prefissata misura di riferimento σ-finita ν su $\mathcal{T}$. Seguendo lo schema di Bayes (Sezione B.2.4), possiamo allora introdurre la **probabilità predittiva** $P^{(\mathrm{pr})}$ individuata dalla μ-densità:

$$p(x) = \int_\Theta f_\theta(x)\pi(\theta)\,\nu(d\theta)$$

e, per ogni campione x, la **probabilità finale** $\mathrm{P}(\cdot|x)$ generata dalla ν-densità:

$$\pi(\theta|x) = \begin{cases} \frac{f_\theta(x)}{p(x)}\,\pi(\theta) & \text{se } 0 < p(x) < +\infty \\ \pi(\theta) & \text{altrimenti} \end{cases}.$$

Supposto, come peraltro già fatto, che la funzione L_d sia $\mathcal{T}$-Borel misurabile, consideriamo inoltre, per ogni decisione d e campione x, il **danno medio finale di d relativo a x**:

$$\rho_{\mathrm{P}}(d|x) = \int_\Theta \mathrm{L}(\theta, d)\mathrm{P}(d\theta|x) = \int_\Theta \mathrm{L}(\theta, d)\pi(\theta|x)\,\nu(d\theta).$$

Siamo così in grado di introdurre, ricorrendo al criterio del valor medio, un particolare tipo di "comportamento" da parte di DM: quello di associare ad ogni campione una decisione che minimizza il danno medio finale relativo a quel campione.

Definizione 2.6.1 Una **funzione di decisione** è una qualsiasi applicazione $\phi : \mathfrak{X} \mapsto \mathfrak{D}$ appartenente o no a Δ. Inoltre, una funzione di decisione ϕ è una **regola bayesiana formale** (**per** P) se $\rho_{\mathrm{P}}(\phi(x)|x) = \inf_{d\in\mathfrak{D}} \rho_{\mathrm{P}}(d|x) < +\infty$ per ogni campione x[36].

Il prossimo esempio mette in evidenza che l'appartenenza di una regola bayesiana formale a Δ non implica necessariamente che sia bayesiana e neppure che sia ammissibile.

Esempio 2.6.2 (Berger, 1985) Posto $\Theta = \mathfrak{D} = \mathbb{R}$, siano l'osservabile costituito da una v.a. X distribuita secondo la normale $\mathsf{N}(\theta, 1)$ in corrispondenza ad ogni stato θ e la densità iniziale π quella della normale $\mathsf{N}(0, 1)$. Allora, per l'Esempio B.2.18(ii),

$$\pi(\theta|x) = \frac{1}{\sqrt{\pi}} \exp\Big[-\Big(\theta - \frac{x}{2}\Big)^2\Big]$$

per ogni stato θ e campione x. Conseguentemente, scelta come funzione di danno la trasformata del danno quadratico:

$$\mathrm{L}(\theta, d) = (\theta - d)^2 \exp\Big[\frac{3\theta^2}{4}\Big],$$

[36]All'inizio della sezione precedente, abbiamo appurato che il criterio del valor medio è l'unico coerente con l'impostazione bayesiana della statistica. Conseguentemente, supposto che DM "aggiorni" le sue opinioni pre-sperimentali – basandosi unicamente sulla conoscenza di quale campione si è realizzato *e null'altro in più* – sostituendo la probabilità iniziale con quella finale, DM sceglierà come decisione una che minimizzi il danno medio finale relativo al campione osservato perseguendo perciò, in ultima analisi, un "comportamento" prescritto da una regola bayesiana formale.

In realtà è bene tenere presente che l'opinione post-sperimentale del decisore, indotta dall'aver osservato il campione x, non sarà rappresentabile, salvo casi sostanzialmente artificiosi, mediante la sola densità finale $\pi(\cdot|x)$. In effetti, l'informazione che DM acquisisce dall'osservazione di x non è, in generale, limitata alla *sola conoscenza della realizzazione del campione x, ma risulterà arricchita da ulteriori conoscenze.* Ad esempio, supponiamo che DM sia tifoso di una squadra di calcio e che ne segua settimanalmente gli impegni di campionato. Essendo interessato all'esito dell'incontro previsto per la squadra tra due settimane, potrebbe, per migliorare il suo attuale pronostico, decidere di osservare quale delle tre alternative, "vittoria", "sconfitta" e "pareggio", si realizzerà per la squadra del cuore nella partita relativa alla prossima settimana. È evidente che, al termine dell'incontro, DM apprenderà non solo quale alternativa si è verificata (informazione campionaria), ma anche con quale punteggio la partita è terminata (2-1, 2-4, 3-3, ...), come pure i commenti dei giocatori e degli allenatori; informazioni che si aggiungono a quella campionaria e che sono indubbiamente utili ai fini del pronostico.

Queste considerazioni mettono in evidenza che, in ultima analisi, anche l'impostazione bayesiana svolge, analogamente a quella classico-frequentista, un'analisi del comportamento del decisore di natura essenzialmente pre-sperimentale.

otteniamo, qualunque sia la decisione d,

$$\rho_{\mathrm{P}}(d|x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} (\theta - d)^2 \exp\Big[\frac{3\theta^2}{4} - (\theta - \frac{x}{2})^2\Big] d\theta.$$

Ora, dalla

$$\frac{3\theta^2}{4} - (\theta - \frac{x}{2})^2 = -\frac{\theta^2 - 4x\theta + x^2}{4} = -\frac{(\theta^2 - 4x\theta + 4x^2) - 3x^2}{4} = \frac{3x^2 - (\theta - 2x)^2}{4}$$

risulta

$$\rho_{\mathrm{P}}(d|x) = 2 \exp\Big[\frac{3x^2}{4}\Big] \int_{-\infty}^{+\infty} (\theta - d)^2 \frac{1}{\sqrt{4\pi}} \exp\Big[-\frac{(\theta - 2x)^2}{4}\Big] d\theta.$$

Osservato che nell'integrale compare la densità della normale $\mathsf{N}(2x, 2)$, la funzione $\rho_{\mathrm{P}}(\cdot|x)$ assume, per il Corollario B.2.11, valore minimo nel punto $2x$. Esiste dunque una sola regola bayesiana formale, precisamente la regola di decisione $\delta'(x) = 2x$ avente funzione di rischio

$$R_{\delta'}(\theta) = \int_{-\infty}^{+\infty} (\theta - 2x)^2 \exp\Big[\frac{3\theta^2}{4}\Big] \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{(x - \theta)^2}{2}\Big] dx = \exp\Big[\frac{3\theta^2}{4}\Big] \mathrm{E}_\theta\big[(2X - \theta)^2\big].$$

Ora, dalla $(2X - \theta)^2 = \big[(2X - 2\theta) + \theta\big]^2 = 4(X - \theta)^2 + 4\theta(X - \theta) + \theta^2$ otteniamo

$$\mathrm{E}_\theta\big[(2X - \theta)^2\big] = 4\mathrm{E}_\theta\big[(X - \theta)^2\big] + 4\theta\,[\mathsf{E}_\theta(X) - \theta] + \theta^2 = 4\mathrm{Var}_\theta(X) + \theta^2 = 4 + \theta^2$$

e quindi $R_{\delta'}(\theta) = \exp\Big[\frac{3\theta^2}{4}\Big](4 + \theta^2)$. Allora,

$$\begin{aligned}
\mathrm{E}_{\mathrm{P}}(R_{\delta'}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (4 + \theta^2) \exp\Big[\frac{3\theta^2}{4} - \frac{\theta^2}{2}\Big] d\theta = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (4 + \theta^2) \exp\Big[\frac{\theta^2}{4}\Big] d\theta \\
&\geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \theta^2 \exp\Big[\frac{\theta^2}{4}\Big] d\theta = \frac{2}{\sqrt{2\pi}} \int_{0}^{+\infty} \theta^2 \exp\Big[\frac{\theta^2}{4}\Big] d\theta \\
&\geq \frac{2}{\sqrt{2\pi}} \int_{1}^{+\infty} \theta \exp\Big[\frac{\theta^2}{4}\Big] d\theta = \frac{4}{\sqrt{2\pi}} \exp\Big[\frac{\theta^2}{4}\Big]\Big|_1^{+\infty} = +\infty.
\end{aligned}$$

Conseguentemente, δ' non è una regola bayesiana pur essendo una regola bayesiana formale. Inoltre, notato che

$$R_{\delta'}(\theta) = \exp\Big[\frac{3\theta^2}{4}\Big](4 + \theta^2) > \exp\Big[\frac{3\theta^2}{4}\Big]$$

per ogni stato θ, non è neppure una regola ammissibile in quanto, come facilmente si verifica, è dominata strettamente dall'identità. ▲

Consideriamo ora una regola bayesiana formale ϕ tale che $\mathrm{L}(\cdot, \phi)$ sia $\mathcal{T} \otimes \aleph$-Borel misurabile. Allora, per il teorema di Tonelli, la funzione di rischio

$R_\phi : \theta \mapsto \mathrm{E}_\theta(\mathrm{L}(\theta, \phi))$ di ϕ è $\mathcal{T}$-Borel misurabile e quindi ammette valor medio $\rho_{\mathrm{P}}(\phi) = \mathrm{E}_{\mathrm{P}}(R_\phi)$; inoltre, ricorrendo anche al Teorema B.2.13(ii), otteniamo

$$\begin{aligned}
\rho_{\mathrm{P}}(\phi) &= \int_\Theta \Big(\int_{\mathfrak{X}} \mathrm{L}(\theta, \phi(x)) f_\theta(x)\, \mu(dx) \Big) \pi(\theta)\, \nu(d\theta) \\
&= \int_{\Theta \times \mathfrak{X}} \mathrm{L}(\theta, \phi(x)) f_\theta(x) \pi(\theta)\, \nu \times \mu(d\theta \times dx) \\
&= \int_{\Theta \times \mathfrak{X}} \mathrm{L}(\theta, \phi(x)) \pi(\theta|x) p(x)\, \nu \times \mu(d\theta \times dx) \\
&= \int_{\mathfrak{X}} \Big(\int_\Theta \mathrm{L}(\theta, \phi(x)) \pi(\theta|x)\, \nu(d\theta) \Big) p(x)\, \mu(dx)
\end{aligned}$$

e quindi la formula di disintegrazione:

$$\rho_{\mathrm{P}}(\phi) = \int_{\mathfrak{X}} \rho_{\mathrm{p}}(\phi(x)|x)\, p(x)\, \mu(dx) = \int_{\mathfrak{X}} \rho_{\mathrm{p}}(\phi(x)|x)\, P^{(\mathrm{pr})}(dx) \tag{2.15}$$

che consente di calcolare il danno medio "misturando" i danni medi finali relativi ai campioni con la probabilità predittiva. Infine, data una regola di decisione δ, si ha $\rho_{\mathrm{P}}(\phi(x)|x) \le \rho_{\mathrm{P}}(\delta(x)|x)$ per ogni campione x. Allora, per (2.15), $\rho_{\mathrm{P}}(\phi) \le \int_{\mathfrak{X}} \rho_{\mathrm{P}}(\delta(x)|x)\, P^{(\mathrm{pr})}(dx) = \rho_{\mathrm{P}}(\delta)$ e quindi, tenuto conto dell'arbitrarietà di δ,

$$\rho_{\mathrm{P}}(\phi) \le \inf_{\delta \in \Delta} \rho_{\mathrm{P}}(\delta) = \rho(\mathrm{P}). \tag{2.16}$$

Conseguentemente, nei problemi statistico-decisionali che ammettono una regola bayesiana formale ϕ tale che $\rho_{\mathrm{P}}(\phi) = +\infty$ (come avviene nell'esempio precedente), il rischio di Bayes di P è infinito e quindi non possono esistere regole bayesiane per P.

Qualora invece il rischio di Bayes di P risulti finito, ogni regola bayesiana formale che sia anche una regola di decisione è, per (2.16), una regola bayesiana; inoltre, come messo in evidenza dal prossimo teorema (di notevole importanza sia teorica che applicativa), tale regola "caratterizza" le regole bayesiane, nel senso che ognuna le è uguale a meno di un insieme di campioni di probabilità predittiva nulla.

Teorema 2.6.3 *Sia il rischio di Bayes di* P *finito. Sussistono allora le seguenti proposizioni:*

(i) *Sia ϕ una regola bayesiana formale per* P*. Allora, è bayesiana per* P *ogni regola di decisione che coincide con ϕ a meno di un insieme di campioni di probabilità predittiva nulla;*

(ii) *Esista una regola di decisione che sia anche una regola bayesiana formale per* P*. Allora, qualunque sia la regola bayesiana δ^* per* P*, ha probabilità predittiva nulla l'insieme $\{x : \rho_{\mathrm{P}}(\delta^*(x)|x) > \inf_{d\in\mathfrak{D}} \rho_{\mathrm{P}}(d|x)\}$ dei campioni ai quali δ^* associa una decisione che non minimizza il relativo danno medio finale.*

Dimostrazione (i) Sia δ^* una regola di decisione tale che l'insieme $\mathfrak{X}_0 = \{\delta^* \neq \phi\}$ abbia probabilità predittiva nulla. Allora, da (2.15), (2.16) si ha

$$\begin{aligned}
\rho_{\mathrm{P}}(\delta^*) &= \int_{\mathfrak{X}} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) = \int_{\mathfrak{X}_0^c} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) \\
&= \int_{\mathfrak{X}_0^c} \rho_{\mathrm{P}}(\phi(x)|x)\, P^{(\mathrm{pr})}(dx) = \int_{\mathfrak{X}} \rho_{\mathrm{P}}(\phi(x)|x)\, P^{(\mathrm{pr})}(dx) \\
&= \rho_{\mathrm{P}}(\phi) \leq \rho(\mathrm{P}) < +\infty
\end{aligned}$$

e quindi δ^* è una regola bayesiana per P.

(ii) Supposto che $\delta' \in \Delta$ sia una regola bayesiana formale per P, consideriamo una regola bayesiana δ^* per P. Proviamo intanto che l'insieme $\mathfrak{X}_1 = \{x : \rho_{\mathrm{P}}(\delta^*(x)|x) > \inf_{d\in\mathfrak{D}} \rho_{\mathrm{P}}(d|x)\}$ appartiene alla σ-algebra $\aleph$. Ricordato che le funzioni $\mathrm{L}(\cdot, \delta^*)$, $\mathrm{L}(\cdot, \delta')$ sono $\mathcal{T} \otimes \aleph$-Borel misurabili, dal teorema di Tonelli otteniamo la $\aleph$-Borel misurabilità delle funzioni:

$$\begin{aligned}
x &\mapsto \int_{\Theta} \mathrm{L}(\theta, \delta^*(x))\pi(\theta|x)\, \nu(d\theta) = \rho(\delta^*(x)|x) \\
x &\mapsto \int_{\Theta} \mathrm{L}(\theta, \delta'(x))\pi(\theta|x)\, \nu(d\theta) = \rho(\delta'(x)|x).
\end{aligned}$$

Conseguentemente, $\mathfrak{X}_1 = \{x : \rho_{\mathrm{P}}(\delta^*(x)|x) > \rho_{\mathrm{P}}(\delta'(x)|x)\} \in \aleph$.

Ora, supposto (per asssurdo) $P^{(\mathrm{pr})}(\mathfrak{X}_1) > 0$, consideriamo la funzione di decisione:

$$\phi(x) = \begin{cases} \delta^*(x) & \text{se } x \notin \mathfrak{X}_1 \\ \delta'(x) & \text{se } x \in \mathfrak{X}_1 \end{cases}.$$

Poichè $\mathrm{L}(\cdot,\phi(x)) = \mathrm{L}(\cdot,\delta'(x))\, I_{\mathfrak{X}_1}(x) + \mathrm{L}(\cdot,\delta^*(x))\, I_{\mathfrak{X}_1^c}(x)$ per ogni campione x, la funzione $\mathrm{L}(\cdot,\phi)$ è $\mathcal{T}\otimes\aleph$-Borel misurabile (Lemma A.5.1(iii)) e inoltre

$$\begin{aligned} R_\phi(\theta) &= \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta')\, I_{\mathfrak{X}_1}\big) + \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta^*)\, I_{\mathfrak{X}_1^c}\big) \\ &\le \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta')\big) + \mathrm{E}_\theta\big(\mathrm{L}(\theta,\delta^*)\big) = R_{\delta'}(\theta) + R_{\delta^*}(\theta) < +\infty \end{aligned}$$

per ogni stato θ. Dunque, ϕ è una regola di decisione. Passando a considerare il relativo danno medio, da (2.15) si ha

$$\begin{aligned} \rho_{\mathrm{P}}(\phi) &= \int_{\mathfrak{X}_1} \rho_{\mathrm{P}}(\phi(x)|x)\, P^{(\mathrm{pr})}(dx) + \int_{\mathfrak{X}_1^c} \rho_{\mathrm{P}}(\phi(x)|x)\, P^{(\mathrm{pr})}(dx) \\ &= \int_{\mathfrak{X}_1} \rho_{\mathrm{P}}(\delta'(x)|x)\, P^{(\mathrm{pr})}(dx) + \int_{\mathfrak{X}_1^c} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) \\ &< \int_{\mathfrak{X}_1} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) + \int_{\mathfrak{X}_1^c} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) \\ &= \int_{\mathfrak{X}} \rho_{\mathrm{P}}(\delta^*(x)|x)\, P^{(\mathrm{pr})}(dx) = \rho_{\mathrm{P}}(\delta^*), \end{aligned}$$

ove la disuguaglianza sussiste in forza del Teorema A.3.6(vi), notato che $\rho_{\mathrm{P}}(\delta^*(x)|x) > \rho_{\mathrm{P}}(\delta'(x)|x)$ per ogni $x \in \mathfrak{X}_1$ e $\int_{\mathfrak{X}_1} \rho_{\mathrm{P}}(\delta'(x)|x)\, P^{(\mathrm{pr})}(dx) < +\infty$. Ricordato infine che $\phi \in \Delta$ e che δ^* è una regola bayesiana per P, otteniamo $\rho_{\mathrm{P}}(\phi) < \rho_{\mathrm{P}}(\delta^*) \le \rho_{\mathrm{P}}(\phi)$ (Contraddizione!). ■

Data la loro importanza pratica, nelle sottosezioni seguenti (ove $\widetilde{\mathrm{P}}$ denota una probabilità iniziale qualsiasi) vengono riportate, indipendentemente dalla specifica forma funzionale della densità iniziale, le regole bayesiane formali relative ad alcune funzioni di danno standard.

2.6.1 Regole bayesiane formali nella stima puntuale

Supposto $\Theta \subseteq \mathfrak{D} \subseteq \mathbb{R}$ con $\mathfrak{D}$ intervallo (limitato o no), assumiamo che il valor medio $\mathrm{E}_{\widetilde{\mathrm{P}}}(Z) = \int_\Theta \theta\, \widetilde{\mathrm{P}}(d\theta)$ sia finito. Allora, per l'internalità della speranza matematica, $\mathrm{E}_{\widetilde{\mathrm{P}}}(Z) \in \mathfrak{D}$. Ciò precisato, andiamo a determinare le regole bayesiane formali relative ai danni quadratico, lineare e assoluto.

• Danno quadratico Supposto che anche il momento secondo $\mathrm{E}_{\widetilde{\mathrm{P}}}(Z^2) = \int_\Theta \theta^2\, \widetilde{\mathrm{P}}(d\theta)$ sia finito, dal Corollario B.2.11, otteniamo che il valor medio $\mathrm{E}_{\widetilde{\mathrm{P}}}(Z)$ è l'unico punto di minimo della funzione $d \mapsto \mathrm{E}_{\widetilde{\mathrm{P}}}(L_d)$. Ciò osservato, proviamo il teorema seguente.

Teorema 2.6.4 *Il momento secondo finale* $\mathrm{E}_{\mathrm{P}}(Z^2|x) = \mathrm{E}_{\mathrm{P}(\cdot|x)}(Z^2)$ *sia finito per ogni campione* x. *Allora, la funzione* $\phi^{(\mathrm{P})}(x) = \mathrm{E}_{\mathrm{P}}(Z|x) = \mathrm{E}_{\mathrm{P}(\cdot|x)}(Z)$, *che associa ad ogni campione il valor medio finale dello stato di natura relativo a quel campione, è l'unica regola bayesiana formale. Inoltre, posto* $\mathrm{Var}_{\mathrm{P}}(Z|x) = \int_\Theta [\theta - \phi^{(\mathrm{P})}(x)]^2\, \mathrm{P}(d\theta|x)$ *per ogni campione* x, *si ha*

$$\rho_{\mathrm{P}}(\phi^{(\mathrm{P})}) = \int_{\mathfrak{X}} \mathrm{Var}_{\mathrm{P}}(Z|x)\, P^{(pr)}(dx)$$

e quindi il danno medio relativo alla regola bayesiana formale si ottiene "misturando" le varianze finali dello stato di natura con la probabilità predittiva.

Dimostrazione Ponendo $\widetilde{\mathrm{P}} = \mathrm{P}(\cdot|x)$ al variare del campione x, otteniamo intanto che $\phi^{(\mathrm{P})}$ è l'unica regola bayesiana formale. Inoltre, osservato che, per il teorema di Fubini A.5.5, $\phi^{(\mathrm{P})}$ è $\aleph$-Borel misurabile, dal Teorema A.5.1(iii) risulta la $\mathcal{T} \otimes \aleph$-Borel misurabilità di $\mathrm{L}(\cdot, \phi^{(\mathrm{P})})$. Da (2.15) si ha allora

$$\begin{aligned}\rho_{\mathrm{P}}(\phi^{(\mathrm{P})}) &= \int_{\mathfrak{X}} \rho_{\mathrm{P}}(\phi^{(\mathrm{P})}(x)|x)\, P^{(\mathrm{pr})}(dx) = \int_{\mathfrak{X}} \Big[\int_\Theta \mathrm{L}(\theta, \phi^{(\mathrm{P})}(x))\, \mathrm{P}(d\theta|x) \Big] P^{(\mathrm{pr})}(dx) \\ &= \int_{\mathfrak{X}} \Big[\int_\Theta \big[\theta - \phi^{(\mathrm{P})}(x)\big]^2 \mathrm{P}(d\theta|x) \Big] P^{(\mathrm{pr})}(dx).\end{aligned}$$

La dimostrazione è così conclusa. ∎

• Danno lineare Sia d' un **quantile di livello** k **relativo a** $\widetilde{\mathrm{P}}$, cioè tale che $\widetilde{\mathrm{P}}(]-\infty, d']) \geq k$ e $\widetilde{\mathrm{P}}([d', +\infty[) \geq 1-k$. Per provare che d' minimizza la funzione $d \mapsto \mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d)$ basta verificare, essendo i danni medi $\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d)$, $\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_{d'})$ finiti ($\mathrm{E}_{\widetilde{\mathrm{P}}}(Z)$ finita!), la disuguaglianza:

$$\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d) - \mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_{d'}) = \mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d - \mathrm{L}_{d'}) \geq 0. \tag{2.17}$$

Sia intanto $d > d'$. Riesce allora

$$\begin{aligned}\mathrm{L}_d(\theta) - \mathrm{L}_{d'}(\theta) &= \begin{cases} (1-k)(d-d') & \text{se } \theta \leq d' \\ k(d'-d) + d - \theta & \text{se } d' < \theta \leq d \\ k(d'-d) & \text{se } \theta > d \end{cases} \\ &= k(d'-d) + \begin{cases} d - d' & \text{se } \theta \leq d' \\ d - \theta & \text{se } d' < \theta \leq d \\ 0 & \text{se } \theta > d \end{cases}\end{aligned}$$

qualunque sia lo stato θ. Ne segue

$$\begin{aligned}
\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d - \mathrm{L}_{d'}) &= k(d'-d) + (d-d')\widetilde{\mathrm{P}}(]-\infty, d']) + \int_{]d',d\,]} (d-\theta)\, \widetilde{\mathrm{P}}(d\theta) \\
&\geq k(d'-d) + (d-d')\widetilde{\mathrm{P}}(]-\infty, d']) \\
&\geq k(d'-d) + k(d-d') = 0,
\end{aligned}$$

ricordato che $\widetilde{\mathrm{P}}(]-\infty, d']) \geq k$. Assunto ora $d < d'$ si ha

$$\begin{aligned}
\mathrm{L}_d(\theta) - \mathrm{L}_{d'}(\theta) &= \begin{cases} (1-k)(d-d') & \text{se } \theta \leq d \\ k(d'-d) + \theta - d' & \text{se } d < \theta < d' \\ k(d'-d) & \text{se } \theta \geq d' \end{cases} \\
&= k(d'-d) + \begin{cases} d-d' & \text{se } \theta \leq d \\ \theta - d' & \text{se } d < \theta < d' \\ 0 & \text{se } \theta \geq d' \end{cases}
\end{aligned}$$

qualunque sia lo stato θ. Ne segue

$$\begin{aligned}
\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d - \mathrm{L}_{d'}) &= k(d'-d) + (d-d')\widetilde{\mathrm{P}}(]-\infty, d]) + \int_{]d,d'[} (\theta-d')\, \widetilde{\mathrm{P}}(d\theta) \\
&\geq k(d'-d) + (d-d')\widetilde{\mathrm{P}}(]-\infty, d]) + (d-d')\widetilde{\mathrm{P}}(]d, d'[) \\
&= k(d'-d) + (d-d')\widetilde{\mathrm{P}}(]-\infty, d'[)
\end{aligned}$$

e quindi

$$\begin{aligned}
\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_d - \mathrm{L}_{d'}) &\geq k(d'-d) + (d-d')\big[1 - \widetilde{\mathrm{P}}([d', +\infty[)\big] \\
&\geq k(d'-d) + k(d-d') = 0,
\end{aligned}$$

ricordato che $\widetilde{\mathrm{P}}([d', +\infty[) \geq 1-k$ e $d - d' < 0$. La dimostrazione della disuguaglianza (2.17) è così conclusa. Sussiste pertanto il risultato seguente.

Teorema 2.6.5 *Il valor medio finale* $\mathrm{E}_{\mathrm{P}}(Z|x)$ *sia finito per ogni campione* x. *Allora, una funzione di decisione che associa ad ogni campione* x *un quantile di livello* k *della distribuzione finale relativa a* x *è una regola bayesiana formale.*

• Danno assoluto Ponendo, in particolare, $k = \frac{1}{2}$ nella funzione di danno lineare, dal teorema precedente otteniamo che sono regole bayesiane formali tutte le funzioni di decisione che associano ad ogni campione x una mediana della distribuzione finale relativa a x.

Nell'esempio seguente troviamo le regole bayesiane e il rischio di Bayes relativi ai campionamenti di Bernoulli e normale (con media incognita), nell'ipotesi che vengano scelte densità iniziali appartenenti alle classi coniugate delle rispettive densità di campionamento e si adotti, per il primo campionamento, il danno quadratico e per il secondo sia quello quadratico che assoluto.

Esempio 2.6.6 (i) Posto $\Theta = \mathfrak{D} = [0,1]$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la distribuzione di Bernoulli *Ber*(θ) in corrispondenza ad ogni stato θ. Scelta come densità iniziale quella della distribuzione *Beta*(α, β), dall'Esempio B.2.18(i) otteniamo che la densità finale relativa a $\mathbf{x}$ è quella della distribuzione *Beta*$(\alpha + n\bar{\mathbf{x}}, \beta + n(1-\bar{\mathbf{x}}))$. Allora, la regola di decisione:

$$\delta_n^{(\alpha,\beta)}(\mathbf{x}) = \frac{n\,\bar{\mathbf{x}} + \alpha}{\alpha + \beta + n} \tag{2.18}$$

è, per il Teorema 2.6.4, l'unica regola bayesiana formale[37]. Inoltre, è anche uno stimatore. Infatti, posto $\delta' = \delta_n^{(\alpha,\beta)}$, si ha[38]

$$\mathrm{Var}_\theta(\delta') = \frac{n^2}{(n\alpha+\beta+1)^2}\,\mathrm{Var}_\theta(\bar{\mathbf{X}}) = \frac{n^2}{(n\alpha+\beta+1)^2}\;\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$
$$\mathrm{Bias}_\theta(\delta') = \mathrm{E}_\theta(\delta') - \theta = \frac{n\,\mathrm{E}_\theta(\bar{\mathbf{X}}) + \alpha}{\alpha+\beta+1} = \frac{n\,\frac{\alpha}{\alpha+\beta} + \alpha}{\alpha+\beta+1}$$

e quindi, per (2.2), $\mathrm{MSE}_{\delta'}(\theta) = \mathrm{Var}_\theta(\delta') + \mathrm{Bias}_\theta(\delta')^2 \in \mathbb{R}$.

Conseguentemente, per il Teorema 2.6.3, $\delta_n^{(\alpha,\beta)}$ è anche uno stimatore bayesiano; inoltre, ogni altro stimatore bayesiano può essere ottenuto modificando opportunamente $\delta_n^{(\alpha,\beta)}$ su un insieme di probabilità predittiva nulla (come lo sono, ad esempio, gli insiemi discreti e l'insieme continuo di Cantor). Infine, osservato che gli errori quadratici medi

$$\mathrm{MSE}_\delta(\theta) = \sum_{\mathbf{x}\in \mathrm{X}} \mathrm{L}(\theta, \delta(\mathbf{x}))\,\theta^{n\,\bar{\mathbf{x}}}\,(1-\theta)^{n(1-\bar{\mathbf{x}})}$$

sono delle funzioni continue, che l'indice di preferibilità individuato dalla speranza matematica è P-strettamente monotono (Teorema 2.4.7) e che la probabilità P verifica la situazione (b) del Teorema 2.4.6, possiamo concludere che tutti gli stimatori bayesiani sono delle regole ammissibili (e quindi, in particolare, anche $\delta_n^{(\alpha,\beta)}$).

Passando a considerare infine il rischio di Bayes $\rho(\alpha,\beta)$ della probabilità iniziale, dal Teorema 2.6.4 otteniamo

$$\rho(\alpha,\beta) = \mathrm{E}_{P(\mathrm{pr})}\left(\frac{(\alpha+n\bar{\mathbf{X}})\left[\beta+n(1-\bar{\mathbf{X}})\right]}{(\alpha+\beta+n+1)(\alpha+\beta+n)^2}\right) \le \frac{(\alpha+n)(\beta+n)}{(\alpha+\beta+n+1)(\alpha+\beta+n)^2}$$

[37] Ricordiamo che la distribuzione *Beta*(α_1, β_1) ha valor medio $\frac{\alpha_1}{\alpha_1+\beta_1}$ e varianza $\frac{\alpha_1\beta_1}{(\alpha_1+\beta_1+1)(\alpha_1+\beta_1)^2}$.

[38] Ricordiamo che, considerate le v.a. $Y_1, \ldots, Y_n$ indipendenti e equidistribuite, la media campionaria $\bar{\mathbf{Y}}$ ha valor medio $\mathrm{E}(\bar{\mathbf{Y}}) = \mathrm{E}(Y_1)$ e varianza $\mathrm{Var}(\bar{\mathbf{Y}}) = \frac{\mathrm{Var}(Y_1)}{n}$.

e quindi il danno medio ineliminabile è un infinitesimo al divergere del numero delle osservazioni.

(ii) Posto $\Theta = \mathfrak{D} = \mathbb{R}$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ. Scelta come densità iniziale quella della normale $\mathsf{N}(\alpha, \beta^2)$, dall'Esempio B.2.18(ii) otteniamo che la densità finale relativa a $\mathbf{x}$ è quella della normale $\mathsf{N}(\alpha_n(\mathbf{x}), \beta_n^2)$, con

$$\alpha_n(\mathbf{x}) = \Big[\frac{\alpha}{\beta^2} + \frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]\beta_n^2, \quad \beta_n^2 = \Big(\frac{1}{\beta^2} + \frac{n}{\sigma^2}\Big)^{-1}.$$

Allora, la funzione di decisione:

$$\delta_n^{(\alpha,\beta^2)}(\mathbf{x}) = \alpha_n(\mathbf{x}) = \frac{\frac{n}{\sigma^2}\,\bar{\mathbf{x}} + \frac{1}{\beta^2}\alpha}{\frac{n}{\sigma^2} + \frac{1}{\beta^2}} = \frac{n\beta^2\,\bar{\mathbf{x}} + \sigma^2\alpha}{n\beta^2 + \sigma^2}, \tag{2.19}$$

ottenuta facendo la media aritmetica ponderata della media campionaria e del valor medio iniziale dello stato di natura, è l'unica regola bayesiana formale (Teorema 2.6.4). Inoltre, è anche uno stimatore. Infatti, posto $\delta' = \delta_n^{(\alpha,\beta^2)}$, si ha

$$\begin{aligned}\mathrm{Var}_\theta(\delta') &= \frac{n^2\beta^4}{(n\beta^2+\sigma^2)^2}\,\mathrm{Var}_\theta(\bar{\mathbf{X}}) = \frac{n^2\beta^4}{(n\beta^2+\sigma^2)^2}\,\frac{\sigma^2}{n}\\ \mathrm{Bias}_\theta(\delta') &= \mathrm{E}_\theta(\delta') - \theta = \frac{n\beta^2\,\mathrm{E}_\theta(\bar{\mathbf{X}}) + \sigma^2\alpha}{n\beta^2+\sigma^2} = \frac{n\beta^2\,\theta + \sigma^2\alpha}{n\beta^2+\sigma^2}\end{aligned}$$

e quindi, per (2.2), $\mathrm{MSE}_{\delta'}(\theta) = \mathrm{Var}_\theta(\delta') + \mathrm{Bias}_\theta(\delta')^2 \in \mathbb{R}$.

Conseguentemente, per il Teorema 2.6.3, $\delta_n^{(\alpha,\beta^2)}$ è anche uno stimatore bayesiano; inoltre, ogni altro stimatore bayesiano può essere ottenuto modificando opportunamente $\delta_n^{(\alpha,\beta^2)}$ su un insieme di probabilità predittiva nulla. Infine, ricordato che l'indice di preferibilità individuato dalla speranza matematica è P-strettamente monotono (Teorema 2.4.7), che le funzioni di rischio degli stimatori sono funzioni continue (Esempio 2.1.3(ii)) e che la probabilità P verifica la situazione (b) del Teorema 2.4.6, possiamo concludere che tutti gli stimatori bayesiani sono ammissibili (e quindi, in particolare, anche $\delta_n^{(\alpha,\beta^2)}$).

Passando a considerare infine il rischio di Bayes $\rho(\alpha, \beta^2)$ della probabilità iniziale, dal Teorema 2.6.4 otteniamo

$$\rho(\alpha, \beta^2) = \beta_n^2 = \frac{\beta^2\sigma^2}{\sigma^2 + n\beta^2} \tag{2.20}$$

e quindi il danno medio ineliminabile è un infinitesimo al divergere del numero delle osservazioni.

(iii) Con riferimento al campionamento considerato in (ii), assumiamo al posto del danno quadratico quello assoluto. Allora, per la simmetria della distribuzione normale, $\delta' = \delta_n^{(\alpha,\beta^2)}$ è una regola bayesiana che è anche l'unica regola bayesiana formale. Ciò osservato, dato un campione $\mathbf{x}$, risulta

$$\rho_{\mathrm{P}}(\delta'(\mathbf{x})|\mathbf{x}) = \int_\Theta \big|\theta - \alpha_n(\mathbf{x})\big|\,\mathrm{P}(d\theta|\mathbf{x}) = \frac{1}{\sqrt{2\pi\beta_n^2}}\int_{-\infty}^{+\infty} \big|\theta - \alpha_n(\mathbf{x})\big| \exp\Big[-\frac{(\theta - \alpha_n(\mathbf{x}))^2}{2\beta_n^2}\Big]\,d\theta.$$

Considerata la traslazione $\tau : \theta \mapsto \theta - \alpha_n(\mathbf{x})$ della retta reale, dal Corollario A.4.2 si ha

$$\rho_{\mathrm{P}}(\delta'(\mathbf{x})|\mathbf{x}) = \frac{1}{\sqrt{2\pi\beta_n^2}} \int_{-\infty}^{+\infty} |\theta| \exp\Big[-\frac{\theta^2}{2\beta_n^2}\Big] d\theta$$

e quindi, per la simmetria della distribuzione normale,

$$\begin{aligned} \rho_{\mathrm{P}}(\delta'(\mathbf{x})|\mathbf{x}) &= \frac{2}{\sqrt{2\pi\beta_n^2}} \int_{0}^{+\infty} \theta \exp\Big[-\frac{\theta^2}{2\beta_n^2}\Big] d\theta = -\frac{2\beta_n^2}{\sqrt{2\pi\beta_n^2}} \exp\Big[-\frac{\theta^2}{2\beta_n^2}\Big]\Big|_0^{+\infty} \\ &= \frac{2\beta_n^2}{\sqrt{2\pi\beta_n^2}} = \sqrt{\frac{2\beta_n^2}{\pi}} = \sqrt{\frac{2\beta^2\sigma^2}{\pi(\sigma^2+n\beta^2)}}. \end{aligned}$$

Ne segue, tramite (2.15), che $\sqrt{\frac{2\beta^2\sigma^2}{\pi(\sigma^2+n\beta^2)}}$ è il rischio di Bayes della probabilità iniziale. ▲

Nei problemi di stima puntuale appare naturale richiedere che lo stimatore non produca deviazioni sistematiche rispetto allo stato di natura. Ne viene che, negli usuali problemi di stima di un parametro reale, si considerano di norma stimatori δ **non distorti**, cioè tali che $\mathrm{E}_\theta(\delta) = \theta$ per ogni stato θ[39]. Il prossimo teorema mostra che l'esistenza di uno stimatore che sia bayesiano formale e non distorto è una situazione ideale, nell'ambito del danno quadratico, in quanto assicura un rischio di Bayes nullo.

Teorema 2.6.7 *Sia δ_1 uno stimatore $\aleph$-Borel misurabile, non distorto e bayesiano formale per* P *rispetto al danno quadratico. Inoltre, sia* $\mathrm{E}_{\mathrm{P}}(Z^2)$ *finita. Allora,* $\rho(\mathrm{P}) = 0$.

Dimostrazione Indicata con P la probabilità dello schema di Bayes $P^{(\mathrm{sb})}$, dal teorema di Tonelli e (2.16) otteniamo

$$\begin{aligned} \rho(\mathrm{P}) = \rho_{\mathrm{P}}(\delta_1) &= \int_\Theta \Big(\int_{\mathfrak{X}} \mathrm{L}(\theta, \delta_1(x)) f_\theta(x)\, \mu(dx)\Big) \pi(\theta)\, \nu(d\theta) \\ &= \int_{\Theta\times\mathfrak{X}} \mathrm{L}(\theta, \delta_1(x)) f_\theta(x) \pi(\theta)\, \nu\times\mu(d\theta\times dx) \\ &= \int_{\Theta\times\mathfrak{X}} (\theta - \delta_1(x))^2\, P(d\theta\times dx) = \mathrm{E}_P((Z-\delta_1)^2). \end{aligned}$$

[39]Ottenendo così sia un buon comportamento in media che una semplificazione dell'errore quadratico medio che viene, per (2.2), a coincidere su ogni stato con la varianza dello stimatore relativa a quello stato.

Proviamo ora che esiste finita $\mathrm{E}_P(Z\delta_1)$. Per i teoremi di Tonelli e B.2.13(ii),

$$\begin{aligned}\mathrm{E}_{\mathrm{P}}(Z^2) &= \int_\Theta \Big(\int_{\mathfrak{X}} f_\theta(x)\,\mu(dx)\Big)\theta^2\pi(\theta)\,\nu(d\theta)\\ &= \int_{\mathfrak{X}}\Big(\int_\Theta \theta^2\pi(\theta|x)\,\nu(d\theta)\Big)p(x)\,\mu(dx) = \int_{\mathfrak{X}} \mathrm{E}_{\mathrm{P}}(Z^2|x)\,P^{(\mathrm{pr})}(dx)\end{aligned}$$

e quindi, per il Teorema A.3.6(vii), $P^{(\mathrm{pr})}(\mathfrak{X}_0) = 1$, avendo posto $\mathfrak{X}_0 = \{x : \mathrm{E}_{\mathrm{P}}(Z^2|x) \text{ finita}\}$. Notato che $\delta_1(x) = \mathrm{E}_{\mathrm{P}}(Z|x)$ per ogni $x \in \mathfrak{X}_0$ (δ_1 stimatore bayesiano formale!), dal teorema di Jensen B.2.5 otteniamo $\delta_1^2(x) = \mathrm{E}_{\mathrm{P}}(Z|x)^2 \leq \mathrm{E}_{\mathrm{P}}(Z^2|x)$ per ogni $x \in \mathfrak{X}_0$ e quindi

$$\begin{aligned}\int_{\mathfrak{X}} \delta_1^2\,dP^{(\mathrm{pr})} &= \int_{\mathfrak{X}_0} \delta_1^2\,dP^{(\mathrm{pr})}\\ &\leq \int_{\mathfrak{X}_0} \mathrm{E}_{\mathrm{P}}(Z^2|x)\,P^{(\mathrm{pr})}(dx) = \int_{\mathfrak{X}} \mathrm{E}_{\mathrm{P}}(Z^2|x)\,P^{(\mathrm{pr})}(dx) < +\infty.\end{aligned}$$

Allora, osservato infine che per i teoremi di Tonelli e B.2.13(ii), $\mathrm{E}_P(Z^2) = \mathrm{E}_{\mathrm{P}}(Z^2)$ e $\mathrm{E}_P(\delta_1^2) = \mathrm{E}_{P^{(\mathrm{pr})}}(\delta_1^2)$, esiste ed è finita, per il Teorema A.3.5(iv), la speranza matematica $\mathrm{E}_P(Z\delta_1)$.

Riesce pertanto

$$\rho(\mathrm{P}) = \mathrm{E}_P((Z-\delta_1)^2) = \big[\mathrm{E}_P(Z^2) - \mathrm{E}_P(Z\delta_1)\big] + \big[\mathrm{E}_P(\delta_1^2) - \mathrm{E}_P(Z\delta_1)\big]$$

e quindi basta verificare che sono nulli i due addendi. Risulta

$$\begin{aligned}\mathrm{E}_P(Z\delta_1) &= \int_{\Theta\times\mathfrak{X}} \theta\delta_1(x) f_\theta(x)\pi(\theta)\,\nu\times\mu(d\theta\times dx)\\ &= \int_\Theta\Big(\int_{\mathfrak{X}} \delta_1(x) f_\theta(x)\,\mu(dx)\Big)\theta\,\mathrm{P}(d\theta) = \int_\Theta \mathrm{E}_\theta(\delta_1)\,\theta\,\mathrm{P}(d\theta)\end{aligned}$$

e quindi $\mathrm{E}_P(Z\delta_1) = \int_\Theta \theta^2\,\mathrm{P}(d\theta) = \mathrm{E}_P(Z^2)$ (δ_1 non distorto!). Inoltre,

$$\begin{aligned}\mathrm{E}_P(Z\delta_1) &= \int_{\Theta\times\mathfrak{X}} \theta\delta_1(x)\pi(\theta|x)p(x)\,\nu\times\mu(d\theta\times dx)\\ &= \int_{\mathfrak{X}}\Big(\int_\Theta \theta\pi(\theta|x)\,\nu(d\theta)\Big)\delta_1(x)\,P^{(\mathrm{pr})}(dx) = \int_{\mathfrak{X}} \mathrm{E}_{\mathrm{P}}(Z|x)\,\delta_1(x)\,P^{(\mathrm{pr})}(dx)\end{aligned}$$

e quindi $\mathrm{E}_P(Z\delta_1) = \int_{\mathfrak{X}} \delta_1^2\,dP^{(\mathrm{pr})} = \mathrm{E}_P(\delta_1^2)$. ■

Nel corso della dimostrazione si è provato che $\rho(\mathrm{P}) = \mathrm{E}_P((Z - \delta_1)^2)$. L'annullarsi del rischio di Bayes implica quindi, per il Teorema A.3.6(v), l'uguaglianza quasi certa di δ_1 e Z, cioè che la stima dello stato di natura è quasi certamente perfetta. Poichè una situazione così ottimale non è realizzabile nei problemi statistico-decisionali reali, il teorema mette in evidenza l'incompatibilità pratica delle due nozioni di non distorsione e di bayesianeità. A titolo d'esempio, mostriamo che la media campionaria nel campionamento di Bernoulli è uno stimatore bayesiano solamente nel caso piuttosto "estremo" che la probabilità iniziale concentri l'intera massa sull'insieme $\{0, 1\}$.

Esempio 2.6.8 Posto $\Theta = \mathfrak{D} = [0, 1]$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la distribuzione di Bernoulli $Ber(\theta)$ in corrispondenza ad ogni stato θ. Osservato che la media campionaria è uno stimatore non distorto, andiamo a determinare le probabilità iniziali per le quali è uno stimatore bayesiano formale. Per (2.2), $\rho_{\mathrm{P}}(\overline{\mathbf{X}}) = \int_{[0,1]} \mathrm{Var}_\theta(\overline{\mathbf{X}})\, \mathrm{P}(d\theta) = \frac{1}{n} \int_{[0,1]} \theta(1-\theta)\, \mathrm{P}(d\theta)$. Osservato che l'integrale, per il Teorema A.3.6(vi), si annulla se e solo se $\mathrm{P}(]0, 1[) = 0$, la probabilità P deve depositare l'intera massa su $\{0, 1\}$. In questo caso, data una regola bayesiana δ' (formale o no), da (2.16) otteniamo $\rho_{\mathrm{P}}(\delta') = \mathrm{MSE}_{\delta'}(0)\mathrm{P}(0) + \mathrm{MSE}_{\delta'}(1)\mathrm{P}(1) = 0$ e quindi $\mathrm{MSE}_{\delta'}(i) = 0$, se $\mathrm{P}(i) > 0$ $(i = 0, 1)$. Osservato che $\mathrm{MSE}_{\delta'}(0) = \delta'(\mathbf{0})^2$ e $\mathrm{MSE}_{\delta'}(1) = (1 - \delta'((1, \ldots, 1)))^2$, otteniamo $\delta'(\mathbf{0}) = 0$, se $\mathrm{P}(0) > 0$, e $\delta'((1, \ldots, 1)) = 1$, se $\mathrm{P}(1) > 0$. Poichè, viceversa, per questo tipo di probabilità ogni stimatore che verifica queste due condizioni è banalmente bayesiano, possiamo concludere che lo è anche la media campionaria. ▲

Passando al problema della stima di un parametro vettoriale, non si incontrano novità sostanziali in quanto le tre funzioni di danno considerate per il caso unidimensionale si possono estendere opportunamente a quello multidimensionale. Mostriamolo per il danno quadratico considerando il caso bivariato[40]. Identifichiamo dunque lo stato di natura con la coppia aleatoria (Z_1, Z_2) e assumiamo che ν_i sia una misura σ-finita di riferimento sui boreliani dell'insieme $\Theta_i \subseteq \mathbb{R}$ dei valori possibili di Z_i $(i = 1, 2)$. Posto $\Theta = \Theta_1 \times \Theta_2$ e $\mathcal{T} = \mathcal{B}^{(2)} \cap \Theta$, supponiamo che $\tilde{\pi}$ sia una densità della probabilità congiunta iniziale $\widetilde{\mathrm{P}}$ (rispetto la misura prodotto $\nu = \nu_1 \times \nu_2$) e che il valor medio $\mathrm{E}_{\widetilde{\mathrm{P}}}(Z_i) = \int_{\Theta_i} \theta_i \tilde{\pi}(\theta_1, \theta_2)\, \nu(d\theta_1 \times d\theta_2)$ sia finito $(i = 1, 2)$. Per quanto riguarda la parte decisionale, identificate le m-ple con i vettori-colonna, assumiamo $\Theta \subseteq \mathfrak{D} \subseteq \mathbb{R}^m$, con $\mathfrak{D}$ insieme convesso chiuso, e come funzione di danno la forma quadratica (nelle componenti del vettore $\boldsymbol{\theta} - \mathbf{d}$):

$$\mathrm{L}(\boldsymbol{\theta}, \mathbf{d}) = (\boldsymbol{\theta} - \mathbf{d})^T\, \mathrm{M}\, (\boldsymbol{\theta} - \mathbf{d}), \tag{2.21}$$

[40] Che non comporta alcuna restrizione sostanziale, in quanto le relative argomentazioni sono immediatamente estendibili a dimensioni superiori.

ove $\mathrm{M} = (a_{ij})_{1\leq i,j\leq 2}$ è una matrice simmetrica definita positiva[41]. Posto infine $\mathbf{d}' = \binom{\mathrm{E}_{\widetilde{\mathrm{P}}}(Z_1)}{\mathrm{E}_{\widetilde{\mathrm{P}}}(Z_2)}$, dalla bilinearità del prodotto righe per colonne si ha

$$\begin{aligned}\mathrm{L}(\boldsymbol{\theta}, \mathbf{d}) &= \big[(\mathbf{d}' - \mathbf{d}) + (\boldsymbol{\theta} - \mathbf{d}')\big]^T \mathrm{M} \big[(\mathbf{d}' - \mathbf{d}) + (\boldsymbol{\theta} - \mathbf{d}')\big] \\ &= \big[(\mathbf{d}' - \mathbf{d})^T + (\boldsymbol{\theta} - \mathbf{d}')^T\big] \mathrm{M} \big[(\mathbf{d}' - \mathbf{d}) + (\boldsymbol{\theta} - \mathbf{d}')\big] \\ &= (\mathbf{d}' - \mathbf{d})^T \mathrm{M} (\mathbf{d}' - \mathbf{d}) \\ &\qquad + 2(\mathbf{d}' - \mathbf{d})^T \mathrm{M} (\boldsymbol{\theta} - \mathbf{d}') + (\boldsymbol{\theta} - \mathbf{d}')^T \mathrm{M} (\boldsymbol{\theta} - \mathbf{d}') \\ &= (\mathbf{d}' - \mathbf{d})^T \mathrm{M} (\mathbf{d}' - \mathbf{d}) \\ &\qquad + 2 \sum_{i=1}^{2} \sum_{j=1}^{2} a_{ij}(d'_i - d_i)(\theta_j - d'_j)) + (\boldsymbol{\theta} - \mathbf{d}')^T \mathrm{M} (\boldsymbol{\theta} - \mathbf{d}').\end{aligned}$$

Osservato che

$$\begin{aligned}\int_\Theta (\theta_j - d'_j)\, \widetilde{\mathrm{P}}(d\theta_1 \times d\theta_2) &= \int_\Theta \theta_j \tilde{\pi}(\theta_1, \theta_2)\, \nu(d\theta_1 \times d\theta_2) - d'_j \\ &= \mathrm{E}_{\widetilde{\mathrm{P}}}(Z_j) - d'_j = 0,\end{aligned}$$

otteniamo

$$\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}(\cdot, \mathbf{d})) = (\mathbf{d}' - \mathbf{d})^T \mathrm{M} (\mathbf{d}' - \mathbf{d}) + \mathrm{E}_{\widetilde{\mathrm{P}}}\big((\cdot - \mathbf{d}')^T \mathrm{M} (\cdot - \mathbf{d}')\big).$$

Conseguentemente, tenuto conto che la matrice M è definita positiva, $\mathbf{d}'$ è l'unico punto di minimo della funzione $\mathbf{d} \mapsto \mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}(\cdot, \mathbf{d}))$, indipendentemente dalla forma particolare della matrice M. Per provare infine che $\mathbf{d}'$ è un elemento di $\mathfrak{D}$, assumiamo (per assurdo) che non vi appartenga. Esistono allora, per il teorema di separazione punto-insieme convesso chiuso[42], una coppia $\mathbf{a} \neq \mathbf{0}$ e un numero reale α tali che $a_1 d'_1 + a_2 d'_2 > \alpha$ e $a_1 d_1 + a_2 d_2 \leq \alpha$ per ogni $\mathbf{d} \in \mathfrak{D}$. Considerata la v.a. $Y = a_1 Z_1 + a_2 Z_2$, dalla seconda disuguaglianza otteniamo $\widetilde{\mathrm{P}}(Y \leq \alpha) = 1$ e quindi $\mathrm{E}_{\widetilde{\mathrm{P}}}(Y) \leq \alpha$; ne segue $a_1 d'_1 + a_2 d'_2 \leq \alpha$ contraddicendo così la prima disuguaglianza.

Sussiste dunque il risultato seguente.

[41] Cioè, tale che $\mathbf{x}^T \mathrm{M}\, \mathbf{x} > 0$ per ogni vettore non nullo $\mathbf{x}$, avendo indicato con $\mathbf{x}^T$ il vettore-riga trasposto di $\mathbf{x}$ e considerato, come prodotto di matrici, il prodotto righe per colonne.

[42] Per *teorema di separazione punto-insieme convesso chiuso* intendiamo il seguente risultato: Siano $B \neq \emptyset$ un insieme convesso chiuso di $\mathbb{R}^m$ e $\mathbf{x}' \in \mathbb{R}^m$ tali che $\mathbf{x}' \notin B$. Esistono allora una m-pla $\mathbf{a} \neq \mathbf{0}$ e un numero reale α tali che $a_1 x'_1 + \cdots + a_m x'_m > \alpha$ e $a_1 y_1 + \cdots + a_m y_m \leq \alpha$ per ogni $\mathbf{y} \in B$ (per una dimostrazione si veda il teorema 2.4.4 del testo citato nella nota 35 di pag. 76).

Teorema 2.6.9 *Il valor medio finale* $\mathrm{E_P}(Z_i|x)$ *sia finito per ogni campione* x $(i = 1, 2)$*. Allora, considerata la funzione di danno (2.21), esiste una sola regola bayesiana formale; precisamente, la funzione di decisione:*

$$\phi^{(\mathrm{P})}(x) = \begin{pmatrix} \mathrm{E_P}(Z_1|x) \\ \mathrm{E_P}(Z_2|x) \end{pmatrix}$$

che associa ad ogni campione il vettore dei valori medi finali dello stato di natura relativi a quel campione.

Esempio 2.6.10 Posto $\Theta = \mathfrak{D} = \mathbb{R}\times\,]0, +\infty]$, sia l'osservabile costituito dalle v.a. X_1, $\ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta_1, \theta_2)$ in corrispondenza ad ogni stato $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Scelta come densità iniziale quella della normale-gamma inversa N-$\Gamma^{-1}(\alpha, \beta, \gamma, \delta)$, dall'Esempio B.2.18(iii) otteniamo che la densità finale relativa a $\mathbf{x}$ è quella della distribuzione normale-gamma inversa N-$\Gamma^{-1}(\alpha_n(\mathbf{x}), \beta_n, \gamma_n, \delta_n(\mathbf{x}))$, con

$$\alpha_n(\mathbf{x}) = \frac{n\bar{\mathbf{x}} + \alpha\beta}{\beta + n}, \quad \beta_n = \beta + n, \quad \gamma_n = \gamma + \frac{n}{2},$$

$$\delta_n(\mathbf{x}) = \delta + \frac{1}{2}\Big(\sum_{i=1}^{n}(x_i - \bar{\mathbf{x}})^2 + \frac{n\beta(\bar{\mathbf{x}} - \alpha)^2}{\beta + n}\Big).$$

Conseguentemente, per la densità marginale finale $\pi_1(\cdot|\mathbf{x})$ della prima componente si ha

$$\begin{aligned}
\pi_1(\theta_1|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)} \int_{]0,+\infty[} \theta_2^{-(\gamma_n+\frac{3}{2})} \exp\Big[-\frac{\beta_n(\theta_1 - \alpha_n(\mathbf{x}))^2 + 2\delta_n(\mathbf{x})}{2\theta_2}\Big]\, d\theta_2 \\
&= \frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)}\, \Gamma\big(\gamma_n + \frac{1}{2}\big) \Big[\frac{\beta_n(\theta_1 - \alpha_n(\mathbf{x}))^2 + 2\delta_n(\mathbf{x})}{2}\Big]^{-(\gamma_n+\frac{1}{2})} \\
&= \frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)}\, \Gamma\big(\gamma_n + \frac{1}{2}\big)\, \delta_n(\mathbf{x})^{-(\gamma_n+\frac{1}{2})} \Big[\frac{(\theta_1 - \alpha_n(\mathbf{x}))^2}{2\gamma_n\,\frac{\delta_n(\mathbf{x})}{\beta_n\gamma_n}} + 1\Big]^{-\frac{2\gamma_n+1}{2}}
\end{aligned}$$

e quindi la distribuzione finale della prima componente è una distribuzione di Student generalizzata StudentGen$(\alpha_1, \beta_1, \gamma_1^2)$ con $\alpha_1 = 2\gamma_n > 1$, $\beta_1 = \alpha_n(\mathbf{x})$ e $\gamma_1^2 = \frac{\delta_n(\mathbf{x})}{\beta_n\gamma_n}$.

Per quanto riguarda la densità marginale finale $\pi_2(\cdot|\mathbf{x})$ della seconda componente otteniamo

$$\begin{aligned}
\pi_2(\theta_2|\mathbf{x}) &= \frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)}\, \theta_2^{-(\gamma_n+\frac{3}{2})} \exp\Big[-\frac{\delta_n(\mathbf{x})}{\theta_2}\Big] \int_{-\infty}^{+\infty} \exp\Big[-\frac{\beta_n(\theta_1 - \alpha_n(\mathbf{x}))^2}{2\theta_2}\Big]\, d\theta_1 \\
&= \frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)}\, \theta_2^{-(\gamma_n+\frac{3}{2})} \exp\Big[-\frac{\delta_n(\mathbf{x})}{\theta_2}\Big] \sqrt{\frac{2\pi\theta_2}{\beta_n}} \\
&= \frac{\delta_n(\mathbf{x})^{\gamma_n}}{\Gamma(\gamma_n)}\, \theta_2^{-(\gamma_n+1)} \exp\Big[-\frac{\delta_n(\mathbf{x})}{\theta_2}\Big]
\end{aligned}$$

e quindi la distribuzione finale della seconda componente è una distribuzione gamma-inversa $\Gamma^{-1}(\alpha_1, \beta_1)$ con $\alpha_1 = \gamma_n > 0$ e $\beta_1 = \delta_n(\mathbf{x})$.

Possiamo quindi concludere che la funzione di decisione

$$\phi^{(\mathrm{P})}(\mathbf{x}) = \begin{pmatrix} \alpha_n(\mathbf{x}) \\ \frac{\delta_n(\mathbf{x})}{\gamma_n - 1} \end{pmatrix}$$

è l'unica regola bayesiana formale, ogniqualvolta risulti $\gamma + \frac{n}{2} = \gamma_n > 1$. ▲

2.6.2 Regole bayesiane formali nella stima intervallare

Identificata la misura di riferimento ν con la misura di Lebesgue, supponiamo che la probabilità $\widetilde{\mathrm{P}}$ ammetta una ν-densità $\widetilde{\pi}$ che sia continua e unimodale (in senso forte)[43]. Allora, il seguente problema di massimo vincolato:

$$\begin{cases} \max \displaystyle\int_a^b \tilde{\pi}(\theta)\, d\theta \\ b - a = \alpha \end{cases} \tag{2.22}$$

ammette una sola soluzione per ogni numero reale $\alpha > 0$. Osservato che sia la funzione obiettivo che quella relativa al vincolo hanno derivate parziali continue, usiamo il metodo dei moltiplicatori di Lagrange per individuarne le soluzioni. Considerata la funzione $F(a, b, \lambda_1) = \int_a^b \tilde{\pi}(\theta)\, d\theta + \lambda_1(b - a - \alpha)$, poniamo dunque

$$\begin{aligned} \frac{\partial}{\partial a} F(a, b, \lambda_1) &= -\tilde{\pi}(a) - \lambda_1 = 0 \\ \frac{\partial}{\partial b} F(a, b, \lambda_1) &= \tilde{\pi}(b) + \lambda_1 = 0 \\ \frac{\partial}{\partial \lambda_1} F(a, b, \lambda_1) &= b - a - \alpha = 0 \end{aligned}$$

da cui otteniamo

$$\begin{cases} \tilde{\pi}(a) = \tilde{\pi}(b) \\ b - a = \alpha \end{cases}$$

[43]Ricordiamo che una funzione g su Θ è unimodale (in senso forte) se esiste un punto di massimo θ' tale che $g(\theta_1) < g(\theta_2)$, se $\theta_1 < \theta_2 < \theta'$, e $g(\theta_1) > g(\theta_2)$, se $\theta' < \theta_1 < \theta_2$.

e quindi la densità $\tilde{\pi}$ assume gli stessi valori sugli estremi dell'intervallo. Conseguentemente, per la unimodularità della densità, l'intervallo $[a, a + \alpha]$ tale che $\tilde{\pi}(a) = \tilde{\pi}(a + \alpha)$ è l'unica soluzione del problema (2.22)[44].

Ciò osservato, siamo ora in grado di provare il teorema seguente.

Teorema 2.6.11 *Siano ν la misura di Lebesgue e la densità finale $\pi(\cdot\,|x)$ una funzione continua e unimodale (in senso forte) per ogni campione x. Inoltre, con riferimento al danno lineare, sia ϕ' una regola bayesiana formale per* P. *Allora, la funzione di decisione che associa ad ogni campione x la soluzione del problema (2.22) con $\tilde{\pi} = \pi(\cdot\,|x)$ e $\alpha = \mathrm{lg}(\phi'(x))$ è una regola bayesiana formale.*

Dimostrazione Indicata con ϕ'' tale funzione di decisione, risulta

$$\begin{aligned}\rho_{\mathrm{P}}(\phi'(x)|x) \leq \rho_{\mathrm{P}}(\phi''(x)|x) &= k\mathrm{lg}(\phi''(x)) + 1 - \int_{\phi''(x)} \pi(\theta|x)\, d\theta \\ &\leq k\mathrm{lg}(\phi'(x)) + 1 - \int_{\phi'(x)} \pi(\theta|x)\, d\theta = \rho_{\mathrm{P}}(\phi'(x)|x)\end{aligned}$$

per ogni campione x. ■

Le regole bayesiane formali vanno quindi ricercate tra le funzioni di decisione ϕ che associano ad ogni campione x l'intervallo di massima probabilità finale tra tutti quelli di lunghezza $\mathrm{lg}(\phi(x))$.

Esempio 2.6.12 Posto $\Theta = \mathbb{R}$ e $\mathfrak{D}$ l'insieme degli intervalli reali limitati, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ. Scelta come densità iniziale quella della normale $\mathsf{N}(\alpha, \beta^2)$,

[44] Osservato che l'unica soluzione del problema di minimo vincolato:

$$\begin{cases} \min(b-a) \\ \displaystyle\int_a^b \tilde{\pi}(\theta)\, d\theta = \beta \end{cases}$$

con $\beta \in\,]0,1]$ verifica la condizione $\tilde{\pi}(a) = \tilde{\pi}(b)$, gli intervalli $[a,b]$ con $\tilde{\pi}(a) = \tilde{\pi}(b)$ possono essere intesi come gli intervalli che sono:
− di massima probabilità tra quelli di lunghezza α;
− di minima lunghezza tra quelli di probabilità β.
Inoltre, per la unimodularità, possono anche essere visti come insiemi del tipo $\{\tilde{\pi} \geq \vartheta\}$ con $\vartheta \in [0,1]$.

dall'Esempio B.2.18(ii) otteniamo che la densità finale relativa a $\mathbf{x}$ è quella della normale $\mathsf{N}(\alpha_n(\mathbf{x}), \beta_n^2)$, con

$$\alpha_n(\mathbf{x}) = \Big[\frac{\alpha}{\beta^2} + \frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]\beta_n^2, \quad \beta_n^2 = \Big(\frac{1}{\beta^2} + \frac{n}{\sigma^2}\Big)^{-1}.$$

Ponendo $\pi(a|\mathbf{x}) = \pi(b|\mathbf{x})$ $(a < b)$ si ha $(a - \alpha_n(\mathbf{x}))^2 = (b - \alpha_n(\mathbf{x}))^2$ da cui otteniamo $(a-b)(a+b-2\alpha_n(\mathbf{x})) = 0$ e quindi $b = 2\alpha_n(\mathbf{x}) - a$. Individuata la relazione intercorrente tra a e b, andiamo a determinare il valore del corrispondente integrale

$$\frac{1}{\sqrt{2\pi\beta_n^2}} \int_a^{2\alpha_n(\mathbf{x})-a} \exp\Big[-\frac{(\theta - \alpha_n(\mathbf{x}))^2}{2\beta_n^2}\Big]\, d\theta.$$

Posto uguale a $1 - \xi(\mathbf{x})$ tale valore e considerata la trasformazione "standard" $\tau : \theta \mapsto \frac{\theta - \alpha_n(\mathbf{x})}{\beta_n}$ della retta reale, dal Corollario A.4.2 risulta

$$\frac{1}{\sqrt{2\pi}} \int_{-\frac{\alpha_n(\mathbf{x})-a}{\beta_n}}^{\frac{\alpha_n(\mathbf{x})-a}{\beta_n}} \exp\Big[-\frac{\theta^2}{2}\Big]\, d\theta = 1 - \xi(\mathbf{x}).$$

Indicato infine con $\theta_{\frac{\xi(\mathbf{x})}{2}}$ lo stato tale che

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{\xi(\mathbf{x})}{2}}^{+\infty} \exp\Big[-\frac{\theta^2}{2}\Big]\, d\theta = \frac{\xi(\mathbf{x})}{2},$$

si ha $\frac{\alpha_n(\mathbf{x})-a}{\beta_n} = \theta_{\frac{\xi(\mathbf{x})}{2}}$ e quindi $a = \alpha_n(\mathbf{x}) - \beta_n\, \theta_{\frac{\xi(\mathbf{x})}{2}}$. Pertanto, $[\alpha_n(\mathbf{x}) - \beta_n\, \theta_{\frac{\xi(\mathbf{x})}{2}}, \alpha_n(\mathbf{x}) + \beta_n\, \theta_{\frac{\xi(\mathbf{x})}{2}}]$ è l'intervallo di lunghezza $2\beta_n\, \theta_{\frac{\xi(\mathbf{x})}{2}}$ di massima probabilità finale relativa a $\mathbf{x}$. ▲

2.6.3 Regole bayesiane formali nella verifica di ipotesi

Osservato che, adottando il danno (2.1), si ha $\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_0) = \int_{\Theta_1} k_0\, d\widetilde{\mathrm{P}}$ e $\mathrm{E}_{\widetilde{\mathrm{P}}}(\mathrm{L}_1) = \int_{\Theta_0} k_1\, d\widetilde{\mathrm{P}}$, la decisione d' minimizza il danno medio se e solo se:

$$d' = \begin{cases} d_0 & \text{se } \int_{\Theta_1} k_0\, d\widetilde{\mathrm{P}} < \int_{\Theta_0} k_1\, d\widetilde{\mathrm{P}} \\ d_0 \text{ o } d_1 & \text{se } \int_{\Theta_1} k_0\, d\widetilde{\mathrm{P}} = \int_{\Theta_0} k_1\, d\widetilde{\mathrm{P}} \\ d_1 & \text{se } \int_{\Theta_1} k_0\, d\widetilde{\mathrm{P}} > \int_{\Theta_0} k_1\, d\widetilde{\mathrm{P}} \end{cases}.$$

Sussistono pertanto i risultati seguenti.

Teorema 2.6.13 *Ogni funzione di decisione ϕ tale che:*

$$\phi(x) = \begin{cases} d_0 & \textit{se } \int_{\Theta_1} k_0(\theta)\, \mathrm{P}(d\theta|x) < \int_{\Theta_0} k_1(\theta)\, \mathrm{P}(d\theta|x) \\ d_1 & \textit{se } \int_{\Theta_1} k_0(\theta)\, \mathrm{P}(d\theta|x) > \int_{\Theta_0} k_1(\theta)\, \mathrm{P}(d\theta|x) \end{cases}$$

è un test bayesiano formale per P.

Corollario 2.6.14 *Sia la funzione k_i una costante di valore K_i $(i = 0, 1)$. Allora, ogni funzione di decisione ϕ tale che:*

$$\phi(x) = \begin{cases} d_0 & \textit{se } \mathrm{P}(\Theta_0|x) > \frac{K_1}{K_0+K_1} \\ d_1 & \textit{se } \mathrm{P}(\Theta_0|x) < \frac{K_1}{K_0+K_1} \end{cases} .$$

è un test bayesiano formale per P.

DIMOSTRAZIONE Basta osservare che si ha $\int_{\Theta_0} k_1(\theta)\, \mathrm{P}(d\theta|x) = K_1 \mathrm{P}(\Theta_0|x)$ e $\int_{\Theta_1} k_0(\theta)\, \mathrm{P}(d\theta|x) = K_0 \mathrm{P}(\Theta_1|x) = K_0\big[1 - \mathrm{P}(\Theta_0|x)\big]$ per ogni campione x. ∎

Nel prossimo esempio determiniamo le regole bayesiane formali per decidere se un dato parametro reale è minore o uguale ad un prefissato valore.

Esempio 2.6.15 Posto $\Theta = \mathbb{R}$, $\nu = \lambda$ e $\Theta_0 =]-\infty, \theta_0]$ (θ_0 fissato), sia $k_0(\theta) = K_0(\theta - \theta_0)$ e $k_1(\theta) = K_1(\theta_0 - \theta)$[45]. Allora, supposto $\mathrm{E}_{\mathrm{P}}(Z|x)$ finita per ogni campione x, otteniamo $\int_{\Theta_0} k_1(\theta)\, \mathrm{P}(d\theta|x) = K_1 \int_{]-\infty,\theta_0]} (\theta_0 - \theta)\, \mathrm{P}(d\theta|x)$ e

$$\begin{aligned} \int_{\Theta_1} k_0(\theta)\, \mathrm{P}(d\theta|x) &= K_0 \int_{]\theta_0,+\infty[} (\theta - \theta_0)\, \mathrm{P}(d\theta|x) \\ &= K_0 \mathrm{E}_{\mathrm{P}}(Z - \theta_0|x) - K_0 \int_{]-\infty,\theta_0]} (\theta - \theta_0)\, \mathrm{P}(d\theta|x) \\ &= K_0 \mathrm{E}_{\mathrm{P}}(Z|x) - K_0\theta_0 + K_0 \int_{]-\infty,\theta_0]} (\theta_0 - \theta)\, \mathrm{P}(d\theta|x) \end{aligned}$$

qualunque sia x. Conseguentemente, ogni funzione di decisione ϕ tale che:

$$\phi(x) = \begin{cases} d_0 & \text{se } \mathrm{E}_{\mathrm{P}}(Z|x) < \theta_0 + \frac{K_1 - K_0}{K_0} \int_{]-\infty,\theta_0]} (\theta_0 - \theta)\, \mathrm{P}(d\theta|x) \\ d_1 & \text{se } \mathrm{E}_{\mathrm{P}}(Z|x) > \theta_0 + \frac{K_1 - K_0}{K_0} \int_{]-\infty,\theta_0]} (\theta_0 - \theta)\, \mathrm{P}(d\theta|x) \end{cases}$$

è un test bayesiano formale per P. Nel caso particolare $K_0 = K_1$, risulta dunque che è bayesiana formale ogni funzione di decisione ϕ tale che:

$$\phi(x) = \begin{cases} d_0 & \text{se } \mathrm{E}_{\mathrm{P}}(Z|x) < \theta_0 \\ d_1 & \text{se } \mathrm{E}_{\mathrm{P}}(Z|x) > \theta_0 \end{cases} .$$

[45] La funzione k_1 dovrebbe essere, come richiesto da (2.1), positiva su Θ_0 e quindi, in particolare, dovrebbe risultare $k_1(\theta_0) > 0$, contrariamente a quanto qui avviene per la particolare scelta funzionale effettuata. D'altra parte, nell'individuazione dei test bayesiani, k_1 interviene solamente nell'integrale $\int_{\Theta_0} k_1(\theta)\, \mathrm{P}(d\theta|x)$ che è indipendente dal valore $k_1(\theta_0)$, essendo $\mathrm{P}(\theta_0|x) = \int_{\{\theta_0\}} \pi(\theta|x)\, d\theta = 0$ per ogni campione x.

Concludiamo determinando l'espressione dell'integrale $\int_{]-\infty,\theta_0]}(\theta_0-\theta)\,\mathrm{P}(d\theta|x)$ nel caso del campionamento normale relativo all'Esempio 2.6.12. Considerata la trasformazione "standard" $\tau : \theta \mapsto \frac{\theta-\alpha_n(\mathbf{x})}{\beta_n}$ della retta reale, dal Corollario A.4.2 otteniamo

$$\begin{aligned}\int_{]-\infty,\theta_0]}(\theta_0-\theta)\,\mathrm{P}(d\theta|x) &= \frac{1}{\sqrt{2\pi\beta_n^2}}\int_{-\infty}^{\theta_0}(\theta_0-\theta)\,\exp\Big[-\frac{(\theta-\alpha_n(\mathbf{x}))^2}{2\beta_n^2}\Big]\,d\theta\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\theta_0-\alpha_n(\mathbf{x})}{\beta_n}}\big(\theta_0-\alpha_n(\mathbf{x})-\beta_n\,\theta\big)\exp\Big[-\frac{\theta^2}{2}\Big]\,d\theta\\ &= \big(\theta_0-\alpha_n(\mathbf{x})\big)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\theta_0-\alpha_n(\mathbf{x})}{\beta_n}}\exp\Big[-\frac{\theta^2}{2}\Big]\,d\theta\\ &\qquad+\frac{\beta_n}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\theta_0-\alpha_n(\mathbf{x})}{\beta_n}}-\theta\,\exp\Big[-\frac{\theta^2}{2}\Big]\,d\theta\end{aligned}$$

e quindi

$$\int_{]-\infty,\theta_0]}(\theta_0-\theta)\,\mathrm{P}(d\theta|x) = \big(\theta_0-\alpha_n(\mathbf{x})\big)\Phi\Big(\frac{\theta_0-\alpha_n(\mathbf{x})}{\beta_n}\Big)+\frac{\beta_n}{\sqrt{2\pi}}\exp\Big[-\frac{(\theta_0-\alpha_n(\mathbf{x}))^2}{2\beta_n^2}\Big],$$

avendo indicato con Φ la funzione di ripartizione della normale $\mathsf{N}(0,1)$. ▲

2.7 Regole di decisione minimax

Ritornando all'Esempio 2.5.2(i), osserviamo che la decisione $d = 0$, oltre ad essere l'unica ammissibile, è pure l'unica decisione ottima secondo il criterio del minimax (introdotto nell'Esempio 2.4.2(i))[46]. Ricordando che tale decisione non è una decisione bayesiana, possiamo allora concludere che, in generale, il criterio del valor medio non è in grado di "generare" tutte le **regole** (**di decisione**) **minimax**, cioè quelle particolari regole di decisione aventi funzione di rischio *limitata* che sono ottime per il criterio del minimax. D'altra parte, qualora lo spazio parametrico sia finito, la convessità dell'insieme di rischio assicura (come metterà in evidenza il teorema successivo) che tutte le regole minimax sono bayesiane per un particolare tipo di probabilità che ora introduciamo.

Definizione 2.7.1 Una probabilità P su $\mathcal{T}$ è **massimamente sfavorevole** (**per** DM) se $\rho(\mathrm{P}') = \sup\{\rho(\mathrm{P}) : \mathrm{P}$ probabilità su $\mathcal{T}\}$, cioè se comporta un

[46]Criterio che, a causa del suo carattere prudenziale, è di uso comune nelle impostazioni della statistica decisionale che, a differenza di quella bayesiana, si basano solamente sulla relazione di dominanza tra regole di decisione escludendo qualsiasi riferimento a valutazioni probabilistiche relative allo stato di natura.

danno medio ineliminabile non inferiore a quello di un'altra qualsiasi probabilità su $\mathcal{T}$.

Viene così fornita una giustificazione dell'uso del criterio del minimax nell'ambito dell'impostazione bayesiana della statistica: se, in base al suo stato d'informazione, DM esprime la sua incertezza sullo stato di natura tramite una probabilità massimamente sfavorevole (e quindi ritiene la Natura a lui avversa almeno per quanto concerne i danni medi ineliminabili), allora egli sarà indotto a scegliere una regola minimax.

Precisato che, in questa sezione, P (dotata o no di apici o pedici) denota una generica probabilità su $\mathcal{T}$ (relativa o no ad una densità iniziale), proviamo il lemma seguente.

Lemma 2.7.2 *Risulta:*

(i) $\sup_{\mathrm{P}} \rho(\mathrm{P}) \leq \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)$*;*

(ii) $\sup_\theta R_\delta(\theta) = \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)$ *per ogni regola di decisione* δ*.*

Dimostrazione (i) Dati δ' e P', si ha $\inf_\delta \rho_{\mathrm{P}'}(\delta) \leq \rho_{\mathrm{P}'}(\delta') \leq \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta')$. Dall'arbitrarietà di δ' segue allora $\rho(\mathrm{P}') = \inf_\delta \rho_{\mathrm{P}'}(\delta) \leq \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)$. Infine, per l'arbitrarietà di P', otteniamo $\sup_{\mathrm{P}} \rho(\mathrm{P}) \leq \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)$.

(ii) Poichè la disuguaglianza $\rho_{\mathrm{P}}(\delta) = \mathrm{E}_{\mathrm{P}}(R_\delta) \leq \sup_\theta R_\delta(\theta)$ sussiste per ogni probabilità P, si ha intanto $\sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) \leq \sup_\theta R_\delta(\theta)$. Per verificare la disuguaglianza opposta, sia $(\theta_n)_{n\geq 1}$ una successione di stati tale che $R_\delta(\theta_n) \uparrow \sup_\theta R_\delta(\theta)$. Considerata allora la misura di Dirac $\mathit{1}_{\theta_n}$ su $\mathcal{T}$ per ogni n, otteniamo $\sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) \geq \rho_{\mathit{1}_{\theta_n}}(\delta) = \mathrm{E}_{\mathit{1}_{\theta_n}}(R_\delta) = R_\delta(\theta_n)$. Passando infine al limite per $n \to +\infty$, riesce $\sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) \geq \sup_\theta R_\delta(\theta)$. ∎

Teorema 2.7.3 *Siano* Θ *finito e l'insieme di rischio convesso. Esiste allora una probabilità* P' *massimamente sfavorevole. Inoltre, ogni regola minimax è bayesiana per* P'*. Infine, per ogni regola minimax* δ'*, si ha* $R_{\delta'} \leq \rho_{\mathrm{P}'}(\delta')$ *e* $R_{\delta'}(\theta) = \rho_{\mathrm{P}'}(\delta')$ *per ogni stato* θ *tale che* $\mathrm{P}'(\theta) > 0$*.*

Dimostrazione Posto $\Theta = \{\theta_1, \ldots, \theta_m\}$, proviamo intanto l'uguaglianza:

$$\sup_{\mathrm{P}} \rho(P) = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta). \tag{2.23}$$

Notato che, per il Lemma 2.7.2(i), basta provare la disuguaglianza $\alpha' = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) \leq \sup_{\mathrm{P}} \rho(\mathrm{P}) = \alpha''$, consideriamo l'insieme:

$$B_t = \{\mathbf{y} : y_i \leq t \, (i = 1, \ldots, m)\} \subset \mathbb{R}^m$$

per ogni numero reale t e poniamo $\alpha = \sup\{t : B_t \cap \mathrm{R} = \emptyset\}$. Allora, poichè $\mathrm{R} \subseteq [0, +\infty[^m$, si ha $\alpha \geq 0$. Riesce inoltre $\alpha < +\infty$; infatti, data una regola di decisione δ e posto $t' = \max\{R_\delta(\theta_1), \ldots, R_\delta(\theta_m)\}$, otteniamo $\mathbf{y}^{(\delta)} \in \mathrm{R} \cap B_t$ per ogni $t \geq t'$ e quindi $\alpha \leq t' < +\infty$.

Osservato che, per ottenere la disuguaglianza $\alpha' \leq \alpha''$, basta constatare che riesce $\alpha' \leq \alpha \leq \alpha''$, iniziamo col verificare che risulta $\alpha' \leq \alpha$. Per la definizione di α si ha $B_{\alpha+\frac{1}{n}} \cap \mathrm{R} \neq \emptyset$ e quindi esiste δ_n tale che $R_{\delta_n}(\theta_i) \leq \alpha + \frac{1}{n}$ per ogni $i \leq m$. Allora, per il Lemma 2.7.2(ii),

$$\alpha' = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) \leq \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta_n) = \sup_\theta R_{\delta_n}(\theta) \leq \alpha + \frac{1}{n}.$$

Ne segue, passando al limite per $n \to +\infty$, la disuguaglianza $\alpha' \leq \alpha$.

Proviamo ora la disuguaglianza $\alpha \leq \alpha''$. Ricordato che l'insieme di rischio R è convesso e osservato che l'insieme $B = \{\mathbf{y} : y_i < \alpha\,(i = 1, \ldots, m)\}$ dei punti interni di B_α è un insieme convesso disgiunto da R, per il teorema dell'iperpiano separatore (nota 35 a p. 76) esistono una m-pla $\mathbf{a} \neq \mathbf{0}$ e un numero reale β tali che

$$a_1 y_1^{(\delta)} + \cdots + a_m y_m^{(\delta)} \geq \beta, \quad \text{per ogni } \delta \in \Delta \tag{2.24}$$

$$a_1 y_1 + \cdots + a_m y_m \leq \beta, \quad \text{per ogni } \mathbf{y} \ll \boldsymbol{\alpha} = (\alpha, \ldots, \alpha). \tag{2.25}$$

Allora, $\mathbf{a} \geq \mathbf{0}$. Infatti, sia (per assurdo) $a_j < 0$ per qualche j. Considerata la m-pla $\mathbf{y}^{(n)}$ tale che $y_i^{(n)} = \alpha - 1$, se $i \neq j$, e $y_j^{(n)} = \alpha - n$, si ha $\mathbf{y}^{(n)} \ll \boldsymbol{\alpha}$ e quindi, per (2.25), $(\alpha - 1)\sum_{i \neq j} a_i + a_j(\alpha - n) \leq \beta$. Passando al limite per $n \to +\infty$ e tenendo conto che $a_j < 0$, risulta

$$\beta \geq \lim_{n \to +\infty} \Big[(\alpha - 1)\sum_{i \neq j} a_i + a_j(\alpha - n)\Big] = +\infty$$

ottenendo così una contraddizione.
Per provare la disuguaglianza:

$$\frac{\beta}{a_1 + \cdots + a_m} \geq \alpha,$$

consideriamo un numero reale $\epsilon > 0$. Allora $(\alpha - \epsilon, \ldots, \alpha - \epsilon) \ll \boldsymbol{\alpha}$ e quindi, per (2.25), $(\alpha - \epsilon)(a_1 + \cdots + a_m) \leq \beta$. Passando al limite per $\epsilon \to 0^+$, risulta $\alpha(a_1 + \cdots + a_m) \leq \beta$.
Sia ora P' la distribuzione di probabilità sugli stati tale che:

$$\pi_i' = \mathrm{P}'(\theta_i) = \frac{a_i}{a_1 + \cdots + a_m} \quad (i = 1, \ldots, m).$$

Poichè, per (2.24), riesce

$$\begin{aligned}\rho_{\mathrm{P}'}(\delta) &= \pi_1' R_\delta(\theta_1) + \cdots + \pi_m' R_\delta(\theta_m)\\ &= \frac{a_1 y_1^{(\delta)} + \cdots + a_m y_m^{(\delta)}}{a_1 + \cdots + a_m} \geq \frac{\beta}{a_1 + \cdots + a_m} \geq \alpha\end{aligned}$$

per ogni regola di decisione δ, si ha $\rho(\mathrm{P}') = \inf_\delta \rho_{\mathrm{P}'}(\delta) \geq \alpha$.

Dunque $\alpha'' = \sup_{\mathrm{P}} \rho(\mathrm{P}) \geq \alpha$ da cui otteniamo $\alpha' = \alpha = \alpha''$, cioè (2.23).

Ne segue $\rho(\mathrm{P}') \geq \alpha = \alpha'' = \sup_{\mathrm{P}} \rho(\mathrm{P})$ e quindi la probabilità P' è massimamente sfavorevole.

Passando all'ultima parte della tesi, sia δ' una regola minimax. Allora, $\sup_\theta R_{\delta'}(\theta) = \inf_\delta \sup_\theta R_\delta(\theta)$ da cui, per il Lemma 2.7.2(ii), risulta

$$\sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta') = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta).$$

Considerata ora una regola di decisione δ, da (2.23) otteniamo

$$\begin{aligned}\rho_{\mathrm{P}'}(\delta') &\leq \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta') = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)\\ &= \sup_{\mathrm{P}} \rho(\mathrm{P}) = \rho(\mathrm{P}') = \inf_\delta \rho_{\mathrm{P}'}(\delta) \leq \rho_{\mathrm{P}'}(\delta),\end{aligned}$$

ove la penultima uguaglianza sussiste in quanto P' è massimamente sfavorevole. Conseguentemente, δ' è una regola bayesiana per P'. Inoltre, scegliendo in particolare $\delta = \delta'$, otteniamo $\rho_{\mathrm{P}'}(\delta') = \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta)$. Per il Lemma 2.7.2(ii) si ha allora

$$R_{\delta'}(\theta) \leq \sup_\theta R_{\delta'}(\theta) = \inf_\delta \sup_\theta R_\delta(\theta) = \inf_\delta \sup_{\mathrm{P}} \rho_P(\delta) = \rho_{\mathrm{P}'}(\delta')$$

qualunque sia lo stato θ.
Supponiamo infine che esista (per assurdo) j tale che $R_{\delta'}(\theta_j) < \rho_{\mathrm{P}'}(\delta')$ con $\pi_j' > 0$. Allora, $\pi_j' R_{\delta'}(\theta_j) < \pi_j' \rho_{\mathrm{P}'}(\delta')$ da cui otteniamo

$$\rho_{\mathrm{P}'}(\delta') = \pi_1' R_{\delta'}(\theta_1) + \cdots + \pi_m' R_{\delta'}(\theta_m) < \pi_1' \rho_{\mathrm{P}'}(\delta') + \cdots + \pi_m' \rho_{\mathrm{P}'}(\delta') = \rho_{\mathrm{P}'}(\delta')$$

pervenendo così ad una contraddizione. ■

Il teorema appena provato assicura che esiste una probabilità (massimamente sfavorevole) P' tale che le regole minimax vanno ricercate tra le regole bayesiane δ' per P' per le quali la funzione di rischio $R_{\delta'}$ ha valori che non superano il rischio di Bayes $\rho(\mathrm{P}')$. Il prossimo risultato evidenzia che quest'ultima proprietà consente di individuare, indipendentemente dalla numerosità degli stati, sia regole bayesiane minimax che probabilità massimamente sfavorevoli.

Teorema 2.7.4 *Sia δ' una regola bayesiana per P' tale che $R_{\delta'} \leq \rho(\mathrm{P}')$. Allora, δ' è una regola minimax e P' è una probabilità massimamente sfavorevole.*

Dimostrazione Risulta

$$\begin{aligned}\inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta) &= \inf_\delta \sup_\theta R_\delta(\theta)\\ &\leq \sup_\theta R_{\delta'}(\theta) \leq \rho(\mathrm{P}') \leq \sup_{\mathrm{P}} \rho(\mathrm{P}) \leq \inf_\delta \sup_{\mathrm{P}} \rho_{\mathrm{P}}(\delta),\end{aligned}$$

ove la prima uguaglianza discende dal Lemma 2.7.2(ii) e l'ultima disuguaglianza dal Lemma 2.7.2(i). Allora, tutte le disuguaglianze sono delle uguaglianze e quindi $\sup_\theta R_{\delta'}(\theta) = \inf_\delta \sup_\theta R_\delta(\theta)$ e $\rho(\mathrm{P}') = \sup_{\mathrm{P}} \rho(\mathrm{P})$. ∎

In molti problemi statistico-decisionali, un metodo efficiente per trovare regole minimax consiste nell'individuare regole di decisione con funzione di rischio costante (dette **regole (di decisione) equalizzanti**) e provare poi, come suggerito dal Teorema 2.7.4, che sono bayesiane. L'esempio seguente, oltre a fornire un'applicazione di tale metodo, mostra che ci può essere più di una probabilità massimamente sfavorevole per il medesimo problema decisionale.

Esempio 2.7.5 Posto $\Theta = \mathfrak{D} = [0,1]$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la distribuzione di Bernoulli $Ber(\theta)$ in corrispondenza ad ogni stato θ. Con riferimento al danno quadratico, vediamo se si possono trovare regole equalizzanti tra le trasformate affini della media campionaria, cioè del tipo $\delta'(\mathbf{x}) = a\overline{\mathbf{x}} + b$ $(a, b \geq 0, a + b \leq 1)$. Da $\mathrm{E}_\theta(\overline{\mathbf{X}}) = \theta$, $\mathrm{Var}_\theta(\overline{\mathbf{X}}) = \frac{\theta(1-\theta)}{n}$ e (2.2) otteniamo

$$\begin{aligned}\mathrm{MSE}_{\delta'}(\theta) &= \mathrm{Var}_\theta(\delta') + \mathrm{Bias}_\theta(\delta')^2 = a^2 \mathrm{Var}_\theta\big(\overline{\mathbf{X}}\big) + \big(a\mathrm{E}_\theta(\overline{\mathbf{X}}) + b - \theta\big)^2\\ &= a^2\,\frac{\theta(1-\theta)}{n} + [b + (a-1)\theta]^2 = \Big[(a-1)^2 - \frac{a^2}{n}\Big]\theta^2 + \Big[\frac{a^2}{n} + 2b(a-1)\Big]\theta + b^2.\end{aligned}$$

Affinchè la funzione di rischio sia una costante, a e b devono essere soluzioni del sistema:

$$\begin{cases}(a-1)^2 - \frac{a^2}{n} = 0\\ \frac{a^2}{n} + 2b(a-1) = 0\end{cases}.$$

Risolvendolo (tenendo conto dei vincoli imposti ad a e b) si ha

$$a = \frac{\sqrt{n}}{1+\sqrt{n}}, \qquad b = \frac{1}{2(1+\sqrt{n})}$$

e quindi

$$\delta'(\mathbf{x}) = \frac{\sqrt{n}\,\overline{\mathbf{x}} + \frac{1}{2}}{1+\sqrt{n}} = \frac{n\,\overline{\mathbf{x}} + \frac{\sqrt{n}}{2}}{\sqrt{n}+n} = \frac{n\,\overline{\mathbf{x}} + \frac{\sqrt{n}}{2}}{\frac{\sqrt{n}}{2} + \frac{\sqrt{n}}{2} + n}.$$

Conseguentemente, per (2.18), δ' coincide con la regola bayesiana corrispondente alla distribuzione $Beta(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$. Possiamo allora concludere, per il Teorema 2.7.4, che δ' è una regola minimax e che la distribuzione $Beta(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$ è massimamente sfavorevole[47].

Al fine di individuare ulteriori probabilità massimamente sfavorevoli basta evidentemente trovare densità iniziali π (rispetto a qualche misura di riferimento ν) che rendano δ' bayesiana per la probabilità P determinata da π. Affinchè ciò avvenga, per il Teorema 2.6.4, deve essere

$$\delta'(\mathbf{x}) = \mathsf{E}_{\mathrm{P}}(Z|\mathbf{x}) = \frac{\displaystyle\int_0^1 \theta^{n\overline{\mathbf{x}}+1}\,(1-\theta)^{n(1-\overline{\mathbf{x}})}\pi(\theta)\,\nu(d\theta)}{\displaystyle\int_0^1 \theta^{n\overline{\mathbf{x}}}\,(1-\theta)^{n(1-\overline{\mathbf{x}})}\pi(\theta)\,\nu(d\theta)}$$

o, equivalentemente,

$$\delta'(\mathbf{x}) = \frac{\displaystyle\sum_{k=0}^{n(1-\overline{\mathbf{x}})} (-1)^k \binom{n(1-\overline{\mathbf{x}})}{k} \int_0^1 \theta^{n\overline{\mathbf{x}}+k+1}\,\pi(\theta)\,\nu(d\theta)}{\displaystyle\sum_{k=0}^{n(1-\overline{\mathbf{x}})} (-1)^k \binom{n(1-\overline{\mathbf{x}})}{k} \int_0^1 \theta^{n\overline{\mathbf{x}}+k}\,\pi(\theta)\,\nu(d\theta)},$$

tenuto conto della formula di Newton:

$$(1-\theta)^{n(1-\overline{\mathbf{x}})} = \sum_{k=0}^{n(1-\overline{\mathbf{x}})} (-1)^k \binom{n(1-\overline{\mathbf{x}})}{k} \theta^k.$$

Conseguentemente, affinchè δ' sia una regola bayesiana per P, basta che tale distribuzione abbia i primi $n+1$ momenti coincidenti con quelli della distribuzione $Beta(\frac{\sqrt{n}}{2}, \frac{\sqrt{n}}{2})$. Dunque, nel caso particolare di una sola osservazione ($n = 1$), è massimamente sfavorevole, come facilmente si verifica, ogni probabilità $\mathrm{P} = \pi_0 \mathit{1}_{\theta_0} + (1-\pi_0)\mathit{1}_{\theta_1}$ con $\frac{1}{3} \le \pi_0 \le \frac{2}{3}$ e

$$\theta_0 = \frac{1}{2} - \frac{\sqrt{2\pi_0(1-\pi_0)}}{4\pi_0}, \qquad \theta_1 = \frac{1}{2} + \frac{\sqrt{2\pi_0(1-\pi_0)}}{4(1-\pi_0)},$$

notato che il momento secondo della distribuzione $Beta(\frac{1}{2}, \frac{1}{2})$ è pari a $\frac{3}{8}$. ▲

Osservato che nell'esempio precedente le regole bayesiane sono anche ammissibili (Esempio 2.6.6(i)), viene spontaneo chiedersi se, in generale, le regole equalizzanti ammissibili siano anche regole minimax. Risposta positiva al quesito viene fornita dal risultato seguente.

[47] Ricordato che $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ è la varianza della distribuzione $Beta(\alpha, \beta)$, otteniamo che la varianza della distribuzione considerata è $\frac{1}{4(1+\sqrt{n})}$. Pertanto, quando il numero delle osservazioni è elevato, l'adozione di questa distribuzione massimamente sfavorevole delinea una forte credenza nella concentrazione degli stati attorno al valore $\frac{1}{2}$ (valor medio della distribuzione). Osserviamo infine che, nei casi particolari $n = 1$ e $n = 4$, tale distribuzione diviene, rispettivamente, quella non informativa di Jeffreys e quella uniforme.

Teorema 2.7.6 *Ogni regola equalizzante ammissibile è una regola minimax.*

DIMOSTRAZIONE Sia $\delta' \in \Delta^+$ una regola equalizzante. Allora, $R_{\delta'}(\theta) = \alpha$ per ogni stato θ. Supposto (per assurdo) che non sia minimax, esiste δ tale che $\sup_\theta R_\delta(\theta) < \sup_\theta R_{\delta'}(\theta) = \alpha$. Conseguentemente, $R_\delta(\theta) < R_{\delta'}(\theta)$ per ogni stato θ e quindi $\delta \succ \delta'$ (Contraddizione!). ■

Nel prossimo esempio mostriamo, con riferimento al campionamento normale con media incognita e danno quadratico, che la media campionaria è uno stimatore minimax constatando che è uno stimatore sia equalizzante che ammissibile.

Esempio 2.7.7 Posto $\Theta = \mathfrak{D} = \mathbb{R}$, sia l'osservabile costituito dalle v.a. $X_1, \ldots, X_n$ indipendenti e distribuite secondo la normale $\mathsf{N}(\theta, \sigma^2)$ in corrispondenza ad ogni stato θ. Posto, $\delta' = \overline{\mathbf{X}}$, da (2.2) otteniamo $\text{MSE}_{\delta'}(\theta) = \frac{\sigma^2}{n}$ per ogni stato θ e quindi la media campionaria è uno stimatore equalizzante. Per provare che è anche ammissibile supponiamo (per assurdo) che esista una regola di decisione δ tale che $\text{MSE}_\delta \le \text{MSE}_{\delta'}$ e $\text{MSE}_\delta(\tilde{\theta}) < \text{MSE}_{\delta'}(\tilde{\theta})$ per qualche stato $\tilde{\theta}$. Poichè, per l'Esempio 2.1.3(ii), le due funzioni di rischio sono continue, esistono $\epsilon, \eta > 0$ tali che $\text{MSE}_{\delta'}(\theta) - \text{MSE}_\delta(\theta) \ge \eta$ per ogni $\theta \in [\tilde{\theta} - \epsilon, \tilde{\theta} + \epsilon]$.

Ciò osservato, scegliamo come distribuzione iniziale P quella della normale $\mathsf{N}(0, m)$. Allora

$$\rho_{\mathrm{P}}(\delta') - \rho_{\mathrm{P}}(\delta) = \int_{-\infty}^{+\infty} (\text{MSE}_{\delta'}(\theta) - \text{MSE}_\delta(\theta))\, \mathrm{P}(d\theta) \ge \int_{\tilde{\theta}-\epsilon}^{\tilde{\theta}+\epsilon} (\text{MSE}_{\delta'}(\theta) - \text{MSE}_\delta(\theta))\, \mathrm{P}(d\theta)$$

e quindi

$$\rho_{\mathrm{P}}(\delta') - \rho_{\mathrm{P}}(\delta) \ge \frac{\eta}{\sqrt{2\pi m}} \int_{\tilde{\theta}-\epsilon}^{\tilde{\theta}+\epsilon} \exp\Big[-\frac{\theta^2}{2m}\Big]\, d\theta.$$

Inoltre, considerato il relativo rischio di Bayes $\rho(0, m)$, per (2.20), si ha

$$\rho(0, m) - \rho_{\mathrm{P}}(\delta') = \frac{m\sigma^2}{\sigma^2 + mn} - \frac{\sigma^2}{n} = -\frac{\sigma^4}{n(mn + \sigma^2)}.$$

Riesce pertanto

$$\begin{aligned} 0 \ge \rho(0, m) - \rho_{\mathrm{P}}(\delta) &= [\rho(0, m) - \rho_{\mathrm{P}}(\delta')] + [\rho_{\mathrm{P}}(\delta') - \rho_{\mathrm{P}}(\delta)] \\ &\ge -\frac{\sigma^4}{n(mn + \sigma^2)} + \frac{\eta}{\sqrt{2\pi m}} \int_{\tilde{\theta}-\epsilon}^{\tilde{\theta}+\epsilon} \exp\Big[-\frac{\theta^2}{2m}\Big]\, d\theta \end{aligned}$$

da cui otteniamo (moltiplicando per $\sqrt{m}$ il primo e l'ultimo termine)

$$0 \ge -\frac{\sqrt{m}\sigma^4}{n(mn + \sigma^2)} + \frac{\eta}{\sqrt{2\pi}} \int_{\tilde{\theta}-\epsilon}^{\tilde{\theta}+\epsilon} \exp\Big[-\frac{\theta^2}{2m}\Big]\, d\theta.$$

Poichè la disuguaglianza sussiste per ogni m, passiamo al limite per $m \to +\infty$. Osservato che $\left(\exp\left[-\frac{\theta^2}{2m}\right]\right)_{m\geq 1}$ è una successione crescente convergente a 1 per ogni θ, dal teorema della convergenza monotona A.3.7 otteniamo

$$\lim_{m\to+\infty}\int_{\widetilde{\theta}-\epsilon}^{\widetilde{\theta}+\epsilon}\exp\left[-\frac{\theta^2}{2m}\right]d\theta = \int_{\widetilde{\theta}-\epsilon}^{\widetilde{\theta}+\epsilon}\lim_{m\to+\infty}\exp\left[-\frac{\theta^2}{2m}\right]d\theta = 2\,\epsilon.$$

Risulta dunque

$$0 \geq \lim_{m\to+\infty}\left[-\frac{\sqrt{m}\sigma^4}{n(mn+\sigma^2)}+\frac{\eta}{\sqrt{2\pi}}\int_{\widetilde{\theta}-\epsilon}^{\widetilde{\theta}+\epsilon}\exp\left[-\frac{\theta^2}{2m}\right]d\theta\right]=\frac{2\epsilon\eta}{\sqrt{2\pi}}>0$$

pervenendo così ad una contraddizione. ▲

Il risultato seguente fornisce un criterio, per constatare se una regola di decisione è minimax, basato sul confronto della funzione di rischio non più con un rischio di Bayes (come avviene nel Teorema 2.7.4) ma con il limite di una successione di rischi di Bayes.

Teorema 2.7.8 *Sia* $(\mathrm{P}_n)_{n\geq 1}$ *una successione di probabilità tale che esista finito il limite* $\lim\limits_{n\to+\infty}\rho(\mathrm{P}_n)$ *dei relativi rischi di Bayes. Allora, ogni regola di decisione* δ' *tale che* $R_{\delta'} \leq \lim\limits_{n\to+\infty}\rho(\mathrm{P}_n)$ *è una regola minimax*[48].

Dimostrazione Sia $\alpha = \lim\limits_{n\to+\infty}\rho(\mathrm{P}_n)$. Supposto (per assurdo) che δ' non sia una regola minimax, otteniamo $\sup_\theta R_{\delta'}(\theta) > \inf_\delta \sup_\theta R_\delta(\theta)$. Scelto allora un numero reale $\epsilon > 0$, esiste una regola di decisione δ'' tale che

$$R_{\delta''} \leq \sup_\theta R_{\delta''}(\theta) \leq \sup_\theta R_{\delta'}(\theta) - \epsilon \leq \alpha - \epsilon.$$

Inoltre, per definizione di α, esiste m tale che $\rho(\mathrm{P}_m) > \alpha-\epsilon$. Riesce pertanto

$$\rho_{\mathrm{P}_m}(\delta'') = \int_\Theta R_{\delta''}\,d\mathrm{P}_m \leq \int_\Theta(\alpha-\epsilon)I_\Theta\,d\mathrm{P}_m = \alpha-\epsilon < \rho(\mathrm{P}_m) = \inf_\delta \rho_{\mathrm{P}_m}(\delta)$$

ottenendo così una contraddizione. ■

[48] Alla luce di questo teorema, si poteva provare, con riferimento all'esempio precedente, che la media campionaria è uno stimatore minimax (senza ricorrere alla sua ammissibilità) considerando, per ogni m, la probabilità P_m avente come densità quella della normale $\mathsf{N}(0,m)$ e osservando che $\lim\limits_{m\to+\infty}\rho(\mathrm{P}^{(m)}) = \lim\limits_{m\to+\infty}\frac{m\sigma^2}{mn+\sigma^2} = \frac{\sigma^2}{n}$.

Come applicazione di quest'ultimo criterio, proviamo che nel campionamento normale bivariato con vettore delle medie incognito e varianza unitaria, lo stimatore che associa ad ogni campione il campione medesimo è una regola minimax, qualora si adotti la funzione di danno (2.21) con M matrice identica.

Esempio 2.7.9 Posto $\Theta = \mathfrak{D} = \mathbb{R}^2$, sia l'osservabile costituito dalle v.a. X_1, X_2 indipendenti e distribuite, rispettivamente, secondo la normale $\mathsf{N}(\theta_i, 1)$ $(i = 1, 2)$ in corrispondenza ad ogni stato $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Allora, indicata con f_i la densità della normale $\mathsf{N}(\theta_i, 1)$ $(i = 1, 2)$, per la densità di campionamento si ha $f_{\boldsymbol{\theta}}(\mathbf{x}) = f_1(x_1)f_2(x_2)$ per ogni $\mathbf{x} \in \mathbb{R}^2$. Adottato il danno (2.21) con M matrice identica, otteniamo $\mathrm{L}(\boldsymbol{\theta}, \mathbf{d}) = (\theta_1 - d_1)^2 + (\theta_2 - d_2)^2$. Proviamo ora che $\mathbf{X}$ è uno stimatore minimax.

Al fine di costruire la successione di probabilità $(\mathrm{P}_m)_{m\geq 1}$, indichiamo con $\pi^{(m)}$ la funzione di densità della normale $\mathsf{N}(0, m)$ e assumiamo che P_m abbia densità: $\pi_m(\boldsymbol{\theta}) = \pi^{(m)}(\theta_1)\pi^{(m)}(\theta_2)$ per ogni stato $\boldsymbol{\theta}$.

Dato il campione $\mathbf{x}$, risulta $\pi_m(\boldsymbol{\theta}|\mathbf{x}) = \pi^{(m)}(\theta_1|x_1)\pi^{(m)}(\theta_2|x_2)$. Infatti, indicata con p_i la densità predittiva relativa alle densità $\pi^{(m)}$ e f_i $(i = 1, 2)$, dal teorema di Tonelli otteniamo

$$
\begin{aligned}
p(\mathbf{x}) &= \int_{\mathbb{R}^2} f_{\boldsymbol{\theta}}(\mathbf{x})\pi_m(\boldsymbol{\theta})\, d\theta_1 d\theta_2 \\
&= \int_{-\infty}^{+\infty} \Big[\int_{-\infty}^{+\infty} f_1(x_1)\pi^{(m)}(\theta_1)\, d\theta_1\Big] f_2(x_2)\pi^{(m)}(\theta_2)\, d\theta_2 = p_1(x_1)p_2(x_2)
\end{aligned}
$$

e quindi

$$
\pi_m(\boldsymbol{\theta}|\mathbf{x}) = \frac{f_{\boldsymbol{\theta}}(\mathbf{x})}{p(\mathbf{x})}\, \pi_m(\boldsymbol{\theta}) = \Big[\frac{f_1(x_1)}{p_1(x_1)}\, \pi_m(\theta_1)\Big]\Big[\frac{f_2(x_2)}{p_2(x_2)}\, \pi_m(\theta_2)\Big] = \pi^{(m)}(\theta_1|x_1)\pi^{(m)}(\theta_2|x_2).
$$

Posto $\mathrm{E}_m(Z_i|x_i) = \int_{-\infty}^{+\infty} \theta_i \pi^{(m)}(\theta_i|x_i)\, d\theta_i$ $(i = 1, 2)$, per il teorema di Tonelli, si ha

$$
\begin{aligned}
\mathrm{E}_{\mathrm{P}_m}(Z_i|\mathbf{x}) &= \int_{\mathbb{R}^2} \theta_i \pi_m(\boldsymbol{\theta}|\mathbf{x})\, d\theta_1 d\theta_2 \\
&= \int_{-\infty}^{+\infty} \Big[\int_{-\infty}^{+\infty} \theta_i \pi^{(m)}(\theta_i|x_i)\, d\theta_i\Big] \pi^{(m)}(\theta_j|x_j) d\theta_j \qquad (j \neq i; i = 1, 2) \\
&= \int_{-\infty}^{+\infty} \mathrm{E}_m(Z_i|x_i)\pi^{(m)}(\theta_j|x_j)\, d\theta_j = \mathrm{E}_m(Z_i|x_i).
\end{aligned}
$$

Pertanto, per il Teorema 2.6.9, la funzione di decisione:

$$
\phi^{(m)}(\mathbf{x}) = \begin{pmatrix} \mathrm{E}_m(Z_1|x_1) \\ \mathrm{E}_m(Z_2|x_2) \end{pmatrix} = \frac{m}{m+1}\, \mathbf{x}
$$

è una regola bayesiana formale, ricordato che, per l'Esempio B.2.18(ii) (riferito all'esperimento consistente nella sola osservazione di X_i), la densità finale relativa a x_i è quella della normale di media $\frac{m}{m+1}\, x_i$ e varianza $\frac{m}{m+1}$ $(i = 1, 2)$.

Ciò osservato, data una funzione di decisione $\phi(\mathbf{x}) = \binom{\phi'(x_1)}{\phi''(x_2)}$, andiamo a determinarne la funzione di rischio. Per il teorema di Tonelli si ha

$$\begin{aligned} R_\phi(\boldsymbol{\theta}) &= \int_{\mathbb{R}^2} \mathrm{L}(\boldsymbol{\theta}, \phi(\mathbf{x})) f_{\boldsymbol{\theta}}(\mathbf{x})\, dx_1 dx_2 \\ &= \int_{\mathbb{R}^2} \big(\theta_1 - \phi'(x_1)\big)^2 f_{\boldsymbol{\theta}}(\mathbf{x})\, dx_1 dx_2 + \int_{\mathbb{R}^2} \big(\theta_2 - \phi''(x_2)\big)^2 f_{\boldsymbol{\theta}}(\mathbf{x})\, dx_1 dx_2 \\ &= \int_{-\infty}^{+\infty} \Big[\int_{-\infty}^{+\infty} \big(\theta_1 - \phi'(x_1)\big)^2 f_1(x_1)\, dx_1\Big] f_2(x_2) dx_2 \\ &\qquad + \int_{-\infty}^{+\infty} \Big[\int_{-\infty}^{+\infty} \big(\theta_2 - \phi''(x_2)\big)^2 f_2(x_2)\, dx_2\Big] f_1(x_1) dx_1 \end{aligned}$$

e quindi

$$R_\phi(\boldsymbol{\theta}) = \mathrm{MSE}_{\phi'}(\theta_1) + \mathrm{MSE}_{\phi''}(\theta_2). \tag{2.26}$$

Allora, $\phi^{(m)}$ è uno stimatore in quanto, per l'Esempio 2.6.6(ii), le sue componenti sono stimatori bayesiani. Dunque, $\phi^{(m)}$ è uno stimatore bayesiano e quindi, per (2.20),

$$\begin{aligned} \rho(\mathrm{P}_m) = \rho_{\mathrm{P}_m}(\phi^{(m)}) &= \mathrm{E}_{\mathrm{P}_m}\big((Z_1 - \phi^{(m)}(x_1))^2 + (Z_2 - \phi^{(m)}(x_2))^2\big) \\ &= \mathrm{E}_{\mathrm{P}_m}\big((Z_1 - \phi^{(m)}(x_1))^2\big) + \mathrm{E}_{\mathrm{P}_m}\big((Z_2 - \phi^{(m)}(x_2))^2\big) = \frac{2m}{m+1}. \end{aligned}$$

Inoltre, da (2.26) e (2.2) otteniamo

$$R_{\mathbf{X}}(\boldsymbol{\theta}) = (\mathrm{Var}_{\theta_1}(X_1) + \mathrm{Bias}_{\theta_1}(X_1)^2) + (\mathrm{Var}_{\theta_2}(X_2) + \mathrm{Bias}_{\theta_2}(X_2)^2) = 1 + 1 = 2$$

e quindi $R_{\mathbf{X}} = 2 = \lim\limits_{m\to+\infty} \frac{2m}{m+1} = \lim\limits_{m\to+\infty} \rho^{(m)}$. Ne segue che $\mathbf{X}$ è minimax[49]. ▲

2.8 Regole di decisione randomizzate

La convessità dell'insieme di rischio fornisce, tramite i teoremi 2.5.4, 2.7.3 e il teorema della classe completa 2.5.5, una giustificazione dell'uso del criterio del valore medio nel caso di un numero finito di stati. Descriviamo ora

[49] Chiaramente, le argomentazioni svolte possono essere estese pari pari a dimensioni maggiori di due. Conseguentemente, $\mathbf{X}$ è uno stimatore minimax nel campionamento normale multivariato con media incognita e varianza unitaria.

Osserviamo che fino al 1955 era opinione comune, nell'ambito statistico, che $\mathbf{X}$ fosse uno stimatore ammissibile (e quindi minimax in forza del Teorema 2.7.6). Purtroppo, questa credenza era priva di fondamento in quanto Charles Stein scoprì nel 1956 il seguente celebre risultato (noto come **paradosso di Stein**): pur avendo tutte le componenti ammissibili, $\mathbf{X}$ è ammissibile solamente per $n \leq 2$. Per dimensioni superiori, uno stimatore che domina strettamente $\mathbf{X}$ è, ad esempio, lo **stimatore di James-Stein**: $\big(1 - \frac{n-2}{\sum_{i=1}^n X_i^2}\big)\mathbf{X}$.

un procedimento che consente, qualora R non sia convesso[50], di ottenere - mediante l'inserimento di "nuove" decisioni accanto a quelle "vecchie" - un problema decisionale (estensione di quello di partenza) avente l'insieme di rischio convesso[51]. L'idea guida della metodologia in parola è di consentire la scelta della decisione da adottare non *solamente* a DM (come sinora fatto), ma *anche* al "caso" tramite meccanismi di estrazione "a sorte"[52].

Fissata dunque una σ-algebra di riferimento $\mathcal{D}$ su $\mathfrak{D}$[53] e denotata con Q (dotata o no di apici o pedici) una qualsiasi probabilità su $\mathcal{D}$, introduciamo la nozione chiave di decisione randomizzata.

Definizione 2.8.1 Una **decisione randomizzata** è una qualsiasi probabilità su $\mathcal{D}$. Inoltre, data una decisione d, la **decisione pura relativa a** d è la probabilità 1_d su $\mathcal{D}$ che concentra l'intera massa su d[54].

Supponiamo ora che DM scelga, a differenza di quanto sinora ammesso, invece di una decisione pura una decisione randomizzata Q. Allora DM, per individuare le conseguenze a cui andrebbe incontro qualora θ fosse lo stato

[50] Come, ad esempio, quando sia le decisioni che i campioni sono in numero finito.

[51] L'idea di aggiungere nuovi elementi, ad un dato contesto, per soddisfare una proprietà che non è ivi verificata, sta alla base del cosiddetto *metodo degli elementi ideali* comunemente usato nella matematica. Si pensi, ad esempio, all'introduzione dei "nuovi" numeri irrazionali accanto ai "vecchi" numeri razionali, per fornire di estremi ogni insieme limitato di razionali; dei "nuovi" numeri complessi accanto ai "vecchi" numeri reali, per dare soluzione all'equazione $x^2 = -1$; dei "nuovi" punti impropri accanto ai "vecchi" punti, per consentire che ogni coppia di rette abbia almeno un punto in comune.

[52] L'idea della *casualizzazione* delle decisioni può farsi risalire a James Waldegrave che l'utilizzò nel 1713 nell'analisi di un particolare gioco di carte, chiamato *le Her*, molto noto nel '700. L'uso delle decisioni aleatorie (nella trattazione delle scelte in condizioni d'incertezza) venne però, dopo Waldegrave, abbandonato e rimase "sepolto" per due secoli fino alla sua "riesumazione", avvenuta nel 1921, ad opera di Emile Borel.

È interessante notare che il ricorso alla scelta casuale delle decisioni non ha carattere squisitamente teorico, in quanto non è affatto inusuale nella concreta pratica decisionale. Si pensi, ad esempio, ad un collezionista al quale vengano offerti due francobolli alquanto rari. A causa del loro prezzo, la sua disponibilità finanziaria gli consente di comperare sia l'uno che l'altro, ma non entrambi. Il collezionista deve quindi fare una scelta che si presenta ardua, poichè ognuno dei due francobolli è importante per completare la sua collezione. Per non avere, in seguito, rimpianti sulla scelta effettuata, egli potrebbe ragionevolmente lanciare una moneta e far quindi scegliere al "caso" quale francobollo acquistare.

[53] Che, nel caso di $\mathfrak{D}$ finito, concide con l'insieme delle parti di $\mathfrak{D}$.

[54] Per quanto detto, le decisioni randomizzate vanno intese, da un punto di vista interpretativo, come meccanismi di sorteggio che forniscono "a sorte" la decisione da adottare.

"vero", considererà l'ente aleatorio $\gamma(\theta,\cdot):\mathfrak{D}\mapsto\mathfrak{C}$. Supposto che sia $(\mathcal{D},\mathcal{C})$-misurabile), rimarrà individuata una particolare lotteria; precisamente, la legge dell'ente aleatorio: $\ell_\theta^{(\mathrm{Q})}(C)=\mathrm{Q}(\gamma(\theta,\cdot)\in C)$ per ogni $C\in\mathcal{C}$. Supposto infine che DM ritenga equivalenti due decisioni randomizzate di medesima legge, egli sarà indotto ad esprimere la relazione di preferenza $\succeq'_\theta$ su tali decisioni tramite la relazione di preferenza $\succeq^*$ definita sulle lotterie. Si avrà quindi (con passaggi analoghi a quelli sviluppati a p. 44)

$$Q_1\succeq'_\theta Q_2 \iff \int_{\mathfrak{D}}\mathrm{L}(\theta,\cdot)\,d\mathrm{Q}_1\le\int_{\mathfrak{D}}\mathrm{L}(\theta.\cdot)\,d\mathrm{Q}_2.$$

Ora, essendo in realtà ignoto a DM lo stato "vero", egli sarà portato ad associare ad ogni decisione randomizzata Q il danno medio aleatorio $\int_{\mathfrak{D}}\mathrm{L}(Z,\cdot)\,d\mathrm{Q}$.

Queste considerazioni suggeriscono di assumere la $\mathcal{T}\otimes\mathcal{D}$-Borel misurabilità della funzione di danno L e di formulare la seguente nozione di funzione di danno nell'ambito delle decisioni randomizzate.

Definizione 2.8.2 La **funzione di danno relativa a** Q è la funzione di dominio lo spazio parametrico:

$$\mathrm{L}_{\mathrm{Q}}(\theta)=\int_{\mathfrak{D}}\mathrm{L}(\theta,\cdot)\,d\mathrm{Q}$$

che associa, ad ogni stato θ, il danno medio a cui va incontro DM adottando la decisione randomizzata Q, qualora θ sia lo stato "vero".

Identificando, come è naturale in questo contesto, ogni decisione d con la decisione pura $\mathit{1}_d$, l'insieme $\mathcal{Q}$ delle decisioni randomizzate diviene un ampliamento dell'insieme $\mathfrak{D}$; inoltre, risultando $\mathrm{L}_{\mathit{1}_d}=\mathrm{L}_d$ per ogni $d\in\mathfrak{D}$, le funzioni di danno delle decisioni pure coincidono con quelle delle relative decisioni. Conseguentemente, il problema decisionale $-$ avente $\mathcal{Q}$ come insieme delle decisioni, Θ come spazio parametrico e funzione di danno $\mathrm{L}^{(r)}:(\theta,\mathrm{Q})\mapsto\mathrm{L}_{\mathrm{Q}}(\theta)-$ risulta un'estensione del problema decisionale originario.

Passando infine all'aspetto statistico-decisionale, le regole di decisione divengono, in questo nuovo contesto, quelle particolari applicazioni $\delta^{(r)}:\mathfrak{X}\mapsto\mathcal{Q}$ - dette **regole** (**di decisione**) **randomizzate** - per le quali esiste ed è a valori finiti la funzione $R_{\delta^{(r)}}$ di dominio Θ così definita:

$$R_{\delta^{(r)}}(\theta)=\mathrm{E}_\theta\big(\mathrm{L}^{(r)}(\theta,\delta^{(r)})\big)=\int_{\mathfrak{X}}\Big(\int_{\mathfrak{D}}\mathrm{L}(\theta,\cdot)\,d\delta^{(r)}(x)\Big)\,\mathrm{P}_\theta(dx),\qquad(2.27)$$

detta, per analogia al caso non randomizzato, **funzione di rischio** di $\delta^{(r)}$. Identificando infine ogni regola di decisione $\delta_1 \in \Delta$ con la regola randomizzata $(\delta_1)^{(r)} : x \mapsto 1_{\delta_1(x)}$, l'insieme $\Delta^{(r)}$ delle regole randomizzate diviene un ampliamento dell'insieme Δ; inoltre, si ha $R_{\delta_1} = R^{(r)}_{(\delta_1)^{(r)}}$ per ogni $\delta_1 \in \Delta$.

Il processo di casualizzazione delle decisioni conduce quindi ad una estensione del problema statistico-decisionale originario. Proviamo ora che tale processo assicura, nel caso di spazi parametrici finiti, la convessità dell'insieme di rischio delle regole randomizzate.

Teorema 2.8.3 *Sia $\Theta = \{\theta_1, \ldots, \theta_m\}$. Allora, l'insieme di rischio:*

$$\mathrm{R}^{(r)} = \{\mathbf{y}^{(\delta^{(r)})} = (R_{\delta^{(r)}}(\theta_1), \ldots, R_{\delta^{(r)}}(\theta_m)) : \delta^{(r)} \in \Delta^{(r)}\}$$

relativo alle regole randomizzate è un insieme convesso.

Dimostrazione Dati $\delta_1^{(r)}, \delta_2^{(r)} \in \Delta^{(r)}$ e $\vartheta \in [0,1]$, sia $\psi = \vartheta\delta_1^{(r)} + (1-\vartheta)\delta_2^{(r)}$. Allora, $\psi(x)$ è una decisione randomizzata per ogni campione x; inoltre, dato uno stato θ, dal Teorema A.3.4(ii) otteniamo

$$\int_{\mathfrak{D}} \mathrm{L}(\theta,\cdot)\, d\psi(x) = \vartheta \int_{\mathfrak{D}} \mathrm{L}(\theta,\cdot)\, d\delta_1^{(r)}(x) + (1-\vartheta) \int_{\mathfrak{D}} \mathrm{L}(\theta,\cdot)\, d\delta_2^{(r)}(x)$$

e quindi, per (2.27), $R_\psi(\theta) = \vartheta R_{\delta_1^{(r)}}(\theta) + (1-\vartheta) R_{\delta_2^{(r)}}(\theta)$ è finito. Allora, $\psi \in \Delta^{(r)}$ e quindi $\vartheta\mathbf{y}^{(\delta_1^{(r)})} + (1-\vartheta)\mathbf{y}^{(\delta_2^{(r)})} = \mathbf{y}^{(\psi)} \in \mathrm{R}^{(r)}$. ■

Come esemplificazione della casualizzazione delle decisioni, consideriamo il problema della verifica d'ipotesi semplici. In questo contesto, una regola randomizzata $\delta^{(r)}$ associa ad ogni campione una distribuzione di probabilità su $\mathfrak{D} = \{d_0, d_1\}$ e quindi può essere identificata con la funzione di $\mathfrak{X}$ in $[0,1]$ tale che $x \mapsto \delta^{(r)}(d_1)$[55]. Conseguentemente, chiamiamo **test randomizzato** ogni applicazione di $\mathfrak{X}$ in $[0,1]$ che sia $\aleph$-Borel misurabile. Indicati allora con τ (dotato o no di apici o pedici) i test randomizzati, possiamo introdurre la **funzione di potenza** β_τ di τ ponendo: $\beta_\tau(\theta) = \mathrm{E}_\theta(\tau)$ per ogni stato θ[56].

[55] Cioè, con la funzione che ad ogni campione assegna (da un punto di vista interpretativo) la probabilità che esca la decisione "rifiutare l'ipotesi nulla" in un meccanismo di sorteggio retto da $\delta^{(r)}(x)$.

[56] Ottenendo così una nozione compatibile con quella data nel caso non randomizzato. Infatti, notato che il test non randomizzato δ è rappresentato, nell'ambiente randomizzato,

Per quanto riguarda la funzione di rischio, da (2.27) risulta

$$\begin{aligned}
R_\tau(\theta) &= \int_{\mathfrak{X}} \Big(\int_{\mathfrak{D}} \mathrm{L}(\theta, \cdot)\, d\tau(x) \Big)\, \mathrm{P}_\theta(dx) \\
&= \int_{\mathfrak{X}} \big(\mathrm{L}(\theta, d_0)(1-\tau(x)) + \mathrm{L}(\theta, d_1)\tau(x) \big)\, \mathrm{P}_\theta(dx) \\
&= \mathrm{L}_0(\theta) \int_{\mathfrak{X}} (1-\tau(x))\, \mathrm{P}_\theta(dx) + \mathrm{L}_1(\theta) \int_{\mathfrak{X}} \tau(x)\, \mathrm{P}_\theta(dx) \\
&= \mathrm{L}_0(\theta)(1-\mathrm{E}_\theta(\tau)) + \mathrm{L}_1(\theta)\mathrm{E}_\theta(\tau) = \mathrm{L}_0(\theta)\,(1-\beta_\tau(\theta)) + \mathrm{L}_1(\theta)\beta_\tau(\theta) \\
&= \mathrm{L}_0(\theta) + \beta_\tau(\theta)[\mathrm{L}_1(\theta) - \mathrm{L}_0(\theta)]
\end{aligned}$$

da cui, posto $K_1 = k_1(\theta_0)$ e $K_0 = k_0(\theta_1)$, otteniamo l'uguaglianza:

$$R_\tau(\theta) = \begin{cases} K_1\beta_\tau(\theta_0) & \text{se } \theta = \theta_0 \\ K_0(1-\beta_\tau(\theta_1)) & \text{se } \theta = \theta_1 \end{cases},$$

che è del tutto analoga a quella relativa al caso non randomizzato. Si ha quindi, $\tau \succeq \tau'$ se e solo se $\beta_\tau(\theta_0) \le \beta_{\tau'}(\theta_0)$ e $\beta_\tau(\theta_1) \ge \beta_{\tau'}(\theta_1)$.

Supposto ora che le densità di campionamento siano a valori finiti, proviamo che la famiglia dei test $\tau^{(t,\vartheta)}$ così definiti:

$$\tau^{(t,\vartheta)}(x) = \begin{cases} 1 & \text{se } f_{\theta_1}(x) > t f_{\theta_0}(x) \\ \vartheta & \text{se } f_{\theta_1}(x) = t f_{\theta_0}(x) \\ 0 & \text{se } f_{\theta_1}(x) < t f_{\theta_0}(x) \end{cases} \quad (0 < t < +\infty;\ 0 \le \vartheta \le 1)$$

$$\tau^{(0,0)}(x) = \begin{cases} 1 & \text{se } f_{\theta_1}(x) > 0 \\ 0 & \text{se } f_{\theta_1}(x) = 0 \end{cases}, \quad \tau^{(+\infty,0)}(x) = \begin{cases} 1 & \text{se } f_{\theta_0}(x) = 0 \\ 0 & \text{se } f_{\theta_0}(x) > 0 \end{cases}$$

è una classe essenzialmente completa costituita da test randomizzati ammissibili, senza ricorrere (come fatto nel lemma di Neyman-Pearson 2.3.3) all'ipotesi: $\mathrm{P}_{\theta_0}(f_{\theta_1} = t\, f_{\theta_0}) = 0$ per ogni numero reale $t \ge 0$[57].

---

dal test:

$$\tau^{(\delta)}(x) = \begin{cases} 1 & \text{se } x \in \mathfrak{X}_1^{(\delta)} \\ 0 & \text{se } x \notin \mathfrak{X}_1^{(\delta)} \end{cases},$$

si ha $\beta_{\tau^{(\delta)}}(\theta) = \mathrm{E}_\theta(\tau^{(\delta)}) = \mathrm{P}_\theta(\tau^{(\delta)} = 1) = \mathrm{P}_\theta(\mathfrak{X}_1^{(\delta)}) = \beta_\delta(\theta)$ per ogni stato θ.

[57] I test $\tau^{(0,0)}$, $\tau^{(+\infty,0)}$ sono le versioni randomizzate, rispettivamente, dei test $\delta^{(0)}$ e $\delta^{(+\infty)}$ considerati nel lemma di Neyman-Pearson; inoltre, per ogni campione x, i test $\tau^{(t,\vartheta)}$ $(0 < t < +\infty)$ rifiutano l'ipotesi nulla, se $f_{\theta_1}(x) > t f_{\theta_0}(x)$, accettano l'ipotesi nulla, se $f_{\theta_1}(x) < t f_{\theta_0}(x)$, e accettano l'ipotesi nulla con probabilità $1-\vartheta$, se $f_{\theta_1}(x) = t f_{\theta_0}(x)$.

Teorema 2.8.4 *Sussistono le seguenti proposizioni:*

(i) $\beta_{\tau^{(t,\vartheta)}}(\theta_0) \geq \beta_\tau(\theta_0) \Rightarrow \beta_{\tau^{(t,\vartheta)}}(\theta_1) \geq \beta_\tau(\theta_1)$ *per ogni test randomizzato* τ *e* $t \in [0, +\infty]$, $\vartheta \in [0, 1]$*;*

(ii) Per ogni $\alpha \in [0, 1]$ *esistono* $t^* \in [0, +\infty]$ *e* $\vartheta^* \in [0, 1]$ *tali che* $\beta_{\tau^{(t^*,\vartheta^*)}}(\theta_0) = \alpha$*;*

(iii) La famiglia dei test $\tau^{(t,\vartheta)}$ *è una classe essenzialmente completa costituita da test randomizzati ammissibili.*

Dimostrazione Poniamo, per semplicità, $f_i = f_{\theta_i}$, $\mathrm{P}_i = \mathrm{P}_{\theta_i}$ $(i = 0, 1)$ e indichiamo con $\beta_{t,\vartheta}$ la funzione potenza del test $\tau^{(t,\vartheta)}$.

(i) Basta ripercorrere la dimostrazione della proposizione (i) del lemma di Neyman-Pearson considerando, nel caso $t \neq +\infty$, la funzione $g = (\tau^{(t,\vartheta)} - \tau)(f_1 - tf_0)$.

(ii) Se $\alpha = 0$, allora $\beta_{+\infty,0}(\theta_0) = 0$ e quindi $(t^*, \vartheta^*) = (+\infty, 0)$. Sia dunque $0 < \alpha \leq 1$. Posto $t \neq +\infty$ e considerata la v.a. Y su $\mathfrak{X}$ così definita:

$$Y(x) = \begin{cases} \frac{f_1(x)}{f_0(x)} & \text{se } f_0(x) > 0 \\ 0 & \text{se } f_0(x) = 0 \end{cases},$$

otteniamo

$$\begin{aligned} \beta_{t,\vartheta}(\theta_0) &= \int_{\mathfrak{X}} \tau^{(t,\vartheta)}\, d\mathrm{P}_0 = \mathrm{P}_0(f_1 > tf_0) + \vartheta \mathrm{P}_0(f_1 = tf_0) \\ &= \mathrm{P}_0\Big(\Big\{\frac{f_1}{f_0} > t \wedge f_0 > 0\Big\}\Big) + \vartheta \mathrm{P}_0\Big(\Big\{\frac{f_1}{f_0} = t \wedge f_0 > 0\Big\}\Big) \\ &= \mathrm{P}_0\big(\{Y > t \wedge f_0 > 0\}\big) + \vartheta \mathrm{P}_0\big(\{Y = t \wedge f_0 > 0\}\big) \\ &= \mathrm{P}_0(Y > t) + \vartheta \mathrm{P}_0(Y = t) = 1 - \mathrm{P}_0(Y \leq t) + \vartheta \mathrm{P}_0(Y = t) \end{aligned}$$

e quindi, considerando la funzione di ripartizione $F_0(t) = \mathrm{P}_0(Y \leq t)$,

$$\beta_{t,\vartheta}(\theta_0) = 1 - F_0(t) + \vartheta \mathrm{P}_0(Y = t).$$

Dobbiamo dunque trovare due numeri reali $t \geq 0$, $\vartheta \in [0, 1]$ tali che $F_0(t) - \vartheta \mathrm{P}_0(Y = t) = 1 - \alpha \in [0, 1[$. Ora, se esiste t^* tale che $F_0(t^*) = 1 - \alpha$, poniamo $t = t^*$ e $\vartheta = 0$. Altrimenti, per le proprietà della funzione di ripartizione

(Teorema B.1.1(iii), (iv), (v)), esiste t^* tale che $\mathrm{P}_0(Y < t^*) \leq 1-\alpha < F_0(t^*)$. Allora, poniamo $t = t^*$ e

$$\vartheta = \frac{F_0(t^*) - (1-\alpha)}{\mathrm{P}_0(Y = t^*)},$$

notato che $\mathrm{P}_0(Y = t^*) = F_0(t^*) - \mathrm{P}_0(Y < t^*) > 0$.

(iii) La dimostrazione è analoga a quella della proposizione (iii) del lemma di Neyman-Pearson. ■

Esempio 2.8.5 Con riferimento all'Osservazione 2.3.4, posto $\mathrm{P}_i = \mathrm{P}_{\theta_i}$ $(i = 0, 1)$, si ha

	$\{\mathrm{P}_1 > t\,\mathrm{P}_0\}$	$\{\mathrm{P}_1 = t\,\mathrm{P}_0\}$	$\{\mathrm{P}_1 < t\,\mathrm{P}_0\}$
$0 \leq t < \frac{1}{2}$	$\mathfrak{X}$	$\emptyset$	$\emptyset$
$t = \frac{1}{2}$	$\{x_2\}$	$\{x_1\}$	$\emptyset$
$\frac{1}{2} < t < \frac{7}{6}$	$\{x_2\}$	$\emptyset$	$\{x_1\}$
$t = \frac{7}{6}$	$\emptyset$	$\{x_2\}$	$\{x_1\}$
$\frac{7}{6} < t \leq +\infty$	$\emptyset$	$\emptyset$	$\mathfrak{X}$

e quindi i test:

$$\tau^{(t,\vartheta)} = \begin{cases} I_{\mathfrak{X}} & \text{se } 0 \leq t < \frac{1}{2} \\ I_{\{x_2\}} + \vartheta I_{\{x_1\}} & \text{se } t = \frac{1}{2} \\ I_{\{x_2\}} & \text{se } \frac{1}{2} < t < \frac{7}{6} \\ \vartheta I_{\{x_2\}} & \text{se } t = \frac{7}{6} \\ I_{\emptyset} & \text{se } t > \frac{7}{6} \end{cases}$$

formano una classe essenzialmente completa nell'ambito dei test randomizzati. Il test con regione di rifiuto $\{x_1\}$, che nel caso non randomizzato non era dominato da alcun test di Neyman-Pearson, è ora dominato, ad esempio, dal test $\tau^{(\frac{7}{6},\frac{1}{4})}$. ▲

Appendice A

Richiami di teoria dell'integrazione

A.1 Misure e loro proprietà

Prendendo spunto dai concetti di lunghezza, area e volume della geometria elementare, la nozione astratta di misura è stata introdotta per assegnare una "estensione" agli insiemi. Poichè, fissato un insieme "ambiente" $\Omega \neq \emptyset$, non è possibile, in generale, "misurare" tutti i suoi sottoinsiemi, si sono considerate come possibili collezioni di insiemi "misurabili" quelle famiglie $\mathcal{A} \subseteq 2^{\Omega}$, dette **$\boldsymbol{\sigma}$-algebre (su $\boldsymbol{\Omega}$)**, che contengono l'insieme ambiente e sono chiuse per complementazione e unioni discrete. Convenuto allora di denotare con la lettera $\mathcal{A}$ una generica σ-algebra su Ω e con A (dotata o no di apici o pedici) un suo elemento generico, otteniamo (ricorrendo alle formule di De Morgan) che sussistono le seguenti proprietà:

- $\emptyset, \Omega \in \mathcal{A}$;

- $A_1 \setminus A_2 \in \mathcal{A}$;

- $\bigcup_{i\in I} A_i$, $\bigcap_{i\in I} A_i \in \mathcal{A}$, se I è un insieme discreto.

Chiaramente, $\{\emptyset, \Omega\}$ e l'insieme delle parti 2^{Ω} sono, rispettivamente, la più "piccola" e la più "grande" σ-algebra; inoltre, è pure una σ-algebra l'intersezione di una collezione arbitraria di σ-algebre (su Ω). Conseguentemente, data una famiglia qualsiasi $\mathcal{F} \subseteq 2^{\Omega}$, l'intersezione $\sigma(\mathcal{F})$ di tutte le σ-algebre (su Ω) includenti $\mathcal{F}$ è una σ-algebra, detta **σ-algebra generata da $\mathcal{F}$**[1].

Nell'esempio seguente vengono introdotte alcune σ-algebre generate di notevole interesse sia teorico che applicativo.

Esempio A.1.1 (i) σ-ALGEBRA GENERATA DA UNA PARTIZIONE DISCRETA Data una partizione discreta $\mathcal{F} = (F_i)_{i\in I}$ di Ω, riesce $\sigma(\mathcal{F}) = \{\bigcup_{j\in J} F_j : J \subseteq I\}$. Notato che, prendendo come sottoinsiemi J i singoletti di I, si ha $\mathcal{F} \subset \mathcal{G} = \{\bigcup_{j\in J} F_j : J \subseteq I\}$ e che ogni σ-algebra includente la partizione deve includere anche la famiglia $\mathcal{G}$, basta verificare che $\mathcal{G}$ è una σ-algebra. Ora, $\Omega = \bigcup_{i\in I} F_i \in \mathcal{G}$; inoltre, $(\bigcup_{j\in J} F_j)^c = \bigcup_{i\in I\setminus J} F_i \in \mathcal{G}$ e $\bigcup_{n\geq 1}(\bigcup_{j\in J_n} F_j) = \bigcup_{j\in\bigcup_{n\geq 1} J_n} F_j \in \mathcal{G}$.

(ii) σ-ALGEBRA INDOTTA DA UN SOTTOINSIEME Dato un insieme non vuoto $S \subset \Omega$, la famiglia $\mathcal{A}\cap S = \{A\cap S : A \in \mathcal{A}\}$ è una σ-algebra su S, detta **traccia di $\mathcal{A}$ su S**. Infatti, $S = \Omega \cap S \in \mathcal{A} \cap S$; inoltre, $S \setminus (A \cap S) = (A \cap S)^c \cap S = (A^c \cup S^c) \cap S = A^c \cap S \in \mathcal{A} \cap S$ e $\bigcup_{n\geq 1}(A_n \cap S) = (\bigcup_{n\geq 1} A_n) \cap S \in \mathcal{A} \cap S$.

(iii) σ-ALGEBRA INDOTTA DA UNA APPLICAZIONE Considerata un'applicazione τ di un insieme $\Omega_0 \neq \emptyset$ in Ω, la famiglia $\tau^{-1}(\mathcal{A}) = \{\tau^{-1}(A) : A \in \mathcal{A}\}$ delle controimmagini degli elementi di $\mathcal{A}$ è una σ-algebra su Ω_0, detta **σ-algebra indotta da τ su Ω_0**. Infatti, $\Omega_0 = \tau^{-1}(\Omega) \in \tau^{-1}(\mathcal{A})$; inoltre, $\tau^{-1}(A)^c = \tau^{-1}(A^c) \in \tau^{-1}(\mathcal{A})$ e $\bigcup_{n\geq 1} \tau^{-1}(A_n) = \tau^{-1}(\bigcup_{n\geq 1} A_n) \in \tau^{-1}(\mathcal{A})$.

(iv) σ-ALGEBRA DI BOREL (di $\mathbb{R}$ e di $\mathbb{R}^*$) Con riferimento alla retta reale, la **σ-algebra di Borel (di $\mathbb{R}$)** è la σ-algebra $\mathcal{B}$ su $\mathbb{R}$ generata dalla famiglia degli intervalli limitati inferiormente aperti e superiormente chiusi[2]. Osservato che

$$\begin{gathered}
]a,b[\; = \bigcup_{n\geq 1} \,]a, b - \frac{1}{n}], \qquad [a,b[\; = \bigcap_{n\geq 1} \,]a - \frac{1}{n}, b[, \qquad [a,b] = \bigcap_{n\geq 1} \,]a - \frac{1}{n}, b + \frac{1}{n}[, \\
]a,b] = \bigcap_{n\geq 1} \,]a, b + \frac{1}{n}[, \qquad]a,b[\; = \bigcup_{n\geq 1} [a + \frac{1}{n}, b[, \qquad]a,b[\; = \bigcup_{n\geq 1} [a + \frac{1}{n}, b - \frac{1}{n}],
\end{gathered}$$

$$\begin{gathered}
]-\infty, a] = \bigcup_{n\geq 1} \,]a - n, a], \qquad]a,b] = \,]-\infty, b] \setminus \,]-\infty, a], \qquad]a, +\infty[\; = \,]-\infty, a]^c, \\
]-\infty, a[\; = \bigcup_{n\geq 1} [a - n, a[, \qquad [a,b[\; = \,]-\infty, b[\; \setminus \,]-\infty, a], \qquad [a, +\infty[\; = \,]-\infty, a[^c
\end{gathered}$$

per ogni $a, b \in \mathbb{R}$, la σ-algebra $\mathcal{B}$ può anche essere descritta come la σ-algebra generata da una qualsiasi famiglia d'intervalli (limitati o no) del medesimo tipo. Infine, $\mathcal{B}$ è anche generata dalla famiglia $\mathcal{U}$ degli insiemi aperti di $\mathbb{R}$ (e quindi anche dalla famiglia $\{W^c : W \in \mathcal{U}\}$ degli insiemi chiusi). Infatti, dato $W \in \mathcal{U}$, possiamo associare ad ogni suo elemento w

[1] Che verifica, per definizione, le proprietà: $\sigma(\mathcal{F}) \subseteq \sigma(\mathcal{F}')$, se $\mathcal{F} \subseteq \mathcal{F}'$, e $\sigma(\mathcal{F}) = \mathcal{F}$, se $\mathcal{F}$ è una σ-algebra.

[2] La sua introduzione, fatta nel 1898 da Emile Borel, inaugurò una nuova era dell'analisi matematica fornendo il punto di partenza sia di una classificazione topologica degli insiemi di punti che della formulazione astratta della nozione d'integrale.

un intervallo aperto $]q'_w, q''_w[$ di estremi razionali tale che $w \in]q'_w, q''_w[\subseteq W$; dunque, W è unione discreta di intervalli aperti e quindi appartiene a $\mathcal{B}$. Pertanto, la σ-algebra $\sigma(\mathcal{U})$ generata dagli insiemi aperti è inclusa in $\mathcal{B}$. Ne segue, essendo ogni intervallo aperto un insieme aperto, $\sigma(\mathcal{U}) = \mathcal{B}$.

Passando alla retta reale ampliata, la **σ-algebra di Borel** (**di** $\mathbb{R}^*$) è la σ-algebra $\mathcal{B}^*$ su $\mathbb{R}^*$ generata dalla famiglia degli intervalli di $\mathbb{R}^*$ inferiormente aperti e superiormente chiusi. Osservato che $[-\infty, a[= \{-\infty\} \cup]-\infty, a[=]-\infty, +\infty]^c \cup]-\infty, a[$ per ogni $a \in \mathbb{R}$ e tenute presenti le uguaglianze precedenti, possiamo descrivere $\mathcal{B}^*$ anche come la σ-algebra generata dalla famiglia degli intorni di base della retta reale ampliata: $]a, b[$, $[-\infty, a[$, $]a, +\infty]$ $(a, b \in \mathbb{R})$. Pertanto, con argomentazioni analoghe a quelle relative al caso della retta reale, $\mathcal{B}^*$ è pure generata dalla famiglia degli insiemi aperti della retta reale ampliata. Notato infine che

$$[-\infty, a] =]-\infty, a] \cup]-\infty, +\infty]^c, \quad]a, b] = [-\infty, b] \setminus [-\infty, a], \quad]a, +\infty] = [-\infty, a]^c$$
$$]-\infty, +\infty] = \{-\infty\}^c = \Big[\bigcap_{n\geq 1} [-\infty, -n]\Big]^c$$

per ogni $a, b \in \mathbb{R}$, possiamo affermare che $\mathcal{B}^*$ è anche generata dalla famiglia $\{\, [-\infty, a] : a \in \mathbb{R}\}$ delle semirette inferiori della retta reale ampliata di origine un numero reale. Osservato infine che $\mathbb{R} = \bigcup_{n\geq 1}]-n, n] \in \mathcal{B}^*$ e che la traccia $\mathcal{B}^* \cap \mathbb{R}$ di $\mathcal{B}^*$ su $\mathbb{R}$ include la famiglia degli intervalli di $\mathbb{R}$ inferiormente aperti e superiormente chiusi, otteniamo $\mathcal{B} \subset \mathcal{B}^*$.

Proviamo infine la chiusura di $\mathcal{B}$ per trasformazioni affini della retta reale. Posto $f(x) = \alpha x + \beta$ $(\alpha, \beta \in \mathbb{R})$, consideriamo la famiglia $\mathcal{F} = \{B \in \mathcal{B} : f(B) \in \mathcal{B}\}$. Osservato che $\mathbb{R} \in \mathcal{F}$ e tenuto conto delle uguaglianze $f(B^c) = f(B)^c$ e $f\big(\bigcup_{n\geq 1} B_n\big) = \bigcup_{n\geq 1} f(B_n)$, è facile verificare che $\mathcal{F}$ è una σ-algebra. Inoltre, $\mathcal{F}$ contiene gli intervalli chiusi in quanto $f([a, b]) = [f(a), f(b)]$, se $\alpha \geq 0$, e $f([a, b]) = [f(b), f(a)]$, se $\alpha < 0$. Conseguentemente, $\mathcal{F} = \mathcal{B}$ e quindi ogni boreliano[3] viene mutato dalla trasformazione f in un boreliano. ▲

Identificate le famiglie degli insiemi "misurabili" con le σ-algebre, passiamo alla nozione astratta di misura. Con riferimento alla σ-algebra $\mathcal{A}$, una applicazione m di $\mathcal{A}$ in $[0, +\infty]$ è una **misura** (**su** $\mathcal{A}$) se si annulla sull'insieme vuoto ed è **numerabilmente additiva** (cioè: $\mathrm{m}(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mathrm{m}(A_n)$ per ogni successione disgiunta $(A_n)_{n\geq 1}$).

Nel prossimo esempio illustriamo delle misure di particolare interesse (sia teorico che applicativo): la misura di Lebesgue unidimensionale e le misure di conteggio.

Esempio A.1.2 (i) MISURA DI LEBESGUE (UNIDIMENSIONALE) Considerata la famiglia:

$$\Im = \{\,]\alpha, \beta] \cap \mathbb{R} : -\infty \leq \alpha \leq \beta \leq +\infty\}$$

costituita dall'insieme vuoto, dalla retta reale, dagli intervalli di $\mathbb{R}$ inferiormente aperti e superiormente chiusi (limitati o no) e da quelli aperti e superiormente illimitati, chiamiamo

[3]Chiamiamo, come d'uso, **boreliano** di $\mathbb{R}$ (**di** $\mathbb{R}^*$) ogni elemento di $\mathcal{B}$ (di $\mathcal{B}^*$).

insieme elementare (**di** $\mathbb{R}$) ogni unione finita di elementi di $\Im$ a due a due disgiunti; inoltre, per **lunghezza** dell'insieme elementare E - unione degli insiemi $]\alpha_1, \beta_1], \ldots,]\alpha_n, \beta_n] \in \Im$ a due a due disgiunti - intendiamo l'elemento della retta reale ampliata:

$$\lg(E) = \sum_{i=1}^{n} (\beta_i - \alpha_i).$$

Quindi, solamente l'insieme vuoto ha lunghezza nulla e gli insiemi elementari illimitati sono gli unici di lunghezza infinita.

Indicata con $\mathcal{E}$ la famiglia degli insiemi elementari, possiamo ora introdurre, per ogni insieme S di numeri reali, la sua **misura esterna di Lebesgue**:

$$\lambda^*(S) = \inf\big\{ \sum_{n\geq 1} \lg(I_n) : (I_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } S \subseteq \bigcup_{n\geq 1} I_n \big\}$$

e la sua **misura interna di Lebesgue**:

$$\lambda_*(S) = \sup\big\{ \lg(E) - \lambda^*(E \setminus S) : \lambda^*(E \setminus S) < +\infty \wedge E \in \mathcal{E} \big\}.$$

Considerata allora la famiglia:

$$\mathcal{M} = \big\{ S \subseteq \mathbb{R} : \lambda_*(S) = \lambda^*(S) \big\}$$

dei sottoinsiemi di $\mathbb{R}$ **misurabili secondo Lebesgue**, la funzione d'insieme λ di $\mathcal{M}$ in $[0, +\infty]$ così definita:

$$\lambda(S) = \lambda_*(S) = \lambda^*(S)$$

per ogni $S \in \mathcal{M}$, può essere vista come una naturale estensione ad insiemi più complessi della nozione elementare di lunghezza, essendo $\mathcal{E} \subset \mathcal{M}$ e $\lambda(E) = \lg(E)$ per ogni insieme elementare E.

Per quanto riguarda le proprietà della famiglia $\mathcal{M}$ e della funzione d'insieme λ, è noto che $\mathcal{M}$ è una σ-algebra tale che $\mathcal{B} \subset \mathcal{M} \subset 2^{\mathbb{R}}$ e che λ è una misura, detta **misura di Lebesgue (unidimensionale)**.[4] Un'altra interessante proprietà di λ è che, data una

[4]Per la dimostrazione (alquanto complessa) rimandiamo a Rao, M.M., Measure Theory and Integration, Wiley (1987), sezioni 1.2 e 2.2. Osserviamo solamente che la misura interna di Lebesgue non può essere sostituita dalla seguente misura interna più semplice:

$$\overline{\lambda}_*(S) = \sup\big\{ \sum_{n\geq 1} \lg(I_n) : (I_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } \bigcup_{n\geq 1} I_n \subseteq S \big\}$$

derivata, per analogia, da quella esterna. Infatti, considerato come S l'insieme numerabile dei numeri razionali dell'intervallo $[0, 1]$ e posto $S' = [0, 1] \setminus S$, si avrebbe $\overline{\lambda}_*([0, 1]) = 1 = \lambda^*([0, 1])$, $\overline{\lambda}_*(S) = 0 = \lambda^*(S)$ e $\overline{\lambda}_*(S') = 0 < 1 = \lambda^*(S')$; conseguentemente, con questa differente definizione di misura interna, non sarebbe soddisfatta la richiesta, peraltro molto naturale, che la differenza di insiemi misurabili sia misurabile.

trasformazione affine $f(x) = \alpha x + \beta$ ($\alpha \neq 0$) della retta reale, risulta $\lambda(f(B)) = |\alpha|\lambda(B)$ per ogni $B \in \mathcal{B}$. Infatti, per ogni insieme elementare $E = \bigcup_{i=1}^{n}]a_i, b_i]$, l'insieme

$$f(E) = \begin{cases} \bigcup_{i=1}^{n}]f(a_i), f(b_i)] & \text{se } \alpha > 0 \\ \bigcup_{i=1}^{n}]f(b_i), f(a_i)] & \text{se } \alpha < 0 \end{cases}$$

è elementare e riesce

$$\lg(f(E)) = \sum_{i=1}^{n} |\alpha|\,(b_i - a_i) = |\alpha| \sum_{i=1}^{n} (b_i - a_i) = |\alpha| \lg(E).$$

Conseguentemente, per ogni boreliano B, si ha

$$\begin{aligned} |\alpha|\,\lambda(B) &= |\alpha|\,\lambda^*(B) \\ &= |\alpha| \inf\Big\{\sum_{n\geq 1} \lg(I_n) : (I_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } B \subseteq \bigcup_{n\geq 1} I_n\Big\} \\ &= \inf\Big\{\sum_{n\geq 1} |\alpha| \lg(I_n) : (I_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } B \subseteq \bigcup_{n\geq 1} I_n\Big\} \\ &= \inf\Big\{\sum_{n\geq 1} \lg(f(I_n)) : (I_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } f(B) \subseteq \bigcup_{n\geq 1} f(I_n)\Big\} \\ &= \inf\Big\{\sum_{n\geq 1} \lg(J_n) : (J_n)_{n\geq 1} \text{ successione disgiunta in } \Im \text{ e } f(B) \subseteq \bigcup_{n\geq 1} J_n\Big\} \\ &= \lambda^*(f(B)) = \lambda(f(B)). \end{aligned}$$

(ii) Misure di conteggio Indicato con $\#J$ il numero di elementi di un qualsiasi insieme finito J, consideriamo un sottoinsieme S di Ω. Allora, la funzione d'insieme $\gamma_S : 2^\Omega \mapsto [0, +\infty]$ così definita:

$$\gamma_S(A) = \begin{cases} \#A \cap S & \text{se } A \cap S \text{ finito} \\ +\infty & \text{altrimenti} \end{cases},$$

che conta il numero di elementi comuni ad A e S, è una misura, detta **misura di conteggio indotta da S (su Ω)**. Poichè $\gamma_S(\emptyset) = 0$, per provarlo basta constatare che γ_S è numerabilmente additiva. Sia dunque $(A_n)_{n\geq 1}$ una successione disgiunta in 2^Ω e $A = \bigcup_{n\geq 1} A_n$. Supponiamo intanto $A \cap S$ finito. Esiste allora m tale che $A_n = \emptyset$ per ogni $n > m$. Ne segue $A \cap S = \bigcup_{i=1}^{m} (A_i \cap S)$ e $\gamma_S(A_n) = 0$ per ogni $n > m$. Riesce pertanto

$$\gamma_S(A) = \#A \cap S = \sum_{i=1}^{m} \#A_i \cap S = \sum_{i=1}^{m} \gamma_S(A_i) = \sum_{n\geq 1} \gamma_S(A_n).$$

Assumiamo infine $A \cap S$ infinito, cioè $\gamma_S(A) = +\infty$. Ora, se $A_m \cap S$ è infinito per qualche m, si ha $\sum_{n\geq 1} \gamma_S(A_n) \geq \gamma_S(A_m) = +\infty$. Se invece risulta $A_n \cap S$ finito per ogni $n \geq 1$, esiste una sottosuccessione $(A_{i_n})_{n\geq 1}$ tale che $A_{i_n} \cap S \neq \emptyset$ per ogni n. Riesce quindi

$$\sum_{n\geq 1} \gamma_S(A_n) \geq \sum_{n\geq 1} \gamma_S(A_{i_n}) \geq 1 + 1 + \dots = +\infty.$$

In ogni caso si ha dunque $\sum_{n\geq 1}\gamma_S(A_n) = +\infty = \gamma_S(A)$.

Nel caso particolare che sia $S = \{\omega\}$, la misura di conteggio $\gamma_{\{\omega\}}$ viene detta **misura di Dirac di indice** ω e denotata con il simbolo δ_ω. Si ha quindi

$$\delta_\omega(A) = \begin{cases} 1 & \text{se } \omega \in A \\ 0 & \text{se } \omega \notin A \end{cases}$$

per ogni sottoinsieme A di Ω. ▲

Nel teorema seguente riportiamo alcune importanti proprietà delle misure. In particolare, (i), (iv) e (v) assicurano che una misura è una funzione d'insieme additiva, monotona e subadditiva (in senso sia finito che numerabile) mentre (vi), (vii) che è continua sulle successioni monotone d'insiemi. La (viii) invece consente di calcolare la misura di un'unione finita di elementi di $\mathcal{A}$ tramite i valori che la misura assume su tutte le loro possibili intersezioni. Le (ix), (x) infine che unioni discrete di insiemi di misura nulla sono ancora di misura nulla e che intersezioni discrete di insiemi di misura massima (purchè finita) sono ancora di misura massima.

Teorema A.1.3 *Sia* m *una misura su* $\mathcal{A}$. *Sussistono allora le seguenti proposizioni:*

(i) ADDITIVITÀ: $\mathrm{m}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathrm{m}(A_i)$, *se* $(A_i)_{i\leq n}$ *è disgiunta;*

(ii) $\mathrm{m}(A_1 \cup A_2) + \mathrm{m}(A_1 \cap A_2) = \mathrm{m}(A_1) + \mathrm{m}(A_2)$*;*

(iii) $\mathrm{m}(A_2 \setminus A_1) = \mathrm{m}(A_2) - \mathrm{m}(A_1)$, *se* $A_1 \subseteq A_2$ *e* $\mathrm{m}(A_1) < +\infty$*;*

(iv) MONOTONIA: $\mathrm{m}(A_1) \leq \mathrm{m}(A_2)$, *se* $A_1 \subseteq A_2$*;*

(v) SUBADDITIVITÀ: $\mathrm{m}(\bigcup_{i\in I} A_i) \leq \sum_{i\in I} \mathrm{m}(A_i)$, *se* I *è discreto;*

(vi) CONTINUITÀ DAL BASSO: $\mathrm{m}(A_n) \uparrow \mathrm{m}(\bigcup_{n\geq 1} A_n)$, *se* $(A_n)_{n\geq 1}$ *è una successione non decrescente;*

(vii) CONTINUITÀ DALL'ALTO: $\mathrm{m}(A_n) \downarrow \mathrm{m}(\bigcap_{n\geq 1} A_n)$, *se* $(A_n)_{n\geq 1}$ *è una successione non crescente tale che* $\mathrm{m}(A_m) < +\infty$ *per qualche* m[5]*;*

[5] L'ipotesi di finitezza della misura di qualche termine della successione non può essere, in generale, omessa. Infatti, posto $A_n = [n, +\infty[$ per ogni n, la successione $(A_n)_{n\geq 1}$ è non crescente; inoltre $\bigcap_{n\geq 1} A_n = \emptyset$ e $\lambda(A_n) = +\infty$ per ogni n. Si ha pertanto $\lambda(A_n) \downarrow +\infty \neq 0 = \lambda(\bigcap_{n\geq 1} A_n)$.

(viii) FORMULA D'INCLUSIONE-ESCLUSIONE: *Sia* $\mathrm{m}(A_i) < +\infty$ $(i = 1, \ldots, n)$. *Allora,*

$$\mathrm{m}\big(\bigcup_{i=1}^{n} A_i\big) = \sum_{\emptyset \subset J \subseteq \{1,\ldots,n\}} (-1)^{\#J-1}\, \mathrm{m}\big(\bigcap_{j\in J} A_j\big);$$

(ix) *Sia* $\mathrm{m}(A_i) = 0$ *per ogni* $i \in I$ *con* I *discreto. Allora,* $\mathrm{m}\big(\bigcup_{i\in I} A_i\big) = 0$*;*

(x) *Sia* $\mathrm{m}(\Omega) < +\infty$ *e* $\mathrm{m}(A_i) = \mathrm{m}(\Omega)$ *per ogni* $i \in I$ *con* I *discreto. Allora,* $\mathrm{m}\big(\bigcap_{i\in I} A_i\big) = \mathrm{m}(\Omega)$[6].

DIMOSTRAZIONE (i) Supposto $A_1, \ldots, A_n$ a due a due disgiunti, sia $A_m = \emptyset$ per ogni $m > n$. Allora, da $\mathrm{m}(\emptyset) = 0$ e dalla numerabile additività otteniamo

$$\mathrm{m}\big(\bigcup_{i=1}^{n} A_i\big) = \mathrm{m}\big(\bigcup_{n\geq 1} A_n\big) = \sum_{n\geq 1} \mathrm{m}(A_n) = \sum_{i=1}^{n} \mathrm{m}(A_i).$$

(ii) Poichè $A_1 \cup A_2 = A_1 \cup (A_2 \setminus A_1)$ e $A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2)$, tramite (i), risulta $\mathrm{m}(A_1 \cup A_2) = \mathrm{m}(A_1) + \mathrm{m}(A_2 \setminus A_1)$ e $\mathrm{m}(A_2) = \mathrm{m}(A_2 \setminus A_1) + \mathrm{m}(A_1 \cap A_2)$. Ne segue

$$\begin{aligned}\mathrm{m}(A_1 \cup A_2) + \mathrm{m}(A_1 \cap A_2) &= [\mathrm{m}(A_1) + \mathrm{m}(A_2 \setminus A_1)] + \mathrm{m}(A_1 \cap A_2)\\ &= \mathrm{m}(A_1) + [\mathrm{m}(A_2 \setminus A_1) + \mathrm{m}(A_1 \cap A_2)] = \mathrm{m}(A_1) + \mathrm{m}(A_2).\end{aligned}$$

(iii) + (iv) Sia $A_1 \subseteq A_2$. Allora, $A_2 = A_1 \cup (A_2 \setminus A_1)$ e quindi, per (i), $\mathrm{m}(A_2) = \mathrm{m}(A_1) + \mathrm{m}(A_2 \setminus A_1)$.

(v) Sia intanto I numerabile. Allora, posto $I = \{i_1, i_2, \ldots\}$ e tenuto conto dell'uguaglianza

$$\bigcup_{n\geq 1} A_{i_n} = A_{i_1} \cup \bigcup_{n\geq 1} [A_{i_{n+1}} \setminus (A_{i_1} \cup \cdots \cup A_{i_n})], \tag{A.1}$$

tramite la numerabile additività e (iv) si ha

$$\begin{aligned}\mathrm{m}\big(\bigcup_{i\in I} A_i\big) &= \mathrm{m}(A_{i_1}) + \sum_{n\geq 1} \mathrm{m}(A_{i_{n+1}} \setminus (A_{i_1} \cup \cdots \cup A_{i_n}))\\ &\leq \mathrm{m}(A_{i_1}) + \sum_{n\geq 1} \mathrm{m}(A_{i_{n+1}}) = \sum_{n\geq 1} \mathrm{m}(A_{i_n}) = \sum_{i\in I} \mathrm{m}(A_i).\end{aligned}$$

Sia ora I finito. Allora, posto $I = \{i_1, \ldots, i_k\}$ e

$$A'_n = \begin{cases} A_{i_n} & \text{se } n \leq k \\ \emptyset & \text{se } n > m \end{cases},$$

[6] L'ipotesi di finitezza della misura non può essere, in generale, omessa. Infatti, posto $A_1 =]-\infty, 0]$ e $A_2 =]0, +\infty[$, otteniamo $\lambda(A_1) = \lambda(A_2) = +\infty = \lambda(\mathbb{R})$ e $\lambda(A_1 \cap A_2) = 0$.

da quanto appena provato e da $\mathrm{m}(\emptyset) = 0$ si ha

$$\mathrm{m}\big(\bigcup_{i\in I} A_i\big) = \mathrm{m}\big(\bigcup_{n\geq 1} A'_n\big) \leq \sum_{n\geq 1} \mathrm{m}(A'_n) = \sum_{i=1}^{k} \mathrm{m}(A'_i) = \sum_{i\in I} \mathrm{m}(A_i).$$

(vi) Sia $(A_n)_{n\geq 1}$ non decrescente. Allora, per (iv), $(\mathrm{m}(A_n))_{n\geq 1}$ è una successione numerica non decrescente. Per provare che $\mathrm{m}(A_n) \to \mathrm{m}(\bigcup_{n\geq 1} A_n)$, assumiamo intanto $\mathrm{m}(A_k) = +\infty$ per qualche k. Si ha quindi $\mathrm{m}(A_n) = +\infty$ per ogni $n \geq k$ e, per (iv), $\mathrm{m}(\bigcup_{n\geq 1} A_n) = +\infty$; dunque $\mathrm{m}(A_n) \to \mathrm{m}(\bigcup_{n\geq 1} A_n)$. Supponiamo ora che $\mathrm{m}(A_n)$ sia sempre finito. Da (A.1) otteniamo $\bigcup_{n\geq 1} A_n = A_1 \cup \bigcup_{n\geq 1} (A_{n+1} \setminus A_n)$ e quindi, tramite (i) e la numerabile additività,

$$\begin{aligned}\mathrm{m}\big(\bigcup_{n\geq 1} A_n\big) &= \mathrm{m}(A_1) + \sum_{n\geq 1} \mathrm{m}(A_{n+1} \setminus A_n) = \lim_{n\to+\infty} \Big[\mathrm{m}(A_1) + \sum_{i=1}^{n} \mathrm{m}(A_{i+1} \setminus A_i)\Big] \\ &= \lim_{n\to+\infty} \mathrm{m}\Big(A_1 \cup \big[\bigcup_{i=1}^{n} (A_{i+1} \setminus A_i)\big]\Big) = \lim_{n\to+\infty} \mathrm{m}(A_{n+1}) = \lim_{n\to+\infty} \mathrm{m}(A_n).\end{aligned}$$

(vii) Siano $(A_n)_{n\geq 1}$ non crescente e m tale che $\mathrm{m}(A_m) < +\infty$. Allora, per (iv), $\mathrm{m}(\bigcap_{n\geq 1} A_n) < +\infty$, $\mathrm{m}(A_n) < +\infty$ per ogni $n \geq m$ e la successione numerica $(\mathrm{m}(A_n))_{n\geq 1}$ è non crescente. Inoltre, osservato che la successione $(A_m \setminus A_n)_{n\geq 1}$ è non decrescente, da (iii), (vi) otteniamo

$$\begin{aligned}\mathrm{m}(A_m) - \mathrm{m}\big(\bigcap_{n\geq 1} A_n\big) &= \mathrm{m}\big(A_m \setminus \bigcap_{n\geq 1} A_n\big) = \mathrm{m}\big(A_m \cap \bigcup_{n\geq 1} A_n^c\big) = \mathrm{m}\big(\bigcup_{n\geq 1} (A_m \setminus A_n)\big) \\ &= \lim_{n\to+\infty} \mathrm{m}(A_m \setminus A_n) = \lim_{n\to+\infty} [\mathrm{m}(A_m) - \mathrm{m}(A_n)] \\ &= \mathrm{m}(A_m) - \lim_{n\to+\infty} \mathrm{m}(A_n)\end{aligned}$$

e quindi, essendo $\mathrm{m}(A_m)$ finito, $\mathrm{m}(A_n) \to \mathrm{m}(\bigcap_{n\geq 1} A_n)$.

(viii) Procediamo per induzione su n. Poichè la base dell'induzione "$n = 1$" è ovvia, assumiamo che (viii) sussista per $n = k$ (ipotesi induttiva) e proviamola per $n = k + 1$. Sia dunque $\mathrm{m}(A_i) < +\infty$ $(i = 1, \ldots, k+1)$. Da (iv), (v) otteniamo $\mathrm{m}\big(\bigcup_{i=1}^{k} (A_i \cap A_{k+1})\big) \leq \mathrm{m}(\bigcup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} \mathrm{m}(A_i) < +\infty$ e quindi, per (ii),

$$\mathrm{m}\big(\bigcup_{i=1}^{k+1} A_i\big) = \mathrm{m}\big(A_{k+1} \cup \bigcup_{i=1}^{k} A_i\big) = \mathrm{m}(A_{k+1}) + \mathrm{m}\big(\bigcup_{i=1}^{k} A_i\big) - \mathrm{m}\big(\bigcup_{i=1}^{k} (A_i \cap A_{k+1})\big).$$

Per l'ipotesi induttiva riesce allora

$$\begin{aligned}\mathrm{m}\big(\bigcup_{i=1}^{k+1} A_i\big) &= \mathrm{m}(A_{k+1}) + \sum_{\emptyset\subset J\subseteq\{1,\ldots,k\}} (-1)^{\#J-1}\, \mathrm{m}\big(\bigcap_{j\in J} A_j\big) \\ &\quad - \sum_{\emptyset\subset J\subseteq\{1,\ldots,k\}} (-1)^{\#J-1}\, \mathrm{m}\big(\bigcap_{j\in J} (A_j \cap A_{k+1})\big)\end{aligned}$$

$$= \mathrm{m}(A_{k+1}) + \sum_{\emptyset \subset J \subseteq \{1,\ldots,k\}} (-1)^{\#J-1}\, \mathrm{m}\big(\bigcap_{j\in J} A_j\big)$$
$$+ \sum_{\emptyset \subset J \subseteq \{1,\ldots,k\}} (-1)^{\#[J\cup\{k+1\}]-1}\, \mathrm{m}\big(\bigcap_{j\in J} (A_j \cap A_{k+1})\big).$$

Osservato infine che, con riferimento all'insieme $I = \{1, \ldots, k+1\}$, il primo addendo della somma riguarda il sottoinsieme di I formato solo dall'elemento $k+1$, il secondo i sottoinsiemi di I formati solo con elementi di $\{1, \ldots, k\}$ e l'ultimo i sottoinsiemi di I contenenti $k+1$ e aventi almeno due elementi, otteniamo

$$\mathrm{m}\big(\bigcup_{i=1}^{k+1} A_i\big) = \sum_{\emptyset \subset J \subseteq \{1,\ldots,k+1\}} (-1)^{\#J-1}\, \mathrm{m}\big(\bigcap_{j\in J} A_j\big).$$

(ix) + (x) La proposizione (ix) è conseguenza immediata di (v). Per quanto riguarda (x), da (iii) risulta $\mathrm{m}(A_i^c) = \mathrm{m}(\Omega) - \mathrm{m}(A_i) = 0$ per ogni $i \in I$. Posto $A = \bigcap_{i\in I} A_i$, per (ix) si ha allora $\mathrm{m}(A^c) = \mathrm{m}\big(\bigcup_{i\in I} A_i^c\big) = 0$ e quindi, per (iii), $\mathrm{m}(A) = \mathrm{m}(\Omega \setminus A^c) = \mathrm{m}(\Omega) - \mathrm{m}(A^c) = \mathrm{m}(\Omega)$. ■

Un ruolo chiave (a livello dimostrativo) per ottenere un criterio standard che assicuri l'uguaglianza di due misure viene svolto da quelle particolari famiglie $\mathcal{D} \subseteq 2^\Omega$, dette **classi di Dynkin (su Ω)**, che verificano le proprietà:

- $\Omega \in \mathcal{D}$;
- $D^c \in \mathcal{D}$, se $D \in \mathcal{D}$;
- $\bigcup_{n\geq 1} D_n \in \mathcal{D}$, se $(D_n)_{n\geq 1}$ è una successione disgiunta in $\mathcal{D}$,

cioè contenenti l'insieme ambiente e chiuse per complementazione e unioni disgiunte numerabili.

Chiaramente, ogni σ-algebra è una classe di Dynkin ma non vale il viceversa; per constatarlo basta porre $\Omega = \{1, 2, 3, 4\}$ e considerare la classe di Dynkin $\mathcal{D} = \{\emptyset, \{1, 2\}, \{3, 4\}, \{1, 3\}, \{2, 4\}, \Omega\}$. D'altra parte, sono σ-algebre le classi di Dynkin $\mathcal{D}$ chiuse per intersezioni finite; basta infatti notare che, per (A.1), $\bigcup_{n\geq 1} D_n = D_1 \cup \bigcup_{n\geq 1} (D_{n+1} \cap D_1^c \cap \cdots \cap D_n^c)$ per ogni successione $(D_n)_{n\geq 1}$ in $\mathcal{D}$.

Il prossimo lemma assicura che una classe di Dynkin includente una famiglia chiusa per intersezioni finite è sufficientemente ampia da contenere anche la σ-algebra generata dalla famiglia stessa.

Lemma A.1.4 *Siano $\mathcal{D}$ una classe di Dynkin e $\mathcal{F} \subseteq \mathcal{D}$ una famiglia chiusa per intersezioni finite. Allora, $\sigma(\mathcal{F}) \subseteq \mathcal{D}$.*

Dimostrazione Sia $\mathcal{D}_0$ l'intersezione di tutte le classi di Dynkin (su Ω) includenti $\mathcal{F}$. Allora, $\mathcal{F} \subseteq \mathcal{D}_0 \subseteq \mathcal{D}$ e inoltre, come facilmente si verifica, $\mathcal{D}_0$ è una classe di Dynkin. Basta quindi provare che $\mathcal{D}_0$ è una σ-algebra, cioè che è chiusa per intersezioni finite. A tal fine, data un'arbitraria famiglia non vuota $\mathcal{G} \subseteq \mathcal{D}_0$, proviamo che la famiglia:

$$\mathcal{D}(\mathcal{G}) = \{D \in \mathcal{D}_0 : D \cap G \in \mathcal{D}_0 \text{ per ogni } G \in \mathcal{G}\} \subseteq \mathcal{D}_0$$

è una classe di Dynkin. Poichè $\Omega \in \mathcal{D}_0$ e $\Omega \cap G = G$ per ogni $G \in \mathcal{G}$ si ha $\Omega \in \mathcal{D}(\mathcal{G})$. Sia ora $D \in \mathcal{D}(\mathcal{G})$. Ne segue $D \in \mathcal{D}_0$; inoltre, scelto un arbitrario $G \in \mathcal{G}$, riesce $D \cap G \in \mathcal{D}_0$. Pertanto, $D^c \in \mathcal{D}_0$, $D^c \cap G = (D \cap G)^c \cap G = [(D \cap G) \cup G^c]^c \in \mathcal{D}_0$ e quindi $D \in \mathcal{D}(\mathcal{G})$. Sia infine $(D_n)_{n\geq 1}$ una successione disgiunta in $\mathcal{D}(\mathcal{G})$. Allora, qualunque sia n, risulta $D_n \in \mathcal{D}_0$; inoltre, scelto un arbitrario $G \in \mathcal{G}$, si ha $D_n \cap G \in \mathcal{D}_0$. Pertanto, $\bigcup_{n\geq 1} D_n \in \mathcal{D}_0$ e $(\bigcup_{n\geq 1} D_n) \cap G = \bigcup_{n\geq 1}(D_n \cap G) \in \mathcal{D}_0$ (la successione $(D_n \cap G)_{n\geq 1}$ è disgiunta!). Quindi $\bigcup_{n\geq 1} D_n \in \mathcal{D}(\mathcal{G})$.

Scelto ora $\mathcal{G} = \mathcal{F}$ e osservato che $\mathcal{F} \subseteq \mathcal{D}(\mathcal{F}) \subseteq \mathcal{D}_0$ ($\mathcal{F}$ è chiusa per intersezioni finite!), dalla definizione di $\mathcal{D}_0$ otteniamo $\mathcal{D}_0 = \mathcal{D}(\mathcal{F})$. Allora, $F \cap D \in \mathcal{D}_0$ per ogni $D \in \mathcal{D}_0$ e $F \in \mathcal{F}$. Conseguentemente, scelto infine $\mathcal{G} = \mathcal{D}_0$, si ha $\mathcal{F} \subseteq \mathcal{D}(\mathcal{D}_0) \subseteq \mathcal{D}_0$ e quindi $\mathcal{D}_0 = \mathcal{D}(\mathcal{D}_0) = \{D \in \mathcal{D}_0 : D \cap D' \in \mathcal{D}_0 \text{ per ogni } D' \in \mathcal{D}_0\}$, cioè $\mathcal{D}_0$ è chiusa per intersezioni finite. ■

Siamo ora in grado di provare un criterio per l'uguaglianza di misure che, in particolare, attesta l'identità di due misure finite che risultano coincidenti su un sistema di generatori chiuso per intersezioni finite e contenente l'insieme ambiente.

Teorema A.1.5 (Criterio standard d'identità) *Siano* m_1, m_2 *misure su* $\mathcal{A}$ *e* $\mathcal{F}$ *una famiglia chiusa per intersezioni finite tale che* $\mathcal{A} = \sigma(\mathcal{F})$ *e* $\mathrm{m}_1(F) = \mathrm{m}_2(F)$ *per ogni* $F \in \mathcal{F}$. *Esista inoltre una successione* $(F_n)_{n\geq 1}$ *in* $\mathcal{F}$ *tale che* $\Omega = \bigcup_{n\geq 1} F_n$ *e* $\mathrm{m}_1(F_n) < +\infty$ *per ogni* n. *Allora,* $\mathrm{m}_1 = \mathrm{m}_2$[7].

Dimostrazione Proviamo intanto che, dato $F \in \mathcal{F}$ tale che $\mathrm{m}_1(F) < +\infty$, risulta $\mathrm{m}_1(F \cap A) = \mathrm{m}_2(F \cap A)$ per ogni A. A tal fine, posto

$$\mathcal{D} = \{A \in \sigma(\mathcal{F}) : \mathrm{m}_1(F \cap A) = \mathrm{m}_2(F \cap A)\}$$

e osservato che $\mathcal{F} \subseteq \mathcal{D}$ ($\mathcal{F}$ è chiusa per intersezioni finite!), per il Lemma A.1.4 basta verificare che $\mathcal{D}$ è una classe di Dynkin. La condizione $\Omega \in \mathcal{D}$ è ovvia. Scelto $D \in \mathcal{D}$, si ha $D^c \in \sigma(\mathcal{F})$ e, per la monotonia della misura, $\mathrm{m}_2(F \cap D) = \mathrm{m}_1(F \cap D) \leq \mathrm{m}_1(F) < +\infty$.

[7] Conseguentemente, la misura di Lebesgue unidimensionale è l'unica misura sui boreliani della retta reale che sia compatibile con la lunghezza degli intervalli; infatti, il sistema di generatori $\mathcal{F} = \{]a, b] : a, b \in \mathbb{R}\}$ è chiuso per intersezioni finite e, indicato con $\mathbb{Z}$ l'insieme dei numeri relativi, $\mathbb{R} = \bigcup_{z\in\mathbb{Z}}]z, z + 1]$.

Ne segue, per il Teorema A.1.3(iii),

$$\begin{aligned}\mathrm{m}_1(F\cap D^c) &= \mathrm{m}_1\big(F\setminus(F\cap D)\big) = \mathrm{m}_1(F) - \mathrm{m}_1(F\cap D)\\ &= \mathrm{m}_2(F) - \mathrm{m}_2(F\cap D) = \mathrm{m}_2\big(F\setminus(F\cap D)\big) = \mathrm{m}_2(F\cap D^c)\end{aligned}$$

e quindi $D^c \in \mathcal{D}$. Considerata infine una successione disgiunta $(D_n)_{n\geq 1}$ in $\mathcal{D}$, si ha $\bigcup_{n\geq 1} D_n \in \sigma(\mathcal{F})$, $\mathrm{m}_1(F\cap D_n) = \mathrm{m}_2(F\cap D_n)$ per ogni n e

$$\mathrm{m}_k\big(F\cap \bigcup_{n\geq 1} D_n\big) = \mathrm{m}_k\big(\bigcup_{n\geq 1}(F\cap D_n)\big) = \sum_{n\geq 1}\mathrm{m}_k(F\cap D_n) \qquad (k=1,2).$$

Pertanto $\bigcup_{n\geq 1} D_n \in \mathcal{D}$.

Proviamo ora che $\mathrm{m}_1 = \mathrm{m}_2$. Dato A, consideriamo la successione non decrescente $\big(\bigcup_{i=1}^n (A\cap F_i)\big)_{n\geq 1}$. Notato che, per la monotonia della misura, $\mathrm{m}_k(A\cap F_n) < +\infty$ per ogni n $(k=1,2)$, dalla continuità dal basso della misura e dalla formula d'inclusione-esclusione otteniamo allora

$$\begin{aligned}\mathrm{m}_k(A) &= \mathrm{m}_k(A\cap\Omega) = \mathrm{m}_k\big(A\cap\bigcup_{n\geq 1}F_n\big) = \mathrm{m}_k\big(\bigcup_{n\geq 1}(A\cap F_n)\big) = \lim_{n\to+\infty}\mathrm{m}_k\big(\bigcup_{i=1}^n (A\cap F_i)\big)\\ &= \lim_{n\to+\infty}\sum_{\emptyset\subset J\subseteq\{1,\dots,n\}}(-1)^{\#J-1}\,\mathrm{m}_k\big(\bigcap_{j\in J}(A\cap F_j)\big)\\ &= \lim_{n\to+\infty}\sum_{\emptyset\subset J\subseteq\{1,\dots,n\}}(-1)^{\#J-1}\,\mathrm{m}_k\big(A\cap\bigcap_{j\in J}F_j\big) \qquad (k=1,2).\end{aligned}$$

Ora, dato un arbitrario insieme non vuoto $J\subseteq\{1,\dots,n\}$, risulta $\bigcap_{j\in J}F_j \in \mathcal{F}$ ($\mathcal{F}$ è chiusa per intersezioni finite!) e, per la monotonia della misura, $\mathrm{m}_1(\bigcap_{j\in J}F_j) < +\infty$; allora, per quanto provato all'inizio della dimostrazione, $\mathrm{m}_1(A\cap\bigcap_{j\in J}F_j) = \mathrm{m}_2(A\cap\bigcap_{j\in J}F_j)$. Conseguentemente, $\mathrm{m}_1(A) = \mathrm{m}_2(A)$. ■

A.2 Applicazioni misurabili

Una nozione centrale nella teoria dell'integrazione è certamente quella di applicazione misurabile che ora formuliamo. Considerata, accanto alla σ-algebra $\mathcal{A}$ su Ω, una σ-algebra $\mathcal{A}'$ su un insieme non vuoto Ω', una applicazione $f:\Omega\mapsto\Omega'$ è **$(\mathcal{A},\mathcal{A}')$-misurabile** se la σ-algebra indotta $f^{-1}(\mathcal{A}') = \big\{f^{-1}(A') : A'\in\mathcal{A}'\big\}$ (su Ω da f) è inclusa in $\mathcal{A}$[8].

[8]Da un punto di vista formale, questa nozione presenta una stretta analogia con quella topologica di funzione continua. Infatti, indicata con $\mathcal{U}^{(n)}$ la famiglia degli insiemi aperti di $\mathbb{R}^n$ $(n\geq 1)$, una funzione $f:\mathbb{R}^m\mapsto\mathbb{R}^n$ è continua se e solo se $f^{-1}(\mathcal{U}^{(n)}) = \big\{f^{-1}(U) : U\in\mathcal{U}^{(n)}\big\}\subseteq\mathcal{U}^m$.

Chiaramente, ogni applicazione costante di Ω in Ω' è $(\mathcal{A}, \mathcal{A}')$-misurabile. Inoltre, se $\mathcal{A} = 2^\Omega$ o $\mathcal{A}' = \{\emptyset, \Omega'\}$, sono $(\mathcal{A}, \mathcal{A}')$-misurabili tutte le applicazioni di Ω in Ω'. Infine, la misurabilità si conserva (come si constata facilmente) per composizione, nel senso che - considerata una σ-algebra $\mathcal{A}''$ su un insieme $\Omega'' \neq \emptyset$ - sono $(\mathcal{A}, \mathcal{A}'')$-misurabili tutte le applicazioni ottenute componendo applicazioni $(\mathcal{A}, \mathcal{A}')$-misurabili con applicazioni $(\mathcal{A}', \mathcal{A}'')$-misurabili.

Convenendo che la lettera A' (dotata o no di pedici) denoti sempre un elemento generico di $\mathcal{A}'$, proviamo che - date una partizione discreta $(A_i)_{i\in I}$ e una famiglia $(f_i)_{i\in I}$ di applicazioni $(\mathcal{A}, \mathcal{A}')$-misurabili - la $(\mathcal{A}, \mathcal{A}')$-misurabilità dell'applicazione ottenuta "per rincollamento delle f_i secondo la partizione $(A_i)_{i\in I}$" (cioè, dell'applicazione coincidente, per ogni $i \in I$, con f_i su A_i).

Lemma A.2.1 *Sussistono le seguenti proposizioni:*

(i) *Siano f un'applicazione $(\mathcal{A}, \mathcal{A}')$-misurabile e S un sottoinsieme non vuoto di Ω. Allora, considerata la traccia $\mathcal{A}\cap S$ di $\mathcal{A}$ su S, la restrizione $f\big|_S$ è un'applicazione $(\mathcal{A} \cap S, \mathcal{A}')$-misurabile;*

(ii) *Sia $(A_i)_{i\in I}$ una partizione discreta di Ω. Inoltre, considerata, per ogni $i \in I$, la traccia $\mathcal{A} \cap A_i$ di $\mathcal{A}$ su A_i, sia $f_i : A_i \mapsto \Omega'$ un'applicazione $(\mathcal{A} \cap A_i, \mathcal{A}')$-misurabile. Allora, l'applicazione $f : \Omega \mapsto \Omega'$ così definita:*

$$f(\omega) = f_i(\omega) \ \textit{ per ogni } \omega \in A_i \textit{ e } i \in I$$

è $(\mathcal{A}, \mathcal{A}')$-misurabile.

Dimostrazione (i) Dato A', otteniamo $f^{-1}(A') \in \mathcal{A}$ e quindi $(f\big|_S)^{-1}(A') = \{\omega \in S : f(\omega) \in A'\} = f^{-1}(A') \cap S \in \mathcal{A} \cap S$.

(ii) Dato A', risulta

$$\begin{aligned} f^{-1}(A') &= f^{-1}(A') \cap \bigcup_{i\in I} A_i = \bigcup_{i\in I} (f^{-1}(A') \cap A_i) = \bigcup_{i\in I} \{\omega \in A_i : f(\omega) \in A'\} \\ &= \bigcup_{i\in I} \{\omega \in A_i : f_i(\omega) \in A'\} = \bigcup_{i\in I} f_i^{-1}(A'). \end{aligned}$$

Osservato che, per ogni $i \in I$, si ha $f_i^{-1}(A') \in \mathcal{A} \cap A_i \subseteq \mathcal{A}$ ($A_i \in \mathcal{A}$!), la controimmagine $f^{-1}(A')$ risulta unione discreta di elementi di $\mathcal{A}$ e quindi appartiene ad $\mathcal{A}$. ■

Considerando ora il caso particolare, ma di estrema importanza, delle funzioni su Ω a valori nella retta reale (o nella retta reale ampliata), precisiamo che faremo sempre riferimento, per la misurabilità, alla σ-algebra di

Borel $\mathcal{B}$ (o $\mathcal{B}^*$) e chiameremo **$\mathcal{A}$-Borel misurabile** (in breve **Borel misurabile**) ogni funzione $(\mathcal{A},\mathcal{B})$-misurabile (o $(\mathcal{A},\mathcal{B}^*)$-misurabile). In questo contesto, di notevole valenza applicativa è il risultato seguente che consente di verificare la Borel misurabilità limitandosi a constatare l'appartenenza ad $\mathcal{A}$ delle controimmagini delle semirette inferiori di origine un numero reale. Conseguentemente, possiano sempre supporre che le funzioni Borel misurabili siano a valori nella retta reale ampliata (notato che una funzione a valori in $\mathbb{R}$ è $(\mathcal{A},\mathcal{B})$-misurabile se e solo se, intesa come funzione a valori in $\mathbb{R}^*$, è $(\mathcal{A},\mathcal{B}^*)$-misurabile).

Teorema A.2.2 (Criterio standard di misurabilità) *Se $\mathcal{F}' \subset \mathcal{A}'$ è tale che $\sigma(\mathcal{F}') = \mathcal{A}'$, allora un'applicazione $f : \Omega \mapsto \Omega'$ è $(\mathcal{A},\mathcal{A}')$-misurabile se $f^{-1}(F') \in \mathcal{A}$ per ogni $F' \in \mathcal{F}'$. In particolare, una funzione f di Ω in $\mathbb{R}$ (in $\mathbb{R}^*$) è Borel misurabile se $\{f \le x\} \in \mathcal{A}$ per ogni numero reale x*[9].

DIMOSTRAZIONE Sia $\sigma(\mathcal{F}') = \mathcal{A}'$. Posto $\mathcal{F}'' = \{A' : f^{-1}(A') \in \mathcal{A}\}$, si ha $\Omega' \in \mathcal{F}''$ e $(F'')^c \in \mathcal{F}''$, se $F'' \in \mathcal{F}''$ (poichè $f^{-1}((F'')^c) = [f^{-1}(F'')]^c$); inoltre, se $(F''_n)_{n\ge1}$ è una successione in $\mathcal{F}''$, risulta $\bigcup_{n\ge1} F''_n \in \mathcal{F}''$ (in quanto $f^{-1}(\bigcup_{n\ge1} F''_n) = \bigcup_{n\ge1} f^{-1}(F''_n)$). Pertanto, $\mathcal{F}''$ è una σ-algebra. Allora, notato che $\mathcal{F}' \subset \mathcal{F}''$, si ha $\mathcal{A}' = \sigma(\mathcal{F}') \subseteq \mathcal{F}'' \subseteq \mathcal{A}'$ da cui otteniamo $\mathcal{F}'' = \mathcal{A}'$, cioè la $(\mathcal{A},\mathcal{A}')$-misurabilità di f. La seconda parte della tesi segue ora immediatamente da quanto stabilito nell'Esempio A.1.1(iv). ∎

Il prossimo teorema assicura, in particolare, che trasformate continue o monotone di funzioni Borel misurabili sono ancora Borel misurabili. Precisiamo che, per ogni funzione f a valori nella retta reale ampliata, denotiamo con f^+ e f^-, rispettivamente, la sua parte positiva e la sua parte negativa[10].

Teorema A.2.3 *Sussistono le seguenti proposizioni:*

(i) Se $S \neq \emptyset$ è un boreliano di $\mathbb{R}^*$, sono allora $\mathcal{B}^* \cap S$-Borel misurabili le funzioni di S in $\mathbb{R}^*$ aventi l'insieme dei punti di discontinuità discreto;

[9] Ovviamente, la verifica della Borel misurabilità può anche essere condotta scegliendo uno qualsiasi dei sistemi di generatori considerati nell'Esempio A.1.1(iv) e constatando che le controimmagini dei suoi elementi sono in $\mathcal{A}$.

[10] Riesce quindi $f^+ = \max(0, f)$ e $f^- = \max(0, -f)$; inoltre, $f = f^+ - f^-$ e $|f| = f^+ + f^-$. Infine, date due funzioni f, g a valori in $\mathbb{R}^*$ e di medesimo dominio, si ha

$$(fg)^+ = \begin{cases} f\,g^+ & \text{se } f \ge 0 \\ -f\,g^- & \text{se } f \le 0 \end{cases}, \qquad (fg)^- = \begin{cases} f\,g^- & \text{se } f \ge 0 \\ -f\,g^+ & \text{se } f \le 0 \end{cases}.$$

(ii) Siano $f : \Omega \mapsto \mathbb{R}^*$ una funzione Borel misurabile e S un boreliano di $\mathbb{R}^*$ tale che $f(\Omega) \subseteq S$. Allora, qualunque sia la funzione $g : S \mapsto \mathbb{R}^*$ avente l'insieme dei punti di discontinuità discreto, la funzione composta $g \circ f$ è Borel misurabile. In particolare, sono Borel misurabili $|f|$, f^+, f^- e, considerati due numeri reali α e β qualsiasi, $\alpha f + \beta$ e anche f^α, se definita ovunque.

Dimostrazione (i) Siano $S \neq \emptyset$ un boreliano di $\mathbb{R}^*$ e $g : S \mapsto \mathbb{R}^*$ una funzione avente l'insieme S_0 dei punti di discontinuità discreto. Allora, $S_0 \in \mathcal{B}^*$ e quindi l'insieme $S_1 = S \setminus S_0^c$ dei punti di continuità di g è un boreliano di $\mathbb{R}^*$. Proviamo ora che la restrizione $g_1 = g\big|_{S_1}$ è $\mathcal{B}^* \cap S_1$-Borel misurabile. Per il criterio standard di misurabilità, basta verificare che, dato un numero reale x, risulta $\{g_1 \leq x\}^c = \{g_1 > x\} \in \mathcal{B}^* \cap S_1$. A tal fine, sia $y \in S_1$ tale che $x < g(y) \leq +\infty$. Poichè la semiretta $]x, +\infty]$ è un intorno di $g(y)$ e y è un punto di continuità di g, esiste un intorno aperto W_y di y tale che $y \in W_y \cap S_1 \subseteq \{g_1 > x\}$. Riesce allora

$$\{g_1 > x\} = \bigcup_{y \in \{g_1 > x\}} (W_y \cap S_1) = \Big(\bigcup_{y \in \{g_1 > x\}} W_y \Big) \cap S_1 \in \mathcal{B}^* \cap S_1,$$

notato che $\bigcup_{y \in \{g_1 > x\}} W_y$ è un insieme aperto di $\mathbb{R}^*$.

Per ogni $y \in S_0$, consideriamo infine la restrizione $g_y = g\big|_{\{y\}}$ che è banalmente una funzione $\mathcal{B}^* \cap \{y\}$-Borel misurabile.

Vengono così individuate una partizione discreta S_1, $\{y\}$ $(y \in S_0)$ di S costituita da elementi di $\mathcal{B}^* \cap S$ e una famiglia di funzioni g_1, g_y $(y \in S_0)$ tali che g_1 è $\mathcal{B}^* \cap S_1$-Borel misurabile e g_y è $\mathcal{B}^* \cap \{y\}$-Borel misurabile per ogni $y \in S_0$. Ne segue, per il Lemma A.2.1(ii) (ponendo $\Omega = S$ e $\mathcal{A} = \mathcal{B}^* \cap S$), la $\mathcal{B}^* \cap S$-Borel misurabilità di g.

(ii) Sia $g : S \mapsto \mathbb{R}^*$ con insieme dei punti di discontinuità discreto. Allora, per (i), g è $\mathcal{B}^* \cap S$-Borel misurabile. Conseguentemente, dato $B \in \mathcal{B}^*$, riesce $g^{-1}(B) = W \cap S$ per qualche $W \in \mathcal{B}^*$ e quindi $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B)) = f^{-1}(W \cap S) \in \mathcal{A}$, osservato che $W \cap S \in \mathcal{B}^*$. A questo punto, l'ultima parte della tesi si ottiene immediatamente osservando che le funzioni considerate sono trasformate di f tramite funzioni continue aventi come insieme di definizione dei boreliani di $\mathbb{R}^*$. ■

Una particolare classe di funzioni, che fornirà il punto di partenza per la costruzione dell'integrale di Lebesgue, è quella delle **$\mathcal{A}$-funzioni semplici** (in breve **funzioni semplici**), cioè delle funzioni $\mathcal{A}$-Borel misurabili con insieme-immagine finito. Chiaramente, se f è una funzione semplice, la famiglia di controimmagini $(\{f = y\})_{y \in f(\Omega)}$ è una partizione finita di Ω costituita da elementi di $\mathcal{A}$ tale che $f = \sum_{y \in f(\Omega)} y \, I_{\{f=y\}}$. Viceversa, comunque si considerino una partizione finita $A_1, \ldots, A_n \in \mathcal{A}$ di Ω e $y_i, \ldots, y_n \in \mathbb{R}^*$, la combinazione lineare $f = \sum_{i=1}^n y_i \, I_{A_i}$ è una funzione semplice; infatti,

$f(\Omega) = \{y_1, \ldots, y_n\}$ è finito e, per il criterio standard di misurabilità, f è Borel misurabile in quanto $\{f \leq x\} = \bigcup_{i \in \{i: y_i \leq x\}} A_i \in \mathcal{A}$ per ogni numero reale x. Pertanto, le funzioni semplici sono tutte e sole le funzioni $f : \Omega \mapsto \mathbb{R}^*$ che ammettono una rappresentazione del tipo $f = \sum_{y \in f(\Omega)} y\, I_{\{f=y\}}$ con $\{f = y\} \in \mathcal{A}$ per ogni $y \in f(\Omega)$ e $f(\Omega)$ finito. In particolare quindi la funzione indicatrice di un insieme $S \subseteq \Omega$ è una funzione semplice se e solo se $S \in \mathcal{A}$.

Il lemma successivo rileva che le usuali manipolazioni algebriche mutano funzioni semplici in funzioni ancora semplici (qualora risultino ovunque definite); inoltre, che ogni funzione Borel misurabile è ottenibile come limite di una successione convergente di funzioni semplici a valori finiti.

Lemma A.2.4 (fondamentale) *Sussistono le seguenti proposizioni:*

(i) Somme, quozienti (se definiti ovunque) e prodotti di funzioni semplici sono ancora funzioni semplici;

(ii) Sia f una funzione Borel misurabile non negativa. Esiste allora una successione $(f_n)_{n \geq 1}$ di funzioni semplici non negative a valori finiti tale che $f_n \uparrow f$;

(iii) Sia f una funzione Borel misurabile. Esiste allora una successione $(f_n)_{n \geq 1}$ di funzioni semplici a valori finiti tale che $f_n \to f$ e $|f_n| \leq |f|$ per ogni n.

Dimostrazione (i) Siano f_1, f_2 funzioni semplici. Allora, $f_h = \sum_{i=1}^{n_1} y_i^{(h)} I_{A_i^{(h)}}$, avendo posto $f_h(\Omega) = \{y_1^{(h)}, \ldots, y_{n_h}^{(h)}\}$ e $A_i^{(h)} = \{f_h = y_i^{(h)}\}$ per ogni $i \leq n_h$ $(h = 1, 2)$. Considerata la partizione finita costituita dagli elementi $A_{ij} = A_i^{(1)} \cap A_j^{(2)} \in \mathcal{A}$ $(i \leq n_1;\, j \leq n_2)$, otteniamo (con facili passaggi)

$$
\begin{aligned}
f_1\, f_2 &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} y_i^{(1)}\, y_j^{(2)}\, I_{A_{ij}} \\
f_1 + f_2 &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} (y_i^{(1)} + y_j^{(2)})\, I_{A_{ij}} \quad \text{(se } f_1 + f_2 \text{ è definita ovunque)} \\
\frac{f_1}{f_2} &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{y_i^{(1)}}{y_j^{(2)}}\, I_{A_{ij}} \quad \left(\text{se } \frac{f_1}{f_2} \text{ è definita ovunque}\right).
\end{aligned}
$$

(ii) Diviso l'intervallo reale $[0, n]$ in $n\, 2^n$ parti, consideriamo le controimmagini:

$$
A_i^{(n)} = \left\{ \frac{i-1}{2^n} \leq f < \frac{i}{2^n} \right\} \quad (i = 1, \ldots, n\, 2^n)
$$

e la funzione semplice non negativa a valori finiti:

$$f_n = \sum_{i=1}^{n2^n} \frac{(i-1)}{2^n}\, I_{A_i^{(n)}} + n\, I_{\{f \geq n\}}.$$

Allora, $(f_n)_{n\geq 1}$ è una successione non decrescente. Infatti, fissato n e dato $\omega \in \Omega$, sia intanto $f(\omega) < n$. Esiste allora $i \leq n2^n$ tale che

$$\begin{aligned} f(\omega) \in \Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[&= \Big[\frac{i-1}{2^n}, \frac{2i-1}{2^{n+1}}\Big[\cup \Big[\frac{2i-1}{2^{n+1}}, \frac{i}{2^n}\Big[\\ &= \Big[\frac{(2i-1)-1}{2^{n+1}}, \frac{2i-1}{2^{n+1}}\Big[\cup \Big[\frac{2i-1}{2^{n+1}}, \frac{2i}{2^{n+1}}\Big[= A_{2i-1}^{(n+1)} \cup A_{2i}^{(n+1)} \end{aligned}$$

e quindi

$$f_n(\omega) = \frac{i-1}{2^n} \leq \begin{cases} \frac{i-1}{2^n} & \text{se } f(\omega) \in A_{2i-1}^{(n+1)} \\ \frac{2i-1}{2^{n+1}} & \text{se } f(\omega) \in A_{2i}^{(n+1)} \end{cases} = f_{n+1}(\omega).$$

Sia ora $f(\omega) \geq n$. Allora, $f_n(\omega) = n$ e

$$\begin{aligned} f(\omega) \in [\,n, +\infty\,[\,=\, [\,n, n+1\,[\,\cup\, [\,n+1, +\infty\,[&= \bigcup_{k=1}^{2^{n+1}} \Big[\,n + \frac{k-1}{2^{n+1}}, n + \frac{k}{2^{n+1}}\Big[\cup [\,n+1, +\infty\,[\\ &= \bigcup_{k=1}^{2^{n+1}} A_{n2^{n+1}+k}^{(n+1)} \cup [\,n+1, +\infty\,[. \end{aligned}$$

Quindi, $f_n(\omega) < n+1 = f_{n+1}(\omega)$, se $f(\omega) \geq n+1$, e

$$f_n(\omega) = n \leq \frac{(n2^{n+1}+k)-1}{2^{n+1}} = f_{n+1}(\omega),$$

se $f(\omega) \in A_{n2^{n+1}+k}^{(n+1)}$ $(1 \leq k \leq 2^{n+1})$. Riesce dunque, in ogni caso, $f_n(\omega) \leq f_{n+1}(\omega)$.

Per provare che $f_n \to f$, fissiamo $\omega \in \Omega$ e assumiamo intanto $f(\omega) = +\infty$. Allora, $f_n(\omega) = n$ per ogni n e quindi $f_n(\omega) \to +\infty = f(\omega)$. Supposto infine $f(\omega)$ finito, sia $\epsilon > 0$. Esiste allora m tale che $f(\omega) < m$ e $\frac{1}{m} < \epsilon$. È quindi possibile individuare $i \leq m\,2^m$ tale che $f(\omega) \in A_i^{(m)}$. Riesce pertanto $f(\omega) - f_m(\omega) < \frac{i}{2^m} - \frac{i-1}{2^m} = \frac{1}{2^m} < \epsilon$. Ne segue, essendo $(f_n)_{n\geq 1}$ non decrescente, $f(\omega) - f_n(\omega) < \epsilon$ per ogni $n \geq m$. Per l'arbitrarietà di ϵ, si ha quindi $f_n(\omega) \to f(\omega)$. In ogni caso risulta dunque $f_n(\omega) \to f(\omega)$.

(iii) Per (ii) e il Teorema A.2.3(ii), esistono due successioni $(g'_n)_{n\geq 1}$, $(g''_n)_{n\geq 1}$ di funzioni semplici a valori finiti tali che $0 \leq g'_n \uparrow f^+$ e $0 \leq g''_n \uparrow f^-$. Posto allora $f_n = g'_n - g''_n$ per ogni n, da (i) otteniamo che $(f_n)_{n\geq 1}$ è una successione di funzioni semplici a valori finiti tale che $f_n \to f$ e $|f_n| \leq g'_n + g''_n \leq f^+ + f^- = |f|$ per ogni n. ∎

Concludiamo la sezione fornendo ulteriori proprietà delle funzioni Borel misurabili che, congiuntamente con la proposizione (ii) del Teorema A.2.3, assicurano che le usuali manipolazioni dell'analisi (in particolare il passaggio al limite di successioni) mutano funzioni Borel misurabili in funzioni (se definite ovunque) ancora Borel misurabili.

Teorema A.2.5 *Sussistono le seguenti proposizioni:*

(i) *Siano f, g funzioni Borel misurabili. Sono allora elementi di $\mathcal{A}$ gli insiemi $\{f = g\}$, $\{f < g\}$, $\{f > g\}$, $\{f \leq g\}$ e $\{f \geq g\}$;*

(ii) *Sia $(f_n)_{n\geq 1}$ una successione di funzioni Borel misurabili. Sono allora anche Borel misurabili le funzioni $\inf\limits_{n\geq 1} f_n$, $\sup\limits_{n\geq 1} f_n$, $\liminf\limits_{n\to+\infty} f_n$ e $\limsup\limits_{n\to+\infty} f_n$. Inoltre, gli insiemi di convergenza:*

$$A_{\lim} = \{\omega \in \Omega : \exists \lim_{n\to+\infty} f_n(\omega)\}, \quad A_{\lim}^{(f)} = \{\omega \in A_{\lim} : \lim_{n\to+\infty} f_n(\omega) \in \mathbb{R}\}$$

appartengono ad $\mathcal{A}$ e $\lim\limits_{n\to+\infty} f_n$ è una funzione Borel misurabile, se $A_{\lim} = \Omega$. Infine, se f è una funzione Borel misurabile, allora l'insieme $\{f_n \to f\} = \{\omega \in \Omega : f_n(\omega) \to f(\omega)\}$ è un elemento di $\mathcal{A}$;

(iii) *Siano f, g funzioni Borel misurabili. Sono allora Borel misurabili le funzioni fg e, se definite ovunque, $f + g$ e $\frac{f}{g}$;*

(iv) *Sia $(f_n)_{n\geq 1}$ una successione di funzioni Borel misurabili non negative. Allora, la serie $\sum_{n\geq 1} f_n$ è una funzione Borel misurabile.*

Dimostrazione (i) Sia $q_1, q_2, \ldots$ una numerazione dei numeri razionali. Allora, qualunque sia $\omega \in \Omega$, si ha $f(\omega) < g(\omega)$ se e solo se $f(\omega) < q_n < g(\omega)$ per qualche n. Conseguentemente $\{f < g\} = \bigcup_{n\geq 1}\{f < q_n < g\} = \bigcup_{n\geq 1}[\{f < q_n\} \cap \{q_n < g\}] \in \mathcal{A}$. Evidentemente, in modo analogo si prova $\{f > g\} \in \mathcal{A}$. Allora, $\{f \neq g\} = \{f < g\} \cup \{f > g\} \in \mathcal{A}$ e quindi $\{f = g\} = \{f \neq g\}^c \in \mathcal{A}$. Ne segue $\{f \leq g\} = \{f < g\} \cup \{f = g\} \in \mathcal{A}$ e $\{f \geq g\} = \{f > g\} \cup \{f = g\} \in \mathcal{A}$.

(ii) La Borel misurabilità di $\sup_{n\geq 1} f_n$ discende dal criterio standard di misurabilità e dall'uguaglianza $\left\{\sup_{n\geq 1} f_n \leq x\right\} = \bigcap_{n\geq 1}\{f_n \leq x\}$ valida per ogni numero reale x; quella di $\inf_{n\geq 1} f_n$ dal Teorema A.2.3(ii) e dall'uguaglianza $\inf_{n\geq 1} f_n = -\sup_{n\geq 1}(-f_n)$. Ne segue, tramite le uguaglianze

$$\liminf_{n\to+\infty} f_n = \sup_{k\geq 1}\big(\inf_{n\geq k} f_n\big), \quad \limsup_{n\to+\infty} f_n = \inf_{k\geq 1}\big(\sup_{n\geq k} f_n\big)$$

la Borel misurabilità di $\liminf\limits_{n\to+\infty} f_n$ e $\limsup\limits_{n\to+\infty} f_n$.

L'appartenenza ad $\mathcal{A}$ degli insiemi di convergenza segue allora da (i) e dalle uguaglianze

$$A_{\lim} = \big\{ \liminf_{n\to+\infty} f_n = \limsup_{n\to+\infty} f_n \big\}$$
$$A_{\lim}^{(f)} = A_{\lim} \cap \big\{ \liminf_{n\to+\infty} f_n > -\infty \big\} \cap \big\{ \limsup_{n\to+\infty} f_n < +\infty \big\}.$$

Supposto $A_{\lim} = \Omega$, la Borel misurabilità del limite si ottiene dal criterio standard di misurabilità e dall'uguaglianza $\big\{ \lim_{n\to+\infty} f_n \leq x \big\} = \big\{ \liminf_{n\to+\infty} f_n \leq x \big\}$ valida per ogni numero reale x.

Assunto infine che f sia Borel misurabile, da (i) risulta $\{f = \liminf_{n\to+\infty} f_n\}$, $\{f = \limsup_{n\to+\infty} f_n\} \in \mathcal{A}$ e quindi $\{f_n \to f\} = \{f = \liminf_{n\to+\infty} f_n\} \cap \{ f = \limsup_{n\to+\infty} f_n\} \in \mathcal{A}$.

(iii) Per il lemma fondamentale, esistono due successioni $(f_n)_{n\geq1}$, $(g_n)_{n\geq1}$ di funzioni semplici a valori finiti tali che $f_n \to f$ e $g_n \to g$. Allora, ancora per il lemma fondamentale, $f_n g_n$ e $f_n + g_n$ sono, per ogni n, delle funzioni semplici tali che $f_n g_n \to fg$ e, supposto $f+g$ definita ovunque, $f_n + g_n \to f + g$. Conseguentemente, per (ii), il prodotto e la somma (se definita ovunque) sono delle funzioni Borel misurabili. Supposto infine $\frac{f}{g}$ definito ovunque e osservato che, per (i), $\{g_n = 0\} \in \mathcal{A}$ qualunque sia n, otteniamo, tramite il lemma fondamentale, che $(\frac{f_n}{g_n + I_{\{g_n=0\}}})_{n\geq1}$ è una successione di funzioni semplici avente limite $\frac{f}{g}$. Pertanto, per (ii), pure il quoziente è una funzione Borel misurabile.

(iv) Segue da (ii) e (iii), tenuto conto della definizione di serie di funzioni. ■

Osservazione A.2.6 Con riferimento alle funzioni Borel misurabili, è facile rendersi conto che la somma, il quoziente e il limite di una successione (anche se non definiti ovunque) hanno dominio in $\mathcal{A}$ e sono Borel misurabili rispetto alla traccia di $\mathcal{A}$ sul loro dominio. Proviamolo, ad esempio, per la somma. Indicato con S il suo dominio, otteniamo $S^c = \big(\{f = -\infty\} \cap \{g = +\infty\}\big) \cup \big(\{f = +\infty\} \cap \{g = -\infty\}\big) \in \mathcal{A}$, cioè $S \in \mathcal{A}$. Ne segue, per il Teorema A.2.5(iii), la Borel misurabilità di $fI_S + gI_S$ e quindi, per il Lemma A.2.1(i), $f + g = (fI_S + gI_S)\big|_S$ è $\mathcal{A} \cap S$-Borel misurabile.

Conseguentemente, nel caso che tali funzioni non siano definite ovunque, possiamo sempre estenderle, per il Lemma A.2.1(ii), su tutto Ω in modo da renderle Borel misurabili; basta infatti, come faremo implicitamente nel seguito, porle "d'ufficio" uguali a zero al di fuori del loro dominio[11]. ♦

A.3 Integrale di Lebesgue

La nozione di integrale, per una funzione definita su un intervallo della retta reale, venne introdotta nella seconda metà del '600 (da Isaac Newton e Got-

[11] Naturalmente, adotteremo lo stesso tipo di estensione anche per ogni funzione che abbia dominio in $\mathcal{A}$ e che sia Borel misurabile rispetto alla traccia di $\mathcal{A}$ sul suo dominio.

tfried W. von Leibniz) ricorrendo, in modo più o meno esplicito, agli infinitesimi attuali. A causa della sostanziale vaghezza del concetto d'infinitesimo attuale, questa nozione non era però sufficientemente precisa da evitare risultati che comportavano un campo di validità non chiaramente definito, come pure dimostrazioni prive di rigore. Bisognò attendere il 1867 affinchè Georg F. B. Riemann ne proponesse una definizione (basata sulla nozione di limite e tuttora presente in ogni testo di analisi matematica) che consentisse di eliminare tali inconvenienti. L'evoluzione della nozione di funzione e lo studio approfondito delle funzioni numeriche (di variabili reali o no) iniziato nella seconda metà dell'800 indussero alcuni matematici[12] a sviluppare, tra il 1890 e il 1920, ricerche che condussero ad un nuovo tipo d'integrale (generalizzazione di quello di Riemann) - chiamato **integrale di Lebesgue** - strettamente legato al concetto astratto di misura e riguardante le funzioni Borel misurabili. Tra le varie definizioni proposte abbiamo scelto quella che consente di ottenere rapidamente i risultati di teoria dell'integrazione usati nel testo.

In seguito, m denoterà una misura su $\mathcal{A}$. Inoltre, data una qualsiasi funzione Borel misurabile f, indicheremo con $\int_\Omega f\,d\mathrm{m}$ (o $\int_\Omega f(\omega)\,\mathrm{m}(d\omega)$) il relativo integrale di Lebesgue (qualora esistente). Infine, per non appesantire le dimostrazioni con eccessivi rimandi, non richiameremo di solito i teoremi A.1.3, A.2.3, A.2.5 e il lemma fondamentale.

A.3.1 Costruzione

Per introdurre la nozione d'integrale (di Lebesgue) per le funzioni Borel misurabili procediamo per passi considerando dapprima il caso delle funzioni semplici non negative, poi quello delle funzioni Borel misurabili non negative e infine quello delle funzioni Borel misurabili di segno qualsiasi.

- **Funzioni semplici non negative**

Sia f una funzione semplice non negativa. Allora, $f(\Omega)$ è un insieme finito e $\{f = y\} \in \mathcal{A}$ per ogni $y \in f(\Omega)$. Possiamo dunque porre:

$$\boxed{\int_\Omega f\,d\mathrm{m} = \sum_{y \in f(\Omega)} y\,\mathrm{m}(\{f = y\}) \geq 0.}$$

[12] Tra i quali sono da ricordare Emile Borel, Johann Radon, Thomas J. Stieltjes e, in particolare, Henri Lebesgue per la notevole importanza della sua opera.

In particolare, si ha quindi $\mathrm{m}(A) = \int_\Omega I_A \, d\mathrm{m}$ per ogni $A \in \mathcal{A}$.

Proviamo ora che l'integrale così definito è un funzionale lineare monotono sul cono delle funzioni semplici non negative.

Lemma A.3.1 *Siano f, g funzioni semplici non negative. Sussistono allora le seguenti proposizioni:*

(i) $\int_\Omega \sum_{i=1}^n y_i I_{A_i} \, \mathrm{dm} = \sum_{i=1}^n y_i \, \mathrm{m}(A_i) \quad (0 \leq y_i \leq +\infty;\ i \leq n)$;

(ii) $\int_\Omega (\alpha f + \beta g) \, \mathrm{dm} = \alpha \int_\Omega f \, \mathrm{dm} + \beta \int_\Omega g \, \mathrm{dm} \quad (0 \leq \alpha,\ \beta \leq +\infty)$;

(iii) $\int_\Omega f \, I_A \, \mathrm{dm} = 0$, *se* $\mathrm{m}(A) = 0$;

(iv) Sia $g \leq f$. *Allora,* $\int_\Omega g \, \mathrm{dm} \leq \int_\Omega f \, \mathrm{dm}$. *Inoltre* $\int_\Omega (f - g) \, \mathrm{dm} = \int_\Omega f \, \mathrm{dm} - \int_\Omega g \, \mathrm{dm}$, *se* $\int_\Omega g \, \mathrm{dm} < +\infty$ *e* g *è a valori finiti;*

(v) $\int_\Omega f \, \mathrm{dm} \geq \alpha \, \mathrm{m}(\{f \geq \alpha\}) \geq \alpha \, \mathrm{m}(\{f > \alpha\}) \quad (0 \leq \alpha \leq +\infty)$.

Dimostrazione (i) Siano $f = \sum_{i=1}^n y_i I_{A_i}$ e

$$A_i^{(h_i)} = \begin{cases} A_i & \text{se } h_i = 1 \\ A_i^c & \text{se } h_i = 0 \end{cases} \quad (h_i = 0, 1;\ i = 1, \ldots, n).$$

Posto allora $I = \{0,1\}^n$, otteniamo che gli insiemi $A_{\mathbf{h}} = A_1^{(h_1)} \cap \cdots \cap A_n^{(h_n)}$ ($\mathbf{h} = (h_1, \ldots, h_n) \in I$) individuano una partizione di Ω formata con elementi di $\mathcal{A}$. Posto infine $y_{\mathbf{h}} = \sum_{i=1}^n h_i \, y_i$ per ogni $\mathbf{h} \in I$, si ha

$$f = \sum_{\mathbf{h} \in I} y_{\mathbf{h}} \, I_{A_{\mathbf{h}}}. \tag{A.2}$$

Infatti, dato $\omega \in \Omega$, esiste un solo $\mathbf{h}' = (h'_1, \ldots, h'_n) \in I$ tale che $\omega \in A_{\mathbf{h}'}$. Allora, $I_{A_i}(\omega) = h'_i$ $(i = 1, \ldots, n)$, $I_{A_{\mathbf{h}'}}(\omega) = 1$ e $I_{A_{\mathbf{h}}}(\omega) = 0$, se $\mathbf{h} \neq \mathbf{h}'$. Risulta quindi

$$f(\omega) = \sum_{i=1}^n y_i \, I_{A_i}(\omega) = \sum_{i=1}^n h'_i \, y_i = y_{\mathbf{h}'} = \sum_{\mathbf{h} \in I} y_{\mathbf{h}} \, I_{A_{\mathbf{h}}}(\omega).$$

Inoltre, dall'additività della misura otteniamo

$$\sum_{i=1}^n y_i \, \mathrm{m}(A_i) = \sum_{i=1}^n y_i \, \mathrm{m}\Big(\bigcup_{\mathbf{h} \in I} (A_i \cap A_{\mathbf{h}})\Big) = \sum_{i=1}^n y_i \sum_{\mathbf{h} \in I} \mathrm{m}(A_i \cap A_{\mathbf{h}}) = \sum_{\mathbf{h} \in I} \sum_{i=1}^n y_i \mathrm{m}(A_i \cap A_{\mathbf{h}})$$
$$= \sum_{\mathbf{h} \in I} \sum_{i=1}^n y_i \, h_i \mathrm{m}(A_{\mathbf{h}}) = \sum_{\mathbf{h} \in I} \Big(\sum_{i=1}^n y_i \, h_i\Big) \mathrm{m}(A_{\mathbf{h}}) = \sum_{\mathbf{h} \in I} y_{\mathbf{h}} \, \mathrm{m}(A_{\mathbf{h}}).$$

Sia ora $f(\Omega) = \{y'_1, \ldots, y'_m\}$. Notato che gli insiemi $A'_j = \{f = y'_j\}$ $(j = 1, \ldots, m)$ costituiscono una partizione di Ω, otteniamo $y_{\mathbf{h}}\, \mathrm{m}(A_{\mathbf{h}} \cap A'_j) = y'_j \mathrm{m}(A_{\mathbf{h}} \cap A'_j)$ per ogni $\mathbf{h} \in I$ e $j \leq m$; infatti, supposto $A_{\mathbf{h}} \cap A'_j \neq \emptyset$ e scelto $\omega \in A_{\mathbf{h}} \cap A'_j$, si ha, tramite (A.2), $y_{\mathbf{h}} = f(\omega) = y'_j$. Dall'additività della misura segue allora

$$\begin{aligned}\sum_{i=1}^{n} y_i\, \mathrm{m}(A_i) &= \sum_{\mathbf{h}\in I} y_{\mathbf{h}}\, \mathrm{m}(A_{\mathbf{h}}) = \sum_{\mathbf{h}\in I} y_{\mathbf{h}}\, \mathrm{m}\Big(\bigcup_{j=1}^{m} (A_{\mathbf{h}} \cap A'_j)\Big) = \sum_{\mathbf{h}\in I} \sum_{j=1}^{m} y_{\mathbf{h}}\, \mathrm{m}(A_{\mathbf{h}} \cap A'_j) \\ &= \sum_{\mathbf{h}\in I} \sum_{j=1}^{m} y'_j\, \mathrm{m}(A_{\mathbf{h}} \cap A'_j) = \sum_{j=1}^{m} \sum_{\mathbf{h}\in I} y'_j\, \mathrm{m}(A_{\mathbf{h}} \cap A'_j) \\ &= \sum_{j=1}^{m} y'_j\, \mathrm{m}\Big(\bigcup_{\mathbf{h}\in I} (A_{\mathbf{h}} \cap A'_j)\Big) = \sum_{j=1}^{m} y'_i\, \mathrm{m}(A'_j) = \sum_{y\in f(\Omega)} y\, \mathrm{m}(\{f = y\}) = \int_\Omega f\, d\mathrm{m}.\end{aligned}$$

(ii) Posto $f(\Omega) = \{\alpha_1, \ldots, \alpha_m\}$, $g(\Omega) = \{\beta_1, \ldots, \beta_n\}$ e, per $h = 1, \ldots, m+n$,

$$\gamma_h = \begin{cases} \alpha\alpha_h & \text{se } 1 \leq h \leq m \\ \beta\beta_{h-m} & \text{se } m+1 \leq h \leq m+n \end{cases}, \qquad A_h = \begin{cases} \{f = \alpha_h\} & \text{se } 1 \leq h \leq m \\ \{g = \beta_{h-m}\} & \text{se } m+1 \leq h \leq m+n \end{cases}$$

otteniamo

$$\alpha f + \beta g = \sum_{i=1}^{m} \alpha\alpha_i\, I_{\{f=\alpha_i\}} + \sum_{j=1}^{n} \beta\beta_j\, I_{\{g=\beta_j\}} = \sum_{h=1}^{m+n} \gamma_h\, I_{A_h}$$

e quindi, tramite (i),

$$\begin{aligned}\int_\Omega (\alpha f + \beta g)\, d\mathrm{m} &= \sum_{h=1}^{m+n} \gamma_h\, \mathrm{m}(A_h) = \alpha \sum_{i=1}^{m} \alpha_i\, \mathrm{m}(\{f = \alpha_i\}) + \beta \sum_{j=1}^{n} \beta_j\, \mathrm{m}(\{g = \beta_j\}) \\ &= \alpha \int_\Omega f\, d\mathrm{m} + \beta \int_\Omega g\, d\mathrm{m}.\end{aligned}$$

(iii) Per la monotonia della misura, da (i) otteniamo

$$\begin{aligned}\int_\Omega f\, I_A d\mathrm{m} &= \int_\Omega \Big(\sum_{y\in f(\Omega)} y\, I_{\{f=y\}}\Big) I_A\, d\mathrm{m} = \int_\Omega \sum_{y\in f(\Omega)} y\, I_{\{f=y\}\cap A}\, d\mathrm{m} \\ &= \sum_{y\in f(\Omega)} y\, \mathrm{m}(\{f = y\} \cap A) \leq \sum_{y\in f(\Omega)} y\, \mathrm{m}(A) = 0.\end{aligned}$$

(iv) Supponiamo intanto $\mathrm{m}(\{g = +\infty\}) > 0$. Dalla $\{g = +\infty\} \subseteq \{f = +\infty\}$ otteniamo, per la monotonia della misura, $\mathrm{m}(\{f = +\infty\}) > 0$. Risulta quindi

$$\int_\Omega g\, d\mathrm{m} = \sum_{y\in g(\Omega)} y\, \mathrm{m}(\{g = y\}) \geq +\infty\, \mathrm{m}(\{g = +\infty\}) = +\infty$$

e analogamente $\int_\Omega f\, d\mathrm{m} \geq +\infty$, cioè $\int_\Omega g\, d\mathrm{m} = \int_\Omega f\, d\mathrm{m}$. Sia ora $\mathrm{m}(\{g = +\infty\}) = 0$. Posto $A = \{g < +\infty\}$, riesce $\mathrm{m}(A^c) = \mathrm{m}(\{g = +\infty\}) = 0$. Considerata allora la funzione semplice non negativa a valori finiti gI_A e osservato che $f - gI_A$ è una funzione semplice non negativa, da (ii) si ha

$$\int_\Omega f\, \mathrm{dm} = \int_\Omega [gI_A + (f - gI_A)]\, \mathrm{dm} = \int_\Omega gI_A\, \mathrm{dm} + \int_\Omega (f - gI_A)\, \mathrm{dm}$$

e, da (ii), (iii),

$$\int_\Omega gI_A\, \mathrm{dm} = \int_\Omega gI_A\, \mathrm{dm} + \int_\Omega gI_{A^c}\, \mathrm{dm} = \int_\Omega g(I_A + I_{A^c})\, \mathrm{dm} = \int_\Omega g\, \mathrm{dm}.$$

Riesce dunque $\int_\Omega f\, \mathrm{dm} = \int_\Omega g\, \mathrm{dm} + \int_\Omega (f - gI_A)\, \mathrm{dm}$. Ne segue $\int_\Omega f\, \mathrm{dm} \geq \int_\Omega g\, \mathrm{dm}$ e, supposto g a valori finiti, $\int_\Omega (f - g)\, \mathrm{dm} = \int_\Omega f\, \mathrm{dm} - \int_\Omega g\, \mathrm{dm}$ (notato che $g = gI_A$).

(v) Conseguenza immediata di (iv), osservato che $f \geq \alpha I_{\{f \geq \alpha\}} \geq \alpha I_{\{f > \alpha\}}$. ■

• Funzioni Borel misurabili non negative

Sia f una funzione Borel misurabile non negativa. Poichè la funzione nulla appartiene all'insieme $\{g : 0 \leq g \leq f \text{ e } g \text{ funzione semplice}\}$, possiamo porre:

$$\boxed{\int_\Omega f\, \mathrm{dm} = \sup\Big\{\int_\Omega g\, \mathrm{dm} : 0 \leq g \leq f \text{ e } g \text{ funzione semplice}\Big\} \geq 0}$$

ottenendo così una nozione d'integrale per questo tipo di funzioni che è coerente, per il Lemma A.3.1(iv), con quella data per le funzioni semplici non negative.

Il prossimo risultato (di notevole valenza dimostrativa), consente il passaggio del limite sotto il segno d'integrale nel caso di successioni non decrescenti di funzioni semplici non negative. Conseguentemente, per il lemma fondamentale, l'integrale di una funzione Borel misurabile non negativa è sempre approssimabile con l'integrale di una opportuna funzione semplice non negativa a valori finiti.

Lemma A.3.2 *Sia $(f_n)_{n\geq 1}$ una successione di funzioni semplici non negative tale che $f_n \uparrow f$. Allora, $\int_\Omega f_n\, \mathrm{dm} \uparrow \int_\Omega f\, \mathrm{dm}$.*

Dimostrazione Osservato che, per il Lemma A.3.1(iv), $(\int_\Omega f_n\, \mathrm{dm})_{n\geq 1}$ è una successione non decrescente tale che $\lim_{n\to+\infty} \int_\Omega f_n\, \mathrm{dm} \leq \int_\Omega f\, \mathrm{dm}$, basta verificare che, data una funzione semplice $g \neq 0$ tale che $0 \leq g \leq f$, risulta

$$\int_\Omega g\, \mathrm{dm} \leq \lim_{n\to+\infty} \int_\Omega f_n\, \mathrm{dm}. \tag{A.3}$$

A tal fine, sia α un numero reale positivo. Allora, $\{g > \alpha\} \subseteq \bigcup_{n\geq 1}\{f_n > \alpha\}$; infatti, dato $g(\omega) > \alpha$, si ha $f(\omega) > \alpha$ e quindi esiste m tale che $f_m(\omega) > \alpha$, cioè $\omega \in \{g_m > \alpha\} \subseteq \bigcup_{n\geq 1}\{f_n > \alpha\}$. Ne segue, dalla continuità dal basso e dalla monotonia della misura (la successione $(\{f_n > \alpha\})_{n\geq 1}$ è non decrescente!),

$$\lim_{n\to+\infty} \mathrm{m}(\{f_n > \alpha\}) = \mathrm{m}\big(\bigcup_{n\geq 1}\{f_n > \alpha\}\big) \geq \mathrm{m}(\{g > \alpha\}).$$

Poichè, per il Lemma A.3.1(v), $\int_\Omega f_n\, d\mathrm{m} \geq \alpha\, \mathrm{m}(\{f_n > \alpha\})$ per ogni n, risulta infine

$$\lim_{n\to+\infty} \int_\Omega f_n\, d\mathrm{m} \geq \alpha \lim_{n\to+\infty} \mathrm{m}(\{f_n > \alpha\}) \geq \alpha\, \mathrm{m}(\{g > \alpha\}). \tag{A.4}$$

Procediamo ora considerando i casi: 1) $\mathrm{m}(\{g = +\infty\}) > 0$; 2) $\mathrm{m}(\{g > 0\}) = +\infty$; 3) $\mathrm{m}(\{g = +\infty\}) = 0$ e $\mathrm{m}(\{g > 0\}) < +\infty$.

• Caso 1. Notato che $\{g = +\infty\} \subseteq \{g > \alpha\}$, si ha, per (A.4) e la monotonia della misura, $\lim\limits_{n\to+\infty} \int_\Omega f_n\, d\mathrm{m} \geq \alpha\, \mathrm{m}(\{g = +\infty\})$. Passando al limite per $\alpha \to +\infty$, otteniamo allora $\lim\limits_{n\to+\infty} \int_\Omega f_n\, d\mathrm{m} = +\infty$ e quindi (A.3).

• Caso 2. Poichè g è una funzione semplice non ovunque nulla, esiste un numero reale $\alpha > 0$ tale che $\{g > \alpha\} = \{g > 0\}$. Ne segue, per (A.4), $\lim\limits_{n\to+\infty} \int_\Omega f_n\, d\mathrm{m} \geq \alpha\, \mathrm{m}(\{g > 0\}) = +\infty$ e quindi (A.3).

• Caso 3. Posto $A = \{0 < g < +\infty\}$, dalla monotonia della misura si ha $\mathrm{m}(A) \leq \mathrm{m}(\{g > 0\}) < +\infty$. Inoltre, osservato che $g = gI_{\{g>0\}} = g(I_A + I_{\{g=+\infty\}}) = gI_A + gI_{\{g=+\infty\}}$, dal Lemma A.3.1(ii),(iii) otteniamo

$$\int_\Omega g\, d\mathrm{m} = \int_\Omega g\, I_A\, d\mathrm{m} + \int_\Omega g\, I_{\{g=+\infty\}}\, d\mathrm{m} = \int_\Omega g\, I_A\, d\mathrm{m}. \tag{A.5}$$

Infine, per il Lemma A.3.1(iv), $\int_\Omega f_n\, d\mathrm{m} \geq \int_\Omega f_n I_A\, d\mathrm{m}$ per ogni n (poichè $f_n \geq f_n\, I_A$).

Sia ora $0 < \epsilon < \min g(A)$. Allora, posto $A_n = \{f_n \geq g - \epsilon\} \cap A$ ($n \geq 1$), le funzioni gI_{A_n}, $(g-\epsilon)I_{A_n}$ e $f_n I_{A_n}$ sono funzioni semplici non negative tali che $\epsilon\, I_{A_n} \leq g\, I_{A_n}$ e $(g-\epsilon)I_{A_n} \leq f_n I_{A_n} \leq f_n$; inoltre, per la monotonia della misura, $\int_\Omega \epsilon\, I_{A_n}\, d\mathrm{m} = \epsilon\, \mathrm{m}(A_n) \leq \epsilon\, \mathrm{m}(A) < +\infty$. Dal Lemma A.3.1(iv) segue allora

$$\begin{aligned}
\int_\Omega f_n\, d\mathrm{m} &\geq \int_\Omega (g-\epsilon)I_{A_n}\, d\mathrm{m} = \int_\Omega (gI_{A_n} - \epsilon I_{A_n})\, d\mathrm{m} \\
&= \int_\Omega gI_{A_n}\, d\mathrm{m} - \int_\Omega \epsilon\, I_{A_n}\, d\mathrm{m} = \int_\Omega gI_{A_n}\, d\mathrm{m} - \epsilon\, \mathrm{m}(A_n) \\
&\geq \int_\Omega gI_{A_n}\, d\mathrm{m} - \epsilon\, \mathrm{m}(A) = \int_\Omega g(I_A - I_{A\setminus A_n})\, d\mathrm{m} - \epsilon\, \mathrm{m}(A) \\
&= \int_\Omega (gI_A - gI_{A\setminus A_n})\, d\mathrm{m} - \epsilon\, \mathrm{m}(A).
\end{aligned}$$

Posto $\beta = \max g(A)$, risulta $gI_{A\setminus A_n} \leq \beta I_{A\setminus A_n} \leq \beta I_A$ da cui, per il Lemma A.3.1(iv), si ha $\int_\Omega gI_{A\setminus A_n}\, d\mathrm{m} \leq \int_\Omega \beta I_{A\setminus A_n}\, d\mathrm{m} = \beta\, \mathrm{m}(A \setminus A_n) \leq \int_\Omega \beta I_A\, d\mathrm{m} = \beta\, \mathrm{m}(A) < +\infty$. Notato infine che $gI_{A\setminus A_n}$ è a valori finiti, sempre tramite il Lemma A.3.1(iv), otteniamo

$$\int_\Omega f_n \,\mathrm{dm} \geq \int_\Omega gI_A \,\mathrm{dm} - \int_\Omega gI_{A\setminus A_n} \,\mathrm{dm} - \epsilon\,\mathrm{m}(A) \geq \int_\Omega gI_A \,\mathrm{dm} - \beta\,\mathrm{m}(A \setminus A_n) - \epsilon\,\mathrm{m}(A)$$

e quindi, per (A.5),

$$\int_\Omega f_n \,\mathrm{dm} \geq \int_\Omega g \,\mathrm{dm} - \beta\,\mathrm{m}(A \setminus A_n) - \epsilon\,\mathrm{m}(A).$$

Passando al limite per $n \to +\infty$, dalla continuità dall'alto della misura risulta allora

$$\lim_{n\to+\infty} \int_\Omega f_n \,\mathrm{dm} \geq \int_\Omega g \,\mathrm{dm} - \beta\,\mathrm{m}\Big(\bigcap_{n\geq 1}(A \setminus A_n)\Big) - \epsilon\,\mathrm{m}(A),$$

notato che $(A \setminus A_n)_{n\geq 1}$ è una successione non crescente e $\mathrm{m}(A \setminus A_1) \leq \mathrm{m}(A) < +\infty$. Ora, $A = \bigcup_{n\geq 1}(A \setminus A_n)$; infatti, dato $\omega \in A$, si ha $g(\omega) - \epsilon < g(\omega) \leq f(\omega)$ e quindi esiste m tale che $g(\omega) - \epsilon < f_m(\omega)$, cioè $\omega \in A_m \subseteq \bigcup_{n\geq 1} A_n$. Ne segue $\bigcap_{n\geq 1}(A \setminus A_n) = A \cap \big(\bigcup_{n\geq 1} A_n\big)^c = \emptyset$ e quindi $\lim_{n\to+\infty} \int_\Omega f_n \,\mathrm{dm} \geq \int_\Omega g \,\mathrm{dm} - \epsilon\,\mathrm{m}(A)$. Passando infine al limite per $\epsilon \to 0^+$, otteniamo (A.3). ■

Proviamo ora che l'integrale è un funzionale lineare monotono sul cono delle funzioni Borel misurabili non negative.

Lemma A.3.3 *Siano f, g funzioni Borel misurabili non negative. Risulta allora:*

(i) $\int_\Omega (\alpha f + \beta g) \,\mathrm{dm} = \alpha \int_\Omega f \,\mathrm{dm} + \beta \int_\Omega g \,\mathrm{dm} \quad (0 \leq \alpha,\, \beta \leq +\infty)$*;*

(ii) $\int_\Omega g \,\mathrm{dm} \leq \int_\Omega f \,\mathrm{dm}$*, se* $g \leq f$*.*

Dimostrazione (i) Consideriamo due successioni $(f_n)_{n\geq 1}$, $(g_n)_{n\geq 1}$ di funzioni semplici non negative tali che $f_n \uparrow f$ e $g_n \uparrow g$. Allora, la successione $(\alpha f_n + \beta g_n)_{n\geq 1}$ è costituita da funzioni semplici non negative tali che $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$. Conseguentemente, per i lemmi A.3.1(ii) e A.3.2,

$$\begin{aligned}
\alpha \int_\Omega f \,\mathrm{dm} + \beta \int_\Omega g \,\mathrm{dm} &= \alpha \lim_{n\to+\infty} \int_\Omega f_n \,\mathrm{dm} + \beta \lim_{n\to+\infty} \int_\Omega g_n \,\mathrm{dm} \\
&= \lim_{n\to+\infty} \Big[\alpha \int_\Omega f_n \,\mathrm{dm} + \beta \int_\Omega g_n \,\mathrm{dm}\Big] \\
&= \lim_{n\to+\infty} \int_\Omega (\alpha f_n + \beta g_n) \,\mathrm{dm} = \int_\Omega (\alpha f + \beta g) \,\mathrm{dm}.
\end{aligned}$$

(ii) Segue banalmente dalla definizione d'integrale. ■

- **Funzioni Borel misurabili qualsiasi**

Sia f una funzione Borel misurabile qualsiasi. Osservato che, per il Teorema A.2.3(ii), le parti positiva f^+ e negativa f^- sono funzioni Borel misurabili non negative, possiamo considerare i rispettivi integrali $\int_\Omega f^+\,d\mathrm{m}$ e $\int_\Omega f^-\,d\mathrm{m}$. Qualora tali integrali non siano entrambi infiniti, chiamiamo m**-sommabile** (in breve **sommabile**) la funzione f e poniamo:

$$\boxed{\int_\Omega f\,d\mathrm{m} = \int_\Omega f^+\,d\mathrm{m} - \int_\Omega f^-\,d\mathrm{m}}$$

ottenendo così una nozione d'integrale che risulta banalmente coerente con quella data per le funzioni Borel misurabili non negative.

Il valore dell'integrale di una funzione sommabile f è quindi:
- $+\infty$, se $\int_\Omega f^+\,d\mathrm{m} = +\infty$;
- $-\infty$, se $\int_\Omega f^-\,d\mathrm{m} = +\infty$;
- finito, se $\int_\Omega f^+\,d\mathrm{m}$ e $\int_\Omega f^-\,d\mathrm{m}$ sono entrambi finiti.

In quest'ultimo caso diremo che f è m**-integrabile** (in breve **integrabile**).

Sia ora $A \in \mathcal{A}$. Allora, per il Teorema A.2.5(iii), fI_A è una funzione Borel misurabile; inoltre, tenuto conto delle disuguaglianze $(fI_A)^+ = f^+I_A \leq f^+$ e $(fI_A)^- = f^-I_A \leq f^-$, dal Lemma A.3.3(ii) otteniamo che fI_A risulta sommabile (integrabile) se f è sommabile (integrabile). Conseguentemente, nel caso di sommabilità della funzione f, possiamo porre:

$$\boxed{\int_A f\,d\mathrm{m} = \int_\Omega fI_A\,d\mathrm{m}}$$

e chiamare $\int_A f\,d\mathrm{m}$ l'**integrale (di Lebesgue) di f su A**.

Una giustificazione della notazione $\int_A f\,d\mathrm{m}$ può ottenersi constatando (come ora faremo) che l'integrale $\int_\Omega fI_A\,d\mathrm{m}$ coincide (nel caso $A \neq \emptyset$) con l'integrale che si ottiene scegliendo come insieme ambiente A, come σ-algebra di riferimento la traccia $\mathcal{A}_1 = \mathcal{A} \cap A \subseteq \mathcal{A}$ di $\mathcal{A}$ su A, come misura la restrizione $\mathrm{m}_1 = \mathrm{m}\big|_{\mathcal{A}_1}$ e infine come funzione la restrizione $f_1 = f\big|_A$. Osservato che, per il Lemma A.2.1(i), f_1 è una funzione $\mathcal{A}_1$-Borel misurabile, per provare l'uguaglianza $\int_A f_1\,d\mathrm{m}_1 = \int_\Omega f\,I_A d\mathrm{m}$ assumiamo intanto che f sia una funzione semplice non negativa. Allora,

$$fI_A = \Big(\sum_{y\in f(\Omega)} y\,I_{\{f=y\}}\Big)\,I_A = \sum_{y\in f(\Omega)} y\,I_{\{f=y\}\cap A} = \sum_{y\in f(A)} y\,I_{\{f=y\}\cap A}$$

da cui, tramite il Lemma A.3.1(i), otteniamo

$$\int_A f_1\, d\mathrm{m}_1 = \sum_{y\in f_1(A)} y\,\mathrm{m}_1(\{f_1 = y\}) = \sum_{y\in f(A)} y\,\mathrm{m}(\{f = y\}\cap A) = \int_\Omega f I_A\, d\mathrm{m}.$$

Supponiamo ora $f \geq 0$. Allora, per il lemma fondamentale, esiste una successione $(g_n)_{n\geq 1}$ di funzioni semplici non negative tale che $g_n \uparrow f$. Conseguentemente, $(g_n I_A)_{n\geq 1}$ è una successione di funzioni semplici non negative tale che $g_n I_A \uparrow f I_A$ e, per il Lemma A.2.1(i), $(g_n\big|_A)_{n\geq 1}$ è una successione di funzioni $\mathcal{A}_1$-semplici non negative tale che $g_n\big|_A \uparrow f_1$. Per il Lemma A.3.2, si ha quindi

$$\int_A f_1\, d\mathrm{m}_1 = \lim_{n\to+\infty}\int_A g_n\big|_A\, d\mathrm{m}_1 = \lim_{n\to+\infty}\int_\Omega g_n\, I_A\, d\mathrm{m} = \int_\Omega f\, I_A\, d\mathrm{m}.$$

Sia infine f una funzione m-sommabile. Dalle $f_1^+ = f^+\big|_A$ e $f_1^- = f^-\big|_A$ risulta

$$\int_A f_1^+\, d\mathrm{m}_1 = \int_\Omega f^+ I_A\, d\mathrm{m}, \qquad \int_A f_1^-\, d\mathrm{m}_1 = \int_\Omega f^- I_A\, d\mathrm{m}$$

e quindi f_1 è m_1-sommabile. Ne segue, ricordato che $(fI_A)^+ = f^+I_A$ e $(fI_A)^- = f^-I_A$,

$$\int_A f_1\, d\mathrm{m}_1 = \int_A f_1^+\, d\mathrm{m}_1 - \int_A f_1^-\, d\mathrm{m}_1 = \int_\Omega (fI_A)^+\, d\mathrm{m} - \int_\Omega (fI_A)^-\, d\mathrm{m} = \int_\Omega fI_A\, d\mathrm{m}.$$

Qualora m sia la misura di Lebesgue unidimensionale, al posto delle notazioni $\int_{\mathbb{R}} f\, d\lambda$, $\int_{]-\infty,a]} f\, d\lambda$, $\int_{[a,+\infty[} f\, d\lambda$ e $\int_{[a,b]} f\, d\lambda$, useremo, rispettivamente, le $\int_{-\infty}^{+\infty} f(x)\, dx$, $\int_{-\infty}^{a} f(x)\, dx$, $\int_a^{+\infty} f(x)\, dx$ e $\int_a^b f(x)\, dx$. Ricordiamo inoltre che quest'ultimo integrale coincide con quello di Riemann se f è una funzione limitata integrabile secondo Riemann nell'intervallo $[a, b]$[13].

A.3.2 Proprietà elementari

Per provare alcune proprietà dell'integrale di Lebesgue seguiremo la seguente procedura standard:

• Constatare che la proprietà sussiste nell'ambito delle funzioni semplici non negative;

• Verificare che rimane valida anche nell'ambito delle funzioni Borel misurabili non negative tramite il passo precedente, il lemma fondamentale e la proprietà di convergenza descritta nel Lemma A.3.2 (che non richiameremo);

• Provare che rimane valida anche nell'ambito delle funzioni sommabili ricorrendo al passo precedente e alle parti positiva e negativa.

[13] Per una dimostrazione si veda la sezione 1.7 del testo di Ash.

La prima proposizione del teorema seguente mostra che le funzioni integrabili a valori finiti formano uno spazio vettoriale reale e che l'integrale è un funzionale lineare su tale spazio; le ultime due assicurano invece che, data una funzione sommabile f, l'applicazione $A \to \int_A f\, d\mathrm{m}$ è una funzione d'insieme finitamente additiva su $\mathcal{A}$ che si annulla su ogni insieme di misura nulla.

Teorema A.3.4 *Sussistono le seguenti proposizioni:*

(i) LINEARITÀ: *Siano f, g funzioni sommabili e α, β numeri reali tali che $\alpha f + \beta g$ risulti definita ovunque. Allora,*

$$\int_\Omega (\alpha f + \beta g)\, d\mathrm{m} = \alpha \int_\Omega f\, d\mathrm{m} + \beta \int_\Omega g\, d\mathrm{m}$$

ogniqualvolta non siano infiniti di segno opposto gli addendi che compaiono al secondo membro dell'uguaglianza;

(ii) Siano m' *una misura su* $\mathcal{A}$ *e* f *una funzione* m*-sommabile e* m'*-sommabile. Allora, qualunque siano i numeri reali non negativi α e α', si ha*

$$\int_\Omega f\, d(\alpha\mathrm{m} + \alpha'\mathrm{m}') = \alpha \int_\Omega f\, d\mathrm{m} + \alpha' \int_\Omega f\, d\mathrm{m}'$$

ogniqualvolta non siano infiniti di segno opposto gli integrali che compaiono al secondo membro dell'uguaglianza;

(iii) ADDITIVITÀ: *Sia f una funzione sommabile. Allora,*

$$\int_{A_1 \cup A_2} f\, d\mathrm{m} = \int_{A_1} f\, d\mathrm{m} + \int_{A_2} f\, d\mathrm{m}$$

qualunque siano gli insiemi disgiunti A_1 e A_2;

(iv) Sia f una funzione sommabile e $\mathrm{m}(A) = 0$. Allora, $\int_A f\, d\mathrm{m} = 0$ e $\int_\Omega f\, d\mathrm{m} = \int_{A^c} f\, d\mathrm{m}$.

DIMOSTRAZIONE (i) Iniziamo col provare l'uguaglianza $\int_\Omega \alpha f\, d\mathrm{m} = \alpha \int_\Omega f\, d\mathrm{m}$. A tal fine, supponiamo intanto $\alpha \geq 0$. Allora, per il Lemma A.3.3(i),

$$\int_\Omega (\alpha\, f)^+\, d\mathrm{m} = \int_\Omega \alpha f^+\, d\mathrm{m} = \alpha \int_\Omega f^+\, d\mathrm{m}$$
$$\int_\Omega (\alpha\, f)^-\, d\mathrm{m} = \int_\Omega \alpha f^-\, d\mathrm{m} = \alpha \int_\Omega f^-\, d\mathrm{m}.$$

Conseguentemente, αf è una funzione sommabile e riesce

$$\int_\Omega \alpha f\, dm = \alpha \int_\Omega f^+\, dm - \alpha \int_\Omega f^-\, dm = \alpha\Big[\int_\Omega f^+\, dm - \int_\Omega f^-\, dm\Big] = \alpha \int_\Omega f\, dm.$$

Assumiamo infine $\alpha < 0$. Allora, sempre per il Lemma A.3.3(i),

$$\int_\Omega (\alpha f)^+\, dm = \int_\Omega (-\alpha) f^-\, dm = (-\alpha)\int_\Omega f^-\, dm$$
$$\int_\Omega (\alpha f)^-\, dm = \int_\Omega (-\alpha) f^+\, dm = (-\alpha)\int_\Omega f^+\, dm.$$

Quindi, αf è una funzione sommabile e risulta

$$\int_\Omega \alpha f\, dm = (-\alpha)\int_\Omega f^- dm + \alpha \int_\Omega f^+ dm = \alpha\left[\int_\Omega f^+ dm - \int_\Omega f^- dm\right] = \alpha \int_\Omega f\, dm.$$

A questo punto, supposto che gli integrali $\int_\Omega f\, dm$, $\int_\Omega g\, dm$ non siano infiniti di segno opposto, basta verificare l'uguaglianza $\int_\Omega (f+g)\, dm = \int_\Omega f\, dm + \int_\Omega g\, dm$. Osservato che $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$, $(f+g)^+ \le f^+ + g^+$ e $(f+g)^- \le f^- + g^-$, dal Lemma A.3.3 otteniamo

$$\int_\Omega (f+g)^+ dm + \int_\Omega f^- dm + \int_\Omega g^- dm = \int_\Omega (f+g)^- dm + \int_\Omega f^+ dm + \int_\Omega g^+ dm \quad \text{(A.6)}$$
$$\int_\Omega (f+g)^+ dm \le \int_\Omega f^+ dm + \int_\Omega g^+ dm \quad \text{(A.7)}$$
$$\int_\Omega (f+g)^- dm \le \int_\Omega f^- dm + \int_\Omega g^- dm. \quad \text{(A.8)}$$

Se f e g sono integrabili, gli integrali delle parti positiva e negativa della funzione somma sono, per (A.7) e (A.8), finiti e quindi la funzione somma risulta integrabile. Inoltre, per (A.6),

$$\begin{aligned}\int_\Omega (f+g)\, dm &= \int_\Omega (f+g)^+ dm - \int_\Omega (f+g)^- dm \\ &= \Big(\int_\Omega f^+ dm - \int_\Omega f^- dm\Big) + \Big(\int_\Omega g^+ dm - \int_\Omega g^- dm\Big) = \int_\Omega f\, dm + \int_\Omega g\, dm.\end{aligned}$$

Supposto ora, ad esempio, che f non sia integrabile, assumiamo intanto che risulti infinito l'integrale della sua parte positiva. Allora, gli integrali delle parti negative di f e di g devono essere entrambi finiti (in caso contrario, si giungerebbe alla contraddizione $\int_\Omega f\, dm = +\infty$ e $\int_\Omega g\, dm = -\infty$). Conseguentemente, per (A.8), l'integrale $\int_\Omega (f+g)^- dm$ è finito e quindi la funzione somma risulta sommabile. Pertanto, tramite (A.6),

$$\begin{aligned}\int_\Omega (f+g)^+\, dm &= \int_\Omega (f+g)^- dm + \Big(\int_\Omega f^+ dm - \int_\Omega f^- dm\Big) + \Big(\int_\Omega g^+ dm - \int_\Omega g^- dm\Big) \\ &= \int_\Omega (f+g)^- dm + \int_\Omega f\, dm + \int_\Omega g\, dm\end{aligned}$$

da cui otteniamo

$$\begin{aligned}\int_\Omega (f+g)\,dm &= \Big(\int_\Omega (f+g)^- dm + \int_\Omega f dm + \int_\Omega g\,dm\Big) - \int_\Omega (f+g)^- dm\\ &= \int_\Omega f\,dm + \int_\Omega g\,dm.\end{aligned}$$

Se invece è infinito l'integrale della parte negativa di f, si procede in modo analogo (usando (A.7) al posto di (A.8)).

(ii) Considerati due numeri reali α, $\alpha' \geq 0$, poniamo $m'' = \alpha m + \alpha' m'$. Supposto che gli integrali del secondo membro dell'uguaglianza non siano infiniti di segno opposto, assumiamo intanto che f sia una funzione semplice non negativa. Riesce allora

$$\begin{aligned}\int_\Omega f\,dm'' &= \sum_{y\in f(\Omega)} y\,m''(\{f=y\}) = \sum_{y\in f(\Omega)} y\,[\alpha\,m(\{f=y\}) + \alpha' m'(\{f=y\})]\\ &= \alpha \sum_{y\in f(\Omega)} y m(\{f=y\}) + \alpha' \sum_{y\in f(\Omega)} y m'(\{f=y\}) = \alpha \int_\Omega f\,dm + \alpha' \int_\Omega f\,dm'.\end{aligned}$$

Supponiamo ora $f \geq 0$. Considerata una successione $(f_n)_{n\geq 1}$ di funzioni semplici non negative tale che $f_n \uparrow f$, otteniamo

$$\begin{aligned}\int_\Omega f\,dm'' &= \lim_{n\to+\infty} \int_\Omega f_n\,dm'' = \lim_{n\to+\infty} \Big[\alpha \int_\Omega f_n\,dm + \alpha' \int_\Omega f_n\,dm'\Big]\\ &= \alpha \lim_{n\to+\infty} \int_\Omega f_n\,dm + \alpha' \lim_{n\to+\infty} \int_\Omega f_n\,dm' = \alpha \int_\Omega f\,dm + \alpha' \int_\Omega f\,dm'.\end{aligned}$$

Sia infine f qualsiasi. Poichè

$$\int_\Omega f^+ dm'' = \alpha \int_\Omega f^+ dm + \alpha' \int_\Omega f^+ dm' \tag{A.9}$$

$$\int_\Omega f^- dm'' = \alpha \int_\Omega f^- dm + \alpha' \int_\Omega f^- dm', \tag{A.10}$$

f è m''-sommabile. Infatti, se gli integrali $\int_\Omega f^+ dm$, $\int_\Omega f^+ dm'$ sono entrambi finiti, per (A.9) è finito pure l'integrale $\int_\Omega f^+ dm''$. Se invece riesce, ad esempio, $\int_\Omega f^+ dm = +\infty$, sono entrambi finiti gli integrali $\int_\Omega f^- dm$ e $\int_\Omega f^- dm'$ (in caso contrario, si giungerebbe alla contraddizione $\int_\Omega f\,dm = +\infty$ e $\int_\Omega f\,dm' = -\infty$); allora, per (A.10) è pure finito l'integrale $\int_\Omega f^- dm''$. In ogni caso quindi f è m''-sommabile. Inoltre, sempre per (A.9) e (A.10),

$$\begin{aligned}\int_\Omega f dm'' &= \Big(\alpha \int_\Omega f^+ dm + \alpha' \int_\Omega f^+ dm'\Big) - \Big(\alpha \int_\Omega f^- dm + \alpha' \int_\Omega f^- dm'\Big)\\ &= \alpha\Big(\int_\Omega f^+ dm - \int_\Omega f^- dm\Big) - \alpha'\Big(\int_\Omega f^+ dm' - \int_\Omega f^- dm'\Big)\\ &= \alpha \int_\Omega f\,dm + \alpha' \int_\Omega f\,dm'.\end{aligned}$$

(iii) Sia $A_1 \cap A_2 = \emptyset$. Allora, $I_{A_1 \cup A_2} = I_{A_1} + I_{A_2}$ e gli integrali $\int_\Omega f I_{A_1}\, d\mathrm{m}$, $\int_\Omega f I_{A_2}\, d\mathrm{m}$ non possono essere infiniti di segno opposto (infatti, se risultasse, ad esempio, $\int_\Omega f I_{A_1}\, d\mathrm{m} = +\infty$, $\int_\Omega f I_{A_2}\, d\mathrm{m} = -\infty$, si avrebbe, per il Lemma A.3.3(ii), $+\infty = \int_\Omega (f I_{A_1})^+\, d\mathrm{m} \leq \int_\Omega f^+\, d\mathrm{m}$, $+\infty = \int_\Omega (f I_{A_2})^-\, d\mathrm{m} \leq \int_\Omega f^-\, d\mathrm{m}$ e quindi si otterrebbe la contraddizione $\int_\Omega f^+\, d\mathrm{m} = \int_\Omega f^-\, d\mathrm{m} = +\infty$). Da (i) segue allora

$$\int_{A_1 \cup A_2} f\, d\mathrm{m} = \int_\Omega (f I_{A_1} + f I_{A_2})\, d\mathrm{m} = \int_\Omega f I_{A_1}\, d\mathrm{m} + \int_\Omega f I_{A_2}\, d\mathrm{m} = \int_{A_1} f\, d\mathrm{m} + \int_{A_2} f\, d\mathrm{m}.$$

(iv) Per provare l'annullamento dell'integrale, basta evidentemente verificarlo nel caso $f \geq 0$. Considerata una successione $(f_n)_{n \geq 1}$ di funzioni semplici non negative tale che $f_n \uparrow f$, la successione $(f_n I_A)_{n \geq 1}$ è costituita da funzioni semplici non negative tali che $f_n I_A \uparrow f I_A$. Dal Lemma A.3.1(iii) si ha allora $0 = \int_\Omega f_n I_A\, d\mathrm{m} \uparrow \int_\Omega f I_A\, d\mathrm{m}$ e quindi $\int_A f\, d\mathrm{m} = 0$. Ne segue, tramite (iii),

$$\int_\Omega f\, d\mathrm{m} = \int_{A \cup A^c} f\, d\mathrm{m} = \int_A f\, d\mathrm{m} + \int_{A^c} f\, d\mathrm{m} = \int_{A^c} f\, d\mathrm{m}.$$

La dimostrazione è così conclusa. ∎

Il prossimo risultato identifica l'integrabilità di una funzione con quella del suo valore assoluto; inoltre, mette in evidenza che l'integrabilità di una funzione è assicurata qualora il suo valore assoluto sia dominato da qualche funzione integrabile oppure, nel caso che la misura sia **finita**[14], dall'integrabilità del suo quadrato.

Teorema A.3.5 *Siano f, g funzioni Borel misurabili. Sussistono allora le seguenti proposizioni:*

(i) f è sommabile se:
- $f \geq g$ con g funzione sommabile tale che $\int_\Omega g\, d\mathrm{m} > -\infty$;
- $f \leq g$ con g funzione sommabile tale che $\int_\Omega g\, d\mathrm{m} < +\infty$;

(ii) f è integrabile se e solo se lo è il suo valore assoluto;

(iii) f è integrabile se $|f| \leq g$ con g funzione integrabile;

(iv) fg è integrabile se f^2, g^2 sono integrabili;

(v) Sia la misura m *finita. Allora, f è integrabile se lo è il suo quadrato. Inoltre, ogni funzione limitata è integrabile.*

[14] Cioè, a valori finiti o, equivalentemente per la monotonia della misura, che assuma valore finito sull'insieme ambiente.

Dimostrazione (i) Poichè le dimostrazioni nei due casi sono simili, ci limitiamo a considerare quella relativa al primo. Sia dunque $f \geq g$ e $\int_\Omega g\, d\mathrm{m} > -\infty$. Allora, $f^- \leq g^-$ e $\int_\Omega g^-\, d\mathrm{m} < +\infty$. Ne segue, per il Lemma A.3.3(ii), $\int_\Omega f^-\, d\mathrm{m} \leq \int_\Omega g^-\, d\mathrm{m} < +\infty$ e quindi la sommabilità di f.

(ii) Basta osservare che, per la linearità dell'integrale, $\int_\Omega |f|\, d\mathrm{m} = \int_\Omega (f^+ + f^-)\, d\mathrm{m} = \int_\Omega f^+\, d\mathrm{m} + \int_\Omega f^-\, d\mathrm{m}$.

(iii) Segue da (ii), osservato che, per il Lemma A.3.3(ii), $\int_\Omega |f|\, d\mathrm{m} \leq \int_\Omega g\, d\mathrm{m} < +\infty$.

(iv) Siano f^2, g^2 integrabili. Allora, per la linearità dell'integrale, $\int_\Omega (f^2 + g^2)\, d\mathrm{m} = \int_\Omega f^2\, d\mathrm{m} + \int_\Omega g^2\, d\mathrm{m} < +\infty$ e quindi anche $f^2 + g^2$ è integrabile. Ora, dalla $0 \leq (|f| - |g|)^2 = f^2 - 2|fg| + g^2$ si ha $|fg| \leq \frac{f^2+g^2}{2} \leq f^2 + g^2$. Ne segue, tramite (iii), l'integrabilità di fg.

(v) Supposto m finita, sia intanto f^2 integrabile. Notato che $\int_\Omega I_\Omega^2\, d\mathrm{m} = \int_\Omega I_\Omega\, d\mathrm{m} = \mathrm{m}(\Omega) < +\infty$, da (iv) otteniamo l'integrabilità di $f = fI_\Omega$. Sia infine f limitata. Esiste allora n tale che $|f| \leq n = nI_\Omega$. Osservato che $\int_\Omega nI_\Omega\, d\mathrm{m} = n\, \mathrm{m}(\Omega) < +\infty$, da (iii) segue l'integrabilità di f. ■

L'ultima proposizione del Teorema A.3.4 mette in evidenza che gli insiemi di misura nulla sono importanti nella teoria dell'integrazione per la curiosa ragione che possono essere "ignorati" ai fini del calcolo degli integrali. Viene allora naturale chiamare m-**trascurabile** (in breve **trascurabile**) ogni sottoinsieme S di Ω per il quale esista un insieme $A \supseteq S$ tale che $\mathrm{m}(A) = 0$; chiaramente, sono trascurabili, oltre all'insieme vuoto, i sottoinsiemi di un insieme trascurabile e, per il Teorema A.1.3(ix), l'unione di una famiglia discreta di insiemi trascurabili; inoltre, per la monotonia della misura, un elemento di $\mathcal{A}$ è trascurabile se e solo se è di misura nulla. Diremo infine che una proprietà P o una relazione binaria $\asymp$, riguardanti applicazioni di dominio Ω, sussiste m-**quasi ovunque** (in breve **quasi ovunque**) se il suo campo di validità include il complementare di un insieme trascurabile; inoltre, per indicare tale situazione, useremo, rispettivamente, le notazioni P (m-q.o.) e $\asymp$ (m-q.o.). Così, ad esempio, scriveremo:
- f finita (m-q.o.), se l'insieme $\{|f| = +\infty\}$ è trascurabile;
- $f \leq g$ (m-q.o.), se l'insieme $\{f > g\}$ è trascurabile;
- $f < g$ (m-q.o.), se l'insieme $\{f \geq g\}$ è trascurabile;
- $f = g$ (m-q.o.), se l'insieme $\{f \neq g\}$ è trascurabile;[15]

[15] Le relazioni "$=$ (m-q.o.)" di **uguaglianza quasi ovunque** e "$<$ (m-q.o.)" di **quasi ovunque minore di** sono, come facilmente si verifica, rispettivamente la parte simmetrica e quella asimmetrica del preordinamento $\leq$ (m-q.o.). Per quanto riguarda le proprietà dell'uguaglianza quasi ovunque, è facile verificare che, se $f_n = g_n$ (m-q.o.) $(n \geq 1)$, si ha $\sum_{i=1}^n f_i = \sum_{i=1}^n g_i$ (m-q.o.), $\prod_{i=1}^n f_i = \prod_{i=1}^n g_i$ (m-q.o.) per ogni n e $\alpha f_1 = \alpha g_1$ (m-q.o.) per ogni $\alpha \in \mathbb{R}^*$; inoltre, $g_n \to f$ (m-q.o.), se $f_n \to f$ (m-q.o.).

- $\lim\limits_{n\to+\infty} f_n$ (m-q.o.), se l'insieme $A_{\lim}^c$ è trascurabile;
- $f_n \to f$ (m-q.o.), se l'insieme $\{f_n \to f\}^c$ è trascurabile;
- $f_n \uparrow f$ (m-q.o.), se l'insieme $\{\omega \in \Omega : \neg\,(f_n(\omega) \uparrow f(\omega))\}$ è trascurabile.

Il teorema successivo fornisce ulteriori proprietà dell'integrale di Lebesgue. In particolare, la prima proposizione assicura che l'integrale è un funzionale monotono rispetto al preordinamento $\leq$ (m-q.o.); la seconda che funzioni Borel misurabili uguali quasi ovunque sono identificabili dal punto di vista dell'integrazione, nel senso che una è sommabile se e solo se ne è l'altra e, in tal caso, i due integrali coincidono; l'ultima che una funzione integrabile è finita quasi ovunque.

Teorema A.3.6 *Sia f una funzione sommabile. Sussistono allora le seguenti proposizioni:*

(i) MONOTONIA: $\int_\Omega g\,d\mathrm{m} \leq \int_\Omega f\,d\mathrm{m}$, *se* $g \leq f$ (m-q.o.) *e g è sommabile;*

(ii) $\int_\Omega g\,d\mathrm{m} = \int_\Omega f\,d\mathrm{m}$, *se* $f = g$ (m-q.o.) *e g è Borel misurabile;*

(iii) $\left|\int_\Omega f\,d\mathrm{m}\right| \leq \int_\Omega |f|\,d\mathrm{m}$*;*

(iv) $\mathrm{m}(A)\inf f(A) \leq \int_A f\,d\mathrm{m} \leq \mathrm{m}(A)\sup f(A)$ *per ogni* A[16]*;*

(v) Se $f \geq 0$ *e* $\int_\Omega f\,d\mathrm{m} = 0$*, allora* $f = 0$ (m-q.o.)*;*

(vi) Se $\mathrm{m}(A) > 0$ *e* $f(\omega) > 0$ *per ogni* $\omega \in A$*, allora* $\int_A f\,d\mathrm{m} > 0$*;*

(vii) Se f è integrabile, allora f è finita quasi ovunque.

DIMOSTRAZIONE (i) Sia g una funzione sommabile tale che $\mathrm{m}(\{g > f\}) = 0$. Allora, posto $A_1 = \{g \leq f\}$, riesce $gI_{A_1} \leq fI_{A_1}$ e $\mathrm{m}(A_1^c) = 0$. Osservato che $(gI_{A_1})^+ \leq (fI_{A_1})^+$ e $(gI_{A_1})^- \geq (fI_{A_1})^-$, dal Teorema A.3.4(iv) otteniamo

$$\begin{aligned}\int_\Omega g\,d\mathrm{m} &= \int_{A_1} g\,d\mathrm{m} = \int_\Omega (gI_{A_1})^+ d\mathrm{m} - \int_\Omega (gI_{A_1})^- d\mathrm{m}\\ &\leq \int_\Omega (fI_{A_1})^+ d\mathrm{m} - \int_\Omega (fI_{A_1})^- d\mathrm{m} = \int_{A_1} f\,d\mathrm{m} = \int_\Omega f\,d\mathrm{m}.\end{aligned}$$

(ii) Sia g una funzione Borel misurabile tale che $\mathrm{m}(\{f \neq g\}) = 0$. Allora, $f \leq g$ (m-q.o.), $f \geq g$ (m-q.o.) e quindi, per (i), basta verificare la sommabilità di g. Poichè f è sommabile, gli integrali delle relative parti positiva e negativa non sono entrambi infiniti. Supposto finito, ad esempio, quello della parte positiva, proviamo che $\int_\Omega g^+\,dm \leq$

[16]Ricordiamo che questa proposizione è nota come **teorema della media integrale**.

$\int_\Omega f^+ \, d\mathrm{m}$. A tal fine, tenuto conto di (i), basta constatare che $f^+ = g^+$ (m-q.o.). Ora, osservato che $\{f^+ \neq g^+\} = \{\max(0,f) \neq \max(0,g)\} \subseteq \{f \neq g\}$, dalla monotonia della misura otteniamo $\mathrm{m}(\{f^+ \neq g^+\}) \leq \mathrm{m}(\{f \neq g\}) = 0$.

(iii) Dalla $-|f| \leq f \leq |f|$, tramite (i) e la linearità dell'integrale, si ha

$$-\int_\Omega |f| \, d\mathrm{m} = \int_\Omega -|f| \, d\mathrm{m} \leq \int_\Omega f \, d\mathrm{m} \leq \int_\Omega |f| \, d\mathrm{m}$$

e quindi $\left|\int_\Omega f \, d\mathrm{m}\right| \leq \int_\Omega |f| \, d\mathrm{m}$.

(iv) Poichè le disuguaglianze $\mathrm{m}(A) \inf f(A) \leq \int_A f d\mathrm{m}$, $\int_A f d\mathrm{m} \leq \mathrm{m}(A) \sup f(A)$ si provano in modo analogo, ci limitiamo a verificare la prima. Per il Teorema A.3.4(iv) possiamo supporre $\mathrm{m}(A) > 0$ e quindi anche $\inf f(A) > -\infty$. Allora, per la linearità dell'integrale, $\inf f(A)\, \mathrm{m}(A) = \inf f(A) \int_\Omega I_A d\mathrm{m} = \int_\Omega \inf f(A)\, I_A d\mathrm{m}$. Ne segue, tenuto conto di $\inf f(A)\, I_A \leq f\, I_A$ e (i), $\inf f(A)\, \mathrm{m}(A) \leq \int_\Omega f\, I_A \, d\mathrm{m} = \int_A f \, d\mathrm{m}$.

(v) Sia $f \geq 0$ e $\int_\Omega f \, d\mathrm{m} = 0$. Posto $A = \{f > 0\}$ supponiamo (per assurdo) $\mathrm{m}(A) > 0$. Osservato che la successione non decrescente $A_n = \{f > \frac{1}{n}\}$ $(n \geq 1)$ è tale che $A = \bigcup_{n \geq 1} A_n$, dalla continuità dal basso della misura si ha $\mathrm{m}(A_n) \uparrow \mathrm{m}(A)$. Ne segue che esiste k tale che $\mathrm{m}(A_k) > 0$. Considerata allora la funzione semplice $g = \frac{1}{k}\, I_{A_k}$, otteniamo $g \leq f$ da cui, tramite (i), si ha la contraddizione $0 = \int_\Omega f \, d\mathrm{m} \geq \int_\Omega g \, d\mathrm{m} = \frac{1}{k}\, \mathrm{m}(A_k) > 0$.

(vi) Supposto $\mathrm{m}(A) > 0$ e $f(\omega) > 0$ per ogni $\omega \in A$, assumiamo (per assurdo) $\int_\Omega f I_A \, d\mathrm{m} = \int_A f \, d\mathrm{m} = 0$. Poichè $f I_A \geq 0$, da (v) si ha $f I_A = 0$ (m-q.o.). Osservato che $A = \{f I_A \neq 0\}$, otteniamo allora la contraddizione $0 < \mathrm{m}(A) = \mathrm{m}(\{f I_A \neq 0\}) = 0$.

(vii) Sia f integrabile. Allora, per il Teorema A.3.5(ii), è pure integrabile il suo valore assoluto. Posto $A = \{|f| = +\infty\}$, assumiamo (per assurdo) $\mathrm{m}(A) > 0$. Considerata la funzione semplice $g = +\infty\, I_A$, otteniamo $g \leq |f|$ e quindi, tramite (i), la contraddizione $+\infty = +\infty\, \mathrm{m}(A) = \int_\Omega g \, d\mathrm{m} \leq \int_\Omega |f| \, d\mathrm{m} < +\infty$. ■

A.3.3 Proprietà di convergenza

In questa sezione proviamo tre fondamentali teoremi di convergenza per l'integrale di Lebesgue che trovano applicazione in tutti i campi che fanno uso di modelli statistici o probabilistici. Iniziamo con quello che fornisce una generalizzazione del Lemma A.3.2.

Teorema A.3.7 (della convergenza monotona) *Sia $(f_n)_{n \geq 1}$ una successione di funzioni Borel misurabili tale che $f_1 \geq 0$ (m-q.o.) e $f_n \uparrow f$. Allora, $f \geq 0$ (m-q.o.) e $\int_\Omega f_n \, d\mathrm{m} \uparrow \int_\Omega f \, d\mathrm{m}$.*

Dimostrazione Poichè l'insieme $\{f_1 < 0\}$ è trascurabile, si ha $f \geq 0$ (m-q.o.) in quanto $\{f < 0\} \subseteq \{f_1 < 0\}$; inoltre, posto $A = \bigcap_{n \geq 1} \{f_n \geq 0\}$ e osservato che $\{f_1 \geq 0\} \subseteq A$, otteniamo $A^c \subseteq \{f_1 < 0\}$ e quindi anche A^c è un insieme trascurabile. Conseguentemente, posto $\tilde{f} = f I_A$ e $\tilde{f}_n = f_n I_A$ per ogni n, risulta $\tilde{f} = f$ (m-q.o.), $\tilde{f}_n \uparrow \tilde{f}$ e $\tilde{f}_n = f_n$ (m-q.o.), $\tilde{f}_n \geq 0$ per ogni n. Allora, per il Teorema A.3.6(ii), f, f_n $(n \geq 1)$ sono funzioni sommabili tali che $\int_\Omega f\, dm = \int_\Omega \tilde{f}\, dm$ e $\int_\Omega f_n\, dm = \int_\Omega \tilde{f}_n\, dm$ per ogni n. Basta quindi verificare che $\int_\Omega \tilde{f}_n\, dm \uparrow \int_\Omega \tilde{f}\, dm$.

Notato che, per la monotonia dell'integrale, $(\int_\Omega \tilde{f}_n\, dm)_{n\geq 1}$ è una successione non decrescente tale che $\lim_{n\to+\infty} \int_\Omega \tilde{f}_n\, dm \leq \int_\Omega \tilde{f}\, dm$, proviamo la disuguaglianza opposta. A tal fine, per ogni m consideriamo una successione di funzioni semplici non negative $(f_n^{(m)})_{n\geq 1}$ tale che $f_n^{(m)} \uparrow \tilde{f}_m$ come $n \to +\infty$. Posto allora $g_n = \max_{m\leq n} f_n^{(m)}$ per ogni n, otteniamo che $(g_n)_{n\geq 1}$ è una successione non decrescente di funzioni semplici non negative tale che

$$f_n^{(m)} \leq g_n \leq \max_{1\leq h\leq n} \tilde{f}_h = \tilde{f}_n \leq \tilde{f} \qquad (m \leq n;\, n \geq 1). \tag{A.11}$$

Si ha pertanto $\tilde{f}_m = \lim_{n\to+\infty} f_n^{(m)} \leq \lim_{n\to+\infty} g_n \leq \tilde{f}$. Passando al limite per $m \to +\infty$, riesce allora $g_n \uparrow \tilde{f}$. Ne segue, per il Lemma A.3.2, $\int_\Omega g_n\, dm \uparrow \int_\Omega \tilde{f}\, dm$. Notato infine che, per la monotonia dell'integrale e (A.11), $\int_\Omega g_n\, dm \leq \int_\Omega \tilde{f}_n\, dm$ per ogni n, otteniamo $\int_\Omega \tilde{f}\, dm = \lim_{n\to+\infty} \int_\Omega g_n\, dm \leq \lim_{n\to+\infty} \int_\Omega \tilde{f}_n\, dm$. ■

Il prossimo teorema di convergenza assicura l'invertibilità dell'operatore d'integrazione con quello di serie nel caso di funzioni integrande non negative.

Teorema A.3.8 (d'integrazione per serie) *Siano $(f_n)_{n\geq 1}$ una successione di funzioni Borel misurabili non negative e $(\alpha_n)_{n\geq 1}$ una successione di numeri reali non negativi. Allora, $\int_\Omega \sum_{n\geq 1} \alpha_n f_n\, dm = \sum_{n\geq 1} \alpha_n \int_\Omega f_n\, dm$.*

Dimostrazione Procedendo per induzione su n otteniamo che l'uguaglianza

$$\int_\Omega \sum_{n=1}^{n} \alpha_i f_i\, dm = \sum_{n=1}^{n} \alpha_i \int_\Omega f_i\, dm$$

sussiste qualunque sia n (la base dell'induzione "$n = 1$" è banale e il passo induttivo segue dalla linearità dell'integrale). Dalla $0 \leq \sum_{i=1}^n \alpha_i f_i \uparrow \sum_{n\geq 1} \alpha_n f_n$ per $n \to +\infty$ si ha allora

$$\int_\Omega \sum_{n\geq 1} \alpha_n f_n\, dm = \lim_{n\to+\infty} \int_\Omega \sum_{i=1}^{n} \alpha_i f_i\, dm = \lim_{n\to+\infty} \sum_{i=1}^{n} \alpha_i \int_\Omega f_i\, dm = \sum_{n\geq 1} \alpha_n \int_\Omega f_n\, dm$$

tenuto conto del teorema della convergenza monotona. ■

Osservazione A.3.9 Nel caso particolare delle misure di conteggio indotte da insiemi discreti, possiamo ottenere, mediante il Teorema A.3.6(ii) e quello d'integrazione per serie, la seguente interessante e utile connessione tra l'integrale di Lebesgue e le serie numeriche. Dato un insieme discreto $S \subseteq \Omega$ avente

$N \leq +\infty$ elementi, sia $f : \Omega \mapsto \mathbb{R}^*$ una funzione γ_S-sommabile. Considerata allora una numerazione $s_1, s_2, \ldots$ di S, riesce

$$\int_\Omega f\, d\gamma_S = \sum_{i=1}^N f^+(s_i) - \sum_{i=1}^N f^-(s_i) = \sum_{i=1}^N f^+(s_i) + \sum_{i=1}^N (-f^-(s_i)) \qquad \text{(A.12)}$$

e quindi, qualora sia $N = +\infty$, l'integrale può essere calcolato sommando le serie dei valori, rispettivamente, positivi e negativi della restrizione $f\big|_S$. Conseguentemente, sussiste l'uguaglianza $\int_\Omega f\, d\gamma_S = \sum_{n\geq 1} f(s_n)$ ogniqualvolta la serie numerica $\sum_{n\geq 1} f(s_n)$ risulti permutabile, come avviene quando f è di segno costante (serie a termini positivi) oppure è γ_S-integrabile (serie assolutamente convergente).

Per provare (A.12), assumiamo intanto $f \geq 0$. Posto $g = \sum_{n\geq 1} f(s_n) I_{\{s_n\}}$, otteniamo $g = f$ (γ_S-q.o.) e quindi, per il Teorema A.3.6(ii),

$$\begin{aligned}\int_\Omega f\, d\gamma_S &= \int_\Omega g\, d\gamma_S = \int_\Omega \sum_{i=1}^N f(s_i) I_{\{s_i\}}\; d\gamma_S \\ &= \sum_{i=1}^N f(s_i) \int_\Omega I_{\{s_i\}}\, d\gamma_S = \sum_{i=1}^N f(s_i)\, \gamma_S(\{s_i\}) = \sum_{i=1}^N f(s_i).\end{aligned}$$

Sia ora f qualsiasi. Osservato che $\int_\Omega f^+\, d\gamma_S = \sum_{i=1}^N f^+(s_i)$ e $\int_\Omega f^-\, d\gamma_S = \sum_{i=1}^N f^-(s_i)$, dalla definizione dell'integrale si ha (A.12). ♦

Il risultato successivo fornisce due disuguaglianze riguardanti, rispettivamente, l'integrale del minimo-limite e quello del massimo-limite di una successione di funzioni Borel misurabili.

Lemma A.3.10 (di Fatou) *Sia $(f_n)_{n\geq 1}$ una successione di funzioni Borel misurabili. Risulta allora:*

(i) $\int_\Omega \liminf\limits_{n\to+\infty} f_n\, dm \leq \liminf\limits_{n\to+\infty} \int_\Omega f_n\, dm$, *se $f_n \geq 0$ per ogni n;*

(ii) $\int_\Omega \limsup\limits_{n\to+\infty} f_n\, dm \geq \limsup\limits_{n\to+\infty} \int_\Omega f_n\, dm$, *se $f_n \leq g$ per ogni n e g è integrabile.*

Dimostrazione (i) Posto $g_n = \inf\limits_{m\geq n} f_m \geq 0$ per ogni n, otteniamo $f_n \geq g_n$ $(n \geq 1)$ e $g_n \uparrow \liminf\limits_{n\to+\infty} f_n$. Allora, per la monotonia dell'integrale e il teorema della convergenza monotona,

$$\int_\Omega \liminf_{n\to+\infty} f_n\, dm = \lim_{n\to+\infty} \int_\Omega g_n\, dm = \liminf_{n\to+\infty} \int_\Omega g_n\, dm \leq \liminf_{n\to+\infty} \int_\Omega f_n\, dm.$$

(ii) Posto $g_n = g - f_n \geq 0$ per ogni n, otteniamo $\liminf\limits_{n\to+\infty} g_n = g - \limsup\limits_{n\to+\infty} f_n$. Inoltre, per la linearità dell'integrale, $\int_\Omega g_n \, dm = \int_\Omega g \, dm - \int_\Omega f_n \, dm$ per ogni n e quindi

$$\liminf_{n\to+\infty} \int_\Omega g_n \, dm = \int_\Omega g \, dm - \limsup_{n\to+\infty} \int_\Omega f_n \, dm.$$

Usando ancora la linearità dell'integrale, tramite (i), si ha allora

$$\begin{aligned}\int_\Omega g \, dm - \int_\Omega \limsup_{n\to+\infty} f_n \, dm &= \int_\Omega (g - \limsup_{n\to+\infty} f_n) \, dm = \int_\Omega \liminf_{n\to+\infty} g_n \, dm \\ &\leq \liminf_{n\to+\infty} \int_\Omega g_n \, dm = \int_\Omega g \, dm - \limsup_{n\to+\infty} \int_\Omega f_n \, dm\end{aligned}$$

e quindi la tesi. ■

Il seguente celebre teorema di convergenza (dovuto a Henri Lebesgue) è senza alcun dubbio una delle "pietre miliari" della teoria dell'integrazione e viene usualmente citato per evidenziare la "superiorità" dell'integrale di Lebesgue rispetto a quello classico di Riemann.

Teorema A.3.11 (della convergenza dominata) *Sia $(f_n)_{n\geq 1}$ una successione di funzioni Borel misurabili tale che $f_n \to f$. Inoltre, sia g una funzione integrabile tale che $|f_n| \leq g$ (m-q.o.) per ogni n. Allora, f, f_n $(n \geq 1)$ sono integrabili e si ha $\int_\Omega f_n \, dm \to \int_\Omega f \, dm$.*

DIMOSTRAZIONE Per il Teorema A.3.6(vii), g è finita quasi ovunque, cioè $\{|g| = +\infty\}$ è un insieme trascurabile. Posto allora $A = \{|g| < +\infty\} \cap \bigcap_{n\geq 1}\{|f_n| \leq g\}$, anche l'insieme $A^c = \{|g| = +\infty\} \cup \bigcup_{n\geq 1}\{|f_n| > g\}$ è trascurabile, tenito conto che l'insieme $\{|f_n| > g\}$ è trascurabile per ogni n. Conseguentemente, posto $\tilde{f}_n = f_n I_A$ per ogni n e $\tilde{f} = f I_A$, risulta $\tilde{f}_n = f_n$ (m-q.o.) per ogni n e $\tilde{f} = f$ (m-q.o.). Per il Teorema A.3.6(ii), basta quindi verificare che $\tilde{f}$, $\tilde{f}_n$ $(n \geq 1)$ sono funzioni integrabili e che $\int_\Omega \tilde{f}_n dm \to \int_\Omega \tilde{f} dm$.

Ora, posto $\tilde{g} = g I_A$, si ha $\tilde{g} = g$ (m-q.o.) e quindi, per il Teorema A.3.6(ii), $\tilde{g}$ è integrabile; inoltre, dalle $\tilde{f}_n \to \tilde{f}$ e $|\tilde{f}_n| \leq \tilde{g}$ per ogni n, otteniamo $|\tilde{f}| \leq \tilde{g}$. Ne segue, $|\tilde{f} - \tilde{f}_n| \leq |\tilde{f}| + |\tilde{f}_n| \leq 2\tilde{g}$ per ogni n e, per il Teorema A.3.5(iii), l'integrabilità di $\tilde{f}$ e di $\tilde{f}_n$ $(n \geq 1)$. Osservato infine che la funzione $2\tilde{g}$ è integrabile, dalla linearità dell'integrale, dal Teorema A.3.6(iii) e dalla proposizione (ii) del lemma di Fatou, si ha

$$\begin{aligned}0 \leq \liminf_{n\to+\infty} \left|\int_\Omega \tilde{f} \, dm - \int_\Omega \tilde{f}_n \, dm\right| &\leq \limsup_{n\to+\infty} \left|\int_\Omega \tilde{f} \, dm - \int_\Omega \tilde{f}_n \, dm\right| \\ &= \limsup_{n\to+\infty} \left|\int_\Omega (\tilde{f} - \tilde{f}_n) \, dm\right| \leq \limsup_{n\to+\infty} \int_\Omega |\tilde{f} - \tilde{f}_n| \, dm \\ &\leq \int_\Omega \limsup_{n\to+\infty} |\tilde{f} - \tilde{f}_n| \, dm = \int_\Omega \lim_{n\to+\infty} |\tilde{f} - \tilde{f}_n| \, dm = 0\end{aligned}$$

e quindi $\lim\limits_{n\to+\infty} \left|\int_\Omega \tilde{f} \, dm - \int_\Omega \tilde{f}_n \, dm\right| = 0$, cioè $\int_\Omega \tilde{f}_n \, dm \to \int_\Omega \tilde{f} \, dm$. ■

Osservazione A.3.12 Qualora non sussista l'ipotesi "g funzione integrabile e $|f_n| \leq g$ (m-q.o.) per ogni n", può avvenire che la convergenza degli integrali venga a mancare. Posto infatti $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B} \cap \Omega$ e $\mathrm{m} = \lambda\big|_{\mathcal{A}}$, sia f_n la funzione $\mathcal{A}$-Borel misurabile così definita:

$$f_n(x) = \begin{cases} 2n^2 x & \text{se } 0 \leq x \leq \frac{1}{2n} \\ -2n^2 x + 2n & \text{se } \frac{1}{2n} \leq x \leq \frac{1}{n} \\ 0 & \text{se } \frac{1}{n} \leq x \leq 1 \end{cases} \qquad (n \geq 1).$$

Allora, $f_n \to f = I_\emptyset$, $\int_0^1 f_n(x)\,dx = \frac{1}{2}$ per ogni n e $\int_0^1 f(x)\,dx = 0$. ♦

Concludiamo la sezione fornendo delle condizioni che consentono l'invertibilità dell'operatore di derivazione con quello d'integrazione.

Teorema A.3.13 (di derivazione sotto il segno d'integrale) *Siano Ω un intervallo (limitato o no) della retta reale, $\mathcal{A}$ la traccia di $\mathcal{B}$ su Ω e m una misura finita. Sia inoltre $f : \Omega^2 \mapsto \mathbb{R}$ una funzione continua con derivata parziale rispetto alla seconda variabile continua. Riesca infine integrabile la funzione $f(\cdot, t)$ per ogni $t \in \Omega$. Allora,*

$$\Big[\frac{d}{dt}\int_\Omega f(\omega,t)\,\mathrm{m}(d\omega)\Big](t_0) = \int_\Omega \frac{\partial f}{\partial t}(\omega,t_0)\,\mathrm{m}(d\omega)$$

per ogni $t_0 \in \Omega$.

DIMOSTRAZIONE Sia $F(t) = \int_\Omega f(\omega,t)\,\mathrm{m}(d\omega)$ per ogni $t \in \Omega$. Dato $t_0 \in \Omega$, assumiamo intanto $t_0 < \sup\Omega$. Sia allora $(\epsilon_n)_{n\geq 1}$ una successione decrescente di numeri reali positivi tale che $t_0 + \epsilon_n \in \Omega$ per ogni n e $\epsilon_n \to 0$. Considerato il rapporto incrementale di F corrispondente all'incremento ϵ_n e relativo al punto t_0, dalla linearità dell'integrale si ha

$$\begin{aligned} \frac{F(t_0+\epsilon_n) - F(t_0)}{\epsilon_n} &= \frac{\displaystyle\int_\Omega f(\omega,t_0+\epsilon_n)\,\mathrm{m}(d\omega) - \int_\Omega f(\omega,t_0)\,\mathrm{m}(d\omega)}{\epsilon_n} \\ &= \int_\Omega \frac{f(\omega,t_0+\epsilon_n) - f(\omega,t_0)}{\epsilon_n}\,\mathrm{m}(d\omega). \end{aligned} \tag{A.13}$$

Ora, qualunque sia $\omega \in \Omega$, la funzione $g = f(\omega,\cdot)$ è derivabile nell'intervallo $[t_0, t_0+\epsilon_1]$; inoltre, la derivata $g'(t) = \frac{\partial f}{\partial t}(\omega,t)$ è, per l'ipotesi di continuità della derivata parziale, limitata su tale intervallo. Esiste quindi $k > 0$ tale che $|g'(t)| \leq k$ per ogni $t \in [t_0, t_0+\epsilon_1]$. Allora, posto

$$R_n(\omega;t_0) = \frac{f(\omega,t_0+\epsilon_n) - f(\omega,t_0)}{\epsilon_n} = \frac{g(t_0+\epsilon_n) - g(t_0)}{\epsilon_n}$$

e osservato che $[t_0, t_0 + \epsilon_n] \subset [t_0, t_0 + \epsilon_1]$, per il teorema di Lagrange esiste $t' \in [t_0, t_0 + \epsilon_n]$ tale che $|R_n(\omega; t_0)| = |\, g'(t')\,| \leq k$.

Conseguentemente, la successione $(R_n(\cdot; t_0))_{n\geq 1}$ è costituita da funzioni continue tutte dominate dalla funzione costante di valore k. Notato che, per il Teorema A.2.3(i), le funzioni della successione sono Borel misurabili, dal teorema della convergenza dominata otteniamo che la funzione limite $\frac{\partial f}{\partial t}(\cdot, t_0) = \lim_{n\to+\infty} R_n(\cdot; t_0)$ è integrabile e riesce

$$\lim_{n\to+\infty} \int_\Omega R_n(\omega; t_0)\, \mathrm{m}(d\omega) = \int_\Omega \lim_{n\to+\infty} R_n(\omega; t_0)\, \mathrm{m}(d\omega) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0)\, \mathrm{m}(d\omega).$$

Ne segue, tramite (A.13),

$$\lim_{n\to+\infty} \frac{F(t_0 + \epsilon_n) - F(t_0)}{\epsilon_n} = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0)\, \mathrm{m}(d\omega).$$

Tenuto infine conto dell'arbitrarietà della successione $(\epsilon_n)_{n\geq 1}$, otteniamo che per la derivata destra $\left(\frac{dF}{dt}\right)^+$ si ha

$$\left(\frac{dF}{dt}\right)^+(t_0) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0)\, \mathrm{m}(d\omega).$$

Supponiamo ora $t_0 > \inf \Omega$. Possiamo allora provare, con argomentazioni analoghe, che la derivata sinistra $\left(\frac{dF}{dt}\right)^-$ verifica l'uguaglianza

$$\left(\frac{dF}{dt}\right)^-(t_0) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0)\, \mathrm{m}(d\omega).$$

Riesce quindi $F'(t_0) = \int_\Omega \frac{\partial f}{\partial t}(\omega, t_0)\, \mathrm{m}(d\omega)$. ■

A.4 Misure definite tramite funzioni

In questa sezione introduciamo due particolari tipi di misure (di fondamentale importanza per il calcolo delle probabilità e le sue applicazioni) e stabiliamo le regole di calcolo dei relativi integrali.

Iniziamo col considerare un'applicazione $\tau : \Omega \mapsto \Omega'$ che sia $(\mathcal{A}, \mathcal{A}')$-misurabile. Possiamo allora introdurre la misura m_τ su $\mathcal{A}'$ così definita:

$$\mathrm{m}_\tau(A') = \mathrm{m}(\tau^{-1}(A')) = \mathrm{m}(\{\tau \in A'\})$$

per ogni A', detta **misura immagine di** m **mediante** τ. Per constatare che siamo effettivamente in presenza di una misura basta osservare che, data una successione disgiunta $(A'_n)_{n\geq 1}$, si ha

$$\begin{aligned}\mathrm{m}_\tau(\bigcup_{n\geq 1} A'_n) &= \mathrm{m}(\tau^{-1}(\bigcup_{n\geq 1} A'_n)) \\ &= \mathrm{m}(\bigcup_{n\geq 1} \tau^{-1}(A'_n)) = \sum_{n\geq 1} \mathrm{m}(\tau^{-1}(A'_n)) = \sum_{n\geq 1} \mathrm{m}_\tau(A'_n),\end{aligned}$$

notato che $(\{\tau \in A'_n\})_{n\geq 1}$ è una successione disgiunta.

Il prossimo teorema mostra come l'integrazione rispetto alla misura immagine possa essere ricondotta a quella rispetto alla misura originaria. Intendendo la relazione $\omega' = \tau(\omega)$ come un "cambio di variabile" di ω con ω', il risultato può anche essere inteso come un teorema sul cambiamento delle variabili nell'integrale di Lebesgue.

Teorema A.4.1 (della misura immagine) *Sia $f : \Omega' \mapsto \mathbb{R}^*$ una funzione $\mathcal{A}'$-Borel misurabile. Allora, f è m_τ-sommabile se e solo se la funzione composta $f(\tau)$ è m-sommabile. Inoltre, nel caso di sommabilità, si ha*

$$\int_{A'} f\, d\mathrm{m}_\tau = \int_{\{\tau\in A'\}} f(\tau)\, d\mathrm{m}$$

per ogni A'.

DIMOSTRAZIONE Dato A', assumiamo intanto $f = I_{A''}$ con $A'' \in \mathcal{A}'$. Allora

$$f(\tau)(\omega) = f(\tau(\omega)) = \begin{cases} 1 & \text{se } \tau(\omega) \in A'' \\ 0 & \text{se } \tau(\omega) \notin A'' \end{cases} = I_{\{\tau\in A''\}}(\omega)$$

per ogni $\omega \in \Omega$ e quindi

$$\begin{aligned}\int_{A'} f\, d\mathrm{m}_\tau &= \int_{\Omega'} I_{A''} I_{A'}\, d\mathrm{m}_\tau = \mathrm{m}_\tau(A'' \cap A') = \mathrm{m}(\tau^{-1}(A'' \cap A')) \\ &= \mathrm{m}(\tau^{-1}(A'') \cap \tau^{-1}(A')) = \int_{\Omega'} I_{\{\tau\in A''\}} I_{\{\tau\in A'\}}\, d\mathrm{m} = \int_{\{\tau\in A'\}} I_{\{\tau\in A''\}}\, d\mathrm{m} \\ &= \int_{\{\tau\in A'\}} f(\tau)\, d\mathrm{m}.\end{aligned}$$

Sia ora f una funzione semplice non negativa. Dalla linearità dell'integrale otteniamo

$$\begin{aligned}\int_{A'} f\, d\mathrm{m}_\tau &= \int_{\Omega'} (\sum_{y\in f(\Omega')} y I_{\{f=y\}}) I_{A'}\, d\mathrm{m}_\tau = \sum_{y\in f(\Omega')} y \int_{A'} I_{\{f=y\}}\, d\mathrm{m}_\tau \\ &= \sum_{y\in f(\Omega')} y \int_{\{\tau\in A'\}} I_{\{f=y\}}(\tau)\, d\mathrm{m} = \int_{\{\tau\in A'\}} (\sum_{y\in f(\Omega')} y I_{\{f=y\}})(\tau)\, d\mathrm{m} \\ &= \int_{\{\tau\in A'\}} f(\tau)\, d\mathrm{m}.\end{aligned}$$

Sia ora $f \geq 0$. Considerata una successione $(f_n)_{n\geq 1}$ di funzioni $\mathcal{A}'$-semplici non negative tale che $f_n \uparrow f$, si ha $0 \leq f_n I_{A'} \uparrow f I_{A'}$ e $0 \leq f_n(\tau) \uparrow f(\tau)$. Allora, poichè $(f_n(\tau))_{n\geq 1}$ è una successione di funzioni Borel misurabili, dal teorema della convergenza monotona otteniamo

$$\begin{aligned}\int_{A'} f\, \mathrm{dm}_\tau &= \lim_{n\to+\infty}\int_{A'} f_n\, \mathrm{dm}_\tau = \lim_{n\to+\infty}\int_{\{\tau\in A'\}} f_n(\tau)\, \mathrm{dm} = \int_{\{\tau\in A'\}} \lim_{n\to+\infty} f_n(\tau)\, \mathrm{dm}\\ &= \int_{\{\tau\in A'\}} f(\tau)\, \mathrm{dm}.\end{aligned}$$

Sia infine f qualsiasi. Dalle $f(\tau)^+ = f^+(\tau)$ e $f(\tau)^- = f^-(\tau)$ risulta

$$\begin{aligned}\int_{A'} f^+\, \mathrm{dm}_\tau &= \int_{\{\tau\in A'\}} f^+(\tau)\, \mathrm{dm} = \int_{\{\tau\in A'\}} f(\tau)^+\, \mathrm{dm}\\ \int_{A'} f^-\, \mathrm{dm}_\tau &= \int_{\{\tau\in A'\}} f^-(\tau)\, \mathrm{dm} = \int_{\{\tau\in A'\}} f(\tau)^-\, \mathrm{dm}\end{aligned}$$

e quindi, essendo A' del tutto arbitrario, la m_τ-sommabilità di f equivale alla m-sommabilità di $f(\tau)$. Riesce inoltre

$$\begin{aligned}\int_{A'} f\, \mathrm{dm}_\tau &= \int_{A'} f^+ \mathrm{dm}_\tau - \int_{A'} f^- \mathrm{dm}_\tau\\ &= \int_{\{\tau\in A'\}} f(\tau)^+\, \mathrm{dm} - \int_{\{\tau\in A'\}} f(\tau)^-\, \mathrm{dm} = \int_{\{\tau\in A'\}} f \mathrm{dm}\end{aligned}$$

nel caso di sommabilità. ■

Nel corollario seguente forniamo, con riferimento alla misura di Lebesgue unidimensionale, la formula di trasformazione dell'integrale di Lebesgue relativa ai cambi di variabile di tipo affine.

Corollario A.4.2 *Sia $f : \mathbb{R} \mapsto \mathbb{R}^*$ una funzione Borel misurabile. Allora, considerata la trasformazione affine non degenere $\tau : x \mapsto ax + b$ della retta reale, la funzione f è λ_τ-sommabile se e solo se la funzione composta $f(\tau)$ è λ-sommabile. Inoltre, nel caso di sommabilità, risulta*

$$\int_B f(x)\, dx = |a| \int_{\frac{1}{a}(B-b)} f(ax+b)\, dx$$

per ogni boreliano $B \neq \emptyset$[17].

[17]Avendo posto, qualunque siano i numeri reali a, b e il sottoinsieme non vuoto S della retta reale, $a\,S + b = \{ax + b : x \in S\}$. Ricordato che la σ-algebra di Borel è chiusa per trasformazioni affini (Esempio A.1.1(iv)), $a\,B + b \in \mathcal{B}$ per ogni boreliano $B \neq \emptyset$.

Dimostrazione Sia $B \neq \emptyset$ un boreliano. Allora, $\tau^{-1}(B) = \frac{1}{a}(B - b)$ da cui otteniamo, per l'Esempio A.1.2(i), $\lambda_\tau(B) = \lambda(\tau^{-1}(B)) = \frac{1}{|a|}\lambda(B)$ e quindi, per il Teorema A.3.4(ii), $\int_B f\, d\lambda_\tau = \frac{1}{|a|}\int_B f\, d\lambda$. Ne segue, tramite il teorema precedente, $\frac{1}{|a|}\int_B f\, d\lambda = \int_{\frac{1}{a}(B-b)} f(ax+b)\, dx$. ■

Passando all'altro tipo di misura, consideriamo una funzione Borel misurabile $g : \Omega \to [0, +\infty]$. Possiamo allora introdurre la misura $g * \mathrm{m}$ su $\mathcal{A}$ così definita:

$$g * \mathrm{m}(A) = \int_A g\, d\mathrm{m}$$

per ogni A, detta **misura di densità** g **rispetto alla misura** (**di riferimento**) m. Per constatare che siamo effettivamente in presenza di una misura basta osservare che $g * \mathrm{m}(\emptyset) = \int_\emptyset g\, d\mathrm{m} = 0$ e che, data una successione disgiunta $(A_n)_{n\geq 1}$, dal teorema d'integrazione per serie otteniamo

$$\begin{aligned} g * \mathrm{m}(\bigcup_{n\geq 1} A_n) &= \int_{\bigcup_{n\geq 1} A_n} g\, d\mathrm{m} = \int_\Omega g I_{\bigcup_{n\geq 1} A_n}\, d\mathrm{m} = \int_\Omega g \sum_{n\geq 1} I_{A_n}\, d\mathrm{m} \\ &= \sum_{n\geq 1}\int_\Omega g I_{A_n}\, d\mathrm{m} = \sum_{n\geq 1}\int_{A_n} g\, d\mathrm{m} = \sum_{n\geq 1} g * \mathrm{m}(A_n). \end{aligned}$$

Il prossimo teorema mostra come l'integrazione rispetto ad una misura definita da una densità possa essere ricondotta a quella rispetto alla misura di riferimento.

Teorema A.4.3 *Sia f una funzione Borel misurabile. Allora, f è $g * \mathrm{m}$-sommabile se e solo se fg è* m*-sommabile. Inoltre, nel caso di sommabilità, $\int_\Omega f\, dg * \mathrm{m} = \int_\Omega fg\, d\mathrm{m}$.*

Dimostrazione Sia intanto f una funzione semplice non negativa. Dalla linearità dell'integrale risulta allora

$$\begin{aligned} \int_\Omega f\, dg * \mathrm{m} &= \sum_{y\in f(\Omega)} y\, g * \mathrm{m}(\{f = y\}) \\ &= \sum_{y\in f(\Omega)} y \int_\Omega I_{\{f=y\}} g\, d\mathrm{m} = \int_\Omega \Big(\sum_{y\in f(\Omega)} y I_{\{f=y\}}\Big) g\, d\mathrm{m} = \int_\Omega fg\, d\mathrm{m}. \end{aligned}$$

Sia ora $f \geq 0$. Considerata una successione $(f_n)_{n\geq 1}$ di funzioni semplici non negative tale che $f_n \uparrow f$ e osservato che $f_n g \uparrow fg$ e che $(f_n g)_{n\geq 1}$ è una successione di funzioni Borel misurabili non negative, dal teorema della convergenza monotona otteniamo

$$\int_\Omega f\, dg * \mathrm{m} = \lim_{n\to+\infty}\int_\Omega f_n\, dg * \mathrm{m} = \lim_{n\to+\infty}\int_\Omega f_n g\, d\mathrm{m} = \int_\Omega \big(\lim_{n\to+\infty} f_n\big) g\, d\mathrm{m} = \int_\Omega fg\, d\mathrm{m}.$$

Sia infine f qualsiasi. Allora,

$$\int_\Omega f^+ \, dg * \mathrm{m} = \int_\Omega f^+ g \, d\mathrm{m} = \int_\Omega (fg)^+ \, d\mathrm{m}$$
$$\int_\Omega f^- \, dg * \mathrm{m} = \int_\Omega f^- g \, d\mathrm{m} = \int_\Omega (fg)^- \, d\mathrm{m}$$

e quindi f è $g * \mathrm{m}$-sommabile se e solo se fg è m-sommabile. Riesce inoltre

$$\int_\Omega f \, dg * \mathrm{m} = \int_\Omega f^+ \, dg * \mathrm{m} - \int_\Omega f^- \, dg * \mathrm{m} = \int_\Omega (fg)^+ \, d\mathrm{m} - \int_\Omega (fg)^- \, d\mathrm{m} = \int_\Omega fg \, d\mathrm{m}$$

nel caso di sommabilità. ■

Sia ora $\tilde{g} : \Omega \mapsto [0, +\infty]$ una funzione Borel misurabile. Allora, se $\tilde{g} = g$ (m-q.o.), dal Teorema A.3.6(ii) risulta $\tilde{g} * \mathrm{m} = g * \mathrm{m}$. Il teorema seguente assicura che sussiste anche il viceversa se la misura m è ***σ*-finita**, cioè se esiste una partizione discreta $(A_n)_{n\geq 1}$ di Ω con $\mathrm{m}(A_n) < +\infty$ per ogni n[18].

Teorema A.4.4 *Siano* m *una misura σ-finita e f, $\tilde{f}$ due funzioni sommabili tali che $\int_A \tilde{f} \, d\mathrm{m} \leq \int_A f \, d\mathrm{m}$ per ogni A. Allora, $\tilde{f} \leq f$* (m-q.o.). *In particolare, $\tilde{f} = f$* (m-q.o.), *se $\int_A \tilde{f} \, d\mathrm{m} = \int_A f \, d\mathrm{m}$ per ogni A.*

DIMOSTRAZIONE Assumiamo intanto che la misura sia finita. Supposto (per assurdo) $\mathrm{m}(\{f < \tilde{f}\}) > 0$ e considerata una numerazione $q_1, q_2, \ldots$ dei numeri razionali, dalla subadditività della misura otteniamo

$$0 < \mathrm{m}(\{f < \tilde{f}\}) = \mathrm{m}\Big(\bigcup_{n\geq 1} \{f < q_n < \tilde{f}\}\Big) \leq \sum_{n \geq 1} \mathrm{m}(\{f < q_n < \tilde{f}\})$$

e quindi esiste n tale che $\mathrm{m}(\{f < q_n < \tilde{f}\}) > 0$. Posto $A = \{f < q_n < \tilde{f}\}$, si ha $fI_A \leq q_n I_A \leq \tilde{f} I_A$. Dall'additività e monotonia dell'integrale riesce allora

$$-\infty < q_n \mathrm{m}(A) = \int_\Omega q_n I_A \, d\mathrm{m} \leq \int_\Omega \tilde{f} I_A \, d\mathrm{m} = \int_A \tilde{f} \, d\mathrm{m}$$
$$+\infty > q_n \mathrm{m}(A) = \int_\Omega q_n I_A \, d\mathrm{m} \geq \int_\Omega f I_A \, d\mathrm{m} = \int_A f \, d\mathrm{m}.$$

Ne segue, tramite il Teorema A.3.6(vi) e la linearità dell'integrale, la contraddizione $0 \geq \int_A \tilde{f} \, d\mathrm{m} - \int_A f \, d\mathrm{m} = \int_A (\tilde{f} - f) \, d\mathrm{m} > 0$.

Sia ora la misura σ-finita. Esiste dunque una partizione discreta $(A_n)_{n\geq 1}$ di Ω con $\mathrm{m}(A_n) < +\infty$ per ogni n. Conseguentemente, la restrizione m_n di m sulla traccia $\mathcal{A} \cap A_n \subseteq \mathcal{A}$ di $\mathcal{A}$ su A_n è una misura finita tale che

$$\int_{\hat{A}} \tilde{f}\big|_A \, d\mathrm{m}_n = \int_{\hat{A}} \tilde{f} \, d\mathrm{m} \leq \int_{\hat{A}} f \, d\mathrm{m} = \int_{\hat{A}} f\big|_A \, d\mathrm{m}_n$$

[18]Esempi di misure σ-finite sono le misure finite, la misura di Lebesgue unidimensionale e le misure di conteggio indotte da insiemi discreti.

per ogni $\hat{A} \in \mathcal{A} \cap A_n$. Ne segue, per la prima parte della dimostrazione, $\tilde{f}\big|_{A_n} \leq f\big|_{A_n}$ (m_n-q.o.). Esiste quindi $\hat{A}_n \in \mathcal{A} \cap A_n$ tale che $\mathrm{m}(\hat{A}_n) = \mathrm{m}_n(\hat{A}_n) = 0$ e $\tilde{f}(\omega) \leq f(\omega)$ per ogni $\omega \in A_n \setminus \hat{A}_n$. Allora, per l'arbitrarietà di n, l'insieme $A = \bigcup_{n\geq 1} \hat{A}_n$ è trascurabile; inoltre, dato $\omega \notin A$, esiste m tale che $\omega \in A_m \setminus A$ ($(A_n)_{n\geq 1}$ è una partizione!) da cui otteniamo $\omega \in A_m \cap \bigcap_{n\geq 1} \hat{A}_n^c = \bigcap_{n\geq 1}(A_m \cap \hat{A}_n^c) = A_m \setminus \hat{A}_m$ e quindi $\tilde{f}(\omega) \leq f(\omega)$. Conseguentemente, $\tilde{f} \leq f$ (m-q.o.).

L'ultima parte della tesi segue ora immediatamente da quanto appena provato. ■

A.5 Misura prodotto

Dato un insieme non vuoto Ω_i, siano $\mathcal{A}_i$ una σ algebra su Ω_i e m_i una misura σ-finita su $\mathcal{A}_i$ ($i = 1, 2$). Considerate, per ogni sottoinsieme S del prodotto cartesiano $\Omega = \Omega_1 \times \Omega_2$, la **sezione di S relativa a** $\omega_1 \in \Omega_1$:

$$S(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in S\}$$

e la **sezione di S relativa a** $\omega_2 \in \Omega_2$:

$$S(\omega_2) = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in S\},$$

introduciamo in questa sezione, proseguendo nella presentazione di alcune misure notevoli, una misura (definita su una opportuna σ-algebra del prodotto cartesiano) che verifica le seguenti due proprietà (suggerite dalla nozione di area delle figure piane e del relativo Principio di Cavalieri):

(a) la misura di un “rettangolo” $\mathsf{A}_1 \times \mathsf{A}_2$ avente “base” $\mathsf{A}_1 \in \mathcal{A}_1$ e “altezza” $\mathsf{A}_2 \in \mathcal{A}_2$ è uguale al prodotto $\mathrm{m}_1(\mathsf{A}_1)\,\mathrm{m}_2(\mathsf{A}_2)$;

(b) due insiemi misurabili A_1, A_2 di Ω hanno misura uguale se, per qualche $i \in \{1, 2\}$, danno luogo a sezioni di misura uguale in corrispondenza ad ogni scelta dell'elemento che le individua in Ω_i, cioè se $\mathrm{m}_j(A_1(\omega_i)) = \mathrm{m}_j(A_2(\omega_i))$ ($j \neq i$) per ogni $\omega_i \in \Omega_i$.

Per quanto riguarda la σ-algebra di riferimento $\mathcal{A}$ su Ω, la identifichiamo con la **$\boldsymbol{\sigma}$-algebra prodotto $\mathcal{A}_1 \otimes \mathcal{A}_2$ (di $\mathcal{A}_1$ e $\mathcal{A}_2$)**, cioè con la σ-algebra generata dalla famiglia chiusa per intersezioni finite $\mathcal{R} = \{\mathsf{A}_1 \times \mathsf{A}_2 : \mathsf{A}_i \in \mathcal{A}_i\ (i = 1, 2)\}$ dei **rettangoli misurabili (di Ω)**[19].

[19]La chiusura per intersezioni finite è evidente osservato che $(\mathsf{A}_1 \times \mathsf{A}_2) \cap (\mathsf{A}_1' \times \mathsf{A}_2') = (\mathsf{A}_1 \cap \mathsf{A}_1') \times (\mathsf{A}_2 \cap \mathsf{A}_2') \in \mathcal{R}$ per ogni $\mathsf{A}_i, \mathsf{A}_i' \in \mathcal{A}_i$ ($i = 1, 2$).

Il risultato seguente collega la Borel misurabilità rispetto alla σ-algebra prodotto con quelle rispetto alle σ-algebre $\mathcal{A}_1$ e $\mathcal{A}_2$; in particolare, assicura che le restrizioni di una funzione $\mathcal{A}_1 \otimes \mathcal{A}_2$-Borel misurabile su "rette parallele all'asse" Ω_i sono $\mathcal{A}_i$-Borel misurabili ($i = 1, 2$). Per non appesantire l'esposizione, denotiamo con ω_i, A_i elementi generici, rispettivamente, di Ω_i e $\mathcal{A}_i$ ($i = 1, 2$) e assumiamo che i, j rappresentino (limitatamente a questa sezione) elementi *distinti* di $\{1, 2\}$.

Lemma A.5.1 *Sussistono le seguenti proposizioni:*

(i) *Sia* $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. *Allora,* $A(\omega_i) \in \mathcal{A}_j$ *per ogni* ω_i $(i = 1, 2)$;

(ii) *Sia* f *una funzione* $\mathcal{A}_1 \otimes \mathcal{A}_2$*-Borel misurabile. Allora, le restrizioni* $f(\omega_1, \cdot)$, $f(\cdot, \omega_2)$ *sono, rispettivamente,* $\mathcal{A}_2$*-Borel misurabile e* $\mathcal{A}_1$*-Borel misurabile qualunque sia* ω_i $(i = 1, 2)$;

(iii) *Sia* f_i *una funzione* $\mathcal{A}_i$*-Borel misurabile. Allora, la funzione* $(\omega_1, \omega_2) \mapsto f_i(\omega_i)$ *è* $\mathcal{A}_1 \otimes \mathcal{A}_2$*-Borel misurabile* $(i = 1, 2)$.

DIMOSTRAZIONE Osservato preliminarmente che, qualunque sia ω_i ($i = 1, 2$), si ha

$$S^c(\omega_i) = S(\omega_i)^c \tag{A.14}$$

$$\Big(\bigcap_{n\geq 1} S_n\Big)(\omega_i) = \bigcap_{n\geq 1} S_n(\omega_i), \quad \Big(\bigcup_{n\geq 1} S_n\Big)(\omega_i) = \bigcup_{n\geq 1} S_n(\omega_i) \tag{A.15}$$

$$I_{S(\omega_1)}(\cdot) = I_S(\omega_1, \cdot), \qquad I_{S(\omega_2)}(\cdot) = I_S(\cdot, \omega_2) \tag{A.16}$$

per ogni S, S_1, $S_2, \ldots \subseteq \Omega$ e

$$(E_1 \times E_2)(\omega_i) = \begin{cases} E_j & \text{se } \omega_i \in E_i \\ \emptyset & \text{se } \omega_i \notin E_i \end{cases} \tag{A.17}$$

per ogni $E_1 \subseteq \Omega_1$ e $E_2 \subseteq \Omega_2$, consideriamo solo il caso $i = 1$, essendo la dimostrazione relativa a $i = 2$ del tutto analoga.

(i) Dato ω_1, sia $\mathcal{F} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : A(\omega_1) \in \mathcal{A}_2\}$. Allora, per (A.14), (A.15) e (A.17), $\mathcal{F}$ è una σ-algebra includente i rettangoli misurabili. Conseguentemente, $\mathcal{F} \supseteq \sigma(\mathcal{R}) = \mathcal{A}_1 \otimes \mathcal{A}_2$ e quindi $\mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2$.

(ii) Sia intanto f una funzione $\mathcal{A}_1 \otimes \mathcal{A}_2$-semplice. Allora, $f = \sum_{y \in f(\Omega)} y I_{\{f=y\}}$ e quindi, per (A.16), $f(\omega_1, \cdot) = \sum_{y \in f(\Omega)} y I_{\{f=y\}}(\omega_1, \cdot) = \sum_{y \in f(\Omega)} y I_{\{f=y\}(\omega_1)}(\cdot)$. Dunque, $f(\omega_1, \cdot)$ è $\mathcal{A}_2$-Borel misurabile in quanto combinazione lineare, per (i), di funzioni $\mathcal{A}_2$-Borel misurabili. Sia infine f qualsiasi. Considerata una successione $(f_n)_{n\geq 1}$ di funzioni

$\mathcal{A}_1 \otimes \mathcal{A}_2$-semplici tale che $f_n \to f$, si ha $f_n(\omega_1, \cdot) \to f(\omega_1, \cdot)$ e quindi $f(\omega_1, \cdot)$ è $\mathcal{A}_2$-Borel misurabile in quanto limite di funzioni $\mathcal{A}_2$-Borel misurabili.

(iii) Sia $f_1' : (\omega_1, \omega_2) \mapsto f_1(\omega_1)$. Allora, dato un boreliano B della retta reale ampliata, si ha $\{f_1' \in B\} = \{(\omega_1, \omega_2) : f_1'(\omega_1, \omega_2) \in B\} = \{(\omega_1, \omega_2) : f_1(\omega_1) \in B\} = \{f_1 \in B\} \times \Omega_2 \in \mathcal{A}_1 \otimes \mathcal{A}_2$. ∎

Ricordato che la misura m_i è σ-finita, consideriamo ora una partizione $(\mathsf{A}_{in})_{n\geq 1}$ di Ω_i tale che $\mathrm{m}_i(\mathsf{A}_{in}) < +\infty$ per ogni n $(i = 1, 2)$. Allora, posto $A_{nm} = \mathsf{A}_{1n} \times \mathsf{A}_{2m}$ per ogni n e m, otteniamo che $(A_{nm})_{n,m\geq 1}$ è una partizione discreta di Ω formata da rettangoli misurabili. Siamo così in grado di provare il lemma successivo che fornisce la proprietà chiave per la costruzione della misura sulla σ-algebra prodotto verificante le proprietà (a) e (b).

Lemma A.5.2 *Sia $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Allora, $\omega_i \mapsto \mathrm{m}_j(A(\omega_i))$ è una funzione $\mathcal{A}_i$-Borel misurabile $(i = 1, 2)$.*

Dimostrazione Poichè le dimostrazioni per $i = 1$ e $i = 2$ sono simili, sviluppiamo solo quella relativa a $i = 1$. Posto $f_A : \omega_1 \mapsto \mathrm{m}_2(A(\omega_1))$ e considerata, per ogni n e m, l'applicazione $f_A^{(nm)} : \omega_1 \mapsto \mathrm{m}_2((A \cap A_{nm})(\omega_1))$, da (A.15) otteniamo

$$\begin{aligned} f_A(\omega_1) &= \mathrm{m}_2\big(\big(A \cap \bigcup_{n,m\geq 1} A_{nm}\big)(\omega_1)\big) = \mathrm{m}_2\big(\big(\bigcup_{n,m\geq 1} (A \cap A_{nm})\big)(\omega_1)\big) \\ &= \mathrm{m}_2\big(\bigcup_{n,m\geq 1} (A \cap A_{nm})(\omega_1)\big) = \sum_{n,m\geq 1} \mathrm{m}_2((A \cap A_{nm})(\omega_1)) = \sum_{n,m\geq 1} f_A^{(nm)}(\omega_1) \end{aligned}$$

per ogni ω_1. Allora, per il Teorema A.2.5(iv), basta constatare la $\mathcal{A}_1$-Borel misurabilità delle funzioni $f_A^{(nm)}$ per ogni n e m. A tal fine, fissati n, m e tenuto conto del Lemma A.1.4, proviamo che la famiglia $\mathcal{D} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : f_A^{(nm)}$ è $\mathcal{A}_1$-Borel misurabile$\}$ è una classe di Dynkin includente la famiglia $\mathcal{R}$. Supposto $A = \mathsf{A}_1 \times \mathsf{A}_2$, da (A.17) si ha

$$\begin{aligned} f_A^{(nm)}(\omega_1) &= \mathrm{m}_2((A \cap A_{nm})(\omega_1)) = \mathrm{m}_2\big(\big[(\mathsf{A}_1 \cap \mathsf{A}_{1n}) \times (\mathsf{A}_2 \cap \mathsf{A}_{2m})\big](\omega_1)\big) \\ &= \mathrm{m}_2(\mathsf{A}_2 \cap \mathsf{A}_{2m})\, I_{\mathsf{A}_1 \cap \mathsf{A}_{1n}}(\omega_1) \end{aligned}$$

per ogni ω_1. Ne segue, poichè $\mathsf{A}_1 \cap \mathsf{A}_{1n} \in \mathcal{A}_1$, la $\mathcal{A}_1$-Borel misurabilità di $f_A^{(nm)}$ e quindi $A \in \mathcal{D}$.
Sia ora $A \in \mathcal{D}$. Allora, $f_A^{(nm)}$ è una funzione $\mathcal{A}_1$-Borel misurabile. Inoltre, dato ω_1, da (A.14), (A.15) e (A.17) otteniamo

$$\begin{aligned} (A^c \cap A_{nm})(\omega_1) &= A(\omega_1)^c \cap A_{nm}(\omega_1) \\ &= A_{nm}(\omega_1) \setminus [A(\omega_1) \cap A_{nm}(\omega_1)] = \begin{cases} \mathsf{A}_{2m} \setminus (A(\omega_1) \cap \mathsf{A}_{2m}) & \text{se } \omega_1 \in \mathsf{A}_{1n} \\ \emptyset \setminus \emptyset & \text{se } \omega_1 \notin \mathsf{A}_{1n} \end{cases} \end{aligned}$$

e quindi, tramite $\mathrm{m}_2(\mathsf{A}_{2m}) < +\infty$, la monotonia della misura e il Teorema A.1.3(iii),

$$\begin{aligned} f_{A^c}^{(nm)}(\omega_1) &= \mathrm{m}_2\big(\big(A_{2m} \setminus (A(\omega_1) \cap A_{2m})\big) I_{\mathsf{A}_{1n}}(\omega_1)\big) \\ &= \big[\mathrm{m}_2(A_{2m}) - \mathrm{m}_2(A(\omega_1) \cap A_{2m})\big] I_{\mathsf{A}_{1n}}(\omega_1) = \mathrm{m}_2(\mathsf{A}_{2m})\, I_{\mathsf{A}_{1n}}(\omega_1) - f_A^{(nm)}(\omega_1) \end{aligned}$$

ove l'ultima uguaglianza sussiste in forza di (A.17). Ne segue la $\mathcal{A}_1$-Borel misurabilità di $f_{A^c}^{(nm)}$ e quindi $A^c \in \mathcal{D}$.
Sia infine $(A_k)_{k\geq 1}$ una successione disgiunta in $\mathcal{D}$. Allora, $f_{A_k}^{(nm)}$ è una funzione $\mathcal{A}_1$-Borel misurabile qualunque sia $k \geq 1$. Inoltre, per (A.15),

$$\begin{aligned} f_{\bigcup_{k\geq 1} A_k}^{(nm)}(\omega_1) &= \mathrm{m}_2\big((A_{nm} \cap \bigcup_{k\geq 1} A_k)(\omega_1)\big) = \mathrm{m}_2\big(\big(\bigcup_{k\geq 1} (A_k \cap A_{nm})\big)(\omega_1)\big) \\ &= \mathrm{m}_2\big(\bigcup_{k\geq 1} (A_k \cap A_{nm})(\omega_1)\big) = \sum_{k\geq 1} \mathrm{m}_2\big((A_k \cap A_{nm})(\omega_1)\big) = \sum_{k\geq 1} f_{A_k}^{(nm)}(\omega_1) \end{aligned}$$

per ogni ω_1. Dunque, $f_{\bigcup_{k\geq 1} A_k}^{(nm)}$ è $\mathcal{A}_1$-Borel misurabile e quindi $\bigcup_{k\geq 1} A_k \in \mathcal{D}$. ■

Il lemma appena provato consente di considerare sulla σ-algebra prodotto la misura $\mathrm{m}_1 \times \mathrm{m}_2$ così definita:

$$\mathrm{m}_1 \times \mathrm{m}_2(A) = \int_{\Omega_1} \mathrm{m}_2(A(\omega_1))\, \mathrm{m}_1(d\omega_1)$$

per ogni $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, detta **misura prodotto di** m_1 **e** m_2. Per constatare che siamo in presenza di una misura basta osservare che $\mathrm{m}_1 \times \mathrm{m}_2(\emptyset) = 0$ e che, per (A.15) e il teorema d'integrazione per serie,

$$\begin{aligned} \mathrm{m}_1 \times \mathrm{m}_2\Big(\bigcup_{n\geq 1} A_n\Big) &= \int_{\Omega_1} \mathrm{m}_2\Big(\bigcup_{n\geq 1} A_n(\omega_1)\Big)\mathrm{m}_1(d\omega_1) \\ &= \int_{\Omega_1} \sum_{n\geq 1} \mathrm{m}_2(A_n(\omega_1))\ \mathrm{m}_1(d\omega_1) \\ &= \sum_{n\geq 1} \int_{\Omega_1} \mathrm{m}_2(A_n(\omega_1))\, \mathrm{m}_1(d\omega_1) = \sum_{n\geq 1} \mathrm{m}_1 \times \mathrm{m}_2(A_n) \end{aligned}$$

per ogni successione disgiunta $(A_n)_{n\geq 1}$.

Il prossimo teorema assicura che la misura prodotto è l'unica misura sulla σ-algebra prodotto che verifica le proprietà (a), (b) (e pertanto è σ-finita in quanto $\mathrm{m}_1 \times \mathrm{m}_2(A_{nm}) = \mathrm{m}_1(\mathsf{A}_{1n})\, \mathrm{m}_2(\mathsf{A}_{2m}) < +\infty$ per ogni n, m).

Teorema A.5.3 *Risulta:*

(i) $\mathrm{m}_1 \times \mathrm{m}_2(\mathsf{A}_1 \times \mathsf{A}_2) = \mathrm{m}_1(\mathsf{A}_1)\,\mathrm{m}_2(\mathsf{A}_2)$ *per ogni* $\mathsf{A}_1 \times \mathsf{A}_2$*;*

(ii) $\mathrm{m}_1 \times \mathrm{m}_2(A) = \int_{\Omega_2} \mathrm{m}_1(A(\omega_2))\,\mathrm{m}_2(d\omega_2)$ *per ogni* $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$.

Inoltre, qualunque sia la misura m' *sulla* σ*-algebra prodotto tale che* $\mathrm{m}'\big|_{\mathcal{R}} = \mathrm{m}_1 \times \mathrm{m}_2\big|_{\mathcal{R}}$*, riesce* $\mathrm{m}' = \mathrm{m}_1 \times \mathrm{m}_2$.

DIMOSTRAZIONE (i) Da (A.17) otteniamo

$$\begin{aligned}\mathrm{m}_1 \times \mathrm{m}_2(\mathsf{A}_1 \times \mathsf{A}_2) &= \int_{\Omega_1} \mathrm{m}_2\big((\mathsf{A}_1 \times \mathsf{A}_2)(\omega_1)\big)\,\mathrm{m}_1(d\omega_1) = \int_{\Omega_1} \mathrm{m}_2(\mathsf{A}_2)\, I_{\mathsf{A}_1}\,\mathrm{m}_1(d\omega_1)\\ &= \mathrm{m}_1(\mathsf{A}_1)\,\mathrm{m}_2(\mathsf{A}_2).\end{aligned}$$

Proviamo ora l'unicità della misura prodotto. Sia dunque m' una misura sulla σ-algebra prodotto tale che $\mathrm{m}'\big|_{\mathcal{R}} = \mathrm{m}_1 \times \mathrm{m}_2\big|_{\mathcal{R}}$. Allora, ricordato che la famiglia dei rettangoli misurabili è chiusa per intersezioni finite e che, per (i), $\mathrm{m}_1 \times \mathrm{m}_2(A_{nm}) < +\infty$ per ogni n e m, dal criterio standard d'identità otteniamo che m' e la misura prodotto coincidono sulla σ-algebra $\sigma(\mathcal{R}) = \mathcal{A}_1 \otimes \mathcal{A}_2$.

(ii) Per il Lemma A.5.2, possiamo considerare sulla σ-algebra prodotto anche la funzione d'insieme $\mathrm{m}'' : A \to \int_{\Omega_2} \mathrm{m}_1(A(\omega_2))\,\mathrm{m}_2(d\omega_2)$ che, con ragionamenti analoghi a quelli fatti per la misura prodotto, risulta essere una misura coincidente con $\mathrm{m}_1 \times \mathrm{m}_2$ sui rettangoli misurabili. Ne segue, per l'unicità della misura prodotto, $\mathrm{m}'' = \mathrm{m}_1 \times \mathrm{m}_2$. ■

I due teoremi seguenti sono ulteriori "pietre miliari" della teoria dell'integrazione in quanto consentono di ridurre l'**integrale doppio** $\int_\Omega f\, d\mathrm{m}_1 \times \mathrm{m}_2$[20] di una qualsiasi funzione $\mathrm{m}_1 \times \mathrm{m}_2$-sommabile f ad un integrale iterato come pure di invertire l'ordine delle integrazioni successive (analogamente a quanto avviene, nel caso dell'integrale doppio di Riemann, per una funzione continua definita su un rettangolo $[a, b] \times [c, d] \subset \mathbb{R}^2$).

Teorema A.5.4 (di Tonelli) *Sia* f *una funzione* $\mathcal{A}_1 \otimes \mathcal{A}_2$*-Borel misurabile non negativa. Allora, per ogni* A_1 *e* A_2*, le funzioni* $\omega_1 \to \int_{\mathsf{A}_2} f(\omega_1, \cdot)\, d\mathrm{m}_2$ *e* $\omega_2 \to \int_{\mathsf{A}_1} f(\cdot, \omega_2)\, d\mathrm{m}_1$ *sono, rispettivamente,* $\mathcal{A}_1$*-Borel misurabile e* $\mathcal{A}_2$*-Borel misurabile; inoltre, si ha*

$$\begin{aligned}\int_{\mathsf{A}_1 \times \mathsf{A}_2} f\, d\mathrm{m}_1 \times \mathrm{m}_2 &= \int_{\mathsf{A}_1} \Big(\int_{\mathsf{A}_2} f(\omega_1, \cdot)\, d\mathrm{m}_2 \Big)\, \mathrm{m}_1(d\omega_1)\\ &= \int_{\mathsf{A}_2} \Big(\int_{\mathsf{A}_1} f(\cdot, \omega_2)\, d\mathrm{m}_1 \Big)\, \mathrm{m}_2(d\omega_2).\end{aligned}$$

[20] Indicato anche con la notazione $\int_\Omega f(\omega_1, \omega_2)\, \mathrm{m}_1 \times \mathrm{m}_2(d\omega_1 \times d\omega_2)$.

Dimostrazione Sia intanto $\mathsf{A}_i = \Omega_i$ $(i = 1, 2)$. Osservato che, per il Lemma A.5.1(ii), possiamo considerare le funzioni $J_1 : \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \cdot)\, d\mathrm{m}_2$ e $J_2 : \omega_2 \mapsto \int_{\Omega_1} f(\cdot, \omega_2)\, d\mathrm{m}_1$, iniziamo col supporre $f = I_A$. Allora, da (A.16), otteniamo

$$J_i(\omega_i) = \int_{\Omega_j} I_{A(\omega_i)}\, d\mathrm{m}_j = \mathrm{m}_j(A(\omega_i))$$

per ogni ω_i $(i = 1, 2)$. Ne segue, tramite il Lemma A.5.2, che J_i è $\mathcal{A}_i$-Borel misurabile $(i = 1, 2)$. Inoltre, per definizione,

$$\mathrm{m}_1 \times \mathrm{m}_2(A) = \int_{\Omega_1} \mathrm{m}_2(A(\omega_1))\, \mathrm{m}_1(d\omega_1) = \int_{\Omega_1} J_1(\omega_1)\, \mathrm{m}_1(d\omega_1)$$

e, per il Teorema A.5.3(ii),

$$\mathrm{m}_1 \times \mathrm{m}_2(A) = \int_{\Omega_2} \mathrm{m}_1(A(\omega_2))\, \mathrm{m}_2(d\omega_2) = \int_{\Omega_2} J_2(\omega_2)\, \mathrm{m}_2(d\omega_2).$$

Risulta dunque $\int_\Omega f\, d\mathrm{m}_1 \times \mathrm{m}_2 = \int_{\Omega_1} J_1(\omega_1)\, \mathrm{m}_1(d\omega_1) = \int_{\Omega_2} J_2(\omega_2)\, \mathrm{m}_2(d\omega_2)$.

Poichè da questo punto in poi le dimostrazioni per $i = 1$ e $i = 2$ sono simili, sviluppiamo solo quella relativa al caso $i = 1$.

Sia ora f una funzione $\mathcal{A}_1 \otimes \mathcal{A}_2$-semplice non negativa. Dalla linearità dell'integrale risulta

$$J_1(\omega_1) = \int_{\Omega_2} \sum_{y \in f(\Omega)} y I_{\{f=y\}}(\omega_1, \cdot)\; d\mathrm{m}_2 = \sum_{y \in f(\Omega)} y \int_{\Omega_2} I_{\{f=y\}}(\omega_1, \cdot)\, d\mathrm{m}_2$$

per ogni ω_1. Ne segue la $\mathcal{A}_1$-Borel misurabilità in quanto J_1 è combinazione lineare di funzioni $\mathcal{A}_1$-Borel misurabili. Sempre per la linearità dell'integrale si ha inoltre

$$\begin{aligned}
\int_\Omega f\, d\mathrm{m}_1 \times \mathrm{m}_2 &= \sum_{y \in f(\Omega)} y \int_\Omega I_{\{f=y\}}\, d\mathrm{m}_1 \times \mathrm{m}_2 \\
&= \sum_{y \in f(\Omega)} y \int_{\Omega_1} \Big(\int_{\Omega_2} I_{\{f=y\}}(\omega_1, \cdot)\, d\mathrm{m}_2 \Big)\, \mathrm{m}_1(d\omega_1) \\
&= \int_{\Omega_1} \Big(\sum_{y \in f(\Omega)} y \int_{\Omega_2} I_{\{f=y\}}(\omega_1, \cdot)\, d\mathrm{m}_2 \Big)\, \mathrm{m}_1(d\omega_1) = \int_{\Omega_1} J_1(\omega_1)\, \mathrm{m}_1(d\omega_1).
\end{aligned}$$

Sia infine f qualsiasi. Considerata una successione $(f_n)_{n \geq 1}$ di funzioni $\mathcal{A}_1 \otimes \mathcal{A}_2$-semplici non negative tale che $f_n \uparrow f$, si ha $0 \leq f_n(\omega_1, \cdot) \uparrow f(\omega_1, \cdot)$ e quindi, per il teorema della convergenza monotona, $0 \leq \int_{\Omega_2} f_n(\omega_1, \cdot)\, d\mathrm{m}_2 \uparrow \int_{\Omega_2} f(\omega_1, \cdot)\, d\mathrm{m}_2 = J_1(\omega_1)$. Pertanto, essendo ω_1 arbitrario, J_1 è $\mathcal{A}_1$-Borel misurabile in quanto limite di funzioni $\mathcal{A}_1$-Borel misurabili. Sempre per il teorema della convergenza monotona otteniamo inoltre

$$\begin{aligned}
\int_\Omega f\, d\mathrm{m}_1 \times \mathrm{m}_2 &= \lim_{n \to +\infty} \int_\Omega f_n\, d\mathrm{m}_1 \times \mathrm{m}_2 \\
&= \lim_{n \to +\infty} \int_{\Omega_1} \Big(\int_{\Omega_2} f_n(\omega_1, \cdot)\, d\mathrm{m}_2 \Big)\, \mathrm{m}_1(d\omega_1) = \int_{\Omega_1} J_1(\omega_1)\, \mathrm{m}_1(d\omega_1).
\end{aligned}$$

Siano ora A_1, A_2 arbitrari. Dall'uguaglianza $I_{\mathsf{A}'_1\times\mathsf{A}'_2}(\omega_1,\omega_2) = I_{\mathsf{A}'_1}(\omega_1)I_{\mathsf{A}'_2}(\omega_2)$ valida per ogni A'_i e ω_i $(i=1,2)$, risulta che $I_{\mathsf{A}_2}f = I_{\Omega_1\times\mathsf{A}_2}f$ e $I_{\mathsf{A}_1}f = I_{\mathsf{A}_1\times\Omega_2}f$ sono funzioni $\mathcal{A}_1\otimes\mathcal{A}_2$-Borel misurabili. Allora, per la prima parte della dimostrazione, le funzioni $\omega_1\to\int_{\mathsf{A}_2} f(\omega_1,\cdot)\,d\mathrm{m}_2 = \int_{\Omega_2} I_{\mathsf{A}_2}f(\omega_1,\cdot)\,d\mathrm{m}_2$ e $\omega_2\to\int_{\mathsf{A}_1} f(\cdot,\omega_2)\,d\mathrm{m}_1 = \int_{\Omega_1} I_{\mathsf{A}_1}f(\cdot,\omega_2)\,d\mathrm{m}_1$ sono, rispettivamente, $\mathcal{A}_1$-Borel misurabile e $\mathcal{A}_2$-Borel misurabile; inoltre, si ha

$$\begin{aligned}\int_{\mathsf{A}_1\times\mathsf{A}_2} f\,d\mathrm{m}_1\times\mathrm{m}_2 &= \int_{\Omega} I_{\mathsf{A}_1\times\mathsf{A}_2}f\,d\mathrm{m}_1\times\mathrm{m}_2\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2} I_{\mathsf{A}_1\times\mathsf{A}_2}(\omega_1,\cdot)f(\omega_1,\cdot)\,d\mathrm{m}_2\Big)\,\mathrm{m}_1(d\omega_1)\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2} I_{\mathsf{A}_2}f(\omega_1,\cdot)\,d\mathrm{m}_2\Big)\,I_{\mathsf{A}_1}(\omega_1)\,\mathrm{m}_1(d\omega_1)\\ &= \int_{\mathsf{A}_1}\Big(\int_{\mathsf{A}_2} f(\omega_1,\cdot)\,d\mathrm{m}_2\Big)\,\mathrm{m}_1(d\omega_1)\end{aligned}$$

e, in modo analogo, $\int_{\mathsf{A}_1\times\mathsf{A}_2} f\,d\mathrm{m}_1\times\mathrm{m}_2 = \int_{\mathsf{A}_2}\big(\int_{\mathsf{A}_1} f(\cdot,\omega_2)\,d\mathrm{m}_1\big)\,\mathrm{m}_2(d\omega_2)$. ■

Teorema A.5.5 (di Fubini) *Sia f una funzione $\mathrm{m}_1\times\mathrm{m}_2$-sommabile. Esistono allora due insiemi*

$$\begin{aligned}\mathsf{A}_1^{(f)} &\subseteq \{\omega_1 : f(\omega_1,\cdot)\text{ è }\mathrm{m}_2\text{-sommabile}\}\\ \mathsf{A}_2^{(f)} &\subseteq \{\omega_2 : f(\cdot,\omega_2)\text{ è }\mathrm{m}_1\text{-sommabile}\}\end{aligned}$$

tali che $\mathsf{A}_i^{(f)}\in\mathcal{A}_i$ e $\mathrm{m}_i\big((\mathsf{A}_i^{(f)})^c\big)=0$ $(i=1,2)$. Inoltre, per ogni A_i $(i=1,2)$, le funzioni:

$$g_1:\omega_1\mapsto\begin{cases}\int_{\mathsf{A}_2} f(\omega_1,\cdot)\,d\mathrm{m}_2 & \textit{se } \omega_1\in\mathsf{A}_1^{(f)}\\ 0 & \textit{se } \omega_1\notin\mathsf{A}_1^{(f)}\end{cases}$$
$$g_2:\omega_2\mapsto\begin{cases}\int_{\mathsf{A}_1} f(\cdot,\omega_2)\,d\mathrm{m}_1 & \textit{se } \omega_2\in\mathsf{A}_2^{(f)}\\ 0 & \textit{se } \omega_2\notin\mathsf{A}_2^{(f)}\end{cases}$$

sono, rispettivamente, $\mathcal{A}_1$-Borel misurabile e $\mathcal{A}_2$-Borel misurabile e tali che $\int_{\mathsf{A}_1\times\mathsf{A}_2} f\,d\mathrm{m}_1\times\mathrm{m}_2 = \int_{\mathsf{A}_1} g_1\,d\mathrm{m}_1 = \int_{\mathsf{A}_2} g_2\,d\mathrm{m}_2$.

Dimostrazione Essendo le dimostrazioni per $i=1$ e $i=2$ simili, sviluppiamo solo quella relativa al caso $i=1$. Poichè f è $\mathrm{m}_1\times\mathrm{m}_2$-sommabile, gli integrali delle sue parti positiva e negativa non possono essere entrambi infiniti. Supponiamo quindi che sia, ad esempio, finito l'integrale $\int_\Omega f^-\,d\mathrm{m}_1\times\mathrm{m}_2$. Allora, per il teorema di Tonelli,

$$\int_{\Omega_1}\Big(\int_{\Omega_2} f^-(\omega_1,\cdot)\,d\mathrm{m}_2\Big)\mathrm{m}_1(d\omega_1) = \int_\Omega f^-\,d\mathrm{m}_1\times\mathrm{m}_2 < +\infty$$

e quindi, per il Teorema A.3.6(vii), esiste un elemento $\mathsf{A}_1^{(f)}$ di $\mathcal{A}_1$ tale che $\mathrm{m}_1((\mathsf{A}_1^{(f)})^c) = 0$ e $\int_{\Omega_2} f(\omega_1, \cdot)^- \, d\mathrm{m}_2 = \int_{\Omega_2} f^-(\omega_1, \cdot) \, d\mathrm{m}_2 < +\infty$ per ogni $\omega_1 \in \mathsf{A}_1^{(f)}$. Conseguentemente, $f(\omega_1, \cdot)$ è una funzione m_2-sommabile per ogni $\omega_1 \in \mathsf{A}_1^{(f)}$.

Sempre per il teorema di Tonelli, $\omega_1 \to \int_{\mathsf{A}_2} f(\omega_1, \cdot)^+ d\mathrm{m}_2$, $\omega_1 \to \int_{\mathsf{A}_2} f(\omega_1, \cdot)^- d\mathrm{m}_2$ sono funzioni $\mathcal{A}_1$-Borel misurabili. Allora, per il Lemma A.2.1(i), le loro restrizioni su $\mathsf{A}_1^{(f)}$ sono $\mathcal{A}_1 \cap \mathsf{A}_1^{(f)}$-Borel misurabili e quindi lo è pure la loro differenza $g_1\big|_{\mathsf{A}_1^{(f)}}$. Pertanto, per il Lemma A.2.1(ii), g_1 è una funzione $\mathcal{A}_1$-Borel misurabile. Inoltre, dalla linearità dell'integrale e dai teoremi di Tonelli e A.3.4(iv) otteniamo

$$\begin{aligned}
\int_{\mathsf{A}_1 \times \mathsf{A}_2} f \, d\mathrm{m}_1 \times \mathrm{m}_2 &= \int_{\mathsf{A}_1 \times \mathsf{A}_2} f^+ \, d\mathrm{m}_1 \times \mathrm{m}_2 - \int_{\mathsf{A}_1 \times \mathsf{A}_2} f^- \, d\mathrm{m}_1 \times \mathrm{m}_2 \\
&= \int_{\mathsf{A}_1} \Big(\int_{\mathsf{A}_2} f^+(\omega_1, \cdot) \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) - \int_{\mathsf{A}_1} \Big(\int_{\mathsf{A}_2} f^-(\omega_1, \cdot) \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) \\
&= \int_{\mathsf{A}_1 \cap \mathsf{A}_1^{(f)}} \Big(\int_{\mathsf{A}_2} f(\omega_1, \cdot)^+ \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) \\
&\qquad - \int_{\mathsf{A}_1 \cap \mathsf{A}_1^{(f)}} \Big(\int_{\mathsf{A}_2} f(\omega_1, \cdot)^- \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) \\
&= \int_{\mathsf{A}_1 \cap \mathsf{A}_1^{(f)}} \Big(\int_{\mathsf{A}_2} f(\omega_1, \cdot)^+ \, d\mathrm{m}_2 - \int_{\mathsf{A}_2} f(\omega_1, \cdot)^- \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) \\
&= \int_{\mathsf{A}_1 \cap \mathsf{A}_1^{(f)}} \Big(\int_{\mathsf{A}_2} f(\omega_1, \cdot) \, d\mathrm{m}_2 \Big) \mathrm{m}_1(d\omega_1) = \int_{\mathsf{A}_1 \cap \mathsf{A}_1^{(f)}} g_1 \, d\mathrm{m}_1 = \int_{\mathsf{A}_1} g_1 \, d\mathrm{m}_1.
\end{aligned}$$

La dimostrazione è così conclusa. ■

La nozione di misura prodotto ed i relativi teoremi di Tonelli e Fubini possono essere estesi al caso di più di due fattori. Infatti, siano $\mathcal{A}_h$ una σ-algebra su un insieme $\Omega_h \neq \emptyset$ e m_h una misura σ-finita su $\mathcal{A}_h$ $(h = 1, \ldots, m)$. Considerata, analogamente al caso $m = 2$, la famiglia chiusa per intersezioni finite $\mathcal{R} = \{\mathsf{A}_1 \times \cdots \times \mathsf{A}_m : \mathsf{A}_h \in \mathcal{A}_h \, (h = 1, \ldots, m)\}$ dei **rettangoli misurabili (di $\Omega = \Omega_1 \times \cdots \times \Omega_m$)**, la **$\boldsymbol{\sigma}$-algebra prodotto (di $\mathcal{A}_1, \ldots, \mathcal{A}_m$)** è la σ-algebra $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_m$ generata da $\mathcal{R}$. Sfruttando i risultati ottenuti nel caso di due fattori e procedendo per induzione su m è allora possibile constatare che esiste una sola misura $\mathrm{m}_1 \times \cdots \times \mathrm{m}_m$ sulla σ-algebra prodotto - detta **misura prodotto (di** $\mathrm{m}_1, \ldots, \mathrm{m}_m$**)** - tale che $\mathrm{m}_1 \times \cdots \times \mathrm{m}_m(\mathsf{A}_1 \times \cdots \times \mathsf{A}_m) = \prod_{h=1}^m \mathrm{m}_h(\mathsf{A}_h)$ per ogni rettangolo misurabile $\mathsf{A}_1 \times \cdots \times \mathsf{A}_m$. Inoltre, è anche possibile provare la seguente generalizzazione dei teoremi di Tonelli e Fubini che permette di ridurre l'**integrale multiplo** $\int_\Omega f \, d\mathrm{m}_1 \times \cdots \times \mathrm{m}_m$[21] di una qualsiasi funzione $\mathrm{m}_1 \times \cdots \times \mathrm{m}_m$-sommabile f ad un integrale iterato

[21]Indicato anche con la notazione $\int_\Omega f(\omega_1, \ldots, \omega_m) \; \mathrm{m}_1 \times \cdots \times \mathrm{m}_m(d\omega_1 \times \cdots \times d\omega_m)$.

come pure di invertire l'ordine delle integrazioni successive[22].

Teorema A.5.6 *Posto* $\mathrm{m} = \mathrm{m}_1 \times \cdots \times \mathrm{m}_m$*, sia* f *una funzione* m*-sommabile. Allora, data una permutazione* $h_1, \ldots, h_m$ *dell'insieme* $\{1, \ldots, m\}$*, si ha*

$$\int_\Omega f \, d\mathrm{m} = \int_{\Omega_{h_1}} \Big(\cdots \Big(\int_{\Omega_{h_m}} f(\omega_1, \ldots, \omega_m) \, \mathrm{m}_{h_m}(d\omega_{h_m}) \Big) \cdots \Big) \, \mathrm{m}_{h_1}(d\omega_{h_1}),$$

dove, nel caso che non sia $f \geq 0$*, ogni integrale*

$$\int_{\Omega_{h_k}} \Big(\cdots \Big(\int_{\Omega_{h_m}} f(\omega_1, \ldots, \omega_m) \, \mathrm{m}_{\mathrm{h_m}}(d\omega_{h_m}) \Big) \cdots \Big) \, \mathrm{m}_{h_k}(d\omega_{h_k}) \qquad (k > 1)$$

esiste per ogni $(\omega_{h_1}, \ldots, \omega_{h_{k-1}})$ *non appartenente ad un insieme* $\mathrm{m}_{h_{k-1}}$*-trascurabile e viene esteso (per effettuare l'integrazione successiva relativa alla misura* $\mathrm{m}_{h_{k-1}}$*) su tale insieme dandogli valore zero.*

Tramite la σ-algebra e la misura prodotto possiamo ora introdurre le versioni m-dimensionali della σ-algebra di Borel e della misura di Lebesgue[23].

Esempio A.5.7 Considerato lo spazio numerico delle m-ple reali ($m \geq 2$), la **σ-algebra di Borel (di $\mathbb{R}^m$)** è la σ-algebra prodotto: $\mathcal{B}^{(m)} = \overbrace{\mathcal{B} \otimes \cdots \otimes \mathcal{B}}^{m \text{ volte}}$ e la **misura di Lebesgue m-dimensionale** è la misura prodotto: $\lambda^{(m)} = \overbrace{\lambda \times \cdots \times \lambda}^{m \text{ volte}}$. Inoltre, i **boreliani** di $\mathbb{R}^m$ sono gli elementi di $\mathcal{B}^{(m)}$.

Naturalmente, la **σ-algebra di Borel di** $(\mathbb{R}^*)^m$ è, per analogia, la σ-algebra prodotto: $(\mathcal{B}^*)^{(m)} = \overbrace{\mathcal{B}^* \otimes \cdots \otimes \mathcal{B}^*}^{m \text{ volte}} \supset \mathcal{B}^{(m)}$ costituita dai **boreliani** di $(\mathbb{R}^*)^m$.

Analogamente al caso unidimensionale, $\mathcal{B}^{(m)}$ può essere vista come la σ-algebra generata dalla famiglia $\mathcal{U}^{(m)}$ degli insiemi aperti di m-ple. Per provarlo, iniziamo col verificare l'inclusione $\mathcal{U}^{(m)} \subseteq \mathcal{B}^{(m)}$. Dato $W \in \mathcal{U}^{(m)}$, per ogni suo punto $\mathbf{w}$ possiamo trovare m intervalli di estremi razionali $]q_{h1}^{(\mathbf{w})}, q_{h2}^{(\mathbf{w})}]$ $(h = 1, \ldots, m)$ tali che

$$\mathbf{w} \in]q_{11}^{(\mathbf{w})}, q_{12}^{(\mathbf{w})}] \times \cdots \times]q_{m1}^{(\mathbf{w})}, q_{m2}^{(\mathbf{w})}] \subseteq W; \tag{A.18}$$

[22] Per le relative dimostrazioni si veda la sezione 2.6 del testo di Ash.

[23] La σ-algebra di Borel m-dimensionale è, senza alcun dubbio, una delle più importanti σ-algebre che si possono considerare sullo spazio numerico $\mathbb{R}^m$, in quanto vi appartiene sostanzialmente ogni insieme di m-ple che sia di qualche interesse per le applicazioni. Inoltre, la misura di Lebesgue m-dimensionale può essere vista come una naturale estensione ad insiemi più complessi delle nozioni elementari di area ($m = 2$) e di volume ($m = 3$).

dunque W è unione discreta di elementi di $\mathcal{B}^{(m)}$ e quindi appartiene a $\mathcal{B}^m$. A questo punto basta provare che, data una σ-algebra $\mathcal{G}$ su $\mathbb{R}^m$ tale che $\mathcal{U}^{(m)} \subseteq \mathcal{G}$, riesce $\mathcal{B}^{(m)} \subseteq \mathcal{G}$. Scegliendo come famiglia di generatori di $\mathcal{B}$ quella degli insiemi aperti e ricorrendo alla nota 8 di p. 123, otteniamo che ogni funzione continua $f : \mathbb{R}^m \mapsto \mathbb{R}$ è $\mathcal{G}$-Borel misurabile. Lo è quindi, in particolare, la proiezione h-sima $p_h : (x_1, \dots, x_m) \mapsto x_h$ $(h = 1, \dots, m)$. Considerati allora dei boreliani qualsiasi $B_1, \dots, B_m$ di $\mathbb{R}$, si ha $p_h^{-1}(B_i) \in \mathcal{G}$ $(h = 1, \dots, m)$ e quindi $B_1 \times \cdots \times B_m = p_1^{-1}(B_1) \cap \cdots \cap p_m^{-1}(B_m) \in \mathcal{G}$. Ne segue, per definizione, $\mathcal{B}^{(m)} \subseteq \mathcal{G}$.

Un altro interessante sistema di generatori di $\mathcal{B}^{(m)}$ si ottiene considerando gli insiemi $S_{x_1 \dots x_m} =]-\infty, x_1] \times \cdots \times]-\infty, x_m]$ al variare di $x_1, \dots, x_m \in \mathbb{R}$ in tutti i modi possibili. Infatti, posto $\mathcal{S}^{(m)} = \{S_{x_1 \dots x_m} : x_1, \dots, x_m \in \mathbb{R}\}$, si ha $\sigma(\mathcal{S}^{(m)}) \subseteq \mathcal{B}^{(m)}$. Per provare l'inclusione opposta, basta evidentemente constatare che ogni insieme aperto W di $\mathbb{R}^m$ appartiene a $\sigma(\mathcal{S}^{(m)})$. Ora, per (A.18), W è unione discreta di rettangoli del tipo $]q_{11}, q_{12}] \times \cdots \times]q_{m1}, q_{m2}]$. Osservato che, indicata con $\mathbf{h} = (h_1, \dots, h_m)$ la generica m-pla di $\{1, 2\}^m$, si ha

$$]q_{11}, q_{12}] \times \cdots \times]q_{m1}, q_{m2}] = S_{q_{12} \dots q_{m2}} \setminus \bigcup_{\mathbf{h} \in \{\mathbf{h} : \exists k (h_k \neq 2)\}} S_{q_{1h_1} \dots q_{mh_m}} \in \sigma(\mathcal{F}),$$

possiamo concludere che $W \in \sigma(\mathcal{S}^{(m)})$.

Osserviamo infine che, per quanto concerne la Borel misurabilità rispetto a $(\mathcal{B}^*)^{(m)}$, rimane valido il Teorema A.2.3 (ovviamente, per $S \in (\mathcal{B}^*)^m$ e $f : \Omega \mapsto (\mathbb{R}^*)^m$), come facilmente si constata ripercorrendone la dimostrazione. ▲

Appendice B

Richiami di teoria della probabilità

B.1 Nozioni e risultati di base

B.1.1 Eventi, variabili aleatorie, enti aleatori

Nell'impostazione logica degli eventi è usuale considerare, a fronte di una famiglia esaustiva Ω di eventi non impossibili a due a due incompatibili (*partizione dell'evento certo*), la famiglia $\mathcal{E}(\Omega)$ degli eventi *logicamente dipendenti* (dalla partizione), cioè degli eventi di valore logico determinato una volta noto il valore logico di ogni elemento (*caso elementare*) della partizione. Poichè tali eventi possono essere identificati con gli eventi ottenibili come disgiunzione (eventualmente infinita) di casi elementari, è possibile instaurare una corrispondenza biunivoca tra $\mathcal{E}(\Omega)$ e 2^{Ω} che muta le operazioni logiche di negazione, congiunzione, disgiunzione, rispettivamente, in quelle insiemistiche di complementazione, intersezione, unione e la relazione logica di implicazione in quella insiemistica di inclusione[1]. Trova così giustificazione l'usuale identificazione della *logica degli eventi* (logicamente dipendenti) con *l'algebra degli insiemi* (inclusi in Ω). In questo contesto, prendendo spunto da una molteplicità di casi concreti, si identificano gli eventi "di interesse" con gli insiemi appartenenti ad una σ-algebra di riferimento $\mathcal{A}$ su Ω. Inoltre,

[1] Per un'ampia e approfondita analisi dell'impostazione logica degli eventi si veda Crisma, L., Introduzione alla teoria delle probabilità coerenti, Volume 1, EUT Edizioni Università di Trieste (2006), capitoli 1,2 e 3.

per quanto riguarda le valutazioni di probabilità relative agli eventi "di interesse", si suppone che ad ogni elemento $A \in \mathcal{A}$ sia possibile assegnare una probabilità $\Pr(A)$ in modo tale che la funzione d'insieme $\Pr : \mathcal{A} \mapsto [0, 1]$ così ottenuta sia una misura su $\mathcal{A}$.

Passando a considerare le applicazioni aleatorie X di dominio la partizione Ω e determinazioni possibili in un insieme $\mathfrak{X}$, per individuare i relativi eventi "di interesse" si procede di solito scegliendo una σ-algebra $\aleph$ su $\mathfrak{X}$ e identificandoli con gli insiemi del tipo: $\{X \in \mathsf{A}\}$ con $\mathsf{A} \in \aleph$[2]. Se poi si desidera assegnare, sulla base delle conoscenze probabilistiche espresse dalla probabilità Pr, anche a questi eventi un valore di probabilità basta richiedere che $\{X \in \mathsf{A}\} \in \mathcal{A}$ per ogni elemento $\mathsf{A} \in \aleph$ e ricorrere alla misura immagine $\Pr_X$.

Da un punto di vista astratto, le considerazioni appena svolte suggeriscono di adottare - come quadro concettuale di riferimento per lo sviluppo della teoria delle probabilità - la seguente impostazione (assiomatica) introdotta nel 1933 da Andrej N. Kolmogorov[3]. Considerata, come nell'appendice precedente, una σ-algebra $\mathcal{A}$ su un insieme ambiente $\Omega \neq \emptyset$, chiamiamo:
- **caso elementare** (**di** Ω) ogni elemento di Ω;
- **evento** (**di** Ω) ogni elemento di $\mathcal{A}$;
- **probabilità** (**sugli eventi** di Ω) ogni misura P su $\mathcal{A}$ tale che $\mathrm{P}(\Omega) = 1$;
- **variabile aleatoria** (**su** Ω) ogni funzione $\mathcal{A}$-Borel misurabile a valori nella retta reale;
- **variabile aleatoria estesa** (**su** Ω) ogni funzione $\mathcal{A}$-Borel misurabile a valori nella retta reale ampliata;
- **ente aleatorio** (**su** Ω) **a valori in** $\mathfrak{X}$ ogni applicazione $(\mathcal{A}, \aleph)$-misurabile.

Precisiamo che, dato un qualsiasi ente aleatorio X, useremo le notazioni $\mathrm{P}(X \in \mathsf{A})$, $\mathrm{P}(X \leq x)$, $\mathrm{P}(X - Y \geq 0)$, etc. al posto delle $\mathrm{P}(\{X \in \mathsf{A}\})$,

[2]Precisiamo che il termine **aleatorio** è sempre inteso in senso relativo associandolo esclusivamente con ciò che è *non conosciuto (incerto) in un dato stato di conoscenza per carenza d'informazione*. In questo senso, ogni applicazione di dominio una partizione dell'evento certo (avente almeno due casi elementari) è un particolare esempio di entità aleatoria, in quanto rappresenta un elemento ben definito (del codominio) ma non noto fintanto che non sia conosciuto quale caso elementare è quello vero. D'altra parte, limitarsi a considerare solo applicazioni aleatorie non lede la generalità del discorso in quanto ogni entità aleatoria Y può essere descritta tramite un'applicazione aleatoria; basta considerare, indicato con $\mathfrak{X}$ l'insieme delle sue possibili determinazioni, la partizione dell'evento certo costituita dai casi elementari $\omega_x =$ "$Y = x$" ($x \in \mathfrak{X}$) e l'applicazione aleatoria $X : \omega_x \mapsto x$.

[3]Impostazione che, nonostante diverse critiche che si possono formulare nei suoi confronti, è tuttora la base più "popolare" per la costruzione matematica del calcolo delle probabilità.

$P(\{X \leq x\})$, $P(\{X - Y \geq 0\})$, etc.; inoltre, abbrevieremo di solito con "v.a." il nome "variabile aleatoria" inteso sia al singolare che al plurale.

B.1.2 Legge e densità di un ente aleatorio

Dato un ente aleatorio X a valori in $\mathfrak{X}$, la **legge** (o **distribuzione**) di X è la probabilità immagine P_X di P mediante X. Conseguentemente, la legge di X associa ad ogni $\mathsf{A} \in \aleph$ la probabilità $P(X \in \mathsf{A})$ che l'ente aleatorio X assuma, come determinazione, un elemento dell'insieme A.

Nel caso particolare che X sia una v.a., considereremo anche la **funzione di ripartizione** di X, cioè la funzione $F_X : \mathbb{R} \mapsto [0,1]$ così definita: $F_X(x) = P(X \leq x) = P_X(]-\infty, x])$ per ogni numero reale x. Proviamo ora che la funzione di ripartizione è una funzione non decrescente continua a destra che caratterizza la legge della variabile aleatoria[4]. A tal fine, conviene porre $S_x =]-\infty, x]$ per ogni numero reale x e osservare che la famiglia $\mathcal{S} = \{S_x : x \in \mathbb{R}\}$ è chiusa per intersezioni finite, $\mathbb{R} = \bigcup_{n \geq 1} S_n$ e $\mathcal{B} = \sigma(\mathcal{S})$.

Teorema B.1.1 *Sia X una v.a.. Riesce allora:*

(i) $F_X(x) \leq F_X(x')$, *se* $x \leq x'$;

(ii) $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0$;

(iii) $F_X(+\infty) = \lim\limits_{x \to +\infty} F_X(x) = 1$;

(iv) $F_X(x^+) = \lim\limits_{t \to x^+} F_X(t) = F_X(x)$;

(v) $F_X(x^-) = \lim\limits_{t \to x^-} F_X(t) = P(X < x) = P_X(]-\infty, x[)$;

(vi) $F_X(x^+) - F_X(x^-) = P(X = x)$.

Inoltre, se m' *è una misura su* $\mathcal{B}$ *tale che* $m'(]-\infty, x]) = F_X(x)$ *per ogni numero reale* x*, si ha* $m' = P_X$.

[4]Consentendo così di scegliere, come avviene soprattutto nelle applicazioni, la notazione $\int_B f(x)\, dF_X(x)$ al posto della $\int_B f\, dP_X$. Scelta peraltro infelice poichè, essendo già adottata per indicare l'usuale integrale di Riemann-Stieltjes, potrebbe indurre a pericolosi fraintendimenti (se non è ben precisato il campo di applicazione).

DIMOSTRAZIONE (i) Conseguenza immediata della monotonia della misura.

(ii) $\div$ (v) Conseguenza immediata della continuità della misura, una volta notato che la successione:
- $(S_{-n})_{n\geq 1}$ è non crescente e si ha $\bigcap_{n\geq 1} S_{-n} = \emptyset$;
- $(S_n)_{n\geq 1}$ è non decrescente e si ha $\bigcup_{n\geq 1} S_n = \mathbb{R}$;
- $(S_{x+\frac{1}{n}})_{n\geq 1}$ è non crescente e si ha $\bigcap_{n\geq 1} S_{x+\frac{1}{n}} = S_x$;
- $(S_{x-\frac{1}{n}})_{n\geq 1}$ è non decrescente e si ha $\bigcap_{n\geq 1} S_{x-\frac{1}{n}} =]-\infty, x[$.

(vi) Conseguenza immediata di (iv), (v) e del Teorema A.1.3(iii).

Passando all'ultima parte della tesi, per ipotesi $\mathrm{m}'\big|_{\mathcal{S}} = \mathrm{P}_X\big|_{\mathcal{S}}$. Ne segue, tramite il criterio standard d'identità A.1.5 e l'Esempio A.1.1(iv), $\mathrm{m}' = \mathrm{P}_X$. ■

Ritornando al caso generale, supponiamo che sulla σ-algebra $\aleph$ sia definita una misura σ-finita di riferimento μ. Chiamiamo allora μ-**densità** di X ogni funzione $\aleph$-Borel misurabile $f : \mathfrak{X} \mapsto [0, +\infty]$ tale che $\mathrm{P}_X = f * \mu$. Conseguentemente, ogni μ-densità f di X è definita a meno di insiemi μ-trascurabili (teoremi A.3.6(ii) e A.4.4), è finita quasi ovunque (Teorema A.3.6(vii)) e risulta $\mathrm{P}(X \in \mathsf{A}) = \int_{\mathsf{A}} f\, d\mu$ per ogni $\mathsf{A} \in \aleph$. Inoltre, per il Teorema A.4.3 e per quello della misura immagine A.4.1, sussiste (come facilmente si verifica) il risultato seguente che consente, in particolare, di computare integrali riguardanti trasformate di v.a.(rispetto alla probabilità P) ricorrendo al calcolo (in generale più "maneggevole") di integrali di funzioni su $\mathbb{R}$ rispetto a misure definite sui boreliani della retta reale[5].

Teorema B.1.2 (fondamentale del calcolo delle probabilità) *Sussistono le seguenti proposizioni:*

(i) Sia $g : \mathfrak{X} \mapsto \mathbb{R}^*$ una funzione sommabile rispetto alla legge di X. Riesce allora

$$\int_{\{X\in\mathsf{A}\}} g(X)\, d\mathrm{P} = \begin{cases} \int_{\mathsf{A}} g\, d\mathrm{P}_X \\ \int_{\mathsf{A}} g\, f\, d\mu & \textit{se } f \textit{ è una } \mu\textit{-densità di } X \end{cases}$$

per ogni $\mathsf{A} \in \aleph$;

(ii) Siano $\tau : \mathfrak{X} \mapsto \Omega'$ un'applicazione $(\aleph, \mathcal{A}')$-misurabile e $\tilde{g} : \Omega' \mapsto \mathbb{R}^*$ una funzione sommabile rispetto alla legge della v.a. trasformata $\tau(X)$.

[5] Circostanza che suggerisce (come di solito si fa nei casi concreti) di operare direttamente sulla σ-algebra $\mathcal{B}$ tramite la legge P_X della v.a. X e "dimenticare" l'ambiente "originario" (l'insieme Ω, la σ-algebra $\mathcal{A}$ e la probabilità P).

Riesce allora $\mathrm{P}_{\tau(X)} = (\mathrm{P}_X)_\tau$ *e*

$$\int_{A'} \tilde{g}\, d\mathrm{P}_{\tau(X)} = \int_{\{\tau \in A'\}} \tilde{g}(\tau)\, d\mathrm{P}_X$$

per ogni $A' \in \mathcal{A}'$.[6]

Il prossimo esempio rileva che la nozione di μ-densità è una generalizzazione delle ben note nozioni di funzione di probabilità e di funzione di densità relative, rispettivamente, alle variabili aleatorie discrete e a quelle assolutamente continue.

Esempio B.1.3 (i) VARIABILI ALEATORIE DISCRETE Data una v.a. discreta X, sia f la sua funzione di probabilità, cioè $f(x) = \mathrm{P}(X = x)$ per ogni numero reale x. Considerata allora la misura di conteggio γ_B su $\aleph = 2^{\mathbb{R}}$ indotta dall'insieme discreto $B = \{x_1, x_2, \dots\}$ dei possibili valori di X, dall'Osservazione A.3.9 otteniamo

$$\mathrm{P}_X(\mathsf{A}) = \mathrm{P}\big(\bigcup_{n\in\{n:x_n\in\mathsf{A}\}} \{X = x_n\}\big) = \sum_{n\in\{n:x_n\in\mathsf{A}\}} f(x_n) = \sum_{n\geq 1} f(x_n)\, I_{\mathsf{A}}(x_n) = \int_{\mathsf{A}} f\, d\gamma_B$$

per ogni $\mathsf{A} \in \aleph$. Conseguentemente, la funzione di probabilità di X può essere vista come una particolare γ_B-densità di X.

(ii) VARIABILI ALEATORIE ASSOLUTAMENTE CONTINUE Data una v.a. assolutamente continua X, sia f una sua funzione di densità, cioè $f \geq 0$ e $F_X(x) = \int_{-\infty}^{x} f(t)\, dt$ per ogni numero reale x. Considerata allora la misura $f * \lambda$ di densità f rispetto alla misura di Lebesgue unidimensionale λ, otteniamo

$$f * \lambda(S_x) = \int_{S_x} f\, d\lambda = \int_{-\infty}^{x} f(t)\, dt = F_X(x) = \mathrm{P}_X(S_x)$$

per ogni $x \in \mathbb{R}$. Ne segue, tramite il criterio standard d'identità, $f * \lambda = \mathrm{P}_X$. Dunque, ogni funzione di densità di X può essere vista come una particolare λ-densità di X. ▲

Passando a considerare infine due enti aleatori X, Y a valori in $\mathfrak{X}$, diremo che essi sono **equidistribuiti** se $\mathrm{P}_X = \mathrm{P}_Y$, cioè se hanno la medesima legge. Conseguentemente, X e Y sono equidistribuiti se sono:
- uguali quasi certamente[7]. Infatti, indicato con A l'insieme ove coincidono,

[6] Chiaramente, sempre per i due teoremi citati, al posto della P_X-sommabilità di g si può assumere la P-sommabilità di $g(X)$ oppure, nel caso della densità, la μ-sommabilità di gf; al posto della $\mathrm{P}_{\tau(X)}$-sommabilità di $\tilde{g}$ si può assumere la P_X-sommabilità di $\tilde{g}(\tau)$.

[7] Nell'ambito della teoria delle probabilità è usuale sostituire la frase "quasi ovunque" con quella di **quasi certamente**, in quanto una proprietà sussiste a meno di un insieme di probabilità nulla se e solo se è valida su un insieme di probabilità unitaria. Conseguentemente, useremo l'abbreviazione (P-q.c.) al posto della (P-q.o.).

Per quanto riguarda l'implicazione inversa, osserviamo che, in generale, X e Y possono essere equidistribuiti pur non coincidendo in alcun caso elementare. Infatti, basta considerare un evento A tale che $\mathrm{P}(A) = \frac{1}{2}$ e porre $X = I_A$ e $Y = I_{A^c}$.

dall'additività e monotonia della probabilità si ha $\mathrm{P}_X(B) = \mathrm{P}(X \in B) = \mathrm{P}(\{X \in B\} \cap A) = \mathrm{P}(\{Y \in B\} \cap A) = \mathrm{P}(Y \in B) = \mathrm{P}_Y(B)$ per ogni $B \in \mathcal{B}$;
- due v.a. con medesima funzione di ripartizione. Infatti, dalla $\mathrm{P}_X(S_x) = F_X(x) = F_Y(x) = \mathrm{P}_Y(S_x)$ valida per ogni numero reale x, otteniamo, tramite il criterio standard d'identità, $\mathrm{P}_X = \mathrm{P}_Y$.

B.1.3 Speranza matematica

La **speranza matematica** (o **valor medio**) di una v.a. P-sommabile X è l'elemento della retta reale ampliata: $\mathrm{E}(X) = \int_\Omega X \, d\mathrm{P}$. Allora, nelle ipotesi del teorema fondamentale del calcolo delle probabilità, perveniamo all'uguaglianza

$$\mathrm{E}(g(X)) = \begin{cases} \int_{\mathbb{R}} g \, d\mathrm{P}_X \\ \int_{\mathbb{R}} g \, f \, d\mu \quad \text{se } f \text{ è una } \mu\text{-densità di } X \end{cases} \tag{B.1}$$

che risulta molto utile per calcolare la speranza matematica di trasformate di X mediante funzioni P_X-sommabili[8]. Nel caso particolare che g sia l'identità si ha quindi

$$\mathrm{E}(X) = \begin{cases} \int_{\mathbb{R}} x \, \mathrm{P}_X(dx) \\ \int_{\mathbb{R}} x \, f(x) \, \mu(dx) \quad \text{se } f \text{ è una } \mu\text{-densità di } X \end{cases}. \tag{B.2}$$

Ricorrendo all'uguaglianza (B.1) ritroviamo, nell'esempio seguente, le ben note espressioni della speranza matematica di trasformate sommabili di v.a. discrete e di quelle assolutamente continue.

Esempio B.1.4 (i) VARIABILI ALEATORIE DISCRETE Data una v.a. discreta X, sia f la sua funzione di probabilità. Allora, supposto che $B = \{x_1, x_2, \dots\}$ sia l'insieme dei valori possibili di X, dall'Esempio B.1.3(i) e dall'Osservazione A.3.9 otteniamo

$$\begin{aligned} \mathrm{E}(g(X)) &= \int_{\mathbb{R}} g \, f d\gamma_B = \sum_{n \geq 1} g^+(x_n) f(x_n) - \sum_{n \geq 1} g^-(x_n) f(x_n) \\ &= \sum_{n \geq 1} g^+(x_n) \, \mathrm{P}(X = x_n) - \sum_{n \geq 1} g^-(x_n) \, \mathrm{P}(X = x_n) \end{aligned}$$

per ogni funzione $g : \mathbb{R} \mapsto \mathbb{R}^*$ che renda finita una delle due serie.

[8] L'uguaglianza (B.1) è anche nota come "legge dello statistico inconsapevole" (unconscious statistician) poichè viene talvolta intesa (soprattutto nelle applicazioni) come la definizione di speranza matematica piuttosto che una sua conseguenza.

(ii) Variabili aleatorie assolutamente continue Data una v.a. assolutamente continua X, sia f una sua funzione di densità. Allora, ricorrendo all'Esempio B.1.3(ii), otteniamo $\mathrm{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)\, f(x)\, dx$ per ogni funzione $g : \mathbb{R} \mapsto \mathbb{R}^*$ che renda il prodotto gf sommabile rispetto alla misura di Lebesgue unidimensionale[9]. ▲

Essendo la speranza matematica definita come integrale di Lebesgue della v.a. rispetto alla probabilità considerata sugli eventi, valgono per essa tutte le proprietà degli integrali provate nella sezione terza dell'Appendice A. Per comodità, nel prossimo teorema ne riportiamo alcune accanto ad altre che dipendono dal fatto che la misura considerata è una probabilità. In particolare, la proprietà $\mathrm{E}(\alpha 1_\Omega) = \alpha$ assicura che la speranza matematica di una v.a. certa coincide con il suo valore; le proposizioni (i), (ii) e (iii) che la speranza matematica è un funzionale lineare monotono e normalizzato sullo spazio vettoriale delle v.a. P-integrabili che associa medesimo valore a variabili aleatorie quasi certamente uguali.

Teorema B.1.5 *Riesce* $\mathrm{E}(I_A) = \mathrm{P}(A)$ *e* $\mathrm{E}(\alpha I_\Omega) = \alpha$ *per ogni* $A \in \mathcal{A}$ *e* $\alpha \in \mathbb{R}$. *Inoltre, supposto che* X, Y *siano v.a. con speranza matematica*[10], *sussistono le seguenti proposizioni:*

(i) linearità: $\mathrm{E}(\alpha X + \beta Y) = \alpha \mathrm{E}(X) + \beta \mathrm{E}(Y)$, *se* α, β *sono numeri reali tali che* $\alpha\mathrm{E}(X)$ *e* $\beta\mathrm{E}(Y)$ *non sono infiniti di segno opposto;*

(ii) monotonia: $\mathrm{E}(X) \leq \mathrm{E}(Y)$, *se* $X \leq Y$ (P-q.c.)*;*

(iii) $\mathrm{E}(X) = \mathrm{E}(Y)$, *se* $X = Y$ (P-q.c.)*;*

(iv) $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$*;*

(v) internalità: $\inf X \leq \mathrm{E}(X) \leq \sup X$*;*

(vi) internalità stretta: $\alpha < \mathrm{E}(X) < \beta$, *se* $\alpha \ll X \ll \beta$.

[9] Cogliamo l'occasione per rilevare che un semplice esempio di v.a. priva di speranza matematica è fornito dalla funzione tangente pensata definita sull'intervallo aperto $\Omega =]-\frac{\pi}{2}, \frac{\pi}{2}[$. Infatti, posto $\mathcal{A} = \mathcal{B} \cap \Omega$ e $\mathrm{P} = \lambda\big|_{\mathcal{A}}$, la v.a. $X : \omega \mapsto \operatorname{tg} \omega$ ha funzione di densità $f(x) = \frac{1}{\pi(1+x^2)}$ e quindi, data la simmetria, $\mathrm{E}(X^+) = \mathrm{E}(X^-) = \int_0^{+\infty} \frac{x}{\pi(1+x^2)}\, dx = \Big|\frac{\log(1+x^2)}{2\pi}\Big|_0^{+\infty} = +\infty$ (da un punto di vista interpretativo, possiamo pensare che X sia ottenuta trasformando, mediante la funzione tangente, un numero aleatorio scelto "a caso" nell'intervallo Ω).

[10] Come usuale, intendiamo la locuzione "v.a. con speranza matematica" come sinonimo della "v.a. P-sommabile".

Infine, una v.a. Z ammette speranza matematica se $Z \geq X$ e $\mathrm{E}(X) > -\infty$ o $Z \leq X$ e $\mathrm{E}(X) < +\infty$.

DIMOSTRAZIONE Dalla definizione e linearità dell'integrale otteniamo, rispettivamente, $\mathrm{E}(I_A) = \mathrm{P}(A)$ e $\mathrm{E}(\alpha I_\Omega) = \alpha \int_\Omega I_\Omega \, d\mathrm{P} = \alpha \mathrm{P}(\Omega) = \alpha$. Inoltre, (i) $\div$ (v) seguono immediatamente dalla linearità dell'integrale e dalle proposizioni (i) $\div$ (iv) del Teorema A.3.6. Infine, l'ultima parte della tesi discende banalmente dal Teorema A.3.5(i). Rimane allora da provare solamente la proposizione (vi). Supponiamo dunque $\alpha < X(\omega) < \beta$ per ogni ω. Poichè la verifica delle disuguaglianze $\alpha < \mathrm{E}(X)$ e $\mathrm{E}(X) < \beta$ avviene in modo analogo, proviamo solamente la prima. Da (i) otteniamo $\mathrm{E}(X) - \alpha = \mathrm{E}(X) - \mathrm{E}(\alpha I_\Omega) = \mathrm{E}(X - \alpha I_\Omega) = \int_\Omega (X - \alpha I_\Omega) \, d\mathrm{P}$ e quindi, per il teorema A.3.6(vi), $\mathrm{E}(X) - \alpha > 0$. ■

Una conseguenza immediata di questo teorema e di quello della convergenza dominata è il seguente criterio di convergenza riguardante successioni di v.a. che siano limitate quasi certamente da un medesimo valore.

Teorema B.1.6 *Siano $(X_n)_{n \geq 1}$ una successione di v.a. tale che $X_n \to X$ e α un numero reale tale che $|X_n| \leq \alpha$ (P-q.c.) per ogni n. Allora, le v.a. X, X_n $(n \geq 1)$ ammettono speranza matematica e risulta $\mathrm{E}(X_n) \to \mathrm{E}(X)$.*

Osservazione B.1.7 Anche per le v.a. estese si usa la medesima nozione di speranza matematica. Conseguentemente, (B.1) e (B.2) rimangono valide sostituendo (come dominio d'integrazione) la retta reale con quella ampliata; inoltre, sussistono tutte le proposizioni del Teorema B.1.5 con l'avvertenza di supporre, nella proposizione (i), che la combinazione lineare $\alpha X + \beta Y$ sia definita ovunque. ♦

B.1.4 Varianza e covarianza

Data una v.a. X con speranza matematica finita, la **varianza** di X è il *momento centrale secondo*: $\mathrm{Var}(X) = \mathrm{E}\big((X - \mathrm{E}(X))^2\big)$ che presenta le caratteristiche di un parametro di dispersione, nel senso che il suo valore consente di intendere se i valori possibili di X sono più o meno concentrati attorno alla speranza matematica $\mathrm{E}(X)$ (che ha invece le caratteristiche di un parametro di posizione)[11].

[11] Anche per le v.a. estese si usa la medesima nozione di varianza. Conviene comunque osservare che limitarci a considerare, nel caso di finitezza della speranza matematica, solo v.a. non lede la generalità del discorso; infatti, una v.a. estesa con speranza matematica finita è, per il Teorema A.3.6(vii), uguale quasi certamente a una variabile aleatoria.

Nel teorema seguente riportiamo alcune proprietà basilari della varianza. In particolare, per la proposizione (iii), le v.a. con varianza nulla coincidono con quelle che sono certe con probabilità unitaria.

Teorema B.1.8 *Sia X una v.a. con speranza matematica finita. Risulta allora:*

(i) $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$*;*

(ii) $\mathrm{Var}(\alpha X + \beta) = \alpha^2\,\mathrm{Var}(X)$ *qualunque siano i numeri reali α e β;*

(iii) $\mathrm{Var}(X) = 0$ *se e solo se* $X = \mathrm{E}(X)$ (P-q.c.).

Dimostrazione (i) Dalla linearità della speranza matematica otteniamo $\mathrm{Var}(X) = \mathrm{E}\big(X^2 - 2X\mathrm{E}(X) + \mathrm{E}(X)^2\big) = \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

(ii) Per la linearità della speranza matematica, $\mathrm{Var}(\alpha X+\beta) = \mathrm{E}\big([(\alpha X+\beta)-\mathrm{E}(\alpha X+\beta)]^2\big) = \mathrm{E}\big([\alpha X + \beta - (\alpha\,\mathrm{E}(X) + \beta)]^2\big) = \mathrm{E}\big(\alpha^2\,(X - \mathrm{E}(X))^2\big) = \alpha^2\,\mathrm{Var}(X)$.

(iii) Conseguenza immediata del Teorema A.3.6(ii),(v). ∎

Considerate ora due v.a. X, Y con speranza matematica finita e tali che esista finita la speranza matematica $\mathrm{E}(XY)$ del loro prodotto[12], la **covarianza** di X e Y è la differenza: $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)$. Riesce quindi $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ e, in particolare, $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$. Inoltre, per la linearità della speranza matematica, si ha

$$\begin{aligned}\mathrm{E}\big((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big) &= \mathrm{E}\big(XY - X\mathrm{E}(Y) - Y\mathrm{E}(X) + \mathrm{E}(X)\mathrm{E}(Y)\big)\\ &= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\\ &\qquad - \mathrm{E}(Y)\mathrm{E}(X) + \mathrm{E}(X)\mathrm{E}(Y)\end{aligned}$$

e quindi l'uguaglianza

$$\mathrm{Cov}(X,Y) = \mathrm{E}\big((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big) \tag{B.3}$$

che, tra l'altro, induce ad aspettarsi una covarianza positiva quando, con riferimento alle rispettive speranze matematiche, a valori tendenzialmente alti (bassi) di X corrispondano valori alti (bassi) di Y.

Il teorema successivo fornisce ulteriori proprietà della covarianza. Di particolare interesse è l'ultima in quanto assicura che, qualora le v.a. siano a due a due **non correlate** (cioè di covarianza nulla), la varianza della loro somma è uguale alla somma delle rispettive varianze.

[12]Come lo sono, ad esempio, due v.a. a quadrato integrabile (Teorema A.3.5(iv)).

Teorema B.1.9 *Siano X, Y, Z v.a. con speranza matematica finita e tali che esistano finite le speranze matematiche* $\mathrm{E}(XY)$ *e* $\mathrm{E}(YZ)$. *Risulta allora:*

(i) $\mathrm{Cov}(\alpha X, Y) = \alpha\,\mathrm{Cov}(X, Y)$ *per ogni numero reale* α*;*

(ii) $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Z)$*;*

(iii) $\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$ *(Disuguaglianza di Cauchy-Schwarz).*

Inoltre, se $X_1, \ldots, X_n$ sono v.a. con speranza matematica finita e tali che esista finita la speranza matematica $\mathrm{E}(X_iX_j)$ *per ogni $i \neq j$, si ha*

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i} \mathrm{Cov}(X_i, X_j).$$

Dimostrazione (i) Dalla linearità della speranza matematica otteniamo $\mathrm{Cov}(\alpha X, Y) = \mathrm{E}(\alpha XY) - \mathrm{E}(\alpha X)\,\mathrm{E}(Y) = \alpha[\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)] = \alpha\,\mathrm{Cov}(X, Y)$.

(ii) Sempre per la linearità della speranza matematica si ha

$$\begin{aligned}
\mathrm{Cov}(X + Z, Y) &= \mathrm{E}\big[(X + Z)Y\big] - \mathrm{E}(X + Z)\,\mathrm{E}(Y)\\
&= \mathrm{E}(XY) + \mathrm{E}(ZY) - \mathrm{E}(X)\mathrm{E}(Y) - \mathrm{E}(Z)\mathrm{E}(Y)\\
&= \big[\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)\big] + \big[\mathrm{E}(YZ) - \mathrm{E}(Y)\mathrm{E}(Z)\big]\\
&= \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Z).
\end{aligned}$$

(iii) Sia intanto $\mathrm{Var}(X) = 0$. Allora, per il teorema B.1.8(iii), $X = \mathrm{E}(X)$ (P-q.c.) e quindi $(X - \mathrm{E}(X))(Y - \mathrm{E}(Y)) = 0$ (P-q.c.); ne segue, per il Teorema B.1.5(iii) e (B.3), $\mathrm{Cov}(X, Y) = 0$ da cui otteniamo la disuguaglianza in oggetto. Sia ora $\mathrm{Var}(X) > 0$. Posto $X' = X - \mathrm{E}(X)$ e $Y' = Y - \mathrm{E}(Y)$, riesce $\mathrm{Var}(X)\,\mathrm{Var}(Y) = \mathrm{E}(|X'|^2)\,\mathrm{E}(|Y'|^2)$ e, per il Teorema B.1.5(iv) e (B.3), $\mathrm{Cov}(X, Y)^2 = \mathrm{E}(X'Y')^2 = |\mathrm{E}(X'Y')|^2 \leq \mathrm{E}(|X'Y'|)^2$. Basta allora provare la disuguaglianza $\mathrm{E}(|X'|\,|Y'|)^2 \leq \mathrm{E}(|X'|^2)\,\mathrm{E}(|Y'|^2)$ che, coinvolgendo solamente $|X'|$ e $|Y'|$, riguarda v.a. non negative.

Date dunque due v.a. $U \geq 0$, $V \geq 0$ con $\mathrm{E}(U^2) > 0$, verifichiamo che si ha $\mathrm{E}(U\,V)^2 \leq \mathrm{E}(U^2)\,\mathrm{E}(V^2)$. A tal fine, considerate le v.a. “troncate”:

$$U_n = U\,I_{\{U \leq n\}} + n\,I_{\{U > n\}}, \quad V_n = V\,I_{\{V \leq n\}} + n\,I_{\{V > n\}} \qquad (n \geq 1),$$

osserviamo che riesce $0 \leq U_nV_n \uparrow UV$, $0 \leq U_n^2 \uparrow U^2$ e $0 \leq V_n^2 \uparrow V^2$. Ne segue, per il teorema della convergenza monotona A.3.7, $\mathrm{E}(U_nV_n)^2 \uparrow \mathrm{E}(UV)^2$ e $\mathrm{E}(U_n^2) \uparrow \mathrm{E}(U^2)$, $\mathrm{E}(V_n^2) \uparrow \mathrm{E}(V^2)$. Conseguentemente, supposto (senza perdere in generalità) $\mathrm{E}(U_n^2) > 0$ $(n \geq 1)$, basta constatare che la disuguaglianza $\mathrm{E}(U_nV_n)^2 \leq \mathrm{E}(U_n^2)\,\mathrm{E}(V_n^2)$ sussiste per ogni n. A tal fine, dato un numero reale t, consideriamo la v.a. $tU_n + V_n$. Le v.a. U_n^2, V_n^2 e U_nV_n in quanto limitate hanno, per la monotonia della speranza matematica, speranza matematica finita. Dalla linearità e monotonia della speranza matematica risulta allora

$$0 \leq \mathrm{E}\big((tU_n + V_n)^2\big) = \mathrm{E}(t^2U_n^2 + 2t\,U_nV_n + V_n^2) = \mathrm{E}(U_n^2)\,t^2 + 2\,\mathrm{E}(U_nV_n)\,t + \mathrm{E}(V_n^2).$$

Pertanto, data l'arbitrarietà di t, l'equazione di secondo grado a coefficienti reali

$$\mathrm{E}(U_n^2)\, t^2 + 2\,\mathrm{E}(U_n V_n)\, t + \mathrm{E}(V_n^2) = 0$$

ammette al più una sola soluzione. Ne segue che il suo discriminante deve essere minore o uguale a zero e quindi $\mathrm{E}(U_n V_n)^2 - \mathrm{E}(U_n^2)\,\mathrm{E}(V_n^2) \leq 0$, cioè $\mathrm{E}(U_n V_n)^2 \leq \mathrm{E}(U_n^2)\,\mathrm{E}(V_n^2)$.

Per quanto riguarda l'ultima parte della tesi, sia intanto $n = 2$. Dalla linearità della speranza matematica e (B.3) otteniamo

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}\big([(X+Y) - \mathrm{E}(X+Y)]^2\big) = \mathrm{E}\big([(X - \mathrm{E}(X)) + (Y - \mathrm{E}(Y))]^2\big) \\
&= \mathrm{E}\big((X-\mathrm{E}(X))^2 + (Y-\mathrm{E}(Y))^2 + 2(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big) \\
&= \mathrm{E}\big((X-\mathrm{E}(X))^2\big) + \mathrm{E}\big((Y-\mathrm{E}(Y))^2\big) + 2\,\mathrm{E}\big((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).
\end{aligned}$$

Procedendo per induzione su n, supponiamo che l'uguaglianza sussista per $m \geq 2$ (ipotesi induttiva) e proviamola per $n = m+1$. Allora, per il caso $n = 2$ e (ii), si ha

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) &= \mathrm{Var}\Big(\sum_{i=1}^{m} X_i + X_n\Big) = \mathrm{Var}\Big(\sum_{i=1}^{m} X_i\Big) + \mathrm{Var}(X_n) + 2\,\mathrm{Cov}\Big(\sum_{i=1}^{m} X_i, X_n\Big) \\
&= \sum_{i=1}^{m} \mathrm{Var}(X_i) + 2\sum_{i=1}^{m-1}\sum_{j>i} \mathrm{Cov}(X_i, X_j) \\
&\qquad + \mathrm{Var}(X_n) + 2\sum_{i=1}^{m} \mathrm{Cov}(X_i, X_n)
\end{aligned}$$

e quindi la tesi. ■

B.1.5 Leggi congiunte e indipendenza

Considerata una σ-algebra $\aleph_i$ sull'insieme non vuoto $\mathfrak{X}_i$ ($i = 1, 2$), siano $\mathfrak{X}$ il prodotto cartesiano $\mathfrak{X}_1 \times \mathfrak{X}_2$ e $\aleph$ la σ-algebra prodotto $\aleph_1 \otimes \aleph_2$. Chiamato allora **coppia aleatoria** ogni ente aleatorio su Ω a valori in $\mathfrak{X}$, l'applicazione $\mathbf{X} = (X_1, X_2) : \omega \mapsto (X_1(\omega), X_2(\omega))$ di Ω in $\mathfrak{X}$ è una coppia aleatoria se e solo se la sua componente i-sima X_i ($i = 1, 2$) è un ente aleatorio a valori in $\mathfrak{X}_i$ (cioè, un'applicazione $(\mathcal{A}, \aleph_i)$-misurabile)[13].

Ciò osservato, dato l'ente aleatorio X_i a valori in $\mathfrak{X}_i$ ($i = 1, 2$), la **legge** (o **distribuzione**) **congiunta** di X_1 e X_2 è la legge $\mathrm{P}_{\mathbf{X}}$ della coppia aleatoria $\mathbf{X}$.

[13]Per constatarlo, ricordato che $\aleph$ è generata dalla famiglia dei rettangoli misurabili e il criterio standard di misurabilità A.1.5, basta osservare che $X_1^{-1}(\mathsf{A}_1) = \mathbf{X}^{-1}(\mathsf{A}_1 \times \mathfrak{X}_2)$, $X_2^{-1}(\mathsf{A}_2) = \mathbf{X}^{-1}(\mathfrak{X}_1 \times \mathsf{A}_2)$ e $\mathbf{X}^{-1}(\mathsf{A}_1 \times \mathsf{A}_2) = X_1^{-1}(\mathsf{A}_1) \cap X_2^{-1}(\mathsf{A}_2)$ per ogni $\mathsf{A}_1 \in \aleph_1$ e $\mathsf{A}_2 \in \aleph_2$.

Supposto poi che su $\aleph_i$ sia definita una misura σ-finita di riferimento μ_i ($i = 1, 2$), una (μ_1, μ_2)**-densità congiunta** di X_1 e X_2 è ogni funzione $\aleph$-Borel misurabile $f : \mathfrak{X} \mapsto [0, +\infty]$ tale che $\mathrm{P}_{\mathbf{X}} = f{*}(\mu_1 \times \mu_2)$. Conseguentemente, se f è una (μ_1, μ_2)-densità congiunta di X_1 e X_2, si ha $\mathrm{P}(\mathbf{X} \in \mathsf{A}) = \int_{\mathsf{A}} f \, d\mu_1 \times \mu_2$ per ogni $\mathsf{A} \in \aleph$; inoltre, f è finita quasi certamente (Teorema A.3.6(vii)) ed è definita, ricordato che la misura prodotto è σ-finita, a meno di insiemi $\mu_1 \times \mu_2$-trascurabili (teoremi A.3.6(ii) e A.4.4).

Nel caso particolare che X_1, X_2 siano v.a., considereremo anche la **funzione di ripartizione congiunta** di X_1 e X_2, cioè la funzione $F_{\mathbf{X}} : \mathbb{R}^2 \mapsto [0, 1]$ così definita:

$$F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}_{\mathbf{X}}\big(]-\infty, x_1] \times]-\infty, x_2]\big) = \mathrm{P}\big((X_1, X_2) \leq (x_1, x_2)\big)$$

che associa, ad ogni coppia reale (x_1, x_2), la probabilità dell'evento "$X_1 \leq x_1$ e $X_2 \leq x_2$".

Precisiamo infine che la **legge marginale** di X_i ($i = 1, 2$) è la legge P_{X_i}; che una μ_i**-densità marginale** di X_i è una sua qualsiasi μ_i-densità; che, nel caso particolare che X_i sia una v.a., la **funzione di ripartizione marginale** di X_i è la funzione di ripartizione F_{X_i}.

Il prossimo risultato segue immediatamente dalla proposizione (i) del teorema fondamentale del calcolo delle probabilità e ne fornisce una versione bidimensionale che risulta utile qualora si debba calcolare, ad esempio, la speranza matematica di una funzione Borel-misurabile dipendente da due v.a. (come avviene nel caso della covarianza ove è necessario valutare la speranza matematica di un prodotto di v.a.).

Teorema B.1.10 *Sia $g : \mathfrak{X} \mapsto \mathbb{R}^*$ una funzione sommabile rispetto alla legge congiunta di X_1 e X_2. Allora, posto $\mu = \mu_1 \times \mu_2$, si ha*

$$\int_{\{\mathbf{X} \in \mathsf{A}\}} g(X_1, X_2) \, d\mathrm{P} = \begin{cases} \int_{\mathsf{A}} g \, d\mathrm{P}_{\mathbf{X}} \\ \int_{\mathsf{A}} g \, f \, d\mu & \textit{se } f \textit{ è una } \mu\textit{-densità di } X_1 \textit{ e } X_2 \end{cases}$$

per ogni $\mathsf{A} \in \aleph$[14].

Mostriamo ora come la nozione di (μ_1, μ_2)-densità congiunta sia una generalizzazione delle ben note nozioni di funzione di probabilità congiunta e di funzione di densità congiunta relative, rispettivamente, alle coppie aleatorie

[14] Per la nota 6 di p. 169, al posto della $\mathrm{P}_{\mathbf{X}}$-sommabilità di g si può assumere la P-sommabilità di $g(X)$ o, nel caso della densità congiunta, la μ-sommabilità di gf.

discrete e a quelle assolutamente continue. A tal fine, ricordiamo che la famiglia $\mathcal{S}^{(2)} = \{S_{x_1x_2} =]-\infty, x_1] \times]-\infty, x_2] : x_1, x_2 \in \mathbb{R}\}$ è chiusa per intersezioni finite e riesce $\mathcal{B}^{(2)} = \sigma(\mathcal{S}^{(2)})$ (Esempio A.5.7) e $\mathbb{R}^2 = \bigcup_{n\geq 1} S_{nn}$.

Esempio B.1.11 (i) COPPIA ALEATORIA DISCRETA Date le v.a. discrete X_1, X_2, poniamo $\mathfrak{X}_i = X_i(\Omega) = \{x_{i1}, x_{i2}, \dots\}$ e $\aleph_i = 2^{\mathfrak{X}_i}$ $(i = 1, 2)$. Allora, $\mathfrak{X}$ è un insieme discreto e quindi $\aleph = 2^{\mathfrak{X}}$. Indicata con f la funzione di probabilità congiunta della coppia aleatoria $\mathbf{X}$, otteniamo $f(\mathbf{x}) = \mathrm{P}(\mathbf{X} = \mathbf{x})$ per ogni $\mathbf{x} = (x_1, x_2) \in \mathfrak{X}$. Considerata infine come misura di riferimento su $\aleph_i$ la misura di conteggio $\mu_i = \gamma_{\mathfrak{X}_i}$ $(i = 1, 2)$ e fissato un qualsiasi $\mathsf{A} \in \aleph$, per l'Osservazione A.3.9, si ha

$$\begin{aligned}
\mathrm{P}(\mathbf{X} \in \mathsf{A}) &= \mathrm{P}\big(\bigcup_{(n,m)\in\{(n,m):(x_{1n},x_{2m})\in\mathsf{A}\}} \{\mathbf{X} = (x_{1n}, x_{2m})\}\big)\\
&= \sum_{(n,m)\in\{(n,m):(x_{1n},x_{2m})\in\mathsf{A}\}} f(x_{1n}, x_{2m}) = \sum_{n\geq 1}\sum_{m\geq 1} f(x_{1n}, x_{2m})\, I_{\mathsf{A}}(x_{1n}, x_{2m})\\
&= \sum_{n\geq 1}\int_{\mathfrak{X}_2} f(x_{1n}, \cdot)\, I_{\mathsf{A}}(x_{1n}, \cdot)\, d\mu_2 = \int_{\mathfrak{X}_1}\Big(\int_{\mathfrak{X}_2} f(x_1, \cdot)\, I_{\mathsf{A}}(x_1, \cdot)\, d\mu_2\Big)\mu_1(dx_1)
\end{aligned}$$

e quindi, per il teorema di Tonelli A.5.4,

$$\mathrm{P}_{\mathbf{X}}(\mathsf{A}) = \int_{\mathfrak{X}} f\, I_{\mathsf{A}}\, d\mu_1 \times \mu_2 = \int_{\mathsf{A}} f\, d\mu_1 \times \mu_2.$$

Conseguentemente, la funzione di probabilità congiunta di $\mathbf{X}$ può essere vista come una particolare (μ_1, μ_2)-densità congiunta delle v.a X_1 e X_2.

(ii) COPPIA ALEATORIA ASSOLUTAMENTE CONTINUA Sia f una funzione di densità congiunta della coppia aleatoria assolutamente continua $\mathbf{X}$. Risulta allora $f \geq 0$ e

$$f * \lambda^{(2)}(S_{x_1x_2}) = \int_{S_{x_1x_2}} f(x, y)\, dxdy = F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}_{\mathbf{X}}(S_{x_1x_2})$$

qualunque siano i numeri reali x_1, x_2[15]. Ne segue, tramite il criterio standard d'identità, $\mathrm{P}_{\mathbf{X}} = f * \lambda^{(2)}$. Conseguentemente, ogni funzione di densità congiunta di $\mathbf{X}$ può essere vista come una particolare (λ, λ)-densità congiunta delle v.a. X_1 e X_2. ▲

Il risultato seguente mostra come ricavare dalla legge congiunta e dalle densità congiunte, rispettivamente, le leggi marginali e le densità marginali come pure, nel caso particolare che X_1, X_2 siano v.a., dalla funzione di ripartizione congiunta le funzioni di ripartizione marginale.

[15]Avendo denotato, per ogni $m \geq 2$ e $B \in \mathcal{B}^{(m)}$, con $\int_B f(x_1, \dots, x_m)\, dx_1 \cdots dx_m$ l'integrale $\int_B f(x_1, \dots, x_m)\, \lambda^{(m)}(dx_1 \times \cdots \times dx_m)$.

Teorema B.1.12 *Risulta* $\mathrm{P}_{X_1}(\mathsf{A}_1) = \mathrm{P}_{\mathbf{X}}(\mathsf{A}_1 \times \mathfrak{X}_2)$ *e* $\mathrm{P}_{X_2}(\mathsf{A}_2) = \mathrm{P}_{\mathbf{X}}(\mathfrak{X}_1 \times \mathsf{A}_2)$ *per ogni* $\mathsf{A}_i \in \aleph_i$ $(i = 1, 2)$. *Inoltre, se* f *è una* (μ_1, μ_2)*-densità congiunta di* X_1 *e* X_2*, la funzione* $f_{X_i} : \mathfrak{X}_i \mapsto [0, +\infty]$ *così definita:*

$$f_{X_i}(x_i) = \begin{cases} \int_{\mathfrak{X}_2} f(x_1, x_2)\, \mu_2(dx_2) & se\ i = 1 \\ \int_{\mathfrak{X}_1} f(x_1, x_2)\, \mu_1(dx_1) & se\ i = 2 \end{cases}$$

per ogni $x_i \in \mathfrak{X}_i$*, è una* μ_i*-densità marginale di* X_i $(i = 1, 2)$. *Infine, nel caso particolare che* X_1*,* X_2 *siano v.a.,* $F_{X_1}(t) = \lim_{x_2 \to +\infty} F_{\mathbf{X}}(t, x_2)$ *e* $F_{X_2}(t) = \lim_{x_1 \to +\infty} F_{\mathbf{X}}(x_1, t)$ *per ogni numero reale* t.

Dimostrazione Ci limitiamo a provare il teorema con riferimento all'ente aleatorio X_1 in quanto le dimostrazioni relative a X_2 sono del tutto simili. Dato $\mathsf{A}_1 \in \aleph_1$, risulta $\mathrm{P}_{X_1}(\mathsf{A}_1) = \mathrm{P}(X_1 \in \mathsf{A}_1) = \mathrm{P}(\mathbf{X} \in \mathsf{A}_1 \times \mathfrak{X}_2) = \mathrm{P}_{\mathbf{X}}(\mathsf{A}_1 \times \mathfrak{X}_2)$. Sia ora f una (μ_1, μ_2)-densità congiunta di X_1 e X_2. Dal teorema di Tonelli otteniamo allora

$$\mathrm{P}_{X_1}(\mathsf{A}_1) = \mathrm{P}_{\mathbf{X}}(\mathsf{A}_1 \times \mathfrak{X}_2) = \int_{\mathsf{A}_1 \times \mathfrak{X}_2} f\, d\mu_1 \times \mu_2 = \int_{\mathsf{A}_1} \Big(\int_{\mathfrak{X}_2} f(x_1, \cdot)\, d\mu_2 \Big) \mu_1(dx_1)$$

qualunque sia $\mathsf{A}_1 \in \aleph_1$. Conseguentemente, f_{X_1} è una μ_1-densità marginale di X_1. Sia infine X_1 una v.a.. Dato un numero reale t, la successione $(S_{tn})_{n \geq 1}$ è non decrescente e riesce $\bigcup_{n \geq 1} S_{tn} =]-\infty, t] \times \mathbb{R}$. Per la monotonia della funzione di ripartizione congiunta rispetto a x_2 e la continuità della legge congiunta otteniamo allora

$$\begin{aligned} \lim_{x_2 \to +\infty} F_{\mathbf{X}}(t, x_2) &= \lim_{n \to +\infty} F_{\mathbf{X}}(t, n) = \lim_{n \to +\infty} \mathrm{P}_{\mathbf{X}}(S_{tn}) \\ &= \mathrm{P}_{\mathbf{X}}(]-\infty, t] \times \mathbb{R}) = \mathrm{P}_{X_1}(]-\infty, t]) = F_{X_1}(t). \end{aligned}$$

La dimostrazione è così conclusa. ■

Mentre la legge, le densità e la funzione di ripartizione congiunte consentono di ottenere, rispettivamente, le leggi, le densità e le funzioni di ripartizione marginali, non vale in generale il viceversa. In altre parole, la conoscenza marginale non è in grado, senza ulteriori ipotesi, di fornire informazioni sul comportamento congiunto di due enti aleatori. D'altra parte, come il prossimo teorema mette in evidenza, è possibile dedurre dalla conoscenza marginale quella congiunta se i due enti aleatori X_1 e X_2 sono **indipendenti**, cioè: $\mathrm{P}_{\mathbf{X}}(\mathsf{A}_1 \times \mathsf{A}_2) = \mathrm{P}_{X_1}(\mathsf{A}_1)\, \mathrm{P}_{X_2}(\mathsf{A}_2)$ per ogni rettangolo misurabile $\mathsf{A}_1 \times \mathsf{A}_2$.

Teorema B.1.13 *Sono equivalenti le seguenti proposizioni:*

(i) X_1 *e* X_2 *sono indipendenti;*

(ii) $\mathrm{P}_{\mathbf{X}} = \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}$.

Inoltre, se $f_{\mathbf{X}}$ *è una* (μ_1, μ_2)*-densità congiunta di* X_1 *e* X_2*, (i) è equivalente alla proposizione:*

(iii) $f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$ *a meno di un insieme* $\mu_1 \times \mu_2$*-trascurabile.*

Infine, se X_1*,* X_2 *sono v.a., (i) è equivalente alla proposizione:*

(iv) $F_{\mathbf{X}}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2)$ *per ogni* $x_1, x_2 \in \mathbb{R}$*.*

Dimostrazione L'implicazione $(i) \Rightarrow (ii)$ segue immediatamente dal Teorema A.5.3; la sua inversa è invece ovvia.

$(i) \iff (iii)$ Sia $f : (x_1, x_2) \mapsto f_{X_1}(x_1) f_{X_2}(x_2)$. Posto $\mu = \mu_1 \times \mu_2$, consideriamo un rettangolo misurabile $\mathsf{A}_1 \times \mathsf{A}_2$. Allora, dal teorema di Tonelli otteniamo

$$\begin{aligned} f * \mu(\mathsf{A}_1 \times \mathsf{A}_2) &= \int_{\mathsf{A}_1 \times \mathsf{A}_2} f \, d\mu = \int_{\mathsf{A}_1} \Big(\int_{\mathsf{A}_2} f(x_1, x_2) \, \mu_2(dx_2) \Big) \, \mu_1(dx_1) \\ &= \int_{\mathsf{A}_1} \Big(\int_{\mathsf{A}_2} f_{X_2}(x_2) \, \mu_2(dx_2) \Big) \, f_{X_1}(x_1) \, \mu_1(dx_1) \\ &= \Big(\int_{\mathsf{A}_1} f_{X_1} \, d\mu_1 \Big) \Big(\int_{\mathsf{A}_2} f_{X_2} \, d\mu_2 \Big) = \mathrm{P}_{X_1}(\mathsf{A}_1) \, \mathrm{P}_{X_2}(\mathsf{A}_2), \end{aligned}$$

osservato che, per il Teorema B.1.12, f_{X_i} è una μ_i-densità di X_i $(i = 1, 2)$. Conseguentemente, per l'arbitrarietà di $\mathsf{A}_1 \times \mathsf{A}_2$, dal Teorema A.5.3 si ha $f * \mu = \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}$. A questo punto, l'equivalenza in oggetto segue immediatamente dall'equivalenza $(i) \iff (ii)$.

$(i) \iff (iv)$ Siano le due v.a. indipendenti. Allora, $F_{\mathbf{X}}(x_1, x_2) = \mathrm{P}_{\mathbf{X}}(S_{x_1 x_2}) = \mathrm{P}_{X_1}(]-\infty, x_1]) \, \mathrm{P}_{X_2}(]-\infty, x_2]) = \mathrm{F}_{X_1}(x_1) \, \mathrm{F}_{X_2}(x_2)$. Viceversa, assumiamo $F_{\mathbf{X}}(x_1, x_2) = \mathrm{F}_{X_1}(x_1) \, \mathrm{F}_{X_2}(x_2)$ per ogni $x_1, x_2 \in \mathbb{R}$. Allora

$$\mathrm{P}_{\mathbf{X}}(S_{x_1 x_2}) = \mathrm{P}_{X_1}(]-\infty, x_1]) \, \mathrm{P}_{X_2}(]-\infty, x_2]) = \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}(S_{x_1 x_2})$$

per ogni $x_1, x_2 \in \mathbb{R}$. Ne segue, tramite il criterio standard d'identità, $\mathrm{P}_{\mathbf{X}} = \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}$ e quindi, per l'equivalenza $(i) \iff (ii)$, X_1 e X_2 sono indipendenti. ■

Di particolare importanza sono i due risultati successivi in quanto assicurano, rispettivamente, che trasformate misurabili "certe" di enti aleatori indipendenti sono a loro volta indipendenti e che, nell'ambito delle v.a. con speranza matematica finita, l'indipendenza implica la non correlazione[16].

[16]Ma, viceversa, la non correlazione non implica, in generale, l'indipendenza. Infatti, posto $\Omega = [0, 2\pi]$, $\mathcal{A} = \mathcal{B} \cap \Omega$ e $\mathrm{P} = \frac{1}{2\pi} \lambda\big|_{\mathcal{A}}$, consideriamo le v.a $X_1 : \omega \mapsto \cos\omega$ e $X_2 : \omega \mapsto \sin\omega$ (che da un punto di vista interpretativo, possiamo pensare ottenute trasformando, mediante le funzioni coseno e seno rispettivamente, un numero aleatorio scelto "a caso" nell'intervallo Ω). Riesce allora, $\mathrm{E}(X_1) = 0 = \mathrm{E}(X_1 X_2)$ e quindi $\mathrm{Cov}(X_1, X_2) = 0$. D'altra parte, considerati gli insiemi $I_1 = \big[-1, -\frac{1}{\sqrt{2}}\big]$, $I_2 = \big[0, \frac{1}{\sqrt{2}}\big]$, $I_3 = \big[\frac{1}{\sqrt{2}}, 1\big]$ e $B = \big[\frac{1}{4}\pi, \frac{1}{2}\pi\big] \cup \big[\frac{3}{2}\pi, \frac{7}{4}\pi\big]$, è facile constatare che $\mathrm{P}_{X_1}(I_2) = \mathrm{P}_{X_2}(I_1 \cup I_3) = \mathrm{P}_{\mathbf{X}}\big(I_2 \times (I_1 \cup I_3)\big) = \mathrm{P}(B) = \frac{\pi}{2}$ e quindi X_1, X_2 non sono indipendenti.

Teorema B.1.14 *Siano $\mathcal{G}_i$ una σ-algebra su un insieme non vuoto $\mathfrak{Y}_i$ e $g_i : \mathfrak{X}_i \mapsto \mathfrak{Y}_i$ un'applicazione $(\aleph_i, \mathcal{G}_i)$-misurabile $(i = 1, 2)$. Siano inoltre X_1 e X_2 indipendenti. Allora, sono pure indipendenti gli enti aleatori trasformati $g_1(X_1)$ e $g_2(X_2)$.*

Dimostrazione Considerata la coppia aleatoria $\mathbf{Y} = (g_1(X_1), g_2(X_2))$, otteniamo

$$\begin{aligned}
\mathrm{P}_{\mathbf{Y}}(G_1 \times G_2) &= \mathrm{P}\big((g_1(X_1), g_2(X_2)) \in G_1 \times G_2\big) = \mathrm{P}\big(\{g_1(X_1) \in G_1\} \cap \{g_2(X_2) \in G_2\}\big) \\
&= \mathrm{P}\big(\{X_1 \in g_1^{-1}(G_1)\} \cap \{X_2 \in g_2^{-1}(G_2)\}\big) \\
&= \mathrm{P}\big((X_1, X_2) \in g_1^{-1}(G_1) \times g_2^{-1}(G_2)\big) = \mathrm{P}_{\mathbf{X}}\big(g_1^{-1}(G_1) \times g_2^{-1}(G_2)\big) \\
&= \mathrm{P}_{X_1}\big(g_1^{-1}(G_1)\big)\, \mathrm{P}_{X_2}\big(g_2^{-1}(G_2)\big) = \mathrm{P}\big(X_1^{-1}(g_1^{-1}(G_1))\big)\, \mathrm{P}\big(X_2^{-1}(g_2^{-1}(G_2))\big) \\
&= \mathrm{P}\big((g_1 \circ X_1)^{-1}(G_1)\big)\, \mathrm{P}\big((g_2 \circ X_2)^{-1}(G_2)\big) = \mathrm{P}_{g_1(X_1)}(G_1)\, \mathrm{P}_{g_2(X_2)}(G_2)
\end{aligned}$$

per ogni rettangolo misurabile $G_1 \times G_2 \in \mathcal{G}_1 \otimes \mathcal{G}_2$. ■

Teorema B.1.15 *Siano X_1, X_2 v.a. indipendenti. Allora, se $X_i \geq 0$ $(i = 1, 2)$ o $\mathrm{E}(X_i)$ finita $(i = 1, 2)$, esiste la speranza matematica del prodotto e si ha $\mathrm{E}(X_1 X_2) = \mathrm{E}(X_1)\,\mathrm{E}(X_2)$.*

Dimostrazione Dal Teorema B.1.13 risulta $\mathrm{P}_{\mathbf{X}} = \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}$. Ciò osservato, sia intanto $X_i \geq 0$ $(i = 1, 2)$. Allora, $\mathrm{P}_{X_i}([0, +\infty[) = \mathrm{P}_{\mathbf{X}}([0, +\infty[\times [0, +\infty[) = 1$ $(i = 1, 2)$ da cui, tramite i teoremi B.1.10, di Tonelli e (B.2), otteniamo

$$\begin{aligned}
\mathrm{E}(X_1 X_2) &= \int_{\mathbb{R}^2} x_1 x_2\, \mathrm{P}_{\mathbf{X}}(dx_1 \times dx_2) = \int_{[0,+\infty[\, \times [0,+\infty[} x_1 x_2\, \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}(dx_1 \times dx_2) \\
&= \int_{[0,+\infty[} \Big(\int_{[0,+\infty[} x_2\, \mathrm{P}_{X_2}(dx_2) \Big) x_1\, \mathrm{P}_{X_1}(dx_1) \\
&= \Big(\int_{[0,+\infty[} x_1\, \mathrm{P}_{X_1}(dx_1) \Big) \Big(\int_{[0,+\infty[} x_2\, \mathrm{P}_{X_2}(dx_2) \Big) \\
&= \Big(\int_{\mathbb{R}} x_1\, \mathrm{P}_{X_1}(dx_1) \Big) \Big(\int_{\mathbb{R}} x_2\, \mathrm{P}_{X_2}(dx_2) \Big) = \mathrm{E}(X_1)\,\mathrm{E}(X_2).
\end{aligned}$$

Sia ora $\mathrm{E}(X_i)$ finita $(i = 1, 2)$. Per il teorema precedente, $|X_1|$, $|X_2|$ sono indipendenti e quindi, per quanto appena provato, $\mathrm{E}(|X_1 X_2|) = \mathrm{E}(|X_1|\,|X_2|) = \mathrm{E}(|X_1|)\,\mathrm{E}(|X_2|) < +\infty$, ove la disuguaglianza sussiste in forza del Teorema A.3.5(ii). Ne segue, sempre per il medesimo teorema, che $X_1 X_2$ ha speranza matematica finita. Allora, la funzione $f : (x_1, x_2) \mapsto x_1 x_2$ è $\mathrm{P}_{X_1} \times \mathrm{P}_{X_2}$-integrabile e inoltre $\mathrm{P}_{X_1}(\{x_1 \in \mathbb{R} : f(x_1, \cdot) \text{ è } \mathrm{P}_{X_2}\text{-sommabile}\}) = \mathrm{P}_{X_1}(\mathbb{R}) = 1$. Conseguentemente, per il teorema di Fubini A.5.5,

$$\mathrm{E}(X_1 X_2) = \int_{\mathbb{R}^2} x_1 x_2\, \mathrm{P}_{X_1} \times \mathrm{P}_{X_2}(dx_1 \times dx_2) = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} x_2\, \mathrm{P}_{X_2}(dx_2) \Big) x_1\, \mathrm{P}_{X_1}(dx_1)$$

e quindi $\mathrm{E}(X_1 X_2) = \mathrm{E}(X_1)\,\mathrm{E}(X_2)$. ■

Per semplicità di esposizione ci siamo limitati a considerare sin qui solamente il caso delle coppie aleatorie. D'altra parte, dovrebbe essere del tutto evidente che le definizioni date si possono estendere "pari pari" (salvo la numerosità delle componenti) al caso delle m-ple aleatorie ($m \geq 3$) come pure che così facendo rimangono validi tutti i risultati ottenuti. Per quanto riguarda l'indipendenza, osserviamo che se gli enti aleatori $X_1, \ldots, X_m$ sono indipendenti, allora lo sono pure quelli ottenuti operando una selezione tra di essi. Considerata infatti la k-pla aleatoria $\mathbf{Y} = (X_1, \ldots, X_k)$ si ha

$$\mathbf{Y} \in \mathsf{A}_1 \times \cdots \times \mathsf{A}_k \iff \mathbf{X} \in \mathsf{A}_1 \times \cdots \times \mathsf{A}_k \times \mathfrak{X}_{k+1} \times \cdots \times \mathfrak{X}_m$$

e quindi

$$\begin{aligned} \mathrm{P}_{\mathbf{Y}}(\mathsf{A}_1 \times \cdots \times \mathsf{A}_k) &= \mathrm{P}_{\mathbf{X}}(\mathsf{A}_1 \times \cdots \times \mathsf{A}_k \times \mathfrak{X}_{k+1} \times \cdots \times \mathfrak{X}_m) \\ &= \mathrm{P}_{X_1}(\mathsf{A}_1) \cdots \mathrm{P}_{X_k}(\mathsf{A}_k)\, \mathrm{P}_{X_{k+1}}(\mathfrak{X}_{k+1}) \cdots \mathrm{P}_{X_m}(\mathfrak{X}_m) \\ &= \mathrm{P}_{X_1}(\mathsf{A}_1) \cdots \mathrm{P}_{X_k}(\mathsf{A}_k) \end{aligned}$$

per ogni $\mathsf{A}_i \in \aleph_i$ ($i = 1, \ldots, k$). Ne segue dalla simmetria della relazione d'indipendenza quanto dichiarato. Pertanto, nel caso particolare che la m-pla sia costituita da v.a. indipendenti e con speranza matematica finita, le sue componenti sono, per il Teorema B.1.15, a due e due non correlate.

Concludiamo la sezione con un risultato (di notevole interesse per le applicazioni) che riguarda la legge della somma di due v.a. indipendenti.

Teorema B.1.16 *Siano X_1, X_2 v.a. indipendenti e $Y = X_1 + X_2$. Allora, $\mathrm{P}_Y(B) = \int_{\mathbb{R}} \mathrm{P}_{X_2}(B - x_1)\, \mathrm{P}_{X_1}(dx_1)$ per ogni boreliano B. Inoltre, se f_i è una funzione di densità di X_i ($i = 1, 2$), la funzione di dominio la retta reale:*

$$t \mapsto \int_{-\infty}^{+\infty} f_1(x_1) f_2(t - x_1)\, dx_1$$

è una funzione di densità della v.a. somma.

Dimostrazione Per il Teorema B.1.13, la legge congiunta $\mathrm{P}_{\mathbf{X}}$ è il prodotto delle leggi marginali $\mathrm{P}_1 = \mathrm{P}_{X_1}$ e $\mathrm{P}_2 = \mathrm{P}_{X_2}$. Considerato allora $B \in \mathcal{B}$, dal Teorema B.1.10 e da quello di Tonelli si ha

$$\begin{aligned} \mathrm{P}_Y(B) &= \int_\Omega I_{\{Y \in B\}}\, d\mathrm{P} = \int_\Omega I_B(X_1 + X_2)\, d\mathrm{P} = \int_{\mathbb{R}^2} I_B(x_1 + x_2)\, \mathrm{P}_1 \times \mathrm{P}_2(dx_1 \times dx_2) \\ &= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} I_B(x_1 + x_2)\, \mathrm{P}_2(dx_2) \Big) \mathrm{P}_1(dx_1) \end{aligned}$$

da cui, osservato che $I_B(x_1 + x_2) = I_{B-x_1}(x_2)$, otteniamo

$$\mathrm{P}_Y(B) = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} I_{B-x_1}(x_2)\, \mathrm{P}_2(dx_2) \Big) \mathrm{P}_1(dx_1) = \int_{\mathbb{R}} \mathrm{P}_2(B - x_1)\, \mathrm{P}_1(dx_1).$$

Passando alla seconda parte della tesi, risulta

$$\mathrm{P}_Y(B) = \int_{\mathbb{R}} \Big(\int_{B-x_1} f_2(x_2)\, dx_2 \Big) f_1(x_1)\, dx_1.$$

Considerata allora la traslazione $\tau : t \mapsto t + x_1$ della retta reale, dal Corollario A.4.2 otteniamo

$$\int_{B-x_1} f_2(x_2)\, dx_2 = \int_{B-x_1} f_2(\tau(x_2) - x_1)\, dx_2 = \int_B f_2(t - x_1)\, dt$$

e quindi, per il teorema di Tonelli,

$$\begin{aligned} \mathrm{P}_Y(B) &= \int_{\mathbb{R}} \Big(\int_B f_2(t - x_1)\, dt \Big) f_1(x_1)\, dx_1 = \int_{\mathbb{R}\times B} f_1(x_1) f_2(t - x_1)\, dx_1 dt \\ &= \int_B \Big(\int_{\mathbb{R}} f_1(x_1) f_2(t - x_1)\, dx_1 \Big)\, dt. \end{aligned}$$

La dimostrazione è così conclusa. ■

B.2 Speranza matematica condizionata

L'introduzione delle nozioni di probabilità condizionata e di speranza matematica condizionata non presenta alcuna difficoltà quando l'evento subordinante ha probabilità positiva. Infatti, dato un evento H con $\mathrm{P}(H) > 0$, possiamo considerare la probabilità $\mathrm{P}(\cdot|H)$ su $\mathcal{A}$ così definita:

$$\mathrm{P}(A|H) = \frac{\mathrm{P}(A \cap H)}{\mathrm{P}(H)}$$

per ogni A, detta **probabilità condizionata a** H (interpretabile, da un punto di vista intuitivo, come la probabilità che si ottiene aggiornando la probabilità P "sapendo che l'evento H è vero"). Per constatare che siamo effettivamente in presenza di una probabilità basta osservare che $\mathrm{P}(\Omega|H) = 1$ e che, data una successione disgiunta di eventi $(A_n)_{n\geq 1}$,

$$\begin{aligned} \mathrm{P}(\bigcup_{n\geq 1} A_n|H) &= \frac{\mathrm{P}(H \cap \bigcup_{n\geq 1} A_n)}{\mathrm{P}(H)} = \frac{\mathrm{P}(\bigcup_{n\geq 1}(A_n \cap H))}{\mathrm{P}(H)} \\ &= \frac{\sum_{n\geq 1} \mathrm{P}(A_n \cap H)}{\mathrm{P}(H)} = \sum_{n\geq 1} \frac{\mathrm{P}(A_n \cap H)}{\mathrm{P}(H)} = \sum_{n\geq 1} \mathrm{P}(A_n|H). \end{aligned}$$

Mostriamo ora come l'integrazione rispetto alla probabilità condizionata possa essere ricondotta a quella rispetto alla probabilità originaria.

Teorema B.2.1 *Sia X una v.a.. Allora, X è $\mathrm{P}(\cdot|H)$-sommabile se e solo se XI_H è P-sommabile. Inoltre, nel caso di sommabilità,*

$$\int_\Omega X\, d\mathrm{P}(\cdot|H) = \frac{\int_H X\, dP}{\mathrm{P}(H)}.$$

DIMOSTRAZIONE Sia intanto $X \geq 0$. Considerata la misura $\mathrm{m} = \mathrm{P}(H)\,\mathrm{P}(\cdot|H)$, risulta $\mathrm{m}(H^c) = 0$ e quindi, per il Teorema A.3.4(iv), $\int_\Omega X\, d\mathrm{m} = \int_H X\, d\mathrm{m}$; inoltre, osservato che P, m coincidono sulla traccia $\mathcal{A} \cap H$ di $\mathcal{A}$ su H, otteniamo $\int_H X\, d\mathrm{m} = \int_H X\, d\mathrm{P}$. Tramite il Teorema A.3.4(ii) riesce allora

$$\int_\Omega X\, d\mathrm{P}(\cdot|H) = \int_\Omega X\, d\Big(\frac{1}{\mathrm{P}(H)}\,\mathrm{m}\Big) = \frac{1}{\mathrm{P}(H)}\int_\Omega X\, d\mathrm{m} = \frac{1}{\mathrm{P}(H)}\int_H X\, d\mathrm{m} = \frac{1}{\mathrm{P}(H)}\int_H X\, d\mathrm{P}.$$

Sia ora X qualsiasi. Poichè

$$\int_\Omega X^+\, d\mathrm{P}(\cdot|H) = \frac{\int_H X^+\, dP}{\mathrm{P}(H)} = \frac{\int_\Omega (XI_H)^+\, dP}{\mathrm{P}(H)}$$
$$\int_\Omega X^-\, d\mathrm{P}(\cdot|H) = \frac{\int_H X^-\, dP}{\mathrm{P}(H)} = \frac{\int_\Omega (XI_H)^-\, dP}{\mathrm{P}(H)},$$

la $\mathrm{P}(\cdot|H)$-sommabilità di X equivale alla P-sommabilità di XI_H. Inoltre, nel caso di sommabilità,

$$\int_\Omega X\, d\mathrm{P}(\cdot|H) = \int_\Omega X^+\, d\mathrm{P}(\cdot|H) - \int_\Omega X^-\, d\mathrm{P}(\cdot|H) = \frac{\int_H X^+\, d\mathrm{P} - \int_H X^-\, d\mathrm{P}}{\mathrm{P}(H)} = \frac{\int_H X\, d\mathrm{P}}{\mathrm{P}(H)}.$$

La dimostrazione è così conclusa. ∎

Tramite questo teorema possiamo introdurre, data una v.a. X con speranza matematica, la nozione di **speranza matematica condizionata di X a H** identificandola. come è naturale, con l'elemento della retta reale ampliata: $\mathrm{E}(X|H) = \int_\Omega X\, d\mathrm{P}(\cdot|H)$. Allora, sempre per il medesimo teorema,

$$\mathrm{E}(X|H) = \frac{\int_H X\, dP}{\mathrm{P}(H)} = \frac{\mathrm{E}(X\, I_H)}{\mathrm{P}(H)}. \tag{B.4}$$

Nel caso che l'evento subordinante abbia probabilità nulla, si possono ancora introdurre le relative nozioni di probabilità e speranza matematica condizionate ricorrendo, sotto convenienti condizioni di regolarità, ad opportuni passaggi al limite. Tuttavia, volendo introdurle anche in contesti più generali, conviene considerare il condizionamento non solo ad un singolo evento (come sinora fatto) ma alla famiglia degli *eventi osservabili* (cioè eventi aleatori il cui valore di verità è acquisibile tramite un prefissato *processo di osservazione*). Notato che sono osservabili sia la negazione di un evento osservabile che la disgiunzione discreta di eventi osservabili, viene naturale ricorrere alle σ-algebre per descrivere formalmente l'*informazione ottenibile* tramite il processo di osservazione. Prendendo spunto da queste considerazioni, nella sezione seguente introduciamo una nozione di speranza matematica condizionata a una σ-algebra di eventi che può ritenersi uno dei concetti più profondi e potenti (sia dal punto di vista teorico che applicativo) della teoria delle probabilità.

B.2.1 Condizionamento a σ-algebre

Iniziamo con un semplice esempio per enucleare le motivazioni intuitive che conducono a formulare la nozione generale di speranza matematica condizionata all'informazione ottenibile. Assumiamo dunque che tale informazione sia rappresentata dalla σ-algebra $\mathcal{H}$ generata da una partizione finita di Ω costituita dagli eventi $H_1, \ldots, H_n \in \mathcal{A}$. Supponiamo inoltre che X sia una v.a. con speranza matematica finita. Dato un qualsiasi evento $H \in \mathcal{H}$, esistono $i_1, \ldots, i_m \in \{1, \ldots, n\}$ tali che $H = \bigcup_{h=1}^m H_{i_h}$ (Esempio A.1.1(i)). Conseguentemente, posto $J_H = \{h : \mathrm{P}(H_{i_h}) > 0\}$ e notato che $\mathrm{E}(XI_{H_{i_h}}) = \int_{H_{i_h}} X\, d\mathrm{P} = 0$ per ogni $h \notin J_H$ (Teorema A.3.4(iv)), dalla linearità della speranza matematica otteniamo

$$\mathrm{E}(XI_H) = \mathrm{E}\big(XI_{\bigcup_{h=1}^m H_{i_h}}\big) = \mathrm{E}\big(X\sum_{h=1}^m I_{H_{i_h}}\big) = \sum_{h=1}^m \mathrm{E}(XI_{H_{i_h}}) = \sum_{h\in J_H} \mathrm{E}(XI_{H_{i_h}})$$

e quindi, tramite (B.4),

$$\int_H X\, d\mathrm{P} = \sum_{h\in J_H} \mathrm{E}(X|H_{i_h})\, \mathrm{P}(H_{i_h}).$$

Considerata ora la $\mathcal{H}$-funzione semplice:

$$Y_0 = \sum_{i\in\{i:\mathrm{P}(H_i)>0\}} \mathrm{E}(X|H_i) I_{H_i},$$

risulta $Y_0 I_H = \sum_{h \in J_H} \mathrm{E}(X|H_{i_h})\, I_{H_{i_h}}$. Ne segue, per la linearità dell'integrale, $\int_H Y_0\, d\mathrm{P} = \sum_{h \in J_H} \mathrm{E}(X|H_{i_h})\mathrm{P}(H_{i_h})$ e quindi l'uguaglianza

$$\int_H X\, d\mathrm{P} = \int_H Y_0\, d\mathrm{P}. \tag{B.5}$$

Le speranze matematiche condizionate di X ai costituenti non trascurabili permettono dunque di costruire una v.a. $\mathcal{H}$-Borel misurabile Y_0 che verifica l'uguaglianza (B.5) per ogni evento osservabile H. È interessante a questo punto notare che ogni v.a. $\mathcal{H}$-Borel misurabile Y che (analogamente a Y_0) verifica la proprietà "$\mathrm{E}(XI_H) = \mathrm{E}(YI_H)$ per ogni evento osservabile H" coincide quasi certamente con Y_0 (Teorema A.4.4); assume su ogni costituente H_i di probabilità positiva valore costante pari alla speranza matematica condizionata $\mathrm{E}(X|H_i)$[17]; consente di calcolare la speranza matematica condizionata di X a un qualsiasi evento osservabile H di probabilità positiva tramite la relazione $\mathrm{E}(X|H) = \frac{\mathrm{E}(YI_H)}{\mathrm{P}(H)}$.

Appare quindi ragionevole ritenere che l'assunzione congiunta della $\mathcal{H}$-Borel misurabilità e della validità di (B.5) per ogni $H \in \mathcal{H}$ possa essere intesa come caratterizzante la speranza matematica condizionata all'informazione ottenibile (rappresentata, nel caso particolare in esame, dalla σ-algebra generata).

Adottando questo punto di vista, consideriamo una qualsiasi σ-algebra $\mathcal{H} \subseteq \mathcal{A}$ i cui elementi, denotati con la lettera H (dotata o no di apici o pedici), chiamiamo **eventi osservabili**. Data allora una v.a. X con speranza matematica finita, la **speranza matematica condizionata di** X **a** $\mathcal{H}$ è una qualsiasi v.a. estesa $\mathrm{E}(X|\mathcal{H})$ tale che:

1. sia $\mathcal{H}$-Borel misurabile[18];
2. risulti $\int_H X\, d\mathrm{P} = \int_H \mathrm{E}(X|\mathcal{H})\, d\mathrm{P}$ per ogni evento osservabile H.[19]

---

[17] Per constatarlo, fissato $\omega_i \in H_i$, sia $y_i = Y(\omega_i)$. Allora, $\omega_i \in \{Y = y_i\} \in \mathcal{H}$ e quindi esiste, per l'Esempio A.1.1(i), $J \subseteq \{1, \ldots, n\}$ tale che $\{Y = y_i\} = \bigcup_{j \in J} H_j$. Riesce dunque $\omega_i \in H_i \cap \{Y = y_i\} = H_i \cap \bigcup_{j \in J} H_j = \bigcup_{j \in J}(H_i \cap H_j)$ da cui otteniamo $H_i = H_j$ per qualche $j \in J$ e quindi $H_i \subseteq \{Y = y_i\}$. Pertanto, Y assume su H_i valore costante y_i. Supposto infine $\mathrm{P}(H_i) > 0$, dalla $\int_{H_i} X\, d\mathrm{P} = \int_{H_i} Y\, d\mathrm{P} = y_i\, \mathrm{P}(H_i)$, si ha $y_i = \mathrm{E}(X|H_i)$.

[18] E quindi **osservabile** (cioè di valore noto qualora sia conosciuto il valore di verità di ogni evento osservabile) in quanto gli eventi $\{\mathrm{E}(X|\mathcal{H}) = y\}$ sono, al variare di $y \in \mathbb{R}^*$, tutti osservabili.

[19] Una definizione analoga viene data anche nel caso di v.a. estese. Comunque, possiamo

Notato che la probabilità $X * \mathrm{P}\big|_{\mathcal{H}}$ su $\mathcal{H}$ si annulla sull'osservabile H ogniqualvolta risulti $\mathrm{P}(H) = 0$ (Teorema A.3.4(iv)), dal teorema di Radon-Nikodym[20] otteniamo intanto l'esistenza di $\mathrm{E}(X|\mathcal{H})$. Inoltre, per i teoremi A.3.6(vii) e A.4.4, possiamo anche affermare che $\mathrm{E}(X|\mathcal{H})$ è finita quasi certamente e risulta definita a meno di eventi osservabili di probabilità nulla[21]. Infine, il risultato seguente assicura che $\mathrm{E}(X|\mathcal{H})$ assume quasi certamente il valore $\mathrm{E}(X|H)$ su H, se H è un P-**atomo** (in breve **atomo**) di $\mathcal{H}$, cioè un evento osservabile di probabilità positiva non ripartibile in due eventi osservabili ancora di probabilità positiva[22].

Teorema B.2.2 *Sia H un* P-*atomo di $\mathcal{H}$. Esiste allora $H_1 \subseteq H$ tale che $\mathrm{P}(H_1) = \mathrm{P}(H)$ e $\mathrm{E}(X|\mathcal{H})(\omega) = \mathrm{E}(X|H)$ per ogni $\omega \in H_1$*[23]*. In particolare, se $\mathcal{H}$ è la σ-algebra generata da una partizione finita di Ω costituita dagli eventi $H_1, \ldots, H_m$ con $\mathrm{P}(H_i) > 0$ $(i = 1, \ldots, m)$, allora $\mathrm{E}(X|\mathcal{H})(\omega) = \mathrm{E}(X|H_i)$ per ogni $\omega \in H_i$ $(i = 1, \ldots, m)$.*

Dimostrazione Posto $Y = \mathrm{E}(X|\mathcal{H})$, basta verificare che Y è costante su un sottoinsieme $H_1 \subseteq H$ tale che $\mathrm{P}(H_1) = \mathrm{P}(H)$. Infatti, supposto $Y(\omega) = \alpha$ per ogni $\omega \in H_1$, dalla $\mathrm{P}(H \setminus H_1) = 0$ otteniamo $YI_H = \alpha I_H$ (P-q.c.). Ne segue, per il Teorema B.1.5(iii), $\mathrm{E}(XI_H) = \mathrm{E}(YI_H) = \mathrm{E}(\alpha I_H) = \alpha \mathrm{P}(H)$ e quindi, per (B.4), $\alpha = \frac{\mathrm{E}(XI_H)}{\mathrm{P}(H)} = \mathrm{E}(X|H)$, ricordato che $\mathrm{P}(H) > 0$.

---

limitarci a considerare solo v.a. in quanto, essendo richiesta la finitezza della speranza matematica, una v.a. estesa è sempre uguale quasi certamente ad una v.a. (Teorema A.3.6(vii)).

[20]Per *teorema di Radon-Nikodym* intendiamo il seguente risultato (da considerarsi, per le sue conseguenze, come un'altra "pietra miliare" della teoria della misura): Con riferimento alla σ-algebra $\mathcal{A}$, siano m una misura σ-finita e $\varphi : \mathcal{A} \mapsto \mathbb{R}^*$ una funzione d'insieme numerabilmente additiva m-*assolutamente continua* (cioè tale che l'implicazione "$\mathrm{m}(A) = 0 \Rightarrow \varphi(A) = 0$" sussista per ogni A). Esiste allora una funzione $\mathcal{A}$-Borel misurabile f a valori nella retta reale ampliata tale che $\varphi(A) = \int_A f \, d\mathrm{m}$ per ogni A (per una dimostrazione si veda la sezione 2.2 del testo di Ash).

Cogliamo l'occasione per osservare che, data una misura σ-finita μ di riferimento su $\aleph$, gli enti aleatori a valori in $\mathfrak{X}$ che ammettono μ-densità coincidono, per questo teorema e il Teorema A.3.4(iv), con quelli aventi legge μ-assolutamente continua.

[21]Conseguentemente, $\mathrm{E}(X|\mathcal{H})$ non individua una v.a. ben precisata ma denota piuttosto l'elemento generico della classe delle speranze matematiche condizionate di X a $\mathcal{H}$. Per questo motivo useremo, nel seguito, la locuzione "Z è una versione di $\mathrm{E}(X|\mathcal{H})$" per indicare che Z è una di queste speranze matematiche condizionate.

[22]In altri termini, per ogni $H_1 \subset H$ si ha $\mathrm{P}(H_1) = 0$ o $\mathrm{P}(H_1) = \mathrm{P}(H)$.

[23]L'ipotesi di atomicità non è in generale rimovibile. Infatti, posto $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B} \cap \Omega$ e $\mathrm{P} = \lambda\big|_{\mathcal{A}}$, sia $\mathcal{H}$ la σ-algebra generata dalla partizione $H_1 = \left[0, \frac{1}{4}\right[$, $H_2 = \left]\frac{1}{4}, \frac{3}{4}\right]$, $H_3 = \left]\frac{3}{4}, 1\right]$ e X la funzione identica di Ω. Allora, $\mathrm{E}(X|H_1 \cup H_2) = \frac{3}{8}$ e $\mathrm{E}(X|\mathcal{H})(\omega) = \frac{1}{8}$, se $\omega \in H_1$, e $\mathrm{E}(X|\mathcal{H})(\omega) = \frac{1}{2}$, se $\omega \in H_2$.

Per individuare H_1, osservato che $\mathrm{P}(H \cap \{Y < -\infty\}) = 0$, poniamo:

$$\beta = \sup\{x \in \mathbb{R}^* : \mathrm{P}(H \cap \{Y < x\}) = 0\}$$

e verifichiamo intanto che $\mathrm{P}(H \cap \{Y < \beta\}) = 0$. Essendo l'uguaglianza ovvia per $\beta = -\infty$, supponiamo $\beta > -\infty$. Esiste allora una successione crescente di numeri reali $(x_n)_{n\geq 1}$ tale che $x_n \uparrow \beta$. Poichè $x_n < \beta$, esiste un numero reale x tale che $x_n < x$ e $\mathrm{P}(H \cap \{Y < x\}) = 0$. Riesce dunque, per la monotonia della probabilità, $\mathrm{P}(H \cap \{Y < x_n\}) \leq \mathrm{P}(H \cap \{Y < x\}) = 0$. Ne segue, per la subadditività della probabilità,

$$\mathrm{P}(H \cap \{Y < \beta\}) = \mathrm{P}\big(\bigcup_{n\geq 1} \{H \cap \{Y < x_n\}\big) \leq \sum_{n\geq 1} \mathrm{P}(H \cap \{Y < x_n\}) = 0.$$

Proviamo ora che $\mathrm{P}(H \cap \{Y > \beta\}) = 0$. Essendo l'uguaglianza ovvia per $\beta = +\infty$, supponiamo $\beta < +\infty$. Esiste allora una successione decrescente di numeri reali $(x'_n)_{n\geq 1}$ tale che $x'_n \downarrow \beta$. Poichè $x'_n > \beta$, si ha $\mathrm{P}(H \cap \{Y < x'_n\}) > 0$ da cui otteniamo, osservato che $H \cap \{Y < x'_n\} \in \mathcal{H}$ e ricordato che H è un atomo, $\mathrm{P}(H \cap \{Y < x'_n\}) = \mathrm{P}(H)$ e quindi, per l'additività della probabilità, $\mathrm{P}(H \cap \{Y \geq x'_n\}) = 0$. Ne segue, per la subadditività della probabilità,

$$\mathrm{P}(H \cap \{Y > \beta\}) = \mathrm{P}\big(\bigcup_{n\geq 1} \{H \cap \{Y \geq x'_n\}\big) \leq \sum_{n\geq 1} \mathrm{P}(H \cap \{Y \geq x'_n\}) = 0.$$

Riesce dunque $\mathrm{P}(H \cap \{Y \neq \beta\}) = \mathrm{P}(H \cap \{Y < \beta\}) + \mathrm{P}(H \cap \{Y > \beta\}) = 0$. Posto infine $H_1 = H \cap \{Y = \beta\} \in \mathcal{H}$, si ha $\mathrm{P}(H_1) = \mathrm{P}(H)$ e Y costante su H_1.

Per quanto riguarda l'ultima parte della tesi, basta osservare che, per la nota 17 di p. 185, le funzioni $\mathcal{H}$-Borel misurabili sono costanti sui costituenti $H_1, \ldots, H_m$ e che questi ultimi sono banalmente degli atomi di $\mathcal{H}$. ■

Nel teorema successivo elenchiamo ulteriori proprietà della speranza matematica condizionata all'informazione ottenibile. In particolare, le proposizioni (xi) ÷ (xiii) forniscono le versioni condizionate dei teoremi della convergenza monotona A.3.7, d'integrazione per serie A.3.8 e della convergenza dominata A.3.11. Per quanto riguarda le uguaglianze considerate nelle proposizioni (ii) e (iii), possiamo interpretarle, rispettivamente, nel modo seguente: qualora l'informazione ottenibile sia quella:
- minima (consistente solamente nei valori di verità degli eventi certo e impossibile), di X possiamo conoscere unicamente la speranza matematica;
- massima (consistente nel sapere quale caso elementare è vero), di X possiamo conoscere il suo valore "vero".

La proposizione (xiv) assicura invece che le v.a. limitate e osservabili si comportano nei confronti della speranza matematica condizionata come le costanti rispetto alla speranza matematica, nel senso che entrambe possono essere "estratte" dall'operatore di speranza matematica. Infine, la proposizione (xv) mette in evidenza che il risultato del condizionamento sequenziale relativo a due σ-algebre di eventi osservabili $\mathcal{K} \subset \mathcal{H}$ non dipende da quale condizionamento venga effettuato per primo ma solamente dalla più "piccola" informazione ottenibile (rappresentata da $\mathcal{K}$).

Teorema B.2.3 *Siano X, Y, X_1, X_2, etc. v.a. con speranza matematica finita. Riesce allora:*

(i) $\mathrm{E}\big(\mathrm{E}(X|\mathcal{H})\big) = \mathrm{E}(X)$;

(ii) $\mathrm{E}(X|\{\emptyset, \Omega\}) = \mathrm{E}(X)$.

Valgono inoltre le proposizioni seguenti nelle quali le relazioni riguardanti speranze matematiche condizionate devono intendersi sussistere P-*quasi certamente.*

(iii) $\mathrm{E}(X|\mathcal{A}) = X$;

(iv) *Se X è anche $\mathcal{H}$-Borel misurabile, allora* $\mathrm{E}(X|\mathcal{H}) = X$. *In particolare,* $\mathrm{E}(\alpha I_\Omega|\mathcal{H}) = \alpha$ *per ogni numero reale* α;

(v) MONOTONIA: $\mathrm{E}(X|\mathcal{H}) \leq \mathrm{E}(Y|\mathcal{H})$, *se* $X \leq Y$ (P-q.c.);

(vi) $\mathrm{E}(X|\mathcal{H}) = \mathrm{E}(Y|\mathcal{H})$, *se* $X = Y$ (P-q.c.);

(vii) INTERNALITÀ: $a \leq \mathrm{E}(X|\mathcal{H}) \leq b$, *se* $a \leq X \leq b$ (P-q.c.);

(viii) INTERNALITÀ STRETTA: $a < \mathrm{E}(X|\mathcal{H}) < b$, *se* $a < X < b$ (P-q.c.);

(ix) LINEARITÀ: $\mathrm{E}\big(\sum_{i=1}^n \alpha_i X_i|\mathcal{H}\big) = \sum_{i=1}^n \alpha_i \mathrm{E}(X_i|\mathcal{H})$ *qualunque siano i numeri reali* $\alpha_1, \ldots, \alpha_n$;

(x) $\big|\mathrm{E}(X|\mathcal{H})\big| \leq \mathrm{E}(|X|\,|\mathcal{H})$;

(xi) $\mathrm{E}(X_n|\mathcal{H}) \uparrow \mathrm{E}(X|\mathcal{H})$, *se* $X_n \uparrow X$ *e* $X_1 \geq 0$ (P-q.c.);

(xii) $\mathrm{E}\big(\sum_{n\geq 1} \alpha_n X_n|\mathcal{H}\big) = \sum_{n\geq 1} \alpha_n \mathrm{E}(X_n|\mathcal{H})$, *se* $(\alpha_n)_{n\geq 1}$ *è una successione di numeri reali non negativi,* $X_n \geq 0$ *per ogni* n *e la speranza matematica della serie* $\sum_{n\geq 1} \alpha_n X_n$ *è finita;*

(xiii) $\mathrm{E}(X_n|\mathcal{H}) \to \mathrm{E}(X|\mathcal{H})$, *se* $|X_n| \leq Z$ (P-q.c.) *per ogni* n, $\mathrm{E}(Z)$ *finita e* $X_n \to X$;

(xiv) $\mathrm{E}(Z\,X|\mathcal{H}) = Z\,\mathrm{E}(X|\mathcal{H})$, *se* Z *è* $\mathcal{H}$*-Borel misurabile e limitata;*

(xv) *Se* $\mathcal{K}$ *è una sotto* σ*-algebra di* $\mathcal{H}$*, allora* $\mathrm{E}\big(\mathrm{E}(X|\mathcal{H})|\mathcal{K}\big) = \mathrm{E}(X|\mathcal{K})$ *e* $\mathrm{E}\big(\mathrm{E}(X|\mathcal{K})|\mathcal{H}\big) = \mathrm{E}(X|\mathcal{K})$.

Dimostrazione Le proposizioni (i) $\div$ (iv) sono conseguenza immediata della definizione. Dalla proposizione (v) discendono invece (vi), (vii) (tenuto conto di (iv)) come pure (x) (tenuto presente (ix) e la disuguaglianza $-|X| \le X \le |X|$).

(v) Sia $X \le Y$ (P-q.c.). Allora, considerato un evento osservabile H, dalla monotonia dell'integrale si ha $\int_H \mathrm{E}(X|\mathcal{H})\,d\mathrm{P} = \int_H X\,d\mathrm{P} \le \int_H Y\,d\mathrm{P} = \int_H \mathrm{E}(Y|\mathcal{H})\,d\mathrm{P}$. Ne segue, per il Teorema A.4.4 (ponendo $\mathcal{A} = \mathcal{H}$), $\mathrm{E}(X|\mathcal{H}) \le \mathrm{E}(Y|\mathcal{H})$ (P-q.c.).

(viii) Per la linearità dell'integrale, $\mathrm{E}(X|\mathcal{H}) - a$ è una versione di $\mathrm{E}(X - a|\mathcal{H})$. Inoltre, posto $A = \{a < X < b\}$, $H = \{a < \mathrm{E}(X|\mathcal{H}) < b\}$, $H_1 = \{\mathrm{E}(X|\mathcal{H}) \le a\}$ e $H_2 = \{\mathrm{E}(X|\mathcal{H}) \ge b\}$, otteniamo $\mathrm{P}(A) = 1$ e $\mathrm{P}(H^c) = \mathrm{P}(H_1) + \mathrm{P}(H_2)$. Ciò osservato, assumiamo (per assurdo) $\mathrm{P}(H_1) > 0$. Allora, $\mathrm{P}(A^c \cap H_1) = 0$, $\mathrm{P}(A \cap H_1) = \mathrm{P}(H_1) > 0$ da cui, tramite i teoremi A.3.4(iv) e A.3.6(i),(vi), risulta

$$0 \ge \int_{H_1} \big(\mathrm{E}(X|\mathcal{H}) - a\big)\,d\mathrm{P} = \int_{H_1} (X - a)\,d\mathrm{P} = \int_{A\cap H_1} (X - a)\,d\mathrm{P} > 0$$

pervenendo così ad una contraddizione. La prova di $\mathrm{P}(H_2) = 0$ è del tutto analoga.

(ix) Per la linearità della speranza matematica, la v.a. somma ha speranza matematica finita per cui ha senso considerare $\mathrm{E}(\sum_{i=1}^n \alpha_i X_i|\mathcal{H})$. Ora, essendo le speranze matematiche condizionate finite quasi certamente, per l'osservabile $H' = \bigcap_{i=1}^n \big\{\mathrm{E}(X_i|\mathcal{H}) \in \mathbb{R}\big\}$ risulta $\mathrm{P}(H') = 1$. Considerata allora, per ogni n, la v.a. $X_i' = \mathrm{E}(X_i|\mathcal{H}) I_{H'}$, otteniamo $X_i' = \mathrm{E}(X_i|\mathcal{H})$ (P-q.c.) $(i = 1, \ldots, n)$ e $\sum_{i=1}^n \alpha_i X_i' = \sum_{i=1}^n \alpha_i \mathrm{E}(X_i|\mathcal{H})$ (P-q.c.). Ne segue, per le solite proprietà dell'integrale,

$$\begin{aligned}\int_H \sum_{i=1}^n \alpha_i X_i\,d\mathrm{P} &= \sum_{i=1}^n \alpha_i \int_H X_i\,d\mathrm{P} = \sum_{i=1}^n \alpha_i \int_H \mathrm{E}(X_i|\mathcal{H})\,d\mathrm{P}\\ &= \sum_{i=1}^n \alpha_i \int_H X_i'\,d\mathrm{P} = \int_H \sum_{i=1}^n \alpha_i X_i'\,d\mathrm{P} = \int_H \sum_{i=1}^n \alpha_i \mathrm{E}(X_i|\mathcal{H})\,d\mathrm{P}\end{aligned}$$

qualunque sia H. Dunque, $\mathrm{E}(\sum_{i=1}^n \alpha_i X_i|\mathcal{H}) = \sum_{i=1}^n \alpha_i \mathrm{E}(X_i|\mathcal{H})$ (P-q.c.), osservato che la combinazione lineare $\sum_{i=1}^n \alpha_i \mathrm{E}(X_i|\mathcal{H})$ è $\mathcal{H}$-Borel misurabile.

(xi) Sia $X_1 \ge 0$ (P-q.c.) e $X_n \uparrow X$. Allora, $X_n \ge 0$ (P-q.c.) per ogni n. Risulta quindi, per (v), (vii), $\mathrm{P}(\mathrm{E}(X_n|\mathcal{H}) \ge 0) = 1$ e $\mathrm{P}\big(\mathrm{E}(X_n|\mathcal{H}) \le \mathrm{E}(X_{n+1}|\mathcal{H})\big) = 1$ per ogni n. Conseguentemente, posto $H' = \bigcap_{n\ge1} \big\{0 \le \mathrm{E}(X_n|\mathcal{H}) \le \mathrm{E}(X_{n+1}|\mathcal{H})\big\}$, si ha $\mathrm{P}(H') = 1$. Considerata allora, per ogni n, la v.a. $X_n' = \mathrm{E}(X_n|\mathcal{H}) I_{H'}$, otteniamo $0 \le X_n' \le X_{n+1}'$, $X_n' = \mathrm{E}(X_n|\mathcal{H})$ (P-q.c.) per ogni n e $\lim_{n\to+\infty} X_n' = \lim_{n\to+\infty} \mathrm{E}(X_n|\mathcal{H})$ (P-q.c.). Ne segue, per il teorema della convergenza monotona,

$$\begin{aligned}\int_H X\,d\mathrm{P} &= \lim_{n\to+\infty} \int_H X_n\,d\mathrm{P} = \lim_{n\to+\infty} \int_H \mathrm{E}(X_n|\mathcal{H})\,d\mathrm{P}\\ &= \lim_{n\to+\infty} \int_H X_n'\,d\mathrm{P} = \int_H \lim_{n\to+\infty} X_n'\,d\mathrm{P} = \int_H \lim_{n\to+\infty} \mathrm{E}(X_n|\mathcal{H})\,d\mathrm{P}\end{aligned}$$

per ogni H. Dunque, $\mathrm{E}(X_n|\mathcal{H}) \to \mathrm{E}(X|\mathcal{H})$ (P-q.c.), osservato che il limite $\lim\limits_{n\to+\infty} \mathrm{E}(X_n|\mathcal{H})$ è $\mathcal{H}$-Borel misurabile. Allora, $\mathrm{E}(X_n|\mathcal{H}) \uparrow \mathrm{E}(X|\mathcal{H})$ (P-q.c.), essendo $H' \cap \{\mathrm{E}(X_n|\mathcal{H}) \to \mathrm{E}(X|\mathcal{H})\} \subseteq \{\mathrm{E}(X_n|\mathcal{H}) \uparrow \mathrm{E}(X|\mathcal{H})\}$.

(xii) Siano $X_n \geq 0$, $\alpha_n \geq 0$ per ogni n e $\mathrm{E}\big(\sum_{n\geq1} \alpha_n X_n\big) < +\infty$. Allora, per (vii), $\mathrm{P}(\mathrm{E}(X_n|\mathcal{H}) \geq 0) = 1$ per ogni n. Conseguentemente, posto $H' = \bigcap_{n\geq1}\{\mathrm{E}(X_n|\mathcal{H}) \geq 0\}$, si ha $\mathrm{P}(H') = 1$. Considerata allora, per ogni n, la v.a. $X'_n = \mathrm{E}(X_n|\mathcal{H})I_{H'}$, riesce $X'_n \geq 0$, $X'_n = \mathrm{E}(X_n|\mathcal{H})$ (P-q.c.) per ogni n e $\sum_{n\geq1} X'_n = \sum_{n\geq1} \mathrm{E}(X_n|\mathcal{H})$ (P-q.c.). Ne segue, per il teorema d'integrazione per serie,

$$\begin{aligned}\int_H \sum_{n\geq1} \alpha_n X_n \, d\mathrm{P} &= \sum_{n\geq1} \alpha_n \int_H X_n \, d\mathrm{P} = \sum_{n\geq1} \alpha_n \int_H \mathrm{E}(X_n|\mathcal{H}) \, d\mathrm{P} \\ &= \sum_{n\geq1} \alpha_n \int_H X'_n \, d\mathrm{P} = \int_H \sum_{n\geq1} \alpha_n X'_n \, d\mathrm{P} = \int_H \sum_{n\geq1} \alpha_n \mathrm{E}(X_n|\mathcal{H}) \, d\mathrm{P}\end{aligned}$$

per ogni H. Dunque, $\mathrm{E}(\sum_{n\geq1} \alpha_n X_n|\mathcal{H}) = \sum_{n\geq1} \alpha_n \mathrm{E}(X_n|\mathcal{H})$ (P-q.c.), osservato che la serie $\sum_{n\geq1} \mathrm{E}(X_n|\mathcal{H})$ è $\mathcal{H}$-Borel misurabile.

(xiii) Siano $|X_n| \leq Z$ (P-q.c.) per ogni n, $\mathrm{E}(Z)$ finita e $X_n \to X$. Allora, posto $A = \bigcap_{n\geq1}\{|X_n| \leq Z\}$, si ha $\mathrm{P}(A) = 1$, $|X_n|\big|_A \leq Z\big|_A$ per ogni n e quindi $|X|\big|_A \leq Z\big|_A$.

Consideriamo ora, per ogni n, la v.a. $Z_n = \sup_{m\geq n} |X_m - X|$. Allora, $Z_n \downarrow 0$; inoltre, $Z_n\big|_A \leq \sup_{m\geq n}(|X_m|\big|_A + |X|\big|_A) \leq 2Z\big|_A$ da cui otteniamo $Z_n \leq 2Z$ (P-q.c.) e quindi Z_n ha speranza matematica finita. Possiamo dunque considerare la successione $\big(\mathrm{E}(Z_n|\mathcal{H})\big)_{n\geq1}$. Osservato infine che $\mathrm{E}(X_n|\mathcal{H}) - \mathrm{E}(X|\mathcal{H}) = \mathrm{E}(X_n - X|\mathcal{H})$ (P-q.c.) (per (ix)), $\big|\mathrm{E}(X_n - X|\mathcal{H})\big| \leq \mathrm{E}(|X_n - X|\,|\mathcal{H})$ (P-q.c.) (per (x)) e $\mathrm{E}(|X_n - X|\,|\mathcal{H}) \leq \mathrm{E}(Z_n|\mathcal{H})$ (P-q.c.) (per (v)), otteniamo $\big|\mathrm{E}(X_n|\mathcal{H}) - \mathrm{E}(X|\mathcal{H})\big| \leq \mathrm{E}(Z_n|\mathcal{H})$ (P-q.c.). Basta dunque provare che $\mathrm{E}(Z_n|\mathcal{H}) \to 0$ (P-q.c.), notato che $\big\{\mathrm{E}(Z_n|\mathcal{H}) \to 0\big\} \subseteq \big\{|\mathrm{E}(X_n|\mathcal{H}) - \mathrm{E}(X|\mathcal{H})| \to 0\big\} = \big\{\mathrm{E}(X_n|\mathcal{H}) \to \mathrm{E}(X|\mathcal{H})\big\}$.

Posto ora $H' = \bigcap_{n\geq1}\big\{\mathrm{E}(Z_n|\mathcal{H}) \geq \mathrm{E}(Z_{n+1}|\mathcal{H}) \geq 0\big\}$, da (v), (vii) si ha $\mathrm{P}(H') = 1$. Conseguentemente, posto $Z'_n = \mathrm{E}(Z_n|\mathcal{H})I_{H'} \geq 0$ per ogni n, la successione $(Z'_n)_{n\geq1}$ è non crescente e quindi possiamo considerare la v.a. $Z' = \lim\limits_{n\to+\infty} Z'_n \geq 0$. Osservato che $Z'_1 = \mathrm{E}(Z_1|\mathcal{H})$ (P-q.c.), risulta $\mathrm{E}(Z'_1) = \mathrm{E}\big(\mathrm{E}(Z_1|\mathcal{H})\big) = \mathrm{E}(Z_1) \in \mathbb{R}$ e quindi, per il teorema della convergenza dominata e (i),

$$0 \leq \mathrm{E}(Z') = \lim_{n\to+\infty} \mathrm{E}(Z'_n) = \lim_{n\to+\infty} \mathrm{E}\big(\mathrm{E}(Z_n|\mathcal{H})\big) = \lim_{n\to+\infty} \mathrm{E}(Z_n) = \int_\Omega \lim_{n\to+\infty} Z_n \, d\mathrm{P} = 0,$$

ricordato che $Z_n \downarrow 0$. Dunque, $\mathrm{E}(Z') = 0$ e quindi, posto $H'' = \{Z' = 0\}$, si ha $\mathrm{P}(H'') = 1$ (Teorema A.3.6(v)). Allora, $\mathrm{E}(Z_n|\mathcal{H})\big|_{H'\cap H''} \to 0$ con $\mathrm{P}(H' \cap H'') = 1$, cioè $\mathrm{E}(Z_n|\mathcal{H}) \to 0$ (P-q.c.).

(xiv) Sia Z una v.a. limitata $\mathcal{H}$-Borel misurabile. Osservato che $Z\,\mathrm{E}(X|\mathcal{H})$ è $\mathcal{H}$-Borel misurabile, poniamo $|Z| \leq \alpha \in \mathbb{R}$. Allora, $\mathrm{E}(|ZX|) \leq \mathrm{E}(\alpha|X|) = \alpha\mathrm{E}(|X|) < +\infty$ e quindi ZX ha speranza matematica finita (Teorema A.3.5(ii)). Ciò osservato, assumiamo intanto $Z = I_{H'}$. Allora,

$$\int_H ZX \, d\mathrm{P} = \int_{H\cap H'} X \, d\mathrm{P} = \int_{H\cap H'} \mathrm{E}(X|\mathcal{H}) \, d\mathrm{P} = \int_H Z\,\mathrm{E}(X|\mathcal{H}) \, d\mathrm{P}$$

per ogni H; dunque $\mathrm{E}(ZX|\mathcal{H}) = Z\,\mathrm{E}(X|\mathcal{H})$ (P-q.c.).
Sia ora Z una funzione $\mathcal{H}$-semplice a valori finiti. Da (ix) otteniamo

$$\begin{aligned}\mathrm{E}(ZX|\mathcal{H}) &= \mathrm{E}\big(\big(\sum_{z\in Z(\Omega)} zI_{\{Z=z\}}\big)X|\mathcal{H}\big)\\ &= \sum_{z\in Z(\Omega)} z\mathrm{E}(I_{\{Z=z\}}X|\mathcal{H}) = \sum_{z\in Z(\Omega)} zI_{\{Z=z\}}\mathrm{E}(X|\mathcal{H}) = Z\,\mathrm{E}(X|\mathcal{H}) \quad \text{(P-q.c.)}.\end{aligned}$$

Sia infine Z qualsiasi. Per il lemma fondamentale A.2.4, esiste una successione $(Z_n)_{n\geq 1}$ di funzioni $\mathcal{H}$-semplici a valori finiti tale che $Z_n \to Z$ e $|Z_n| \leq |Z| \leq \alpha$ per ogni n. Possiamo allora considerare, per ogni n, la speranza matematica condizionata $\mathrm{E}(Z_nX|\mathcal{H})$ in quanto $\mathrm{E}(Z_nX)$ è finita. Posto $H' = \bigcap_{n\geq 1}\big\{\mathrm{E}(Z_nX|\mathcal{H}) = Z_n\mathrm{E}(X|\mathcal{H})\big\}$ e osservato che $\mathrm{E}(Z_nX|\mathcal{H}) = Z_n\mathrm{E}(X|\mathcal{H})$ (P-q.c.) per ogni n, otteniamo $\mathrm{P}(H') = 1$ e $H' \subseteq \{\mathrm{E}(Z_nX|\mathcal{H}) \to Z\,\mathrm{E}(X|\mathcal{H})\}$. D'altra parte, notato che $|Z_nX| \leq \alpha|X|$ per ogni n e $Z_nX \to ZX$, da (xiii) segue $\mathrm{E}(Z_nX|\mathcal{H}) \to \mathrm{E}(ZX|\mathcal{H})$ (P-q.c.). Allora, posto $H'' = \big\{\mathrm{E}(Z_nX|\mathcal{H}) \to \mathrm{E}(ZX|\mathcal{H})\big\}$, si ha $\mathrm{P}(H' \cap H'') = 1$ e $H' \cap H'' \subseteq \big\{Z\,\mathrm{E}(X|\mathcal{H}) = \mathrm{E}(ZX|\mathcal{H})\big\}$, cioè $\mathrm{E}(ZX|\mathcal{H}) = Z\,\mathrm{E}(X|\mathcal{H})$ (P-q.c.).

(xv) Sia $\mathcal{K}$ una sotto σ-algebra di $\mathcal{H}$. Dato $K \in \mathcal{K}$, si ha $\int_K \mathrm{E}(X|\mathcal{H})\,d\mathrm{P} = \int_K X\,d\mathrm{P} = \int_K \mathrm{E}(X|\mathcal{K})\,d\mathrm{P}$, ove la prima uguaglianza sussiste in quanto $K \in \mathcal{H}$. Dall'arbitrarietà di K risulta allora $\mathrm{E}\big(\mathrm{E}(X|\mathcal{H})|\mathcal{K}\big) = \mathrm{E}(X|\mathcal{K})$ (P-q.c.). Per quanto riguarda invece l'altra uguaglianza, basta osservare che la v.a. $\mathrm{E}(X|\mathcal{K})$ è $\mathcal{H}$-Borel misurabile (in quanto $\mathcal{K}$-Borel misurabile) e usare (iv). ■

Concludiamo la sezione con un risultato (al quale premettiamo un lemma inerente le funzioni convesse) che fornisce alcune celebri e importanti disuguaglianze riguardanti le speranze matematiche (condizionate o no) di trasformate convesse di variabili aleatorie.

Lemma B.2.4 *Siano J un intervallo (limitato o no) della retta reale e $g : J \mapsto \mathbb{R}$ una funzione convessa. Sussistono allora le seguenti proposizioni:*

(i) Qualunque siano $x_1, x_2, x_3 \in J$ tali che $x_1 < x_2 < x_3$ si ha

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} \leq \frac{g(x_3) - g(x_1)}{x_3 - x_1} \leq \frac{g(x_3) - g(x_2)}{x_3 - x_2};$$

(ii) Per ogni punto interno a di J esiste un numero reale b tale che $g(x) \geq b(x - a) + g(a)$ per ogni $x \in J$.

Inoltre, se J è un intervallo aperto, sono valide anche le proposizioni:

(iii) *Indicato con $\mathbb{Q}$ l'insieme dei numeri razionali e, per ogni $q \in J \cap \mathbb{Q}$, con b_q un numero reale tale che $g(x) \geq b_q(x-q) + g(q)$ per ogni $x \in J$ (esistente per (ii)), risulta $g(x) = \sup_{q \in J \cap \mathbb{Q}} [b_q(x-q)+g(q)]$ per ogni $x \in J$;*

(iv) *La funzione g è $\mathcal{B} \cap J$-Borel misurabile.*

Dimostrazione (i) Siano $x_1, x_2, x_3 \in J$ tali che $x_1 < x_2 < x_3$. Essendo le dimostrazioni della prima e della seconda disuguaglianza del tutto simili, proviamo solamente la prima. Per la convessità di g si ha

$$g(x_2) = g\Big(\frac{x_3 - x_2}{x_3 - x_1}\, x_1 + \frac{x_2 - x_1}{x_3 - x_1}\, x_3\Big) \leq \frac{x_3 - x_2}{x_3 - x_1}\, g(x_1) + \frac{x_2 - x_1}{x_3 - x_1}\, g(x_3)$$

da cui otteniamo

$$\begin{aligned} g(x_2) - g(x_1) &\leq \Big(\frac{x_3 - x_2}{x_3 - x_1} - 1\Big) g(x_1) + \frac{x_2 - x_1}{x_3 - x_1}\, g(x_3) \\ &= \frac{x_1 - x_2}{x_3 - x_1} g(x_1) + \frac{x_2 - x_1}{x_3 - x_1}\, g(x_3) = \frac{x_2 - x_1}{x_3 - x_1}\, [g(x_3) - g(x_1)] \end{aligned}$$

e quindi

$$\frac{g(x_2) - g(x_1)}{x_2 - x_1} \leq \frac{g(x_3) - g(x_1)}{x_3 - x_1}.$$

(ii) Considerati due elementi qualsiasi $x', x'' \in J$ tali che $x' < a < x''$, da (i) si ha

$$\frac{g(a) - g(x')}{a - x'} \leq \frac{g(x'') - g(a)}{x'' - a}.$$

Ne segue, per l'arbitrarietà di x',

$$\frac{g(a) - g(x')}{a - x'} \leq b = \sup_{\substack{x<a \\ x \in J}} \frac{g(a) - g(x)}{a - x} \leq \frac{g(x'') - g(a)}{x'' - a}$$

e quindi b è un numero reale. Riesce inoltre, per l'arbitrarietà di x', x'' in J, $g(x) \geq b(x-a) + g(a)$ per ogni $x \in J$.

(iii) Osservato che la disuguaglianza $g(x) \geq \sup_{q \in J \cap \mathbb{Q}} [b_q(x-q)+g(q)]$ sussiste banalmente per ogni $x \in J$, proviamo la disuguaglianza opposta. Sia dunque $x \in J$. Per (ii) (posto $a = x$), esiste un numero reale b tale che $g(t) \geq b(t-x) + g(x)$ per ogni $t \in J$. Ciò osservato, sia $x' \in J$ tale che $x < x'$. Dato allora un arbitrario numero razionale q' tale che $x < q' < x'$, riesce $g(x') \geq b_{q'}(x' - q') + g(q')$, cioè $b_{q'} \leq \frac{g(x')-g(q')}{x'-q'}$ e quindi

$$\begin{aligned} b_{q'}(x - q') + g(q') &= g(q') - b_{q'}(q' - x) \\ &\geq g(q') - \frac{g(x') - g(q')}{x' - q'}(q' - x) = \frac{x' - x}{x' - q'}\, g(q') - \frac{q' - x}{x' - q'}\, g(x') \\ &\geq \frac{x' - x}{x' - q'}\, \big[b(q' - x) + g(x)\big] - \frac{q' - x}{x' - q'}\, g(x'). \end{aligned}$$

Dunque

$$\sup_{q\in J\cap\mathbb{Q}}\big[b_q(x-q)+g(q)\big] \geq \frac{x'-x}{x'-q'}\big[b(q'-x)+g(x)\big]-\frac{q'-x}{x'-q'}\,g(x')$$

per ogni numero razionale q' tale che $x < q' < x'$. Risulta pertanto

$$g(x) = \lim_{\substack{q'\to x^+\\ q'\in\mathbb{Q}}}\Big[\frac{x'-x}{x'-q'}\big[b(q'-x)+g(x)\big]-\frac{q'-x}{x'-q'}\,g(x')\Big] \leq \sup_{q\in J\cap\mathbb{Q}}\big[b_q(x-q)+g(q)\big].$$

(iv) Per ogni numero razionale $q \in J$, la funzione $x \mapsto b_q(x-q)+g(q)$ di dominio J è, per il Teorema A.2.3(i), $\mathcal{B}\cap J$-Borel misurabile. Allora, per (iii) e il Teorema A.2.5(ii), lo è pure g. ∎

Teorema B.2.5 (di Jensen) *Dato un intervallo aperto J (limitato o no) della retta reale, siano $g : J \to \mathbb{R}$ una funzione convessa e $X : \Omega \mapsto J$ una v.a. con speranza matematica finita. Allora,* $\mathrm{E}(X|\mathcal{H}) \in J$ (P-q.c.). *Inoltre, la v.a. trasformata $g(X)$ ammette speranza matematica e risulta* $g(\mathrm{E}(X)) \leq \mathrm{E}(g(X))$. *Infine, nel caso che* $\mathrm{E}(g(X))$ *sia finita,* $g(\mathrm{E}(X|\mathcal{H})) \leq \mathrm{E}(g(X)|\mathcal{H})$ (P-q.c.).

Dimostrazione Dalla disuguaglianza $\inf J \ll X \ll \sup J$ otteniamo, tramite il Teorema B.2.3(viii), $\mathrm{E}(X|\mathcal{H}) \in J$ (P-q.c.). Proviamo ora che $g(X)$ è una v.a. con speranza matematica. Che sia una v.a. è conseguenza immediata del Lemma B.2.4(iv). Per quanto riguarda la sommabilità, osserviamo che, per il Teorema B.1.5(vi), $\mathrm{E}(X) \in J$. Esiste quindi, per il Lemma B.2.4(ii), un numero reale b tale che $g(x) \geq b\,[x-\mathrm{E}(X)]+g(\mathrm{E}(X))$ per ogni $x \in J$. Considerata allora la v.a. $Y = b\,[X-\mathrm{E}(X)]+g(\mathrm{E}(X))$, si ha $g(X) \geq Y$ e

$$\mathrm{E}(Y) = b\,[\mathrm{E}(X)-\mathrm{E}(X)] + \mathrm{E}\big(g(\mathrm{E}(X))I_\Omega\big) = g(\mathrm{E}(X)) \in \mathbb{R}.$$

Ne segue, per il Teorema B.1.5 (ponendo $Z = g(X)$ e $X = Y$), la sommabilità di $g(X)$. A questo punto la disuguaglianza $g(\mathrm{E}(X)) \leq \mathrm{E}(g(X))$ discende immediatamente dalla $g(X) \geq Y$ e dalla monotonia della speranza matematica.

Sia ora $\mathrm{E}(g(X))$ finita. Dato un numero razionale $q \in J$, dal Lemma B.2.4(iii) otteniamo $g(X) \geq b_q(X-q)+g(q)$. Allora, posto $X' = \mathrm{E}(X|\mathcal{H})$, dal Teorema B.2.3(iv),(v),(ix) risulta $\mathrm{E}(g(X)|\mathcal{H}) \geq b_q(X'-q)+g(q)$ (P-q.c.). Dunque, posto $H_q = \{\mathrm{E}(g(X)|\mathcal{H}) \geq b_q(X'-q)+g(q)\}$, si ha $\mathrm{P}(H_q) = 1$. Posto infine $H = \{X' \in J\} \cap \bigcap_{q\in J\cap\mathbb{Q}} H_q$, otteniamo $\mathrm{P}(H) = 1$ e, per il Lemma B.2.4(iii),

$$g(X'(\omega)) = \sup_{q\in J\cap\mathbb{Q}}[b_q(X'(\omega)-q)+g(q)] \leq \mathrm{E}(g(X)|\mathcal{H})(\omega)$$

per ogni $\omega \in H$, cioè $g\big(\mathrm{E}(X|\mathcal{H})\big) \leq \mathrm{E}(g(X)|\mathcal{H})$ (P-q.c.). ∎

B.2.2 Funzione di regressione

In moltissime situazioni pratiche l'informazione ottenibile consiste nell'osservare il valore di un dato ente aleatorio. Al fine di inquadrare questa situazione nell'ambito del condizionamento a σ-algebre assumiamo, con riferimento all'ente aleatorio X, che la σ-algebra $\aleph$ contenga i singoletti. Supposto che l'informazione ottenibile sia l'osservazione di X, viene naturale identificare la σ-algebra degli eventi osservabili $\mathcal{H}$ con la σ-algebra indotta $X^{-1}(\aleph) \subseteq \mathcal{A}$ da X su Ω (Esempio A.1.1(iii))[24]; inoltre, per ogni v.a. Y con speranza matematica finita, considerare la speranza matematica condizionata $\mathrm{E}(Y|X) = \mathrm{E}(Y|X^{-1}(\aleph))$ come la speranza matematica condizionata di Y all'osservazione di X.

Il seguente notevole risultato assicura che la speranza matematica condizionata alla σ-algebra indotta può sempre ottenersi come trasformata certa dell'osservabile X tramite una opportuna funzione Borel misurabile.

Teorema B.2.6 *Siano Y una v.a. con speranza matematica finita e Z una versione di $\mathrm{E}(Y|X)$. Esiste allora una funzione $\aleph$-Borel misurabile $g : \mathfrak{X} \mapsto \mathbb{R}^*$ tale che $Z = g(X)$.*

Dimostrazione Posto $\mathcal{H} = X^{-1}(\aleph)$ e ricordato che Z è $\mathcal{H}$-Borel misurabile, supponiamo intanto che Z sia una funzione $\mathcal{H}$-semplice a valori finiti, cioè $Z = \sum_{i=1}^{n} \alpha_i I_{H_i}$ con $H_i \in \mathcal{H}$ e $\alpha_i \in \mathbb{R}$ $(i = 1, \dots, n)$. Esistono allora $\mathsf{A}_1, \dots \mathsf{A}_n \in \aleph$ tali che $H_i = \{X \in \mathsf{A}_i\}$ $(i = 1, \dots, n)$. Conseguentemente, $g = \sum_{i=1}^{n} I_{\mathsf{A}_i}$ è $\aleph$-Borel misurabile e

$$g(X(\omega)) = \sum_{i=1}^{n} \alpha_i I_{\mathsf{A}_i}(X(\omega)) = \sum_{i=1}^{n} \alpha_i I_{\{X \in \mathsf{A}_i\}}(\omega) = \sum_{i=1}^{n} \alpha_i I_{H_i}(\omega) = Z(\omega)$$

per ogni $\omega \in \Omega$, cioè $Z = g(X)$. Assumiamo infine che Z sia una funzione $\mathcal{H}$-Borel misurabile non necessariamente semplice. Per il lemma fondamentale A.2.4, esiste una successione $(Z_n)_{n\geq 1}$ di funzioni $\mathcal{H}$-semplici a valori finiti tale che $Z_n \to Z$. Per quanto appena provato, esiste inoltre una successione $(g_n)_{n\geq 1}$ di funzioni $\aleph$-Borel misurabili tale che $Z_n = g_n(X)$ per ogni n. Allora, $\aleph_{\lim} = \big\{x \in \mathfrak{X} : \exists \lim_{n\to+\infty} g_n(x)\big\} \in \aleph$ e quindi l'applicazione $g : \mathfrak{X} \mapsto \mathbb{R}^*$ così definita:

$$g(x) = \begin{cases} \lim_{n\to+\infty} g_n(x) & \text{se } x \in \aleph_{\lim} \\ 0 & \text{se } x \notin \aleph_{\lim} \end{cases}$$

è, per l'Osservazione A.2.6, $\aleph$-Borel misurabile. Risulta inoltre $Z(\omega) = \lim_{n\to+\infty} Z_n(\omega) = \lim_{n\to+\infty} g_n(X(\omega)) = g(X(\omega))$ per ogni $\omega \in \Omega$, cioè $Z = g(X)$. ■

[24] In quanto, da un punto di vista interpretativo, la conoscenza dei valori di verità di tutti gli eventi di tale σ-algebra equivale alla conoscenza del "vero" valore di X.

Considerata allora una v.a. Y con speranza matematica finita, possiamo trovare funzioni P_X-integrabili g tali che $\int_{\{X\in\mathsf{A}\}} Y\,d\mathrm{P} = \int_{\mathsf{A}} g\,d\mathrm{P}_X$ per ogni $\mathsf{A}\in\aleph$; infatti, scelta una versione Z di $\mathrm{E}(Y|X)$, esiste (per il teorema precedente) una funzione $\aleph$-Borel misurabile g tale che $Z = g(X)$ e quindi, per il teorema fondamentale del calcolo delle probabilità,

$$\int_{\{X\in\mathsf{A}\}} Y\,d\mathrm{P} = \int_{\{X\in\mathsf{A}\}} Z\,d\mathrm{P} = \int_{\{X\in\mathsf{A}\}} g(X)\,d\mathrm{P} = \int_{\mathsf{A}} g\,d\mathrm{P}_X$$

per ogni $\mathsf{A}\in\aleph$. Ciò constatato, chiamiamo **funzione di regressione di Y su X** ogni funzione $\mathrm{E}(Y|X=\cdot):\mathfrak{X}\mapsto\mathbb{R}^*$ tale che:

1. sia P_X-integrabile;
2. risulti $\int_{\{X\in\mathsf{A}\}} Y\,d\mathrm{P} = \int_{\mathsf{A}} \mathrm{E}(Y|X=x)\,\mathrm{P}_X(dx)$ per ogni $\mathsf{A}\in\aleph$.

Conseguentemente, per i teoremi A.3.6(vii) e A.4.4, $\mathrm{E}(X|T=\cdot)$ è finita P_X-quasi certamente ed è definita a meno di insiemi P_X-trascurabili. Inoltre, $\mathrm{E}\big(Y|X=X(\cdot)\big)$ è una versione di $\mathrm{E}(Y|\,X)$. Infatti, sempre per il teorema fondamentale del calcolo delle probabilità,

$$\int_{\{X\in\mathsf{A}\}} Y\,d\mathrm{P} = \int_{\mathsf{A}} \mathrm{E}(Y|X=x)\,\mathrm{P}_X(dx) = \int_{\{X\in\mathsf{A}\}} \mathrm{E}(Y|X=X(\omega))\,\mathrm{P}(d\omega)$$

per ogni $\mathsf{A}\in\aleph$; inoltre, la $X^{-1}(\aleph)$-Borel misurabilità segue banalmente osservando che $\mathrm{E}\big(Y|X=X(\cdot)\big)$ si ottiene componendo l'applicazione $(X^{-1}(\aleph),\aleph)$-misurabile X con una funzione $\aleph$-Borel misurabile[25]. Infine, il prossimo risultato rileva che la funzione di regressione assume come valore l'usuale speranza matematica condizionata sugli eventi osservabili $\{X = x_0\}$ di probabilità positiva.

[25]Una funzione di regressione consente quindi di legare funzionalmente (con pratica certezza) la media di Y condizionata all'informazione ottenibile con l'osservabile X. Viene così giustificato l'uso del termine "regressione" che nel linguaggio probabilistico-statistico è sinonimo di "relazione funzionale tra due enti aleatori". Ricordiamo che il termine deriva da un famoso lavoro di Francis Galton (pubblicato nel 1885) relativo al confronto delle altezze di 928 adulti con le altezze medie dei loro genitori. Avendo constatato che le altezze medie dei figli tendevano ad allinearsi con l'altezza media (se i genitori erano più bassi (più alti) della media generale, le altezze medie dei figli salivano (calavano) in modo da "recuperare" il divario), scrisse che ciò costituiva una "regression towards mediocrity".

Teorema B.2.7 *Sia* $x_0 \in \mathfrak{X}$ *tale che* $\mathrm{P}(X = x_0) > 0$. *Allora,*

$$\mathrm{E}(Y|X = x_0) = \frac{\int_{\{X=x_0\}} Y \, d\mathrm{P}}{\mathrm{P}(X = x_0)} = \mathrm{E}(Y|\{X = x_0\}).$$

Dimostrazione Proviamo intanto che l'unico sottoinsieme proprio di $H_0 = \{X = x_0\}$ che appartenga alla σ-algebra indotta $\mathcal{H} = X^{-1}(\aleph)$ è l'insieme vuoto. Dato $\{X \in \mathsf{A}\} \subset H_0$, si ha $x_0 \notin \mathsf{A}$ (in caso contrario, si avrebbe la contraddizione $H_0 \subseteq \{X \in \mathsf{A}\} \subset H_0$) e quindi $\{X \in \mathsf{A}\} = H_0 \cap \{X \in \mathsf{A}\} = \emptyset$.

Dunque, H_0 è un atomo di $\mathcal{H}$. Allora, per il Teorema B.2.2, esiste un sottoinsieme $H_1 \in \mathcal{H}$ di H_0 tale che $\mathrm{P}(H_1) = \mathrm{P}(H_0) > 0$ e sul quale $\mathrm{E}(Y|X)$ è la costante di valore $\alpha = \mathrm{E}(Y|H_0)$. Inoltre, poichè $\mathrm{E}(Y|X) = \mathrm{E}(Y|X = X(\cdot))$ (P-q.c.), esiste $H_2 \in \mathcal{H}$ tale che $\mathrm{P}(H_2) = 1$ e $\mathrm{E}(Y|X)(\omega) = \mathrm{E}(Y|X = X(\omega))$ per ogni $\omega \in H_2$. Posto allora $H = H_1 \cap H_2 \subseteq H_0$, si ha $\mathrm{P}(H) = \mathrm{P}(H_1) + \mathrm{P}(H_2) - \mathrm{P}(H_1 \cup H_2) = P(H_1) > 0$ da cui otteniamo $H \neq \emptyset$ e quindi $H = H_0$. Scelto infine $\omega_0 \in H_0$, risulta $\mathrm{E}(Y|X = x_0) = \mathrm{E}(Y|X = X(\omega_0)) = \mathrm{E}(Y|X)(\omega_0) = \alpha$. L'altra uguaglianza si ottiene invece da (B.4) ■

Nell'esempio seguente individuiamo (in alcuni casi semplici) una versione della funzione di regressione $\mathrm{E}(Y|X = \cdot)$[26], partendo da una versione Z della speranza matematica condizionata $\mathrm{E}(Y|X)$ e trovando il legame funzionale sussistente tra Z e l'osservabile X. Viene così messa concretamente in evidenza la profonda differenza concettuale delle due nozioni: funzione dipendente dai valori osservabili la prima; dai casi elementari la seconda.

Esempio B.2.8 Posto $\Omega = [0, 1]$, siano $\mathcal{A} = \mathcal{B} \cap \Omega$ e $\mathrm{P} = \lambda\big|_{\mathcal{A}}$. Supposto inoltre $\mathfrak{X} = \mathbb{R}$ e $\aleph = \mathcal{B}$, consideriamo una funzione continua Y di dominio Ω e andiamo a determinare sia la speranza matematica condizionata di Y a $\mathcal{H} = X^{-1}(\mathcal{B})$ che la funzione di regressione di Y su X, relativamente a quattro diverse ipotesi sulla forma funzionale dell'osservabile.

(i) Sia X una funzione continua strettamente monotona (crescente o decrescente). Allora, dato un qualsiasi intervallo chiuso $[a, b]$ di Ω si ha $[a, b] \in \mathcal{H}$; infatti, $[a, b] = X^{-1}([X(a), X(b)])$, se X è crescente, e $[a, b] = X^{-1}([X(b), X(a)])$, se X è decrescente. Ne segue, $\{[a, b] \cap \Omega : a, b \in \mathbb{R}\} \subset \mathcal{H}$ e quindi $\mathcal{H} = \mathcal{A}$ (ricordato che $\mathcal{B}$ è generata dagli intervalli chiusi)[27]. La v.a. Y risulta dunque $\mathcal{H}$-Borel misurabile e quindi, per il Teorema

[26] Analogamente al caso del condizionamento a σ-algebre, usiamo la locuzione "g è una versione di $\mathrm{E}(Y|X = \cdot)$" per indicare che g è una funzione di regressione di Y su X.

[27] Con riferimento al caso generale, sussiste infatti il risultato seguente: Data una famiglia $\mathcal{F} \subseteq 2^\Omega$ tale che $\mathcal{A} = \sigma(\mathcal{F})$, la traccia $\mathcal{A} \cap S$ è generata dalla famiglia $\mathcal{F} \cap S = \{F \cap S : F \in \mathcal{F}\}$. Infatti, sia $\mathcal{G} \subseteq \mathcal{A} \cap S$ la σ-algebra su S generata da $\mathcal{F} \cap S$ e $\mathcal{A}' = \{A : A \cap S \in \mathcal{G}\} \subseteq \mathcal{A}$. Allora, $\Omega \in \mathcal{A}'$, notato che $S = \Omega \cap S \in \mathcal{G}$; $A^c \in \mathcal{A}'$ se $A \in \mathcal{A}'$, osservato che $A^c \cap S = S \setminus (A \cap S) \in \mathcal{G}$; $\bigcup_{n \geq 1} A_n \in \mathcal{A}'$ se $A_n \in \mathcal{A}'$ per ogni n, in quanto $(\bigcup_{n \geq 1} A_n) \cap S = \bigcup_{n \geq 1}(A_n \cap S) \in \mathcal{G}$. Pertanto, $\mathcal{A}'$ è una σ-algebra su Ω includente $\mathcal{F}$. Ne segue $\mathcal{A}' = \mathcal{A}$ e quindi $\mathcal{G} = \mathcal{A} \cap S$.

B.2.3(iv), Y è una versione di $\mathrm{E}(Y|X)$. Conseguentemente, notato che la funzione:

$$g(x) = \begin{cases} Y(X^{-1}(x)) & \text{se } x \in X(\Omega) \\ 0 & \text{se } x \notin X(\Omega) \end{cases}$$

è Borel misurabile[28] e $Y = g(X)$, possiamo concludere che g è una versione di $\mathrm{E}(Y|X = \cdot)$.

(ii) Sia X tale che $X(\omega) = -1$, se $\omega < \frac{1}{3}$, e $X(\omega) = 1$, se $\omega \geq \frac{1}{3}$. Allora, posto $H_1 = [0, \frac{1}{3}[$ e $H_2 = [\frac{1}{3}, 1]$, otteniamo $\mathcal{H} = \{\emptyset, \Omega, H_1, H_2\}$ e $\mathrm{P}(H_1) = \lambda(H_1) = \frac{1}{3}$, $\mathrm{P}(H_2) = \lambda(H_2) = \frac{2}{3}$. Conseguentemente, per il Teorema B.2.2 e (B.4),

$$\mathrm{E}(Y|X)(\omega) = \begin{cases} \mathrm{E}(Y|H_1) = 3\displaystyle\int_0^{\frac{1}{3}} Y(x)\,dx & \text{se } \omega < \frac{1}{3} \\ \mathrm{E}(Y|H_2) = \frac{3}{2}\displaystyle\int_{\frac{1}{3}}^{1} Y(x)\,dx & \text{se } \omega \geq \frac{1}{3} \end{cases}$$

e quindi la funzione:

$$g(x) = \begin{cases} 3\displaystyle\int_0^{\frac{1}{3}} Y(x)\,dx & \text{se } x = -1 \\ \frac{3}{2}\displaystyle\int_{\frac{1}{3}}^{1} Y(x)\,dx & \text{se } x = 1 \\ 0 & \text{altrimenti} \end{cases}$$

è Borel misurabile e riesce $\mathrm{E}(Y|X) = g(X)$. Pertanto, g è una versione di $\mathrm{E}(Y|X = \cdot)$.

(iii) Con riferimento agli intervalli H_1, H_2 considerati in (ii), sia X tale che $X(\omega) = -1$, se $\omega \in H_1$, e $X(\omega) = \frac{1}{2}(3\omega - 1)$, se $\omega \in H_2$. Allora, posto $\mathcal{B}' = \mathcal{B} \cap H_2$, otteniamo $\mathcal{H} = \mathcal{B}' \cup \{B' \cup H_1 : B' \in \mathcal{B}'\}$ e quindi H_1 è l'unico atomo di $\mathcal{H}$. Ne segue che la v.a. $\mathcal{H}$-Borel misurabile:

$$Z(\omega) = \begin{cases} \mathrm{E}(Y|H_1) = 3\displaystyle\int_0^{\frac{1}{3}} Y(x)\,dx & \text{se } \omega < \frac{1}{3} \\ \big(Y\big|_{H_2}\big)(\omega) & \text{se } \omega \geq \frac{1}{3} \end{cases}$$

è una versione di $\mathrm{E}(Y|X)$. Infatti, dato $H \in \mathcal{H}$, l'uguaglianza $\int_H Y\,d\mathrm{P} = \int_H Z\,d\mathrm{P}$ è ovvia sia per $H = H_1$ che per $H \in \mathcal{B}'$; inoltre, per l'additività dell'integrale,

$$\begin{aligned} \int_{B' \cup H_1} Y\,d\mathrm{P} &= \int_{B'} Y\,d\mathrm{P} + \int_{H_1} Y\,d\mathrm{P} = \int_{B'} Z\,d\mathrm{P} + \frac{\int_{H_1} Y\,d\mathrm{P}}{\mathrm{P}(H_1)}\,\mathrm{P}(H_1) \\ &= \int_{B'} Z\,d\mathrm{P} + \mathrm{E}(Y|H_1)\mathrm{P}(H_1) = \int_{B'} Z\,d\mathrm{P} + \int_{H_1} Z\,d\mathrm{P} = \int_{B' \cup H_1} Z\,d\mathrm{P} \end{aligned}$$

[28] Essendo, per il teorema di connessione delle funzioni continue, $J = X(\Omega)$ un intervallo chiuso della retta reale, la funzione inversa $X^{-1} : J \mapsto \Omega$ è continua (in quanto inversa di una funzione strettamente monotona definita su un intervallo) e quindi, per il Teorema A.2.3(ii), $Y \circ X^{-1} : J \mapsto \mathbb{R}$ è $\mathcal{B} \cap J$-Borel misurabile. Ne segue, per il Lemma A.2.1(ii), la Borel misurabilità di g.

per ogni $H = B' \cup H_1$ con $B' \in \mathcal{B}'$. Conseguentemente, la funzione:

$$g(x) = \begin{cases} 3\displaystyle\int_0^{\frac{1}{3}} Y(x)\,dx & \text{se } x = -1 \\ Y(X^{-1}(x)) = Y(\frac{2x+1}{3}) & \text{se } 0 \le x \le 1 \\ 0 & \text{altrimenti} \end{cases}$$

è Borel misurabile e riesce $Z = g(X)$. Pertanto, g è una versione di $\mathrm{E}(Y|X = \cdot)$.

(iv) Sia $X(\omega) = (\omega - \frac{1}{2})^2$ per ogni ω. Allora, $\mathcal{H} = \{A \cup (1-A) : A \in \mathcal{A} \cap H_0\} \subset \mathcal{A}$[29], avendo posto $H_0 = [0, \frac{1}{2}]$. Infatti, dato $B \in \mathcal{B}$ e posto $A = X^{-1}(B) \cap H_0 \in \mathcal{A} \cap H_0$, otteniamo $X^{-1}(B) = A \cup (1-A)$; viceversa, dato $A \in \mathcal{A} \cap H_0$ e posto $B = \big(X\big|_{H_0}\big)^{-1}(A) \in \mathcal{B}$, si ha $X^{-1}(B) = A \cup (1-A)$. Per individuare una versione di $\mathrm{E}(Y|X)$, fissiamo un arbitrario elemento $A \in \mathcal{A} \cap H_0$ e valutiamo l'integrale $\int_{A\cup(1-A)} Y\,d\mathrm{P}$. Ora, considerata la trasformazione $\tau : t \mapsto 1 - t$ della retta reale, dal Corollario A.4.2 risulta $\int_A Y\,d\lambda = \int_{1-A} Y(1-x)\,dx$ e $\int_{1-A} Y\,d\lambda = \int_A Y(1-x)\,dx$. Ne segue, per l'additività dell'integrale,

$$\begin{aligned}\int_{A\cup(1-A)} Y\,d\mathrm{P} &= \int_A Y\,d\lambda + \int_{1-A} Y\,d\lambda \\ &= \frac{\int_A Y(x)\,dx + \int_A Y(1-x)\,dx}{2} + \frac{\int_{1-A} Y(1-x)\,dx + \int_{1-A} Y(x)\,dx}{2} \\ &= \int_A \frac{Y(x) + Y(1-x)}{2}\,dx + \int_{1-A} \frac{Y(x)+Y(1-x)}{2}\,dx \\ &= \int_{A\cup(1-A)} \frac{Y(x)+Y(1-x)}{2}\,dx.\end{aligned}$$

Allora, la v.a. $\mathcal{H}$-Borel misurabile:

$$Z(\omega) = \frac{Y(\omega) + Y(1-\omega)}{2}$$

è una versione di $\mathrm{E}(Y|X)$. Pertanto, la funzione:

$$g(x) = \begin{cases} \frac{1}{2}\left[Y(\frac{1}{2} + \sqrt{x}) + Y(\frac{1}{2} - \sqrt{x})\right] & \text{se } 0 \le x \le \frac{1}{4} \\ 0 & \text{altrimenti} \end{cases}$$

è Borel misurabile e riesce $Z = g(X)$. Dunque, g è una versione di $\mathrm{E}(Y|X = \cdot)$. ▲

Nel teorema successivo vengono elencate proprietà della funzione di regressione analoghe (eccetto le ultime due) a quelle considerate nel Teorema B.2.3 per la speranza matematica condizionata a σ-algebre.

[29]Essendo $1 - A$ un boreliano in quanto $\mathcal{B}$ è chiusa per trasformazioni affini (Esempio A.1.1(iv)).

Teorema B.2.9 *Siano Y, Y_1, Y_2 etc. v.a. con speranza matematica finita. Riesce allora:*

(i) $\mathrm{E}(Y) = \int_{\mathfrak{X}} \mathrm{E}(Y|X = x)\, \mathrm{P}_X(dx)$.

Valgono inoltre le proposizioni seguenti nelle quali le relazioni riguardanti funzioni di regressione devono intendersi sussistere P_X-quasi certamente.

(ii) *Se $X = x_0 \in \mathfrak{X}$ (P-q.c.), allora $\mathrm{E}(Y|X = \cdot) = \mathrm{E}(Y)$. Inoltre, se X è l'applicazione identica, allora $\mathrm{E}(Y|X = \cdot) = Y$;*

(iii) $\mathrm{E}(\alpha I_\Omega|X = \cdot) = \alpha$;

(iv) MONOTONIA: $\mathrm{E}(Y_1|X = \cdot) \leq \mathrm{E}(Y_2|X = \cdot)$, *se* $Y_1 \leq Y_2$ (P-q.c.);

(v) $\mathrm{E}(Y_1|X = \cdot) = \mathrm{E}(Y_2|X = \cdot)$, *se* $Y_1 = Y_2$ (P-q.c.);

(vi) INTERNALITÀ: $a \leq \mathrm{E}(Y|X = \cdot) \leq b$, *se* $a \leq Y \leq b$ (P-q.c.);

(vii) INTERNALITÀ STRETTA: $a < \mathrm{E}(Y|X = \cdot) < b$, *se* $a < Y < b$ (P-q.c.);

(viii) LINEARITÀ: $\mathrm{E}\big(\sum_{i=1}^n \alpha_i Y_i|X = \cdot\big) = \sum_{i=1}^n \alpha_i \mathrm{E}(Y_i|X = \cdot)$ *qualunque siano i numeri reali $\alpha_1, \ldots, \alpha_n$;*

(ix) $\big|\mathrm{E}(Y|X = \cdot)\big| \leq \mathrm{E}(|Y|\,|X = \cdot)$;

(x) $\mathrm{E}(Y_n|X = \cdot) \uparrow \mathrm{E}(Y|X = \cdot)$, *se* $Y_n \uparrow Y$ *e* $Y_1 \geq 0$ (P-q.c.);

(xi) $\mathrm{E}\big(\sum_{n\geq 1} \alpha_n Y_n|X = \cdot\big) = \sum_{n\geq 1} \alpha_n \mathrm{E}(Y_n|X = \cdot)$, *se $(\alpha_n)_{n\geq 1}$ è una successione di numeri reali non negativi, $Y_n \geq 0$ per ogni n e la speranza matematica della serie $\sum_{n\geq 1} \alpha_n Y_n$ è finita;*

(xii) $\mathrm{E}(Y_n|X = \cdot) \to \mathrm{E}(Y|X = \cdot)$, *se* $|Y_n| \leq Z$ (P-q.c.) *per ogni n,* $\mathrm{E}(Z)$ *finita e $Y_n \to Y$;*

(xiii) *Sia $g : \mathfrak{X} \mapsto \mathbb{R}^*$ una funzione $\aleph$-Borel misurabile tale che la v.a. $g(X)Y$ ammetta speranza matematica finita. Allora,*

$$\mathrm{E}(g(X)Y|X = \cdot) = g(\cdot)\, \mathrm{E}(Y|X = \cdot).$$

In particolare, $\mathrm{E}(g(X)|X = \cdot) = g(\cdot)$;

(xiv) $\mathrm{E}(Y|X = \cdot) = \mathrm{E}(Y)$, *se X e Y sono indipendenti.*

Dimostrazione Le proposizioni (i), (iii) sono conseguenza immediata della definizione. Le dimostrazioni di (iv) $\div$ (xii) si ottengono, con ovvie modifiche, da quelle delle proposizioni (v) $\div$ (xiii) del Teorema B.2.3.

(ii) Sia $X = x_0$ (P-q.c). Dato $\mathsf{A} \in \aleph$, sia intanto $x_0 \in \mathsf{A}$. Allora, $\mathrm{P}_X(\mathsf{A}) = \mathrm{P}(X \in \mathsf{A}) = 1$ e quindi, per il Teorema A.3.4(iv), $\int_{\{X\in\mathsf{A}\}} Y\, d\mathrm{P} = \int_\Omega Y\, d\mathrm{P} = \mathrm{E}(Y) = \int_{\mathsf{A}} \mathrm{E}(Y)\, d\mathrm{P}_{\mathrm{X}}$. Sia ora $x_0 \notin \mathsf{A}$. Ne segue $\mathrm{P}_X(\mathsf{A}) = 0$ e quindi $\int_{\{X\in\mathsf{A}\}} Y\, d\mathrm{P} = \int_\emptyset Y\, d\mathrm{P} = 0 = \int_{\mathsf{A}} \mathrm{E}(Y)\, d\mathrm{P}_{\mathrm{X}}$.

Sia X l'applicazione identica di Ω. Allora, $\aleph = \mathcal{A}$, $\mathcal{H} = \mathcal{A}$ e $\mathrm{P}_X = \mathrm{P}$. Conseguentemente, Y è P_X-integrabile e $\int_{\{X\in A\}} Y\, d\mathrm{P} = \int_A Y\, d\mathrm{P}_X$ per ogni $A \in \mathcal{A}$.

(xiii) Sia intanto $g = I_{\mathsf{A}'}$ con $\mathsf{A}' \in \aleph$. Riesce allora

$$\begin{aligned}\int_{\{X\in\mathsf{A}\}} g(X)Y\, d\mathrm{P} &= \int_{\{X\in\mathsf{A}\}} I_{\{X\in\mathsf{A}'\}} Y\, d\mathrm{P} = \int_{\{X\in\mathsf{A}\cap\mathsf{A}'\}} Y\, d\mathrm{P} \\ &= \int_{\mathsf{A}\cap\mathsf{A}'} \mathrm{E}(Y|X=x)\, \mathrm{P}_X(dx) = \int_{\mathsf{A}} g(x)\, \mathrm{E}(Y|X=x)\, \mathrm{P}_X(dx).\end{aligned}$$

Conseguentemente, la tesi sussiste per g funzione indicatrice. Le dimostrazioni per g semplice e g arbitraria sono del tutto analoghe a quelle sviluppate nella dimostrazione della proposizione (xiv) del Teorema B.2.3. L'ultima parte della tesi segue immediatamente da quanto appena provato e da (iii) (prendendo come Y la costante unitaria).

(xiv) Dato $\mathsf{A} \in \aleph$, per il Teorema B.1.14, le v.a. Y e $I_{\{X\in\mathsf{A}\}}$ sono indipendenti e quindi

$$\mathrm{E}(Y I_{\{X\in\mathsf{A}\}}) = \mathrm{E}(Y)\mathrm{E}(I_{\{X\in\mathsf{A}\}}) = \mathrm{E}(Y)\mathrm{P}(X \in \mathsf{A}) = \int_{\{X\in\mathsf{A}\}} \mathrm{E}(Y)\, d\mathrm{P}_X,$$

tenuto conto del Teorema B.1.15. ■

Concludiamo la sezione mettendo in luce l'importanza fondamentale della funzione di regressione nella soluzione del problema (centrale nella problematica statistica e di grande interesse applicativo) di stimare, a partire dall'osservabile X, il non osservabile Y commettendo un "errore più piccolo possibile". Ovviamente occorre precisare, prima di procedere, cosa dobbiamo intendere per "errore più piccolo possibile". A tal fine, consideriamo una stima $g(X)$ di Y. Notato che, data una funzione continua e crescente φ tale che $\varphi(0) = 0$ e $\lim_{x\to+\infty} \varphi(x) = +\infty$, la speranza matematica $\mathrm{E}\big(\varphi(|Y - g(X)|)\big)$ è una quantità che tende ad essere grande, se $g(X)$ assume valori distanti da Y, e piccola, se $g(X)$ assume valori vicini a Y, viene naturale usarla per misurare la bontà dell'approssimazione di Y con $g(X)$. Tra le varie scelte possibili di φ, adottiamo quella relativa al *metodo dei minimi quadrati*, cioè poniamo $\varphi(x) = x^2$. Giungiamo così ad intendere la frase "errore più piccolo possibile" nel senso di "*errore quadratico medio* più piccolo possibile" e quindi a cercare una funzione g^* tale che

$$\mathrm{E}\big((Y - g^*(X))^2\big) \le \mathrm{E}\big((Y - g(X))^2\big)$$

per ogni funzione g a quadrato P_X-integrabile[30]. Il prossimo risultato collega la funzione di regressione con il metodo dei minimi quadrati assicurando che ogni versione della funzione di regressione di Y su X fornisce un esempio di stima dei minimi quadrati di Y.

Teorema B.2.10 *Sia Y una v.a. a quadrato integrabile. Allora,*

$$\mathrm{E}\big((Y-\mathrm{E}(Y|X))^2\big)=\min\Big\{\mathrm{E}\big((Y-g(X))^2\big)\colon \int_{\mathfrak{X}} g^2\,d\mathrm{P}_X<+\infty\Big\}.$$

Dimostrazione Per il teorema di Jensen B.2.5, $\mathrm{E}(Y|X)^2 \le \mathrm{E}(Y^2|X)$ (P-q.c.) e quindi

$$0\le \int_{\mathfrak{X}} \mathrm{E}(Y|X=x)^2\,\mathrm{P}_X(dx)=\int_\Omega \mathrm{E}(Y|X)^2\,d\mathrm{P}\le \int_\Omega \mathrm{E}(Y^2|X)\,d\mathrm{P}=\int_\Omega Y^2\,d\mathrm{P}<+\infty,$$

cioè $\mathrm{E}(Y|X=\cdot)$ è a quadrato P_X-integrabile. Inoltre, per il Teorema B.2.9(i), $\mathrm{E}(Y^2) = \int_{\mathfrak{X}} \mathrm{E}(Y^2|X=x)\,\mathrm{P}_X(dx)\ge 0$ e quindi anche $\mathrm{E}(Y^2|X=\cdot)$ è a quadrato P_X-integrabile.

Ciò osservato, sia g una qualsiasi funzione a quadrato P_X-integrabile. Allora, $0 \le \int_\Omega g(X)^2\,d\mathrm{P} = \int_{\mathfrak{X}} g^2\,d\mathrm{P}_X < +\infty$ e quindi, per il Teorema A.3.5(iv), la v.a. $g(X)Y$ ha speranza matematica finita.

Considerato ora l'insieme $\mathfrak{X}' = \{x \in \mathfrak{X} : |\mathrm{E}(Y^2|X=x)|, |\mathrm{E}(Y|X=x)|, |g(x)| \in \mathbb{R}\}$, dai teoremi A.3.5(v) e A.5.6(vii) otteniamo $\mathrm{P}_X(\mathfrak{X}')=1$. Ne segue, per il Teorema B.2.9(i),

$$\begin{aligned}\mathrm{E}\big((Y-g(X))^2\big) &= \int_{\mathfrak{X}} \mathrm{E}\big((Y-g(X))^2|X=x\big)\,\mathrm{P}_X(dx)\\ &= \int_{\mathfrak{X}'} \mathrm{E}\big((Y-g(X))^2|X=x\big)\,\mathrm{P}_X(dx)\\ &= \int_{\mathfrak{X}'} \mathrm{E}\big(Y^2-2g(X)Y+g(X)^2|X=x\big)\,\mathrm{P}_X(dx)\end{aligned}$$

e quindi, tramite il Teorema B.2.9(viii),(xiii),

$$\mathrm{E}\big((Y-g(X))^2\big)=\int_{\mathfrak{X}'}\Big[\mathrm{E}(Y^2|X=x)-2g(x)\mathrm{E}(Y|X=x)+g(x)^2\Big]\mathrm{P}_X(dx). \tag{B.6}$$

In particolare, ponendo $g(\cdot)=\mathrm{E}(Y|X=\cdot)$, si ha dunque

$$\mathrm{E}\big((Y-\mathrm{E}(Y|X))^2\big)=\int_{\mathfrak{X}'}\Big[\mathrm{E}(Y^2|X=x)-\mathrm{E}(Y|X=x)^2\Big]\mathrm{P}_X(dx).$$

[30] Il metodo dei minimi quadrati risale a lavori di Adrien-Marie Legendre (che lo introdusse nel 1805 senza alcuna giustificazione) e di Carl F. Gauss (che invece ne fornì una giustificazione nel 1809). In particolare, Gauss ne studiò le proprietà statistiche proponendo anche algoritmi iterativi per la sua soluzione numerica; lo applicò a problemi geodetici, astronomici e al problema della compensazione degli errori casuali (teoria degli errori).

Notato infine che, per ogni $x \in \mathfrak{X}'$, il polinonio $t^2 - 2t\,\mathrm{E}(Y|X=x) + \mathrm{E}(Y^2|X=x)$ assume valore minimo nel punto $t = \mathrm{E}(Y|X=x)$, da (B.6) e dalla monotonia dell'integrale segue allora

$$\begin{aligned}\mathrm{E}\big((Y-g(X))^2\big) &\geq \int_{\mathfrak{X}'} \Big[\mathrm{E}(Y^2|X=x) - 2\mathrm{E}(Y|X=x)^2 + \mathrm{E}(Y|X=x)^2\Big]\mathrm{P}_X(dx) \\ &= \int_{\mathfrak{X}'} \Big[\mathrm{E}(Y^2|X=x) - \mathrm{E}(Y|X=x)^2\Big]\mathrm{P}_X(dx) = \mathrm{E}\big((Y-\mathrm{E}(Y|X))^2\big).\end{aligned}$$

La dimostrazione è così conclusa. ■

Il corollario seguente si ottiene immediatamente dal teorema appena provato assumendo che l'osservabile X sia la costante di valore $\mathrm{E}(Y)$.

Corollario B.2.11 *Sia Y una v.a. a quadrato integrabile. Allora,*

$$\mathrm{Var}(Y) = \mathrm{E}\big((Y-\mathrm{E}(Y))^2\big) = \min\big\{\mathrm{E}\big((Y-t)^2\big) \colon t \in \mathbb{R}\big\}.$$

B.2.3 Legge e densità condizionali

In questa sezione, accanto all'osservabile X, consideriamo una σ-algebra $\mathcal{T}$ su un insieme non vuoto Θ e un ente aleatorio Z a valori in Θ. Osservato che, per la legge P_Z di Z, riesce $\mathrm{P}_Z(T) = \int_\Omega I_{\{Z\in T\}}\,d\mathrm{P} = \mathrm{E}(I_{\{Z\in T\}})$ per ogni $T \in \mathcal{T}$, viene naturale considerare la famiglia di funzioni di regressione parametrizzata sugli elementi di $\mathcal{T}$:

$$\mathrm{Q}_{Z|X}(T|\cdot) = \mathrm{E}(I_{\{Z\in T\}}|X=\cdot) \qquad (T \in \mathcal{T}).$$

Ora, nel caso particolare che $x_0 \in \mathfrak{X}$ sia tale che $\mathrm{P}(X=x_0) > 0$, dal Teorema B.2.7 otteniamo

$$\begin{aligned}\mathrm{Q}_{Z|X}(T|x_0) &= \frac{\int_{\{X=x_0\}} I_{\{Z\in T\}}\,d\mathrm{P}}{\mathrm{P}(X=x_0)} = \frac{\int_\Omega I_{\{Z\in T\}\cap\{X=x_0\}}\,d\mathrm{P}}{\mathrm{P}(X=x_0)} \\ &= \frac{\mathrm{P}(\{Z\in T\}\cap\{X=x_0\})}{\mathrm{P}(X=x_0)} = \mathrm{P}(Z\in T|\{X=x_0\}).\end{aligned}$$

Pertanto, $\mathrm{Q}_{Z|X}(\cdot|x)$ assume necessariamente il significato di legge di Z "sapendo che X ha preso il valore x" ogniqualvolta la probabilità dell'evento "$X = x$" risulti positiva. D'altra parte, al di fuori di questo caso, $\mathrm{Q}_{Z|X}(\cdot|x)$ non presenta un particolare significato intuitivo in quanto non è possibile assicurare, nel contesto generale dato, che sia una probabilità su $\mathcal{T}$ o almeno lo sia a meno di un insieme P_X-trascurabile. Tuttavia, questa situazione è, per molti versi, "patologica" essendo presente in contesti alquanto artificiosi e di scarsa (se non addirittura nulla) importanza per le applicazioni usuali della probabilità.

Motivati da quest'ultima osservazione, abbandoniamo il caso generale considerando solo le **leggi condizionali di Z rispetto** X, cioè quelle particolari famiglie di funzioni di regressione $\big(\mathrm{Q}_{Z|X}(T|\cdot)\big)_{T\in\mathcal{T}}$ parametrizzate su $\mathcal{T}$ che verificano le seguenti proprietà:

1. esiste $\mathfrak{X}' \in \aleph$ tale che $\mathrm{P}_X(\mathfrak{X}') = 1$ e $\mathrm{Q}_{Z|X}(\cdot|x)$ è una probabilità su $\mathcal{T}$ per ogni $x \in \mathfrak{X}'$;

2. $\mathrm{Q}_{Z|X}(T|\,\cdot)$ è P_X-integrabile per ogni $T \in \mathcal{T}$;

3. risulta
$$\mathrm{P}((Z,X) \in T \times \mathsf{A}) = \int_{\mathsf{A}} \mathrm{Q}_{Z|X}(T|x)\,\mathrm{P}_X(dx) = \int_{\mathsf{A}\cap\mathfrak{X}'} \mathrm{Q}_{Z|X}(T|x)\,\mathrm{P}_X(dx)$$
per ogni $\mathsf{A} \in \aleph$ e $T \in \mathcal{T}$.

Prima di constatarne l'esistenza in un caso particolare (ma di indubbia importanza applicativa), proviamo un teorema che le collega alla funzione di regressione di trasformate certe di Z su X.

Teorema B.2.12 *Sia* $(\mathrm{Q}_{Z|X}(T|\,\cdot))_{T\in\mathcal{T}}$ *una legge condizionale di* Z *rispetto* X. *Inoltre, sia* $g : \Theta \mapsto \mathbb{R}$ *una funzione* $\mathrm{Q}_{Z|X}(\cdot|x)$*-sommabile per ogni* $x \in \mathfrak{X}'$ *e tale che la speranza matematica* $\mathrm{E}(g(Z))$ *risulti finita. Allora, la funzione:*

$$x \mapsto \begin{cases} \int_\Theta g(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x) & \textit{se } x \in \mathfrak{X}' \\ 0 & \textit{se } x \notin \mathfrak{X}' \end{cases}$$

è una versione della funzione di regressione $\mathrm{E}(g(Z)|X = \cdot)$.

Dimostrazione Indicata con f tale funzione, sia intanto $g = I_T$. Dato A, si ha

$$\begin{aligned}\mathrm{E}(g(Z)I_{\{X\in\mathsf{A}\}}) &= \mathrm{E}(I_T(Z)\,I_{\{X\in\mathsf{A}\}}) = \mathrm{P}\big(\{Z \in T\} \cap \{X \in \mathsf{A}\}\big) \\ &= \int_{\mathsf{A}\cap\mathfrak{X}'} \mathrm{Q}_{Z|X}(T|x)\,\mathrm{P}_X(dx) = \int_{\mathsf{A}\cap\mathfrak{X}'} \Big(\int_\Theta I_T(\theta)\mathrm{Q}_{Z|X}(d\theta|x)\Big)\,\mathrm{P}_X(dx) \\ &= \int_{\mathsf{A}\cap\mathfrak{X}'} \Big(\int_\Theta g(\theta)\mathrm{Q}_{Z|X}(d\theta|x)\Big)\,\mathrm{P}_X(dx) = \int_{\mathsf{A}\cap\mathfrak{X}'} f\,d\mathrm{P}_X = \int_{\mathsf{A}} f\,d\mathrm{P}_X.\end{aligned}$$

Ne segue, per l'arbitrarietà di A, che f è una versione di $\mathrm{E}(g(Z)|X=\cdot)$.
Sia ora $g=\sum_{i=1}^n y_i I_{T_i}\geq 0$ una funzione semplice. Posto:

$$f_i(x)=\begin{cases}\int_\Theta I_{T_i}(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x) & \text{se } x\in\mathfrak{X}'\\ 0 & \text{se } x\notin\mathfrak{X}'\end{cases}\qquad (i=1,\dots,n),$$

dalla linearità dell'integrale otteniamo $f=\sum_{i=1}^n y_i f_i$. Ne segue, poichè f_i è una versione di $\mathrm{E}(f_i(Z)|X=\cdot)$ $(i=1,\dots,n)$, dalla linearità della funzione di regressione (Teorema B.2.9(viii)) che f è una versione di $\mathrm{E}(\sum_{i=1}^n y_i f_i(Z)|X=\cdot)=\mathrm{E}(g(Z)|X=\cdot)$.
Sia ora $g\geq 0$. Considerata una successione $(g_n)_{n\geq 1}$ di funzioni semplici non negative tale che $g_n\uparrow g$, poniamo:

$$f_n(x)=\begin{cases}\int_\Theta g_n(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x) & \text{se } x\in\mathfrak{X}'\\ 0 & \text{se } x\notin\mathfrak{X}'\end{cases}\qquad (n\geq 1).$$

Allora, per il teorema della convergenza monotona, $0\leq f_n\uparrow f$. Conseguentemente, poichè f_n è una versione di $\mathrm{E}(g_n(Z)|X=\cdot)$ $(n\geq 1)$, dal Teorema B.2.9(x) otteniamo che f è una versione di $\mathrm{E}(\lim_{n\to+\infty} g_n(Z)|X=\cdot)=\mathrm{E}(g(Z)|X=\cdot)$.
Sia infine g qualsiasi. Osservato che $g(Z)^+=g^+(Z)$ e $g(Z)^-=g^-(Z)$, si ha la finitezza delle speranze matematiche $\mathrm{E}(g^+(Z))$ e $\mathrm{E}(g^-(Z))$; inoltre, dall'ipotesi di $\mathrm{Q}_{Z|X}(\cdot|x)$-sommabilità risulta

$$\int_\Theta g(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x)=\int_\Theta g^+(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x)-\int_\Theta g^-(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x)$$

per ogni $x\in\mathfrak{X}'$. Posto allora

$$f_1(x)=\begin{cases}\int_\Theta g^+(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x) & \text{se } x\in\mathfrak{X}'\\ 0 & \text{se } x\notin\mathfrak{X}'\end{cases}$$

$$f_2(x)=\begin{cases}\int_\Theta g^-(\theta)\,\mathrm{Q}_{Z|X}(d\theta|x) & \text{se } x\in\mathfrak{X}'\\ 0 & \text{se } x\notin\mathfrak{X}'\end{cases},$$

otteniamo $f=f_1-f_2$. Ne segue, poichè f_1, f_2 sono, rispettivamente, una versione di $\mathrm{E}(g^+(Z)|X=\cdot)$ e $\mathrm{E}(g^-(Z)|X=\cdot)$, dalla linearità della funzione di regressione che f è una versione di $\mathrm{E}(g^+(Z)-g^-(Z)|X=\cdot)=\mathrm{E}(g(Z)|X=\cdot)$. ■

Per quanto riguarda l'esistenza delle leggi condizionali, consideriamo due misure σ-finite μ, ν di riferimento definite, rispettivamente, su $\aleph$, $\mathcal{T}$ e assumiamo (fino alla fine della sezione) che $f:\Theta\times\mathfrak{X}\mapsto\mathbb{R}^*$ sia una (ν,μ)-densità congiunta della coppia aleatoria (Z,X). Posto allora $\mathfrak{X}_0=\{\,0<f_X<+\infty\}$, introduciamo, per ogni $x\in\mathfrak{X}$, la funzione $\mathcal{T}$-Borel misurabile $f_{Z|X}(\cdot|x)$ su Θ così definita:

$$f_{Z|X}(\theta|x)=\begin{cases}\frac{f(\theta,x)}{f_X(x)} & \text{se } x\in\mathfrak{X}_0\\ f_Z(\theta) & \text{se } x\notin\mathfrak{X}_0\end{cases}\tag{B.7}$$

per ogni $\theta \in \Theta$ (Lemma A.5.1(ii)). Il risultato successivo mette il luce la possibilità di ottenere, mediante la famiglia $(f_{Z|X}(\cdot|x))_{x\in\mathfrak{X}}$, sia una densità congiunta della coppia aleatoria che una marginale di Z.

Teorema B.2.13 *Sussistono le seguenti proposizioni:*

(i) $\mathrm{P}_X(\mathfrak{X}_0) = 1$;

(ii) $(\theta, x) \mapsto f_{Z|X}(\theta|x)f_X(x)$ *è una* (ν, μ)*-densità congiunta di* (Z, X);

(iii) $f_Z(\cdot) = \int_{\mathfrak{X}} f_{Z|X}(\cdot|x)\mathrm{P}_X(dx)$ (ν-q.o.).

Dimostrazione (i) Risulta

$$\begin{aligned}\mathrm{P}_X(f_X \in \{0, +\infty\}) &= \int_{\{f_X=0\}\cup\{f_X=+\infty\}} f_X\, d\mu = \int_{\{f_X=0\}} f_X\, d\mu + \int_{\{f_X=+\infty\}} f_X\, d\mu \\ &= \int_{\{f_X=+\infty\}} f_X\, d\mu = +\infty\, \mu(\{f_X = +\infty\}).\end{aligned}$$

Allora, $+\infty\, \mu(\{f_X = +\infty\}) \leq 1$ da cui otteniamo $\mu(\{f_X = +\infty\}) = 0$ e quindi $\mathrm{P}_X(\mathfrak{X}_0) = \mathrm{P}_X(f_X \notin \{0, +\infty\}) = 1$.

(ii) Osservato che, per il Lemma A.5.1(iii), la funzione $(\theta, x) \mapsto f_{Z|X}(\theta|x)\, f_X(x)$ è $\mathcal{T} \otimes \aleph$-Borel misurabile e che, per (i) e il Teorema B.1.12, $\mathrm{P}_{(Z,X)}(\Theta \times \mathfrak{X}_0) = \mathrm{P}_X(\mathfrak{X}_0) = 1$, dal Teorema A.3.4(iv) otteniamo

$$\begin{aligned}\mathrm{P}_{(Z,X)}(S) &= \int_{S\cap(\Theta\times\mathfrak{X}_0)} f\, d\nu \times \mu = \int_{S\cap(\Theta\times\mathfrak{X}_0)} f_{Z|X}(\theta|x)\, f_X(x)\, \nu \times \mu(d\theta \times dx) \\ &= \int_S f_{Z|X}(\theta|x)\, f_X(x)\, \nu \times \mu(d\theta \times dx)\end{aligned}$$

per ogni $S \in \mathcal{T} \otimes \aleph$.

(iii) Conseguenza immediata di (ii) e del Teorema B.1.12. ■

Tramite le funzioni $f_{Z|X}(\cdot|x)$ possiamo introdurre la famiglia:

$$\mathrm{P}_{Z|X}(T|x) = f_{Z|X}(\cdot|x) * \nu(T) = \int_T f_{Z|X}(\theta|x)\, \nu(d\theta) \qquad (x \in \mathfrak{X}) \tag{B.8}$$

di misure sulla σ-algebra $\mathcal{T}$. Ora, per la linearità dell'integrale,

$$\mathrm{P}_{Z|X}(\Theta|x) = \int_\Theta \frac{f(\theta, x)}{f_X(x)}\, \nu(d\theta) = \frac{\int_\Theta f(\theta, x)\, \nu(d\theta)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1,$$

se $x \in \mathfrak{X}_0$, e $\mathrm{P}_{Z|X}(\Theta|x) = \mathrm{P}_Z(\Theta) = 1$, se $x \notin \mathfrak{X}_0$; inoltre, per i teoremi B.2.13(ii) e di Tonelli,

$$\begin{aligned}
\mathrm{P}((Z,X) \in T \times \mathsf{A}) &= \int_{T\times\mathsf{A}} f_{Z|X}(\theta|x) f_X(x)\, \nu \times \mu(d\theta \times dx) \\
&= \int_{\mathsf{A}} \Big(\int_T f_{Z|X}(\cdot|x)\, d\nu \Big) f_X(x)\, \mu(dx) \\
&= \int_{\mathsf{A}} \mathrm{P}_{Z|X}(T|x)\, \mathrm{P}_X(dx)
\end{aligned}$$

per ogni $\mathsf{A} \in \aleph$ e $T \in \mathcal{T}$. Conseguentemente, $(\mathrm{P}_{Z|X}(\cdot|x))_{x\in\mathfrak{X}}$ è una famiglia di probabilità su $\mathcal{T}$ che fornisce una versione della legge condizionale di Z rispetto X. Tenuto conto della sua costruzione, viene allora naturale chiamare **ν-densità condizionale** (in breve **densità condizionale**) **di Z rispetto X** la famiglia $(f_{Z|X}(\cdot|x))_{x\in\mathfrak{X}}$[31].

Il prossimo esempio mette in evidenza che, in generale, interpretare la probabilità $\mathrm{P}_{Z|X}(\cdot|x)$ come la legge di Z "sapendo che X ha preso il valore x" anche quando l'evento "$X = x$" è trascurabile, oltre a non avere alcun fondamento, può essere fuorviante se non addirittura pericoloso.

Esempio B.2.14 (Paradosso di Borel-Kolmogorov) Sia (Z, Y) una coppia aleatoria distribuita in $]0,1]^2$ con densità congiunta $f(\theta, y) = 4y\theta$. Posto $J = I_{]0,1]}$, otteniamo allora $f(\theta, y) = 4y\theta J(y)J(\theta)$ per ogni $y, z \in \mathbb{R}$. Considerate le v.a. $X = Y - Z$ e $X' = \frac{Y}{Z}$, determiniamo dapprima la densità $f_{Z|X}(\cdot|0)$ e poi la densità $f_{Z|X'}(\cdot|1)$.

Dati i numeri reali a, b e posto $B = \{(\theta, y) \in \mathbb{R}^2 : \theta \le a \wedge y - \theta \le b\}$, dal teorema di Tonelli risulta

$$\begin{aligned}
F_{(Z,X)}(a,b) = \mathrm{P}((Z,Y) \in B) &= \int_B 4\, y\theta J(y)J(\theta)\, d\theta dy = \int_{\mathbb{R}^2} 4\, y\theta J(y) J(\theta) I_B(y,\theta)\, d\theta dy \\
&= \int_{\mathbb{R}^2} 4\, y\theta J(y)J(\theta) I_{]-\infty,b]}(y-\theta) I_{]-\infty,a]}(\theta)\, dy\, d\theta \\
&= \int_{\mathbb{R}} 4 \Big(\int_{\mathbb{R}} yJ(y) I_{]-\infty,b]}(y-\theta)\, dy \Big) \theta J(\theta) I_{]-\infty,a]}(\theta)\, d\theta \\
&= \int_{-\infty}^{a} 4 \Big(\int_{\mathbb{R}} yJ(y) I_{]-\infty,b]}(y-\theta)\, dy \Big) \theta J(\theta)\, d\theta.
\end{aligned}$$

[31] Osserviamo che nella definizione (B.7) avremmo potuto considerare al posto della densità marginale f_Z una qualsiasi altra funzione $\mathcal{T}$-Borel misurabile, senza pregiudicare nè la validità del Teorema B.2.13(ii) nè che la famiglia $(\mathrm{P}_{Z|X}(\cdot|x))_{x\in\mathfrak{X}}$ sia una versione della legge condizionale di Z rispetto X.

Considerata la traslazione $\tau : y \mapsto y + \theta$ della retta reale, dal Corollario A.4.2 otteniamo $\int_{\mathbb{R}} yJ(y)I_{]-\infty,b]}(y-\theta)\,dy = \int_{\mathbb{R}}(x+\theta)J(x+\theta)I_{]-\infty,b]}(x)\,dx = \int_{-\infty}^{b}(x+\theta)J(x+\theta)\,dx$. Ne segue, per il teorema di Tonelli,

$$F_{(Z,X)}(a,b) = \int_{-\infty}^{a} 4\Big(\int_{-\infty}^{b}(x+\theta)J(x+\theta)dx\Big)\,\theta J(\theta)d\theta = \int_{S_{ab}} 4\theta(x+\theta)J(x+\theta)J(\theta)d\theta dx.$$

La coppia aleatoria (Z,X) ha dunque densità congiunta $f_{Z,X}(\theta,x) = 4\theta(x+\theta)J(x+\theta)J(\theta)$. Ne segue $f_X(0) = 4\int_0^1 \theta^2\,d\theta = \frac{4}{3}$ e quindi $f_{Z|X}(\theta|0) = 3\theta^2 J(\theta)$ per ogni θ.

Passando alla seconda densità, dati i numeri reali a e b, poniamo $B = \{(\theta,y) \in \mathbb{R}^2 : \theta \leq a \wedge \frac{y}{\theta} \leq b\}$. Risulta allora

$$\begin{aligned} F_{(Z,X')}(a,b) &= \int_{\mathbb{R}^2} 4\,y\theta J(y)J(\theta)I_B(\theta,y)\,d\theta dy = \int_{\mathbb{R}^2} 4\,y\theta J(y)J(\theta)I_{]-\infty,b]}\big(\frac{y}{\theta}\big)I_{]-\infty,a]}(\theta)\,d\theta dy \\ &= \int_{\mathbb{R}} 4\,\Big(\int_{\mathbb{R}} yJ(y)I_{]-\infty,b]}\big(\frac{y}{\theta}\big)\,dy\Big)\,\theta J(\theta)I_{]-\infty,a]}(\theta)\,d\theta \\ &= \int_{-\infty}^{a} 4\,\Big(\int_{\mathbb{R}} yJ(y)I_{]-\infty,b]}\big(\frac{y}{\theta}\big)\,dy\Big)\,\theta J(\theta)\,d\theta. \end{aligned}$$

Dato $\theta > 0$ e considerata la trasformazione $\tau : y \mapsto \theta y$ della retta reale, dal Corollario A.4.2 si ha $\int_{\mathbb{R}} yJ(y)I_{]-\infty,b]}\big(\frac{y}{\theta}\big)\,dy = \int_{\mathbb{R}} x\theta^2 J(x\theta)I_{]-\infty,b]}(x)\,dx = \int_{-\infty}^{b} x\theta^2 J(x\theta)\,dx$. Ne segue

$$F_{(Z,X')}(a,b) = \int_{-\infty}^{a} 4\,\Big(\int_{-\infty}^{b} x\theta^2 J(x\theta)\,dx\Big)\,\theta J(\theta)\,d\theta = \int_{S_{ab}} 4\,x\theta^3 J(x\theta)J(\theta)\,d\theta dx.$$

La coppia aleatoria (Z,X') ha quindi densità congiunta $f_{Z,X'}(x,\theta) = 4\,x\theta^3 J(x\theta)J(\theta)$. Allora, $f_{X'}(1) = 4\int_0^1 \theta^3\,d\theta = 1$ e quindi $f_{Z|X'}(\theta|1) = 4\theta^3 J(\theta)$ per ogni θ.

Riesce dunque $\mathrm{P}_{Z|X}(\cdot|0) \neq \mathrm{P}_{Z|X'}(\cdot|1)$. Se ora si osserva che le proposizioni "X ha preso il valore 0" e "X' ha preso il valore 1" descrivono il *medesimo* evento (precisamente l'evento trascurabile $\{Y = Z\}$), quanto provato mostra che è piuttosto paradossale interpretare $\mathrm{P}_{Z|X}(\cdot|0)$ e $\mathrm{P}_{Z|X'}(\cdot|1)$, rispettivamente, come le leggi di Z "sapendo che X ha preso il valore 0" e "sapendo che X' ha preso il valore 1". ▲

Le probabilità $\mathrm{P}_{Z|X}(\cdot|x)$ sono di particolare importanza in quanto consentono di ottenere anche versioni della funzione di regressione di trasformate certe della coppia aleatoria su X come pure una formula di disintegrazione per la legge congiunta.

Teorema B.2.15 *Sussistono le seguenti proposizioni:*

(i) *Sia* $g : \Theta \times \mathfrak{X} \mapsto [0. + \infty]$ *una funzione* $\mathcal{T} \otimes \aleph$*-Borel misurabile tale che la funzione composta* $g(Z,X)$ *ammetta speranza matematica finita. Allora, la funzione* $x \mapsto \int_{\Theta} g(\theta,x)\,\mathrm{P}_{Z|X}(d\theta|x)$ *è una versione della funzione di regressione* $\mathrm{E}(g(Z,X)|X = \cdot)$*;*

(ii) $\mathrm{P}_{(Z,X)}(S) = \int_{\mathfrak{X}} \mathrm{P}_{Z|X}(S(x)|x)\,\mathrm{P}_X(dx)$ *per ogni* $S \in \mathcal{T} \otimes \aleph$[32].

[32]Ricordiamo che $S(x) = \{\theta \in \Theta : (\theta,x) \in S\}$ e $S(\theta) = \{x \in \mathfrak{X} : (\theta,x) \in S\}$.

Dimostrazione (i) Dato $\mathsf{A} \in \aleph$, dai teoremi B.1.10, B.2.13(ii) e di Tonelli otteniamo

$$\begin{aligned}\int_{\{X\in\mathsf{A}\}} g(Z,X)\,d\mathrm{P} &= \int_{\Theta\times\mathfrak{X}} g(\theta,x)I_A(x)f_{Z|X}(\theta|x)\,f_X(x)\,\nu\times\mu(d\theta\times dx)\\ &= \int_{\mathfrak{X}}\Big(\int_\Theta g(\theta,x)f_{Z|X}(\theta|x)\,\nu(d\theta)\Big)\,I_A(x)f_X(x)\,\mu(dx)\\ &= \int_A\Big(\int_\Theta g(\theta,x)\,\mathrm{P}_{Z|X}(d\theta|x)\Big)\,\mathrm{P}_X(dx).\end{aligned}$$

(ii) Dato $S \in \mathcal{T}\otimes\aleph$, ponendo $g = I_S$ in (i) risulta

$$\begin{aligned}\mathrm{P}_{(Z,X)}(S) &= \int_\Omega I_S(Z,X)\,d\mathrm{P} = \int_{\mathfrak{X}}\Big(\int_\Theta I_S(\theta,x)\,\mathrm{P}_{Z|X}(d\theta|x)\Big)\,\mathrm{P}_X(dx)\\ &= \int_{\mathfrak{X}}\Big(\int_\Theta I_{S(x)}(\theta)\,\mathrm{P}_{Z|X}(d\theta|x)\Big)\,\mathrm{P}_X(dx) = \int_{\mathfrak{X}}\mathrm{P}_{Z|X}(S(x)|x)\,\mathrm{P}_X(dx).\end{aligned}$$

La dimostrazione è così conclusa. ■

Per la definizione (B.7), la densità $f_{Z|X}(\cdot|x)$ coincide con la densità marginale di Z qualora $x \notin \mathfrak{X}_0$. Facciamo ora vedere come si possa instaurare un legame tra tali densità anche al di fuori di questo caso. A tal fine, posto $\Theta_0 = \{\,0 < f_Z < +\infty\}$, consideriamo, per ogni $\theta\in\Theta$, la funzione $\aleph$-Borel misurabile $f_{X|Z}(\cdot|\theta)$ così definita:

$$f_{X|Z}(x|\theta) = \begin{cases}\frac{f(\theta,x)}{f_Z(\theta)} & \text{se } \theta\in\Theta_0\\ f_X(x) & \text{se } \theta\notin\Theta_0\end{cases} \tag{B.9}$$

per ogni $x\in\mathfrak{X}$, e la corrispondente probabilità sulla σ-algebra $\aleph$:

$$\mathrm{P}_{X|Z}(\mathsf{A}|\theta) = f_{X|Z}(\cdot|\theta) * \mu(\mathsf{A}) = \int_{\mathsf{A}} f_{X|Z}(x|\theta)\,\mu(dx)$$

per ogni $\mathsf{A}\in\aleph$. Allora, con ragionamenti analoghi a quelli fatti in precedenza, otteniamo sia che la famiglia $(\mathrm{P}_{X|Z}(\cdot|\theta))_{\theta\in\Theta}$ è una versione della legge condizionale di X rispetto Z che la validità del prossimo risultato.

Teorema B.2.16 *Sussistono le seguenti proposizioni:*

(i) $\mathrm{P}_Z(\Theta_0) = 1$;

(ii) $(\theta,x)\mapsto f_{X|Z}(x|\theta)f_Z(\theta)$ *è una* (ν,μ)*-densità congiunta di* (Z,X);

(iii) $f_X(\cdot) = \int_\Theta f_{X|Z}(\cdot|\theta)\mathrm{P}_Z(d\theta)$ (μ-q.o.);

(iv) *Sia* $g : \Theta \times \mathfrak{X} \mapsto [0. +\infty]$ *una funzione* $\mathcal{T} \otimes \aleph$*-Borel misurabile tale che la funzione composta* $g(Z, X)$ *ammetta speranza matematica finita. Allora, la funzione* $\theta \mapsto \int_{\mathfrak{X}} g(\theta, x)\,\mathrm{P}_{X|Z}(dx|\theta)$ *è una versione della funzione di regressione* $\mathrm{E}(g(Z, X)|Z = \cdot)$*;*

(v) $\mathrm{P}_{(Z,X)}(S) = \int_\Theta \mathrm{P}_{X|Z}(S(\theta)|\theta)\,\mathrm{P}_Z(d\theta)$ *per ogni* $S \in \mathcal{T} \otimes \aleph$*.*

Sussiste inoltre, come facilmente si constata, il seguente teorema di Bayes che fornisce il collegamento di f_Z con $(f_{Z|X}(\cdot|x))_{x\in\mathfrak{X}_0}$ tramite la famiglia $(f_{X|Z}(\cdot|\theta))_{\theta\in\Theta}$ chiamata, per analogia al caso precedente, **μ-densità condizionale di X rispetto Z**[33].

Teorema B.2.17 (di Bayes) *Risulta:*

$$f_{Z|X}(\theta|x) = \frac{f_{X|Z}(x|\theta)}{\displaystyle\int_\Theta f_{X|Z}(x|\theta) f_Z(\theta)\,\nu(d\theta)}\, f_Z(\theta)$$

a meno di un insieme $\nu \times \mu$*-trascurabile.*

B.2.4 Densità iniziale, finale e predittiva

Gran parte delle applicazioni della statistica (sia inferenziale che decisionale) fanno riferimento alla struttura seguente (detta **esperimento statistico**): considerato un ente aleatorio osservabile X a valori nell'insieme $\mathfrak{X}$ (**spazio campionario**), si suppone che non sia possibile individuarne con precisione (per mancanza d'informazione) la legge ma solo ritenerla appartenente ad una data collezione di probabilità $\mathcal{P} = \{\mathrm{P}_\theta : \theta \in \Theta\}$ su $\aleph$ parametrizzata su un insieme Θ (**spazio parametrico**). Inoltre, interpretati (in accordo con il paradigma bayesiano) gli elementi dell'insieme Θ come possibili determinazioni di un ente aleatorio (**parametro**) Z, si considera, accanto alla famiglia $\mathcal{P}$, la legge di Z.

[33] La denominazione del teorema deriva da un celebre manoscritto del rev. Thomas Bayes (pubblicato postumo nel 1764 a tre anni dalla morte) nel quale si fornisce la probabilità condizionata di un evento E al numero di volte che si verifica in n osservazioni, nel caso che la probabilità $\Pr(E)$ sia distribuita uniformemente nell'intervallo $[0, 1]$. Osserviamo, a tale proposito, che Pierre S. de Laplace fu il primo a comprendere l'importanza cruciale del risultato di Bayes e ad applicarlo sistematicamente (a partire dal 1774) allo studio di numerosi problemi, sia delle scienze fisiche che morali (come, ad esempio, all'eccentricità delle orbite dei pianeti e alla validità delle testimonianze).

In questa situazione, viene naturale chiedersi se sia possibile costruire un contesto che consenta di interpretare $\mathcal{P}$ come una versione della legge condizionale di X rispetto Z. Non affronteremo il problema nella sua generalità, ma ci limiteremo a descrivere il procedimento (noto come **schema di Bayes**) che risolve positivamente il problema nel caso particolare che le probabilità considerate ammettano densità.

Date dunque, come nella sezione precedente, due misure di riferimento σ-finite μ e ν, rispettivamente su $\aleph$ e $\mathcal{T}$, assumiamo che $\pi : \Theta \mapsto [0, +\infty]$ sia una ν-densità (**densità iniziale**) di Z e, per ogni $\theta \in \Theta$, che f_θ sia una μ-densità di X relativa alla legge P_θ (detta **densità di campionamento relativa a** θ). Supponiamo inoltre che la funzione $(\theta, x) \mapsto f_\theta(x)$ sia $\mathcal{T} \otimes \aleph$-Borel misurabile. Allora, per il Lemma A.5.1(iii), la funzione $(\theta, x) \mapsto f_\theta(x)\pi(\theta)$ è $\mathcal{T} \otimes \aleph$-Borel misurabile e quindi possiamo considerare sulla σ-algebra prodotto la probabilità così definita:

$$P^{(\mathrm{sb})}(\hat{A}) = \int_{\hat{A}} f_\theta(x)\pi(\theta)\, \nu \times \mu(d\theta \times dx)$$

per ogni $\hat{A} \in \mathcal{T} \otimes \aleph$[34]. Se ora supponiamo che $P^{(\mathrm{sb})}$ sia la legge congiunta $\mathrm{P}_{(Z,X)}$ della coppia aleatoria (Z, X), la funzione $f' : (\theta, x) \mapsto f_\theta(x)\pi(\theta)$ diviene una $\nu \times \mu$-densità congiunta di (Z, X). Allora, tenuto conto della definizione (B.9), $f_\theta(\cdot) = f'_{X|Z}(\cdot|\theta)$ per ogni $\theta \in \Theta_0 = \{0 < \pi < +\infty\}$ e quindi $\mathrm{P}_\theta(\mathsf{A}) = \int_{\mathsf{A}} f_\theta \, d\mu = \mathrm{P}_{X|Z}(\mathsf{A}|\theta)$ per ogni $\mathsf{A} \in \aleph$ e $\theta \in \Theta_0$. Conseguentemente, l'identificazione $P^{(\mathrm{sb})} = \mathrm{P}_{(Z,X)}$ consente di interpretare la famiglia $\mathcal{P}$ come una versione della legge condizionale di X rispetto Z[35].

Assumiamo dunque $P^{(\mathrm{sb})} = \mathrm{P}_{(Z,X)}$. Allora, la **densità predittiva** p di X così definita:

$$p(x) = \int_\Theta f_\theta(x)\pi(\theta)\, \nu(d\theta)$$

per ogni $x \in \mathfrak{X}$, è una densità marginale di X mentre la **densità finale**

---

[34]Per constatare che $P^{(\mathrm{sb})}$ è una probabilità basta osservare che, per il teorema di Tonelli, $P^{(\mathrm{sb})}(\Theta \times \mathfrak{X}) = \int_\Theta \big(\int_{\mathfrak{X}} f_\theta(x)\, \mu(dx)\big)\pi(\theta)\, \nu(d\theta) = 1$.

[35]D'altra parte, questo è l'unico modo per ottenere tale interpretazione, nel senso che, assunta l'esistenza della densità congiunta f e supposto $f_\theta(\cdot) = f_{X|Z}(\cdot|\theta)$ per ogni $\theta \in \Theta_0$, dal Teorema B.2.16(ii) otteniamo che anche f' è una densità congiunta della coppia aleatoria e quindi deve essere $P^{(\mathrm{sb})} = \mathrm{P}_{(Z,X)}$.

$\pi_{Z|X}(\cdot|x)$ **di** Z **relativa a** x così definita:

$$\pi_{Z|X}(\theta|x) = \begin{cases} \frac{f_\theta(x)}{p(x)}\,\pi(\theta) & \text{se } x \in \mathfrak{X}_0 \\ \pi(\theta) & \text{se } x \notin \mathfrak{X}_0 \end{cases}$$

per ogni $\theta \in \Theta$, fornisce, al variare di $x \in \mathfrak{X}$, la densità condizionale di Z rispetto X.

Il calcolo della densità finale non sempre è agevole in quanto, in un contesto generale, la densità predittiva può essere individuata solamente ricorrendo ai metodi dell'integrazione numerica. Una considerevole semplificazione avviene quando si considerino densità appartenenti ad una **famiglia coniugata per** $(f_\theta)_{\theta\in\Theta}$, cioè ad una famiglia $\mathbb{D}$ di densità iniziali tale che, qualunque siano $x \in \mathfrak{X}$ e $\pi \in \mathbb{D}$, la densità finale $\pi_{Z|X}(\cdot|x)$ appartenga a $\mathbb{D}$[36]. Nell'esempio seguente forniamo famiglie coniugate relative a tre modelli statistico-probabilistici di particolare importanza per le applicazioni; vedremo così che le corrispondenti densità finali si ottengono semplicemente "aggiornando" i parametri delle densità iniziali.

Esempio B.2.18 Supponiamo che l'osservabile X sia una n-pla $\mathbf{X} = (X_1, \ldots, X_n)$ di v.a. equidistribuite e indipendenti in corrispondenza ad ogni valore del parametro; cioè, tali che la densità congiunta relativa a θ sia del tipo $f_\theta(\mathbf{x}) = g_\theta(x_1)\cdots g_\theta(x_n)$ per ogni $\theta \in \Theta$. Inoltre, poniamo, come d'uso, $\bar{\mathbf{x}} = \frac{1}{n}(x_1 + \cdots + x_n)$ per ogni n-pla reale $\mathbf{x}$.

(i) Per ogni $\theta \in \Theta = [0,1]$, sia g_θ la distribuzione di Bernoulli $Ber(\theta)$. Allora, $f_\theta(\mathbf{x}) = \theta^{n\bar{\mathbf{x}}}\,(1-\theta)^{n(1-\bar{\mathbf{x}})}$ per ogni $\mathbf{x} \in \mathfrak{X} = \{0,1\}^n$. Supposto che la densità iniziale sia quella della distribuzione $Beta(\alpha, \beta)$, poniamo

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1},$$

ove Γ denota, come d'uso, la funzione gamma: $\Gamma(t) = \int_0^{+\infty} \theta^{t-1}e^{-\theta}\,d\theta$ $(t > 0)$. Riesce allora

$$f_\theta(\mathbf{x})\pi(\theta) = K\,\theta^{\alpha_n(\mathbf{x})-1}(1-\theta)^{\beta_n(\mathbf{x})-1}$$

avendo posto:

$$\alpha_n(\mathbf{x}) = \alpha + n\bar{\mathbf{x}}, \quad \beta_n(\mathbf{x}) = \beta + n(1-\bar{\mathbf{x}}), \quad K = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

Passando alla densità predittiva, otteniamo

$$p(\mathbf{x}) = K\int_0^1 \theta^{\alpha_n(\mathbf{x})-1}\,(1-\theta)^{\beta_n(\mathbf{x})-1}\,d\theta = K\frac{\Gamma(\alpha_n(\mathbf{x}))\Gamma(\beta_n(\mathbf{x}))}{\Gamma(\alpha_n(\mathbf{x})+\beta_n(\mathbf{x}))} > 0$$

[36] Per maggiori ragguagli si veda, ad esempio, il capitolo nono del testo di DeGroot.

e quindi $\mathfrak{X}_0 = \mathfrak{X}$. Ne segue

$$\pi_{Z|\mathbf{X}}(\theta|\mathbf{x}) = \frac{\Gamma(\alpha_n(\mathbf{x}) + \beta_n(\mathbf{x}))}{\Gamma(\alpha_n(\mathbf{x}))\Gamma(\beta_n(\mathbf{x}))}\, \theta^{\alpha_n(\mathbf{x})-1}\, (1-\theta)^{\beta_n(\mathbf{x})-1}.$$

La densità finale di Z relativa a $\mathbf{x}$ è dunque quella della distribuzione $Beta(\alpha_n(\mathbf{x}), \beta_n(\mathbf{x}))$.

(ii) Per ogni $\theta \in \Theta = \mathbb{R}$, sia g_θ la densità della distribuzione normale $\mathsf{N}(\theta, \sigma^2)$. Allora

$$f_\theta(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}\Big]$$

per ogni $\mathbf{x} \in \mathfrak{X} = \mathbb{R}^n$. Conseguentemente, osservato che

$$\begin{aligned}
\sum_{i=1}^n (x_i-\theta)^2 &= \sum_{i=1}^n \left[(x_i-\bar{\mathbf{x}}) + (\bar{\mathbf{x}}-\theta)\right]^2 = \sum_{i=1}^n \left[(x_i-\bar{\mathbf{x}})^2 + 2(\bar{\mathbf{x}}-\theta)(x_i-\bar{\mathbf{x}}) + (\bar{\mathbf{x}}-\theta)^2\right] \\
&= \sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2 + 2(\bar{\mathbf{x}}-\theta)\sum_{i=1}^n (x_i-\bar{\mathbf{x}}) + n(\bar{\mathbf{x}}-\theta)^2 \\
&= \sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2 + 2(\bar{\mathbf{x}}-\theta)\Big(\sum_{i=1}^n x_i - n\bar{\mathbf{x}}\Big) + n(\bar{\mathbf{x}}-\theta)^2 \\
&= \sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2 + n(\bar{\mathbf{x}}-\theta)^2,
\end{aligned}$$

otteniamo

$$f_\theta(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2}{2\sigma^2}\Big] \exp\Big[-\frac{n(\bar{\mathbf{x}}-\theta)^2}{2\sigma^2}\Big]. \tag{B.10}$$

Ne segue, supposto che la densità iniziale π sia quella della distribuzione normale $\mathsf{N}(\alpha, \beta^2)$,

$$f_\theta(\mathbf{x})\pi(\theta) = K_n(\mathbf{x}) \exp\Big\{-\frac{1}{2}\Big[\frac{(\theta-\alpha)^2}{\beta^2} + \frac{n(\bar{\mathbf{x}}-\theta)^2}{\sigma^2}\Big]\Big\}$$

con

$$K_n(\mathbf{x}) = (2\pi\sigma^2)^{-\frac{n}{2}}\, (2\pi\beta^2)^{-\frac{1}{2}} \exp\Big[-\frac{\sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2}{2\sigma^2}\Big].$$

Posto allora:

$$\beta_n^2 = \Big(\frac{1}{\beta^2} + \frac{n}{\sigma^2}\Big)^{-1}, \quad \alpha_n(\mathbf{x}) = \Big[\frac{\alpha}{\beta^2} + \frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]\beta_n^2, \quad K_n'(\mathbf{x}) = \frac{\alpha_n(\mathbf{x})^2}{\beta_n^2} - \Big[\frac{\alpha^2}{\beta^2} + \frac{n\bar{\mathbf{x}}^2}{\sigma^2}\Big]$$

risulta

$$f_\theta(\mathbf{x})\pi(\theta) = K_n(\mathbf{x}) \exp\Big[\frac{K_n'(\mathbf{x})}{2}\Big] \exp\Big[-\frac{[\theta-\alpha_n(\mathbf{x})]^2}{2\beta_n^2}\Big],$$

osservato che

$$\begin{aligned}
\frac{(\theta-\alpha)^2}{\beta^2} + \frac{n(\bar{\mathbf{x}}-\theta)^2}{\sigma^2} &= \Big[\frac{1}{\beta^2}+\frac{n}{\sigma^2}\Big]\,\theta^2 - 2\Big[\frac{\alpha}{\beta^2}+\frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]\theta + \Big[\frac{\alpha^2}{\beta^2}+\frac{n\bar{\mathbf{x}}^2}{\sigma^2}\Big] \\
&= \frac{1}{\beta_n^2}\Big\{\theta^2 - 2\Big[\frac{\alpha}{\beta^2}+\frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]\beta_n^2\,\theta + \Big[\frac{\alpha}{\beta^2}+\frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]^2\beta_n^2\Big\} \\
&\qquad - \Big[\frac{\alpha}{\beta^2}+\frac{n\bar{\mathbf{x}}}{\sigma^2}\Big]^2\beta_n^2 + \Big[\frac{\alpha^2}{\beta^2}+\frac{n\bar{\mathbf{x}}^2}{\sigma^2}\Big] \\
&= \frac{[\theta^2 - 2\alpha_n(\mathbf{x})\theta + \alpha_n(\mathbf{x})]^2}{\beta_n^2} - \Big\{\frac{\alpha_n(\mathbf{x})^2}{\beta_n^2} - \Big[\frac{\alpha^2}{\beta^2}+\frac{n\bar{\mathbf{x}}^2}{\sigma^2}\Big]\Big\} \\
&= \frac{[\theta-\alpha_n(\mathbf{x})]^2}{\beta_n^2} - K_n'(\mathbf{x}).
\end{aligned}$$

Passando alla densità predittiva, si ha allora

$$p(\mathbf{x}) = K_n(\mathbf{x})\exp\Big[\frac{K_n'(\mathbf{x})}{2}\Big]\int_{-\infty}^{+\infty}\exp\Big[-\frac{[\theta-\alpha_n(\mathbf{x})]^2}{2\beta_n^2}\Big]d\theta = K_n(\mathbf{x})\exp\Big[\frac{K_n'(\mathbf{x})}{2}\Big]\sqrt{2\pi\beta_n^2} > 0$$

e quindi $\mathfrak{X}_0 = \mathfrak{X}$. Ne segue

$$\pi_{Z|\mathbf{X}}(\theta|\mathbf{x}) = \frac{1}{\sqrt{2\pi\beta_n^2}}\exp\Big[-\frac{[\theta-\alpha_n(\mathbf{x})]^2}{2\beta_n^2}\Big].$$

La densità finale di Z relativa a $\mathbf{x}$ è dunque quella della distribuzione normale $\mathsf{N}(\alpha_n(\mathbf{x}), \beta_n^2)$.

(iii) Per ogni $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta = \mathbb{R}\times\,]0,+\infty[$, sia $g_{\boldsymbol{\theta}}$ la densità della distribuzione normale $\mathsf{N}(\theta_1, \theta_2)$. Da (B.10) otteniamo allora

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}\theta_2^{-\frac{n}{2}}\exp\Big[-\frac{\sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2 + n(\bar{\mathbf{x}}-\theta_1)^2}{2\theta_2}\Big]$$

per ogni $\mathbf{x} \in \mathfrak{X} = \mathbb{R}^n$. Supposto che la densità iniziale sia quella della distribuzione normale-gamma inversa $\mathsf{N}\text{-}\Gamma^{-1}(\alpha,\beta,\gamma,\delta)$, poniamo

$$\pi(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}}\frac{\delta^\gamma\sqrt{\beta}}{\Gamma(\gamma)}\theta_2^{-(\gamma+\frac{3}{2})}\exp\Big[-\frac{\beta(\theta_1-\alpha)^2+2\delta}{2\theta_2}\Big].$$

Ne segue, posto:

$$\alpha_n(\mathbf{x}) = \frac{n\bar{\mathbf{x}}+\alpha\beta}{\beta+n}, \quad \beta_n = \beta+n, \quad \gamma_n = \gamma+\frac{n}{2}$$

$$\delta_n(\mathbf{x}) = \delta + \frac{1}{2}\Big(\sum_{i=1}^n (x_i-\bar{\mathbf{x}})^2 + \frac{n\beta(\bar{\mathbf{x}}-\alpha)^2}{\beta+n}\Big), \quad K_n = (2\pi)^{-\frac{n+1}{2}}\frac{\delta^\gamma\sqrt{\beta}}{\Gamma(\gamma)}$$

la seguente espressione della densità congiunta:

$$f_{\boldsymbol{\theta}}(\mathbf{x})\pi(\boldsymbol{\theta}) = K_n\,\theta_2^{-(\gamma_n+\frac{3}{2})}\exp\Big[-\frac{\beta_n(\theta_1-\alpha_n(\mathbf{x}))^2+2\delta_n(\mathbf{x})}{2\theta_2}\Big],$$

osservato che

$$\begin{aligned}
\beta(\theta_1-\alpha)^2+n(\bar{\mathbf{x}}-\theta_1)^2 &= \beta_n\left[\theta_1^2-2\alpha_n(\mathbf{x})\theta_1+\frac{n\bar{\mathbf{x}}^2+\alpha^2\beta}{\beta_n}\right]\\
&= \beta_n\left[\big(\theta_1^2-2\alpha_n(\mathbf{x})\theta_1+\alpha_n(\mathbf{x})^2\big)-\alpha_n(\mathbf{x})^2+\frac{n\bar{\mathbf{x}}^2+\alpha^2\beta}{\beta_n}\right]\\
&= \beta_n\big(\theta_1-\alpha_n(\mathbf{x})\big)^2-\beta_n\alpha_n(\mathbf{x})^2+n\bar{\mathbf{x}}^2+\alpha^2\beta\\
&= \beta_n\big(\theta_1-\alpha_n(\mathbf{x})\big)^2-\frac{(n\bar{\mathbf{x}}+\alpha\beta)^2}{\beta_n}+n\bar{\mathbf{x}}^2+\alpha^2\beta\\
&= \beta_n\big(\theta_1-\alpha_n(\mathbf{x})\big)^2+\frac{n\beta(\bar{\mathbf{x}}-\alpha)^2}{\beta_n}.
\end{aligned}$$

Passando alla densità predittiva, si ha allora

$$\begin{aligned}
p(\mathbf{x}) &= K_n\int_{\mathbb{R}\times\,]0.+\infty]}\theta_2^{-(\gamma_n+\frac{3}{2})}\exp\left[-\frac{\beta_n(\theta_1-\alpha_n(\mathbf{x}))^2+2\delta_n(\mathbf{x})}{2\theta_2}\right]d\theta_1 d\theta_2\\
&= K_n\,\frac{\sqrt{2\pi}\,\Gamma(\gamma_n)}{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}>0
\end{aligned}$$

e quindi $\mathfrak{X}_0=\mathfrak{X}$. Ne segue

$$\pi_{Z|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x})=\frac{1}{\sqrt{2\pi}}\,\frac{\delta_n(\mathbf{x})^{\gamma_n}\sqrt{\beta_n}}{\Gamma(\gamma_n)}\,\theta_2^{-(\gamma_n+\frac{3}{2})}\exp\left[-\frac{\beta_n(\theta_1-\alpha_n(\mathbf{x}))^2+2\delta_n(\mathbf{x})}{2\theta_2}\right].$$

La densità finale di Z relativa a $\mathbf{x}$ è dunque quella della distribuzione normale-gamma inversa $\mathsf{N}\text{-}\Gamma^{-1}(\alpha_n(\mathbf{x}),\beta_n,\gamma_n,\delta_n(\mathbf{x}))$. ▲

Bibliografia essenziale

[1] Ash, R.B. (1972): Real Analysis and Probability, Academic Press, New York.

[2] Berger, J.O. (1985): Statistical Decision Theory and Bayesian Analysis (Second Edition), Springer-Verlag, New York.

[3] Bernardo, J.M. - Smith, A.F.M. (1994): Bayesian Theory, John Wiley & Sons, Chichester.

[4] DeGroot, M.H. (1970): Optimal Statistical Decisions, McGraw-Hill, New York.

[5] Ferguson, T.S. (1967): Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York

[6] Hoffmann-Jorgensen, J. (1994): Probability with a view toward Statistics, Volumi I e II, Chapman & Hall, New York

[7] Lehmann, E.L. (1983): Theory of Point Estimation, John Wiley & Sons, New York.

[8] Lehmann, E.L. (1986): Testing Statistical Hypothesis (Second Edition), John Wiley & Sons, New York.

[9] Piccinato, L. (1996): Metodi per le Decisioni Statistiche, Springer-Verlag Italia, Milano.

[10] Robert, C.P. (1994): The Bayesian Choice: A Decision-Theoretic Motivation, Springer-Verlag, New York.

[11] Schervish, M.J. (1995): Theory of Statistics, Springer-Verlag, New York.