Probabilità Elementare

(Appunti per gli studenti)

Testo provvisorio in fase di revisione

Silvano Holzer

DEAMS

Universita degli Studi di Trieste

20 maggio 2015

Indice

Il testo è un sommario delle lezioni di natura teorica del corso di Calcolo delle Probabilità posto al primo anno del corso di laurea in Statistica e Informatica per l'Azienda, la Finanza e l'Assicurazione. Scopo del corso, nelle sue componenti teorica ed esercitativa, è fornire sia una esposizione elementare dei principali risultati del calcolo delle probabilità che porre le basi per i successivi sviluppi astratti che vengono trattati nel corso di Calcolo delle Probabilità (corso progredito) del terzo anno di corso.

Giudizi, critiche e suggerimenti da parte degli studenti o da altri eventuali lettori saranno accolti con gratitudine.

L'oggetto della teoria delle probabilità appartiene in ugual misura alla scienza del numero e a quella della logica.

George Boole, *An Investigation of the Laws of Thought*, 1853

Capitolo 1

Descrizione dell'incertezza

Impostando il problema della descrizione dell'incertezza in termini logici, si introducono le due nozioni chiave di evento (assoluto e condizionato) e di numero aleatorio ancorandole allo stato d'informazione. Si definiscono pure alcune fondamentali operazioni e relazioni tra eventi studiandone le principali proprietà.

1.1 Enunciati. Forme enunciative

Tra gli elementi linguistici di una lingua naturale (nomi, parole, frasi, ...) si può fare una prima distinzione riguardante il ruolo da essi svolto dal punto di vista del significato. Elementi linguistici:
- che hanno un significato (ad esempio, una proprietà, una relazione, un individuo);
- che, agendo sul significato di altri elementi linguistici, ne generano uno nuovo dotato ancora di significato (ad esempio le particelle "e", "o", "non"e i quantificatori "tutti", "qualche").

Nell'ambito dei primi, chiamiamo **enunciato** ogni elemento linguistico per il quale abbia senso chiedersi se è vero o è falso. Sono quindi esempi di enunciati:

(1) l'ambo "7,45"uscirà nella prossima estrazione del lotto sulla ruota di Venezia

(2) Otello è geloso di Desdemona

(3) $7 + 3 = 10$

(4) $\sqrt{2}$ è un numero razionale

(5) nei due lanci consecutivi di un dado a sei facce il 5 uscirà prima del 4

(6) Carlo presenta Maria a Bruno

mentre non lo sono:

- Alto la!

- Carlo suona la tromba?

Gli esempi (1), (4) enunciano il fatto che una certa proprietà si addice ad un certo individuo; (2), (5) enunciano il fatto che una certa relazione binaria sussiste tra una certa coppia di individui; (3), (6) enunciano il fatto che una certa relazione ternaria sussiste tra una certa terna di individui.

Se ora s'introducono, come da secoli viene fatto nelle scienze deduttive, quelle entità segniche chiamate *variabili*, possiamo esprimere le proprietà e le relazioni binarie e ternarie che intervengono negli enunciati sopra considerati nel modo seguente:
- l'ambo "x,y" uscirà nella prossima estrazione del lotto sulla ruota di Venezia
- x è geloso di y
- x + y = z
- x è un numero razionale
- nei due lanci consecutivi di un dado a sei facce il numero x uscirà prima del numero y
- x presenta y a z

individuando così particolari **forme enunciative**, cioè configurazioni linguistiche che pur non essendo degli enunciati lo diventano non appena si sostituiscano le variabili in esse occorrenti con individui.

Gli enunciati "Socrate è il filosofo greco che nel 399 a.C. bevve in Atene la cicuta", "Socrate era il maestro di Platone" e "Socrate è Socrate" sono, pur essendo tutti veri, profondamente diversi dal punto di vista dell'interesse conoscitivo. Ciò comporta che di un enunciato dobbiamo distinguere due aspetti cruciali: il suo valore di verità (*estensione*) e il suo contenuto concettuale (*intensione*)[1].

[1] Aspetti che possono prevalere l'uno rispetto all'altro, a seconda del contesto nel quale è inserito l'enunciato. Ad esempio, nel caso dell'enunciato diretto "Alessandro Manzoni è l'autore dei Promessi Sposi" risulta interessante il suo aspetto estensionale, mentre nel caso dell'enunciato indiretto "Carlo sa che Alessandro Manzoni è l'autore dei Promessi Sposi", dell'enunciato "Alessandro Manzoni è l'autore dei Promessi Sposi" risulta interessante il suo aspetto intensionale (cioè il concetto che esprime).

Passando a considerare una forma enunciativa, identifichiamo la sua estensione con la classe degli individui che, sostituiti alle variabili in essa occorrenti, la trasformano in un enunciato vero e la sua intensione con la proprietà da essa descritta. Così , ad esempio, le forme enunciative:
- x è un triangolo con due angoli uguali
- x è un triangolo con due lati uguali

hanno intensione diversa (avere due angoli uguali è concettualmente diverso da avere due lati uguali) ma medesima estensione, cioè la classe dei triangoli isosceli.

Precisiamo che *nel seguito considereremo solo l'aspetto estensionale* degli enunciati e delle forme enunciative, tralasciando completamente il loro aspetto intensionale.

1.2 Connettivi

Due enunciati possono sempre venir combinati in modo da ottenere un nuovo enunciato tramite i *connettivi*, tra cui quelli più comuni sono: il **bicondizionale** “se e solo se”, la **congiunzione** “e”, la **disgiunzione** “o”(nel senso del latino “vel”) e il **condizionale** “se..., allora”(nel senso della *implicazione materiale*[2]). Indicandoli, in ordine, con le notazioni $\leftrightarrow$, $\wedge$, $\vee$ e $\rightarrow$, le estensioni dei corrispondenti enunciati vengono riportate nella tabella seguente:

p	q	$p \leftrightarrow q$	$p \wedge q$	$p \vee q$	$p \rightarrow q$
V	V	V	V	V	V
V	F	F	F	V	F
F	V	F	F	V	V
F	F	V	F	F	V

ove “V”significa vero e “F”falso. Conseguentemente:
- il bicondizionale di due enunciati è vero se e solo se entrambi gli enunciati sono veri oppure entrambi sono falsi;
- la congiunzione di due enunciati è vera se e solo se entrambi gli enunciati sono veri;
- la disgiunzione di due enunciati è vera se e solo se almeno uno degli enunciati è vero;

[2]Nota dai tempi di Filone di Megara (IV sec. a.C.) e quindi detta anche *implicazione filoniana*.

- il condizionale $p \to q$ è vero se e solo se il suo *antecedente* p è falso oppure il suo *conseguente* q è vero.

Osservazione 1.2.1 Per convincersi che sia abbastanza naturale ritenere il condizionale vero in tutti i casi tranne quello in cui l'antecedente è vero e il conseguente è falso, si pensi ai condizionali "se 7 è un numero pari, allora 8 è un numero dispari"(antecedente e conseguente falsi) e "se $\sqrt{2}$ è un numero razionale, allora $\sqrt{2}\sqrt{2} = 2$ è un numero razionale"(antecedente falso e conseguente vero). D'altra parte si rifletta sul fatto che, avendo scelto una trattazione basata solamente sull'estensione degli enunciati, risultano veri anche condizionali apparentemente paradossali come, ad esempio, "se $3 > 7$, allora Socrate era il maestro di Platone".

Osserviamo infine che solo l'implicazione materiale assicura sia la validità dell'usuale procedura di "andata e ritorno":

"Per dimostrare che due enunciati hanno medesima estensione basta verificare che il primo implica il secondo e che il secondo implica il primo"

che la diversità, dal punto di vista estensionale, del bicondizionale e del condizionale. Infatti, indicati con A, A', B, B', C valori generici di verità, dalla tabella:

p	q	$p \leftrightarrow q$	$p \to q$	$q \to p$	$(p \to q) \wedge (q \to p)$
V	V	V	V	V	V
V	F	F	F	A'	F
F	V	F	A	F	F
F	F	V	B	B'	C

risulta che gli enunciati $p \leftrightarrow q$ e $(p \to q) \wedge (q \to p)$ hanno, in ogni caso, medesima estensione solamente se C = V, cioè se $B = B' = V$; inoltre, per distinguere estensionalmente, il bicondizionale dal condizionale, deve essere $A = A' = V$. △

Dato infine un enunciato p è sempre possibile ottenerne un altro tramite la particella "non": la sua **negazione**. Indicata con $\neg p$ detta negazione, la tabella seguente ne fornisce l'estensione.

p	$\neg p$
V	F
F	V

Dunque: la negazione di un enunciato è vera se e solo se l'enunciato è falso.

1.3 Quantificatori

Considerata una forma enunciativa $P(x)$ (ove x rappresenta le variabili occorrenti in essa), sia U l'**universo del discorso** (cioè la classe degli individui con i quali intendiamo sostituire la variabile x). Tra i vari operatori che agiscono sulla forma enunciativa trasformandola in un enunciato, di particolare importanza sono i *quantificatori* che trasformano $P(x)$ in un enunciato specificando la quantità (almeno uno, uno solo, un numero finito, tutti, ...) degli individui di U che verificano la proprietà espressa dalla forma enunciativa.

Di notevole interesse sono il **quantificatore universale** $\forall$ e il **quantificatore esistenziale** $\exists$. Il primo quantifica *universalmente* la variabile trasformando la forma enunciativa nell'enunciato "**tutti** gli individui di U godono della proprietà in questione"; enunciato che indicheremo con la notazione $\forall x \in UP(x)$. Il secondo esistenziale invece quantifica *esistenzialmente* la variabile trasformando la forma enunciativa nell'enunciato "**almeno** un individuo di U gode della proprietà in questione"; enunciato che indicheremo con la notazione $\exists x \in UP(x)$.

Passando alla loro estensione, è del tutto naturale farla dipendere dalle estensioni degli enunciati ottenuti sostituendo alla variabile x gli individui u di U. Considerata allora l'applicazione $\text{ext}^{(P(x))} : U \to \{V, F\}$ che associa ad ogni individuo $u \in U$ l'estensione dell'enunciato $P(u)$, le estensioni dei due quantificatori in esame sono date dalla tabella:

$\text{ext}^{(P(x))}(U)$	$\forall x \in UP(x)$	$\exists x \in UP(x)$
{V}	V	V
{V, F}	F	V
{F}	F	F

Conseguentemente:
- la quantificazione universale della forma enunciativa $P(x)$ è vera se e solo se l'enunciato $P(u)$ è vero qualunque sia l'individuo u in U;
- la quantificazione esistenziale della forma enunciativa $P(x)$ è vera se e solo se l'enunciato $P(u)$ è vero per qualche individuo u in U.

Se ora indichiamo con p_u l'enunciato $P(u)$ - ottenuto sostituendo in $P(x)$ la variabile x con l'elemento $u \in U$ - possiamo considerare la famiglia di enunciati $(p_u)_{u \in U}$ e quindi esprimere gli enunciati $\forall x \in UP(x)$ e $\exists x \in UP(x)$, in forma equivalente, rispettivamente tramite gli enunciati "per ogni $u \in U$ vale p_u" e "per qualche $u \in U$ vale p_u".

Generalizzando il discorso, queste considerazioni suggeriscono di considerare, per ogni famiglia di enunciati $\mathcal{P} = (p_i)_{i\in I}$ con $I \neq \emptyset$, la:
- **congiunzione multipla** di $\mathcal{P}$ (in simboli $\bigwedge_{i\in I} p_i$), cioè l'enunciato "per ogni $i \in I$ vale p_i";
- **disgiunzione multipla** di $\mathcal{P}$ (in simboli $\bigvee_{i\in I} p_i$), cioè l'enunciato "per qualche $i \in I$ vale p_i"
e, per quanto riguarda la loro estensione, assumere (in analogia al caso dei quantificatori) che:
- la **congiunzione multipla** di $\mathcal{P}$ è vera se e solo se ogni enunciato della famiglia $\mathcal{P}$ è vero;
- la **disgiunzione multipla** di $\mathcal{P}$ è vera se e solo se qualche enunciato della famiglia $\mathcal{P}$ è vero.

Per quanto riguarda infine il caso $I = \emptyset$, adottiamo l'usuale convenzione di assumere la congiunzione multipla $\bigwedge_{i\in\emptyset} p_i$ vera e la disgiunzione multipla $\bigvee_{i\in\emptyset} p_i$ falsa[3].

1.4 Alcune importanti leggi logiche

Chiamato **tautologia** un qualsiasi enunciato composto (cioè costruito mediante connettivi a partire da altri enunciati) che è vero qualunque sia l'estensione degli enunciati che lo compongono, per **legge logica** intendiamo una qualsiasi proposizione - costruita tramite connettivi a partire da un numero finito di variabili $p, q, r, s, \ldots$ denotanti enunciati - che diviene una tautologia comunque si sostituiscano le variabili con enunciati. Riportiamo ora alcune leggi logiche (di verifica immediata) che saranno utili nel seguito. Per evitare di usare troppe parentesi, assumiamo che i connettivi $\wedge$, $\vee$ leghino più strettamente dei connettivi $\leftrightarrow$, $\rightarrow$. Infine precisiamo che nella prima legge logica π denota una qualsiasi permutazione dell'insieme degli indici I.

(1) $\bigwedge_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I} p_{\pi(i)}, \quad \bigvee_{i\in I} p_i \leftrightarrow \bigvee_{i\in I} p_{\pi(i)}$ (legge commutativa)

(2) $(p \wedge q) \wedge r \leftrightarrow p \wedge (q \wedge r), \quad (p \vee q) \vee r \leftrightarrow p \vee (q \vee r)$ (l. associativa)

(3) $p \wedge \bigvee_{i\in I} p_i \leftrightarrow \bigvee_{i\in I}(p \wedge p_i), \quad p \vee \bigwedge_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I}(p \vee p_i)$ (l. distributiva)

(4) $p \leftrightarrow \bigwedge_{i\in I} p_i, \quad p \leftrightarrow \bigvee_{i\in I} p_i$, se $p_i \leftrightarrow p$ per ogni $i \in I$ (l. di idempotenza)

[3] Per giustificarla basta osservare che se fosse $\bigwedge_{i\in\emptyset} p_i$ falso ($\bigvee_{i\in\emptyset} p_i$ vero), esisterebbe un enunciato della famiglia falso (vero) e quindi $I \neq \emptyset$.

(5) $p \wedge (p \vee q) \leftrightarrow p, \quad p \vee (p \wedge q) \leftrightarrow p$ (l. di assorbimento)

(6) $\neg \bigwedge_{i\in I} p_i \leftrightarrow \bigvee_{i\in I} \neg p_i, \quad \neg \bigvee_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I} \neg p_i$ (leggi di De Morgan)

(7) $\neg(\neg p) \leftrightarrow p$ (l. della doppia negazione)

(8) $\neg p \vee p$ (l. del terzo escluso)

(9) $\neg(\neg p \wedge p)$ (l. di non contraddizione)

(10) $(p \to q) \leftrightarrow \neg p \vee q$ (l. di Filone di Megara)

(11) $\bigwedge_{i\in I}(p_{i0} \vee p_{i1}) \leftrightarrow \bigvee_{f\in\{0,1\}^I}(\bigwedge_{i\in I} p_{if(i)})$

(12) $\bigwedge_{i\in I_1\cup I_2} p_i \leftrightarrow (\bigwedge_{i\in I_1} p_i)\wedge(\bigwedge_{i\in I_2} p_i), \quad \bigvee_{i\in I_1\cup I_2} p_i \leftrightarrow (\bigvee_{i\in I_1} p_i)\vee(\bigvee_{i\in I_2} p_i)$

(13) $\bigwedge_{i\in I}(p_i\wedge q_i) \leftrightarrow (\bigwedge_{i\in I} p_i)\wedge(\bigwedge_{i\in I} q_i), \quad \bigvee_{i\in I}(p_i\vee q_i) \leftrightarrow (\bigvee_{i\in I} p_i)\vee(\bigvee_{i\in I} q_i)$

1.5 Eventi

Considerato un **stato di conoscenza** $\mathcal{C}$ (cioè una famiglia di enunciati ritenuti veri (in via ipotetica o effettiva)), denotiamo, per ogni enunciato p, con $\mathcal{C} \models p$ la proposizione "lo stato di conoscenza $\mathcal{C}$ *costringe* l'enunciato p *ad essere vero*"[4].

Dato un insieme non vuoto di enunciati $\mathcal{L}$, diciamo che gli enunciati $p, q \in \mathcal{L}$ sono **equivalenti sub** $\mathcal{C}$ (in simboli $p \sim_{\mathcal{C}} q$) se $\mathcal{C} \models p \leftrightarrow q$, cioè se lo stato di conoscenza $\mathcal{C}$ costringe p e q ad avere la medesima estensione. Osservato che tale relazione è una relazione di equivalenza, possiamo introdurre la nozione di evento relativo allo stato di conoscenza $\mathcal{C}$ (ricorrendo all'usuale *metodo di definizione per astrazione*) chiamando **$\mathcal{C}$-eventi** di $\mathcal{L}$ le relative classi di equivalenza, cioè gli elementi dell'insieme quoziente $\mathcal{E}_{\mathcal{C}} = \mathcal{L}/\sim_{\mathcal{C}}$; inoltre, per

[4]In altri termini, nello stato di conoscenza $\mathcal{C}$ si è in grado di affermare la verità di p; conseguentemente, se $\mathcal{C}$ è uno stato di conoscenza effettivo, p è certamente vero; se invece è uno stato di conoscenza ipotetico, nulla si può dire sulla verità di p, salvo affermarla qualora capiti di apprendere che tutti gli enunciati di $\mathcal{C}$ assunti veri in via ipotetica sono in realtà veri. Osserviamo infine che la differenza fondamentale tra il concetto di "verità"e quello di "costringe ad essere vero"è rappresentata dal fatto che mentre la verità gode del principio del terzo escluso, il secondo concetto in generale non ne gode; infatti, la frase "nella prossima estrazione del lotto sulla ruota di Venezia uscirà il numero 45" è vera o falsa ma non siamo in grado, nell'attuale stato di conoscenza, di affermare nè che è vera, nè che è falsa.

ogni enunciato p di $\mathcal{L}$, indichiamo con $[p]_\mathcal{C}$ il relativo $\mathcal{C}$-evento - cioè poniamo $[p]_\mathcal{C} = \{q \in \mathcal{L} : q \sim_\mathcal{C} p\}$ - e chiamiamo p una sua *descrizione*[5]. Poichè lo stato di conoscenza $\mathcal{C}$ costringe ogni descrizione dell'evento $E = [p]_\mathcal{C}$ ad avere la medesima estensione di p, viene naturale assumere che l'estensione dell'evento E coincida con quella di p.

Supposto che esistano due enunciati $p', p'' \in \mathcal{L}$ tali che $\mathcal{C} \models p'$ e $\mathcal{C} \models \neg p''$, chiamiamo **$\mathcal{C}$-evento certo** di $\mathcal{L}$ e **$\mathcal{C}$-evento impossibile** di $\mathcal{L}$, rispettivamente, gli eventi $\Omega_\mathcal{C} = [p']_\mathcal{C}$ e $\emptyset_\mathcal{C} = [p'']_\mathcal{C}$; dunque l'evento certo (impossibile) colleziona gli enunciati di $\mathcal{L}$ che si è in grado di affermare veri (falsi) nello stato di conoscenza $\mathcal{C}$ e quindi il suo valore di verità (in tale stato di conoscenza) è "vero"("falso"). Infine, chiameremo **$\mathcal{C}$-evento possibile** di $\mathcal{L}$ ogni $\mathcal{C}$-evento che sia diverso da $\Omega_\mathcal{C}$ e da $\emptyset_\mathcal{C}$.

Osservazione 1.5.1 Qualora lo stato di conoscenza venga sottinteso, elimineremo dai $\mathcal{C}$-eventi il prefisso $\mathcal{C}$. Conseguentemente parleremo di **eventi** di $\mathcal{L}$, di **eventi possibili** di $\mathcal{L}$, di **evento certo** di $\mathcal{L}$ e di **evento impossibile** di $\mathcal{L}$ e denoteremo questi ultimi due, rispettivamente con Ω e $\emptyset$. Inoltre, denoteremo sempre gli eventi con lettere latine maiuscole. $\triangle$

Esempio 1.5.2 Sia $\mathcal{L}$ costituito dai seguenti enunciati riguardanti il lancio di un dado con sei facce:

(a) esce il numero 2;	(f) esce un numero minore di 5;
(b) esce un numero primo;	(g) esce un numero primo pari;
(c) esce un numero con due cifre;	(h) esce uno dei numeri 2, 4, 6;
(d) esce un numero minore di 50;	(i) esce un numero con una sola cifra;
(e) esce un numero pari;	(l) esce un numero dispari.

Per gli stati di conoscenza seguenti si suppone che non includano alcuna informazione sul lancio del dado eccetto quella dichiarata.

[5]Osserviamo che, per l'impostazione estensionale della nozione di evento quì adottata, le descrizioni di un evento possono avere intensioni profondamente diverse; ad esempio, gli enunciati "Maria si sposò ed ebbe un figlio" e "Maria ebbe un figlio e si sposò" determinano il medesimo evento pur essendo, nel linguaggio comune, di significato totalmente diverso.

L'idea di associare l'evento ad uno stato d'informazione, che può sembrare strana, è presente, in modo più o meno velato, sin dagli albori del calcolo delle probabilità. Infatti, in *La logique, ou l'art de penser* di Port Royal (1662) possiamo leggere: "Al fine di stabiiire la verità di un certo evento, e di decidere se credere o no nel suo verificarsi, è necessario considerare l'evento non isolatamente, come si farebbe per una proposizione di geometria, ma in relazione a tutte le circostanze, sia interne che esterne, che lo accompagnano."

Lo stato di conoscenza $\mathcal{C}_1$ include l'informazione "il dado viene lanciato". Allora, gli eventi sono: $\Omega_{\mathcal{C}_1} = \{(\mathrm{d}), (\mathrm{i})\}$, $\emptyset_{\mathcal{C}_1} = \{(\mathrm{c})\}$, $E_1 = \{(\mathrm{a}), (\mathrm{g})\}$, $E_2 = \{(\mathrm{e}), (\mathrm{h})\}$, $E_3 = \{(\mathrm{b})\}$, $E_4 = \{(\mathrm{f})\}$ e $E_5 = \{(\mathrm{l})\}$.

Lo stato di conoscenza $\mathcal{C}_2$ include l'informazione "il dado viene lanciato e esce un numero pari". Allora, gli eventi sono: $\Omega_{\mathcal{C}_2} = \{(\mathrm{d}), (\mathrm{e}), (\mathrm{h}), (\mathrm{i})\}$, $\emptyset_{\mathcal{C}_2} = \{(\mathrm{c}), (\mathrm{l})\}$, $F_1 = \{(\mathrm{a}), (\mathrm{b}), (\mathrm{g})\}$ e $F_2 = \{(\mathrm{f})\}$.

Lo stato di conoscenza $\mathcal{C}_3$ include l'informazione "il dado viene lanciato e esce il 2 oppure il 4". Allora gli eventi sono: $\Omega_{\mathcal{C}_3} = \{(\mathrm{d}), (\mathrm{e}), (\mathrm{f}), (\mathrm{h}), (\mathrm{i})\}$, $\emptyset_{\mathcal{C}_3} = \{(\mathrm{c}), (\mathrm{l})\}$ e $G_1 = \{(\mathrm{a}), (\mathrm{b}), (\mathrm{g})\}$.

Lo stato di conoscenza $\mathcal{C}_4$ include l'informazione "il dado viene lanciato e esce il 2 oppure il 3". Allora gli eventi sono: $\Omega_{\mathcal{C}_4} = \{(\mathrm{b}), (\mathrm{d}), (\mathrm{f}), (\mathrm{i})\}$, $\emptyset_{\mathcal{C}_4} = \{(\mathrm{c})\}$, $H_1 = \{(\mathrm{a}), (\mathrm{g})\}$, $H_2 = \{(\mathrm{e}), (\mathrm{h})\}$ e $H_3 = \{(\mathrm{l})\}$.

Lo stato di conoscenza $\mathcal{C}_5$ include l'informazione "il dado viene lanciato e esce il numero 4". Allora ci sono due soli eventi: quello certo $\Omega_{\mathcal{C}_5} = \{(\mathrm{d}), (\mathrm{e}), (\mathrm{f}), (\mathrm{h}), (\mathrm{i})\}$ e quello impossibile $\emptyset_{\mathcal{C}_5} = \{(\mathrm{a}), (\mathrm{b}), (\mathrm{c}), (\mathrm{g}), (\mathrm{l})\}$. $\diamondsuit$

1.6 Operazioni con eventi

Al fine d'introdurre per gli eventi delle operazioni desunte dalla negazione, congiunzione e disgiunzione (multiple o no), *supponiamo che l'insieme di enunciati $\mathcal{L}$ sia tale che*:

- $\neg p, p \wedge q, p \vee q, p \to q, p \leftrightarrow q \in \mathcal{L}$, per ogni $p, q \in \mathcal{L}$;
- $\bigwedge_{i\in I} p_i, \bigvee_{i\in I} p_i \in \mathcal{L}$, per ogni famiglia $(p_i)_{i\in I}$ in $\mathcal{L}$.

Conseguentemente, preso $p \in \mathcal{L}$, risulta $\neg p \vee p, \neg p \wedge p \in \mathcal{L}$ e quindi, per le leggi del terzo escluso e di non contraddizione, l'evento certo $\Omega = [\,\neg p \vee p\,]_{\mathcal{C}}$ e quello impossibile $\emptyset = [\,\neg p \wedge p\,]_{\mathcal{C}}$ sono eventi di $\mathcal{L}$.

Dati due eventi $E = [p]_{\mathcal{C}}$, e $F = [q]_{\mathcal{C}}$, chiameremo:
- **negazione di** E, l'evento $\overline{E} = [\neg p\,]_{\mathcal{C}}$;
- **congiunzione di** E e F, l'evento $E \wedge F = [\,p \wedge q\,]_{\mathcal{C}}$;
- **disgiunzione di** E e F, l'evento $E \vee F = [\,p \vee q\,]_{\mathcal{C}}$.

Inoltre, data una famiglia di eventi $(E_i = [\,p_i\,]_{\mathcal{C}})_{i\in I}$, chiameremo:
- **congiunzione multipla** della famiglia, l'evento $\bigwedge_{i\in I} E_i = [\,\bigwedge_{i\in I} p_i\,]_{\mathcal{C}}$;
- **disgiunzione multipla** della famiglia, l'evento $\bigvee_{i\in I} E_i = [\,\bigvee_{i\in I} p_i\,]_{\mathcal{C}}$.

Che le definizioni date siano ben fondate è conseguenza della proposizione seguente[6] che assicura la compatibilità dell'equivalenza $\sim_{\mathcal{C}}$ sia con la nega-

[6]La cui facile dimostrazione viene lasciata allo studente.

zione che con la congiunzione e disgiunzione (multiple o no). Precisiamo che, al fine di snellire l'esposizione, *useremo le abbreviazioni metalinguistiche: "$\Rightarrow$"per "se ...,* allora" *e "$\Leftrightarrow$"per* "se e solo se"*; inoltre, assumeremo implicitamente che le famiglie di enunciati che considereremo siano in $\mathcal{L}$.*

Proposizione 1.6.1 *Sussistono le proposizioni seguenti:*

(i) $p \sim_{\mathcal{C}} p' \Rightarrow \neg p \sim_{\mathcal{C}} \neg p'$;

(ii) $p \sim_{\mathcal{C}} p'$ *e* $q \sim_{\mathcal{C}} q' \Rightarrow p \wedge q \sim_{\mathcal{C}} p' \wedge q'$ *e* $p \vee q \sim_{\mathcal{C}} p' \vee q'$;

(iii) $p_i \sim_{\mathcal{C}} p'_i$ *per ogni* $i \in I \Rightarrow \bigwedge_{i\in I} p_i \sim_{\mathcal{C}} \bigwedge_{i\in I} p'_i$ *e* $\bigvee_{i\in I} p_i \sim_{\mathcal{C}} \bigvee_{i\in I} p'_i$.

Osserviamo che, per definizione, $\bigwedge_{i\in\emptyset} E_i = \Omega$, $\bigvee_{i\in\emptyset} E_i = \emptyset$ e $\bigwedge_{i\in\{1\}} E_i = E_1 = \bigvee_{i\in\{1\}} E_i$; risulta inoltre, come facilmente si verifica, $\bigwedge_{i\in\{1,,2\}} E_i = E_1 \wedge E_2$ e $\bigvee_{i\in\{1,,2\}} E_i = E_1 \vee E_2$.

Il teorema seguente assicura, in particolare, che l'insieme degli eventi - con l'operazione unaria di negazione e quelle binarie di congiunzione e disgiunzione - è un'algebra di Boole.

Teorema 1.6.2 *Sussistono le seguenti proprietà algebriche per la negazione e le operazioni di congiunzione e disgiunzione (multiple o no). Precisiamo che nella prima proposizione π denota una qualsiasi permutazione dell'insieme degli indici I.*

(i) $\bigwedge_{i\in I} E_i = \bigwedge_{i\in i} E_{\pi(i)}, \quad \bigvee_{i\in I} E_i = \bigvee_{i\in i} E_{\pi(i)}$ *(proprietà commutativa);*

(ii) $(E\wedge F)\wedge G = E\wedge(F\wedge G), \quad (E\vee F)\vee G = E\vee(F\vee G)$ *(p. associativa);*

(iii) $E\wedge\bigvee_{i\in I} E_i = \bigvee_{i\in I}(E\wedge E_i), \; E\vee\bigwedge_{i\in I} E_i = \bigwedge_{i\in I}(E\vee E_i)$ *(p. distributiva);*

(iv) $\bigwedge_{i\in I} E_i = E = \bigvee_{i\in I} E_i$, *se* $E_i = E$ *per ogni* $i \in I$ *(p. di idempotenza);*

(v) $E = E \wedge (E \vee F), \quad E = E \vee (E \wedge F)$ *(p. di assorbimento);*

(vi) $\overline{\bigwedge_{i\in I} E_i} = \bigvee_{i\in i} \overline{E_i}, \quad \overline{\bigvee_{i\in I} E_i} = \bigwedge_{i\in i} \overline{E_i}$ *(leggi di De Morgan);*

(vii) $E \wedge \overline{E} = \emptyset, \quad E \vee \overline{E} = \Omega$;

(viii) $\overline{(\overline{E})} = E, \quad \overline{\Omega} = \emptyset, \quad \overline{\emptyset} = \Omega$;

(ix) $E \wedge \Omega = E = E \vee \emptyset$;

(x) $\bigwedge_{i\in I}(E_{i0} \vee E_{i1}) = \bigvee_{f\in\{0,1\}^I} \bigwedge_{i\in I} E_{if(i)}$;

(xi) $\bigwedge_{i\in I_1\cup I_2} E_i = (\bigwedge_{i\in I_1} E_i)\wedge(\bigwedge_{i\in I_2} E_i)$, $\bigvee_{i\in I_1\cup I_2} E_i = (\bigvee_{i\in I_1} E_i)\vee(\bigvee_{i\in I_2} E_i)$;

(xii) $\bigwedge_{i\in I} E_i = \bigwedge_{i\in\{i\in I: E_i\neq\Omega\}} E_i$, $\quad \bigvee_{i\in I} E_i = \bigvee_{i\in\{i\in I: E_i\neq\emptyset\}} E_i$;

(xiii) $\bigwedge_{i\in I}(E_i \wedge F_i) = (\bigwedge_{i\in I} E_i) \wedge (\bigwedge_{i\in I} F_i)$,
$\bigvee_{i\in I}(E_i \vee F_i) = (\bigvee_{i\in I} E_i) \vee (\bigvee_{i\in I} F_i)$;

(xiv) $E \wedge \bigwedge_{i\in I} F_i = \bigwedge_{i\in I}(E \wedge F_i)$, $\quad E \vee \bigvee_{i\in I} F_i = \bigvee_{i\in I}(E \vee F_i)$;

(xv) $\bigwedge_{i\in I} E_i = \emptyset$, *se* $E_i = \emptyset$ *per qualche* $i \in I$;

(xvi) $\bigvee_{i\in I} E_i = \Omega$, *se* $E_i = \Omega$ *per qualche* $i \in I$.

Dimostrazione Le proposizioni (i)÷(vi) seguono banalmente dalle leggi logiche omonime; la (vii) dalle (8), (9) mentre le (x), (xi) e (xiii) dalle (11), (12) e (13), rispettivamente.

(viii) La prima uguaglianza si ottiene dalla legge della doppia negazione. Per quanto riguarda la seconda, da quanto appena provato e da (vii), (vi), si ha $\overline{\Omega} = \overline{\Omega \vee \overline{\Omega}} = \overline{\Omega} \wedge \overline{\overline{\Omega}} = \overline{\Omega} \wedge \Omega = \emptyset$. Ne segue, passando alla terza uguaglianza, $\Omega = \overline{(\overline{\Omega})} = \overline{\emptyset}$.

(ix) Da (vii), (v) otteniamo $E \wedge \Omega = E \wedge (E \vee \overline{E}) = E$. Ne segue, tramite (viii), (vi), $E = \overline{\overline{E}} = \overline{\overline{E} \wedge \Omega} = \overline{\overline{E}} \vee \overline{\Omega} = E \vee \emptyset$.

(xii) Posto $I_1 = \{i \in I : E_i \neq \emptyset\}$ e $I_2 = \{i \in I : E_i \neq \Omega\}$, da (xi), (iv), (ix) risulta $\bigvee_{i\in I} E_i = \bigvee_{i\in I_1\cup I_1^c} E_i = \bigvee_{i\in I_1} E_i \vee \bigvee_{i\in I_1^c} E_i = \bigvee_{i\in I_1} E_i \vee \emptyset = \bigvee_{i\in I_1} E_i$ e $\bigwedge_{i\in I} E_i = \bigwedge_{i\in I_2\cup I_2^c} E_i = \bigwedge_{i\in I_2} E_i \wedge \bigwedge_{i\in I_2^c} E_i = \bigwedge_{i\in I_2} E_i \wedge \Omega = \bigwedge_{i\in I_1} E_i$.

(xiv) Posto $E_i = E$ per ogni $i \in I$, da (iv) otteniamo $E = \bigwedge_{i\in I} E_i = \bigvee_{i\in I} E_i$ e quindi, per (xiii), la tesi.

(xv) Sia $E_{i_0} = \emptyset$. Posto $I_2 = \{i_0\}$, $I_1 = I \setminus \{i_0\}$ e $E = \bigwedge_{i\in I_1} E_i$, da (xi), (vii), (ii) e (iv) otteniamo $\bigwedge_{i\in I} E_i = \bigwedge_{i\in I_1\cup I_2} E_i = \bigwedge_{i\in I_1} E_i \wedge \bigwedge_{i\in I_2} E_i = E \wedge E_{i_0} = E \wedge (E \wedge \overline{E}) = (E \wedge E) \wedge \overline{E} = E \wedge \overline{E} = \emptyset$.

(xvi) Da (viii), (xv) risulta $\bigwedge_{i\in I} \overline{E}_i = \emptyset$ da cui, per (viii), (vi), $\Omega = \overline{\bigwedge_{i\in I} \overline{E}_i} = \bigvee_{i\in I} \overline{\overline{E}}_i = \bigvee_{i\in I} E_i$. □

1.7 Implicazione tra eventi

Introduciamo ora la relazione di implicazione tra gli eventi che esprime, da un punto di vista interpretativo, la possibilità di trasferire la verità da un evento ad un'altro, nel senso che se l'evento E implica l'evento F, allora la verità di E forza la verità di F.

Dati gli eventi $E = [p]_{\mathcal{C}}$ e $F = [q]_{\mathcal{C}}$, diremo che E **implica** F (in simboli $E \to F$) se $\mathcal{C} \models p \to q$. Osserviamo che la definizione è ben fondata, in quanto, dati gli enunciati p', q' tali che $p' \sim_{\mathcal{C}} p$ e $q' \sim_{\mathcal{C}} q$, si ha $\mathcal{C} \models p \to q \Leftrightarrow \mathcal{C} \models p' \to q'$.

Esempio 1.7.1 Con riferimento all'Esempio 1.5.2, nello stato di conoscenza $\mathcal{C}_1$ risulta $E_1 \to E_2$, $E_1 \to E_3$ e $E_1 \to E_4$; nello stato di conoscenza $\mathcal{C}_2$ riesce $F_1 \to F_2$; nello stato di conoscenza $\mathcal{C}_4$ si ha $H_1 \to H_2$. Inoltre, nello stato di conoscenza $\mathcal{C}_1$, nessuno dei due eventi E_2, E_3 implica l'altro. $\Diamond$.

Il prossimo teorema assicura, tra l'altro, che la relazione d'implicazione è una relazione d'ordine (in generale parziale) nell'insieme degli eventi avente l'evento impossibile come elemento minimo, l'evento certo come elemento massimo e che, data una famiglia arbitraria di eventi, la relativa congiunzione multipla è il più "grande" evento che implica ogni evento della famiglia mentre la relativa disgiunzione multipla è il più "piccolo" evento che è implicato da ogni evento della famiglia.

Teorema 1.7.2 *Sussistono le proposizioni seguenti:*

(i) $E \to F \Leftrightarrow \overline{E} \vee F = \Omega \Leftrightarrow E = E \wedge F \Leftrightarrow F = E \vee F \Leftrightarrow \overline{F} \to \overline{E}$;

(ii) $\to$ *è un ordinamento nell'insieme degli eventi avente* $\emptyset$ *come elemento minimo e* Ω *come elemento massimo;*

(iii) $\bigwedge_{i\in I} E_i \to \bigwedge_{i\in I} F_i$ *e* $\bigvee_{i\in I} E_i \to \bigvee_{i\in I} F_i$, *se* $E_i \to F_i$ *per ogni* $i \in I$;

(iv) $E \to \bigwedge_{i\in I} E_i$ *se* $E \to E_i$ *per ogni* $i \in I$;

(v) $\bigvee_{i\in I} E_i \to E$ *se* $E_i \to E$ *per ogni* $i \in I$;

(vi) $\bigwedge_{i\in I} E_i \to E_{i_0} \to \bigvee_{i\in I} E_i$ *per ogni* $i_0 \in I$.

DIMOSTRAZIONE (i) La dimostrazione si basa sui passi seguenti.

$\diamond\, E \to F \Leftrightarrow \overline{E} \vee F = \Omega$. Posto $E = [p]_{\mathcal{C}}$ e $F = [q]_{\mathcal{C}}$, dalla legge di Filone di Megara otteniamo $[p \to q]_{\mathcal{C}} = [\neg p \vee q]_{\mathcal{C}} = [\neg p]_{\mathcal{C}} \vee [q]_{\mathcal{C}} = \overline{E} \vee F$ e quindi $E \to F \Leftrightarrow \mathcal{C} \models p \to q \Leftrightarrow \Omega = [p \to q]_{\mathcal{C}} \Leftrightarrow \Omega = \overline{E} \vee F$.

$\diamond E \to F \Leftrightarrow \overline{F} \to \overline{E}$ discende da quanto appena provato e dalla $\overline{E} \vee F = \overline{(\overline{F})} \vee \overline{E}$ (conseguenza della proprietà commutativa e del Teorema 1.6.2(viii)).

$\diamond\, \overline{E} \vee F = \Omega \Rightarrow E = E \wedge F$. Supposto $\overline{E} \vee F = \Omega$, tramite la proprietà distributiva e il Teorema 1.6.2(ix),(vii), otteniamo $E = E \wedge \Omega = E \wedge (\overline{E} \vee F) = (E \wedge \overline{E}) \vee (E \wedge F) = \emptyset \vee (E \wedge F) = E \wedge F$.

$\diamond\, E = E \wedge F \Rightarrow F = E \vee F$. Supposto $E = E \wedge F$, per le proprietà commutativa e di assorbimento, si ha $E \vee F = (E \wedge F) \vee F = F \vee (F \wedge E) = F$.

$\diamond\, F = E \vee F \Rightarrow \overline{E} \vee F = \Omega$. Supposto $F = E \vee F$, per la proprietà associativa e il Teorema 1.6.2(vii),(xvi), risulta $\overline{E} \vee F = \overline{E} \vee (E \vee F) = (\overline{E} \vee E) \vee F = \Omega \vee F = \Omega$.

(ii) Dalla proprietà di idempotenza otteniamo $E \wedge E = E$ da cui, tramite (i), risulta $E \to E$. Poi, dalle $E \to F$ e $F \to E$ segue, per (i), $E = E \wedge F$, $F = F \wedge E = E \wedge F$ e quindi $E = F$. Infine, dalle $E \to F$ e $F \to G$ segue, per (i), $E = E \wedge F$, $F = F \wedge G$ da cui $E = E \wedge (F \wedge G) = (E \wedge F) \wedge G = E \wedge G$ e quindi, per (i), $E \to G$. Dunque $\to$ è un ordinamento. Che $\emptyset, \Omega$ siano, rispettivamente, il minimo e il massimo deriva, tramite (i), dalle $\emptyset = \emptyset \wedge E$ e $E = E \wedge \Omega$ (Teorema 1.6.2(xv),(ix)).

(iii) Da (i) otteniamo $E_i = E_i \wedge F_i$ e $F_i = E_i \vee F_i$ per ogni $i \in I$. Allora, per il Teorema 1.6.2(xiii), $(\bigwedge_{i \in I} E_i) \wedge (\bigwedge_{i \in I} F_i) = \bigwedge_{i \in I} (E_i \wedge F_i) = \bigwedge_{i \in I} E_i$ e $(\bigvee_{i \in I} E_i) \vee (\bigvee_{i \in I} F_i) = \bigvee_{i \in I} (E_i \vee F_i) = \bigvee_{i \in I} F_i$ e quindi, per (i), la tesi.

(iv) + (v) Seguono da (iii) e dalla proprietà di idempotenza, una volta considerata la famiglia costante $(G_i)_{i \in I}$ di valore E.

(vi) Posto $G_{i_0} = E_{i_0}$ e $G_i = \Omega$ per ogni $i \in I \setminus \{i_0\}$, tramite (ii), (iii) e il Teorema 1.6.2(xii), otteniamo $\bigwedge_{i \in I} E_i \to \bigwedge_{i \in I} G_i = \bigwedge_{i \in \{i_0\}} G_i = G_{i_0} = E_{i_0}$. Analogamente, posto $H_{i_0} = E_{i_0}$ e $H_i = \emptyset$ per ogni $i \in I \setminus \{i_0\}$, risulta $E_{i_0} = H_{i_0} = \bigvee_{i \in \{i_0\}} H_{i_0} = \bigvee_{i \in I} H_i \to \bigvee_{i \in I} E_i$. □

1.8 Dipendenza logica da una partizione

.

Introduciamo ora le partizioni dell'evento certo che sono lo strumento usuale con cui descrivere l'incertezza connessa con una data situazione aleatoria, in quanto forniscono una descrizione dell'incertezza "essenziale" tramite eventi a due a due incompatibili di cui uno solo è quello vero (che risulta sconosciuto per mancanza d'informazione). A tal fine, ricordiamo che due eventi sono detti **incompatibili** se la loro congiunzione è l'evento impossibile.

Una famiglia di eventi $(E_i)_{i \in I}$ viene detta **esaustiva** se $\bigvee_{i \in I} E_i = \Omega$; viene chiamata una **partizione** (**dell'evento certo**) se è esaustiva e i suoi elementi sono a due a due incompatibili (cioè $E_i \wedge E_j = \emptyset$ per ogni $i \neq j$). Nel seguito, data una partizione $\mathcal{P}$ dell'evento certo, chiameremo **costituenti** di $\mathcal{P}$ i suoi elementi mentre **casi elementari** di $\mathcal{P}$ quei costituenti che sono diversi dall'evento impossibile; inoltre ω (con o senza apici o pedici) denoterà sempre il generico caso elementare della partizione e $\mathcal{P}^{(ce)} = \mathcal{P} \setminus \{\emptyset_{\mathcal{C}}\}$ la famiglia dei casi elementari.

Esempio 1.8.1 (i) La più "piccola"partizione dell'evento certo è la famiglia costituita solamente dall'evento certo. Dato un evento possibile, si può considerare la partizione costituita dall'evento stesso e dalla sua negazione. Se $\mathcal{P}$ è una partizione, allora è una partizione, per il Teorema 1.6.2(xii), anche la famiglia $\mathcal{P}^{(ce)}$.

(ii) Con riferimento al lancio di un dado con sei facce, fissiamo $m \in \{1, \ldots, 6\}$ e consideriamo gli eventi:[7]

$$\begin{aligned} E_{<m} &: \text{ nel lancio è uscito un numero minore di } m \\ E_{=m} &: \text{ nel lancio è uscito il numero } m \\ E_{>m} &: \text{ nel lancio è uscito un numero maggiore di } m \\ E_i &: \text{ nel lancio è uscito il numero } i \quad (1 \leq i \leq 6). \end{aligned}$$

Allora, sono partizioni dell'evento certo le famiglie $(E_{<m}, E_{=m}, E_{>m})$ e $(E_i)_{i=1,\ldots,6}$.

(iii) Con riferimento al gioco del lotto, consideriamo - relativamente alle estrazioni su una data ruota - gli eventi:

$$\begin{aligned} E_{(n_1,n_2,n_3,n_4,n_5)} &: \text{ primo estratto } n_1, \ldots, \text{ quinto estratto } n_5 \\ E_{\{n_1,n_2,n_3,n_4,n_5\}} &: \text{ i numeri estratti sono } n_1, \ldots, n_5. \end{aligned}$$

Allora, le famiglie:

$$\begin{aligned} \mathcal{P}_{sec} &= \{E_{(n_1,n_2,n_3,n_4,n_5)} : 1 \leq n_i \leq 90\,(i = 1, \ldots, 5)\,\mathrm{e}\, n_i \neq n_j (i \neq j)\} \\ \mathcal{P}_{sem} &= \{E_{\{n_1,n_2,n_3,n_4,n_5\}} : 1 \leq n_i \leq 90\,(i = 1, \ldots, 5)\,\mathrm{e}\, n_i \neq n_j (i \neq j)\} \end{aligned}$$

sono partizioni dell'evento certo costituite, rispettivamente da $\binom{90}{5}\, 5! = \frac{90!}{85!}$ e da $\binom{90}{5}$ casi elementari[8].

(iv) Con riferimento ad una partita di calcio tra due squadre A e B, consideriamo le famiglie di eventi:

$$E_{vi} : A \text{ vince}, \quad E_{pe} : A \text{ perde}, \quad E_{pa} : A \text{ pareggia}$$

[7]Qui e nel seguito, per snellire l'esposizione, useremo, dati un evento E e un enunciato p, la notazione "$E : p$" per indicare che p è una descrizione di E nello stato di conoscenza considerato (usualmente sottinteso).

[8]Adottando l'usuale terminologia del gioco del lotto, la prima partizione descrive l'incertezza tramite le *cinquine secche*, la seconda invece mediante le *cinquine semplici*.

$$E_{ij} : A \text{ segna } i \text{ reti e } B \text{ segna } j \text{ reti} \quad (i, j \geq 0).$$

Se si assume che la partita si svolga regolarmente e non venga annullata (per impraticabilità del campo, per invasione, ...), sono partizioni dell'evento certo le famiglie (E_{vi}, E_{pe}, E_{pa}) e $(E_{ij})_{i,j\geq 0}$. $\diamondsuit$

Dati un evento E e un caso elementare ω della partizione $\mathcal{P}$, otteniamo

$$\emptyset \neq \omega = \omega \wedge (E \vee \overline{E}) = (\omega \wedge E) \vee (\omega \wedge \overline{E})$$

e quindi, non potendo gli eventi $\omega \wedge E$, $\omega \wedge \overline{E}$ essere entrambi impossibili (Teorema 1.6.2(ix)), si può verificare solamente uno dei tre casi seguenti:
- $\omega = \omega \wedge E$, cioè $\omega \to E$ (Teorema 1.7.2(i)), quando $\omega \wedge \overline{E} = \emptyset$;
- $\omega = \omega \wedge \overline{E}$, cioè $\omega \to \overline{E}$, quando $\omega \wedge E = \emptyset$;
- $\omega \wedge E \neq \emptyset$ e $\omega \wedge \overline{E} \neq \emptyset$.

Conseguentemente, data una partizione, viene naturale considerare la seguente classificazione degli eventi, relativamente alla possibilità di conoscere la loro estensione una volta risolta l'incertezza descritta dalla partizione (cioè, una volta individuato il caso elementare vero).

Dati un evento E e una partizione dell'evento certo $\mathcal{P}$, diremo che E è:
- **logicamente dipendente** da $\mathcal{P}$ se ogni caso elementare di $\mathcal{P}$ implica E oppure $\overline{E}$;
- **logicamente semidipendente** da $\mathcal{P}$ se esiste un caso elementare ω' di $\mathcal{P}$ che implica E oppure $\overline{E}$ ed esiste un caso elementare ω'' di $\mathcal{P}$ tale che $\omega'' \wedge E \neq \emptyset$ e $\omega'' \wedge \overline{E} \neq \emptyset$;
- **logicamente indipendente** da $\mathcal{P}$ se, per ogni caso elementare ω di $\mathcal{P}$, risulta $\omega \wedge E \neq \emptyset$ e $\omega \wedge \overline{E} \neq \emptyset$.[9]

Esempio 1.8.2 (i) Con riferimento alle estrazioni del gioco del lotto su una data ruota (Esempio 1.8.1(iii)), consideriamo gli eventi:

A : il primo numero estratto ha una sola cifra
B : esce l'ambo 20, 50
C : i numeri sono estratti in modo crescente.

[9] La risoluzione dell'incertezza descritta da $\mathcal{P}$ consente di risolvere l'incertezza anche per l'evento solo se E è logicamente dipendente da $\mathcal{P}$ mentre, nel caso che E sia logicamente semidipendente, solo qualora risulti vero un caso elementare del tipo ω'; infine, nel caso di logica indipendenza, nulla si può dire dell'estensione di E.

Allora, i tre eventi sono logicamente dipendenti dalla partizione $\mathcal{P}_{sec}$. Per quanto riguarda invece la partizione $\mathcal{P}_{sem}$, risulta che A è logicamente semidipendente, che B è logicamente dipendente e che C è logicamente indipendente.

(ii) Con riferimento ad una partita di calcio tra due squadre A e B (Esempio 1.8.1(iv)), consideriamo gli eventi:

$$\begin{array}{lcl} E & : & A \text{ segna qualche rete} \\ F & : & A \text{ e } B \text{ segnano complessivamente 4 reti.} \\ G & : & A \text{ e } B \text{ segnano complessivamente 5 reti.} \end{array}$$

Allora, i tre eventi sono logicamente dipendenti dalla partizione $(E_{ij})_{i,j\geq 0}$ mentre E, G sono logicamente semidipendenti e F è logicamente indipendente dalla partizione (E_{vi}, E_{pe}, E_{pa}). $\diamondsuit$

Il prossimo teorema fornisce una fondamentale caratterizzazione della logica dipendenza che consente, come vedremo, di gettare un ponte tra la logica degli eventi e quella degli insiemi. Al fine di snellire l'esposizione, indichiamo con $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ l'insieme degli eventi logicamente dipendenti dalla partizione $\mathcal{P}$ e, dato un evento E, con $\bigvee_{\omega\to E}\omega$ la disgiunzione multipla (della famiglia) dei casi elementari di $\mathcal{P}$ che implicano E.

Teorema 1.8.3 (di caratterizzazione) *Sussistono le proposizioni seguenti:*

(i) $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P}) \Rightarrow E = \bigvee_{\omega\to E}\omega$*;*

(ii) Le disgiunzioni multiple di casi elementari di $\mathcal{P}$ *appartengono a* $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$*.*

Dimostrazione[10] (i) Considerato un evento E logicamente dipendente da $\mathcal{P}$, poniamo $F = \bigvee_{\omega\to E}\omega$ e $G = \bigvee_{\omega\to \overline{E}}\omega$. Allora

$$E \wedge F = \bigvee_{\omega\to E}(E\wedge\omega) = \bigvee_{\omega\to E}\omega = F.$$

Inoltre, osservato che, per ogni $\omega \to \overline{E}$, si ha $\omega = \omega \wedge \overline{E}$ e quindi $E \wedge \omega = \omega \wedge E = (\omega \wedge \overline{E}) \wedge E = \omega \wedge (\overline{E} \wedge E) = \omega \wedge \emptyset = \emptyset$, otteniamo

$$E \wedge G = \bigvee_{\omega\to \overline{E}}(E\wedge\omega) = \emptyset.$$

[10]Non verranno richiamate esplicitamente le proposizioni (i)÷(ix), (xv), (xvi) del Teorema 1.6.2 come pure la proposizione (i) del Teorema 1.7.2

Infine, per il Teorema 1.6.2(xi) e per la logica dipendenza di E da $\mathcal{P}$, risulta

$$F \vee G = \bigvee_{\omega \in \{\omega : \omega \to E\} \cup \{\omega : \omega \to \overline{E}\}} \omega = \bigvee_{\omega \in \mathcal{P}} \omega = \Omega.$$

Ne segue $E = E \wedge \Omega = E \wedge (F \vee G) = (E \wedge F) \vee (E \wedge G) = F \vee \emptyset = F$.

(ii) Dato $I \subset \mathcal{P}^{(ce)}$, poniamo $E = \bigvee_{\omega \in I} \omega$. Considerato $\omega_0 \in \mathcal{P}$, sia intanto $\omega_0 \in I$. Allora, per il Teorema 1.6.2(xi), risulta

$$\begin{aligned} \omega_0 \wedge E &= \bigvee_{\omega \in I} (\omega_0 \wedge \omega) = (\bigvee_{\omega \in \{\omega_0\}} (\omega_0 \wedge \omega)) \vee (\bigvee_{\omega \in I \setminus \{\omega_0\}} (\omega_0 \wedge \omega)) \\ &= \bigvee_{\omega \in \{\omega_0\}} (\omega_0 \wedge \omega) = \omega_0 \wedge \omega_0 = \omega_0 \end{aligned}$$

e quindi $\omega_0 \to E$. Sia ora $\omega_0 \notin I$. Allora, $\omega_0 \wedge E = \bigvee_{\omega \in I} (\omega_0 \wedge \omega) = \emptyset$ e quindi $\omega_0 = \omega_0 \wedge (E \vee \overline{E}) = (\omega_0 \wedge E) \vee (\omega_0 \wedge \overline{E}) = \omega_0 \wedge \overline{E}$, cioè $\omega_0 \to \overline{E}$. Ne segue, data l'arbitrarietà di ω_0, che l'evento E è logicamente dipendente da $\mathcal{P}$, cioè $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. □

Tramite il teorema di caratterizzazione è ora facile verificare che l'insieme degli eventi logicamente dipendenti da una partizione è chiuso per negazione e per congiunzioni e disgiunzioni (multiple o no).

Teorema 1.8.4 *Siano E e E_i $(i \in I)$ eventi logicamente dipendenti da $\mathcal{P}$. Allora, lo sono anche gli eventi $\overline{E}$, $\bigwedge_{i \in I} E_i$ e $\bigvee_{i \in I} E_i$.*

Dimostrazione Che $\overline{E} \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ segue dalla definizione di dipendenza logica; infatti, ogni caso elementare implica $E = \overline{(\overline{E})}$ o $\overline{E}$. Dalla logica dipendenza di E_i per ogni i, otteniamo, per il Teorema 1.8.3(i), $\bigvee_{i \in I} E_i = \bigvee_{i \in I} (\bigvee_{\omega \to E_i} \omega) = \bigvee_{\omega \in \{\omega : \exists i \in I (\omega \to E_i)\}} \omega$ e quindi, tramite il Teorema 1.8.3(ii), $\bigvee_{i \in I} E_i \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. Per quanto riguarda la congiunzione $\bigwedge_{i \in I} E_i$, basta osservare che $\bigwedge_{i \in I} E_i = \overline{\bigvee_{i \in I} \overline{E}_i}$ e tenere presente quanto appena provato. □

Consideriamo ora l'applicazione set : $\mathcal{E}_{\mathrm{L}}(\mathcal{P}) \to 2^{\mathcal{P}^{(ce)}}$ che associa ad ogni evento E logicamente dipendente da $\mathcal{P}$ l'insieme $\mathrm{set}(E) = \{\omega \in \mathcal{P} : \omega \to E\}$ dei casi elementari che lo implicano. Il teorema seguente assicura che tale applicazione è un isomorfismo booleano tra $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ e l'insieme delle parti di

$\mathcal{P}^{(ce)}$ che muta la relazione di implicazione (tra eventi) in quella di inclusione (tra insiemi) [11].

Teorema 1.8.5 (di rappresentazione) *L'applicazione* set *è un'applicazione biunivoca. Inoltre, se E, F e E_i ($i \in I$) sono elementi di $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$, risulta:*

(i) $\mathrm{set}(\bigvee_{\omega \in I} \omega) = I$, *se* $I \subset \mathcal{P}^{(ce)}$;

(ii) $\mathrm{set}(\Omega_{\mathcal{C}}) = \mathcal{P}$ *e* $\mathrm{set}(\emptyset_{\mathcal{C}}) = \emptyset$;

(iii) $\mathrm{set}(\bigwedge_{i \in I} E_i) = \bigcap_{i \in I} \mathrm{set}(E_i)$;

(iv) $\mathrm{set}(\bigvee_{i \in I} E_i) = \bigcup_{i \in I} \mathrm{set}(E_i)$;

(v) $\mathrm{set}(\overline{E}) = \mathrm{set}(E)^c$;

(vi) $\mathrm{set}(E) \subset \mathrm{set}(F) \Leftrightarrow E \to F$.

Dimostrazione Poichè la suriettività è conseguenza immediata di (i), proviamo l'iniettività. Da $\mathrm{set}(E) = \mathrm{set}(F)$ segue, per il Teorema 1.8.3(i), $E = \bigvee_{\omega \in \mathrm{set}(E)} \omega = \bigvee_{\omega \in \mathrm{set}(F)} \omega = F$.

(i) Sia $I \subset \mathcal{P}^{(ce)}$ e $E = \bigvee_{\omega \in I} \omega$. Allora, per il Teorema 1.8.3(ii), $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ e quindi, sempre per il medesimo teorema, $E = \bigvee_{\omega \in \mathrm{set}(E)} \omega$. Ne segue $\mathrm{set}(E) = I$. Infatti, se $\omega_0 \in I$, allora, per il Teorema 1.7.2(vi), $\omega_0 \to E$ e quindi $\omega_0 \in \mathrm{set}(E)$; viceversa, se, per assurdo, fosse $\omega_0 \in \mathrm{set}(E) \setminus I$, si avrebbe la contraddizione $\emptyset \neq \omega_0 = \omega_0 \wedge E = \omega_0 \wedge \bigvee_{\omega \in I} \omega = \bigvee_{\omega \in I}(\omega_0 \wedge \omega) = \emptyset$.

(ii) Basta ricordare che gli eventi certo e impossibile sono, rispettivamente, l'elemento massimo e quello minimo dell'implicazione (Teorema 1.7.2(ii)).

(iii) Sia intanto $\omega \in \mathrm{set}(\bigwedge_{i \in I} E_i)$. Allora, per il Teorema 1.7.2(vi), $\omega \to E_i$ per ogni $i \in I$ e quindi $\omega \in \mathrm{set}(E_i)$ per ogni $i \in I$, cioè $\omega \in \bigcap_{i \in I} \mathrm{set}(E_i)$. Viceversa, se $\omega \in \bigcap_{i \in I} \mathrm{set}(E_i)$, allora $\omega \to E_i$ per ogni $i \in I$ e quindi, per il Teorema 1.7.2(iv), $\omega \to \bigwedge_{i \in I} E_i$, cioè $\omega \in \mathrm{set}(\bigwedge_{i \in I} E_i)$.

(iv) Sia intanto $\omega_0 \in \mathrm{set}(\bigvee_{i \in I} E_i)$. Allora $\emptyset \neq \omega_0 = \omega_0 \wedge \bigvee_{i \in I} E_i = \bigvee_{i \in I}(\omega_0 \wedge E_i)$ e quindi $\omega_0 \wedge E_{i_0} \neq \emptyset$ per qualche $i_0 \in I$. Pertanto, per il Teorema 1.8.3(i) ($E_{i_0} \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$!), si ha $\emptyset \neq \omega_0 \wedge \bigvee_{\omega \to E_{i_0}} \omega = \bigvee_{\omega \to E_{i_0}} (\omega_0 \wedge \omega)$ e quindi esiste $\omega \to E_{i_0}$ tale che $\omega = \omega_0$, cioè $\omega_0 \in \mathrm{set}(E_{i_0}) \subset \bigcup_{i \in I} \mathrm{set}(E_i)$.

[11] Giustificando così la scelta che viene usualmente fatta, nei testi di calcolo delle probabilità, del linguaggio insiemistico (più noto allo studente) a scapito di quello logico (certamente più riposto ma anche più inerente alla logica dell'incertezza).

Viceversa, supposto $\omega_0 \in \bigcup_{i\in I} \mathrm{set}(E_i)$, sia, ad esempio, $\omega_0 \in \mathrm{set}(E_{i_0})$. Allora, $\omega_0 \to E_{i_0}$ e quindi, per il Teorema 1.7.2(vi), $\omega_0 \to \bigvee_{i\in I} E_i$, cioè $\omega_0 \in \mathrm{set}(\bigvee_{i\in I} E_i)$.

(v) Tramite (iii), (ii) otteniamo $\mathrm{set}(\overline{E}) \cap \mathrm{set}(E) = \mathrm{set}(\overline{E} \wedge E) = \mathrm{set}(\emptyset_{\mathcal{C}}) = \emptyset$ e, tramite (iv), (ii), $\mathrm{set}(\overline{E}) \cup \mathrm{set}(E) = \mathrm{set}(E \vee \overline{E}) = \mathrm{set}(\Omega_{\mathcal{C}}) = \mathcal{P}$. Ne segue la tesi.

(vi) Sia intanto $\mathrm{set}(E) \subset \mathrm{set}(F)$. Allora, posto $G = \bigvee_{\omega \in \mathrm{set}(F)\setminus \mathrm{set}(E)} \omega$, si ha, per il Teorema 1.6.2(xi), $F = E \vee G$ e quindi $E \vee F = F$, cioè $E \to F$ (Teorema1.7.2(i)). L' implicazione rimanente segue banalmente dalla proprietà transitiva dell'implicazione tra eventi. □

1.9 Partizione generata

Dare due partizioni dell'evento certo $\mathcal{P}$ e $\mathcal{P}'$, diremo che $\mathcal{P}$ è **più fine di** $\mathcal{P}'$ se ogni caso elementare di $\mathcal{P}'$ è logicamente dipendente da $\mathcal{P}$. Conseguentemente, la relazione "$\mathcal{P}$ è più fine di $\mathcal{P}'$"esprime, da un punto di vista interpretativo, che la risoluzione dell'incertezza descritta da $\mathcal{P}$ consente di risolvere anche quella descritta da $\mathcal{P}'$; inoltre che $\mathcal{P}$ fornisce, alla luce del teorema di caratterizzazione, un dettaglio descrittivo dell'incertezza maggiore (o al più uguale) di quello di $\mathcal{P}'$.

Il prossimo teorema, al quale premettiamo un lemma, fornisce un criterio utile per verificare se una partizione è più fine di un'altra.

Lemma 1.9.1 *Siano $\mathcal{P}$, $\mathcal{P}'$ due partizioni dell'evento certo. Allora, per ogni caso elementare $\omega \in \mathcal{P}$ esiste un caso elementare $\omega' \in \mathcal{P}'$ tale che $\omega \wedge \omega' \neq \emptyset$.*

Dimostrazione Dato $\omega \in \mathcal{P}$, risulta $\emptyset \neq \omega = \omega \wedge \Omega = \omega \wedge \bigvee_{\omega' \in \mathcal{P}'} \omega' = \bigvee_{\omega' \in \mathcal{P}'} (\omega \wedge \omega')$ e quindi esiste $\omega' \in \mathcal{P}'$ tale che $\omega \wedge \omega' \neq \emptyset$. □

Teorema 1.9.2 *Date le partizioni $\mathcal{P}$ e $\mathcal{P}'$, sono equivalenti le proposizioni:*

(i) $\mathcal{P}$ è più fine di $\mathcal{P}'$;

(ii) Ogni caso elementare di $\mathcal{P}$ implica un (unico) caso elementare di $\mathcal{P}'$.

Dimostrazione (i) $\Rightarrow$ (ii). Dato $\omega \in \mathcal{P}$, esiste, per il Lemma 1.9.1, un caso elementare ω' di $\mathcal{P}'$ tale che $\omega \wedge \omega' \neq \emptyset$. Ora, poichè ω' è logicamente dipendente da $\mathcal{P}$, riesce $\omega \to \omega'$ oppure $\omega \to \overline{\omega'}$. Ma la seconda implicazione non può sussistere, perchè altrimenti si avrebbe $\omega = \omega \wedge \overline{\omega'}$ da cui seguirebbe la contraddizione $\emptyset \neq \omega \wedge \omega' = (\omega \wedge \overline{\omega'}) \wedge \omega' = \omega \wedge (\overline{\omega'} \wedge \omega') = \omega \wedge \emptyset = \emptyset$. Provata l'esistenza verifichiamo l'unicità. A tal fine, siano $\omega \in \mathcal{P}$ e $\omega', \overline{\omega}' \in \mathcal{P}'$ tali che $\omega \to \omega'$ e $\omega \to \overline{\omega}'$. Risulta allora, per il Teorema 1.7.2(i), $\emptyset \neq \omega = \omega \wedge \omega' = (\omega \wedge \overline{\omega}') \wedge \omega' = \omega \wedge (\overline{\omega}' \wedge \omega')$ e quindi $\overline{\omega}' \wedge \omega' \neq \emptyset$, cioè $\overline{\omega}' = \omega'$.

(ii) $\Rightarrow$ (i). Sia ω' un caso elementare di $\mathcal{P}'$. Considerato allora il caso elementare ω di $\mathcal{P}$, esiste un caso elementare ω'_0 di $\mathcal{P}'$ tale che $\omega \to \omega'_0$, cioè $\omega = \omega \wedge \omega'_0$. Quindi, se $\omega' = \omega'_0$, risulta $\omega \to \omega'$; se invece $\omega' \neq \omega'_0$, riesce $\omega = \omega \wedge \Omega = \omega \wedge (\omega' \vee \overline{\omega'}) = (\omega \wedge \omega') \vee (\omega \wedge \overline{\omega'}) = [(\omega \wedge \omega'_0) \wedge \omega'] \vee (\omega \wedge \overline{\omega'}) = [\omega \wedge (\omega'_0 \wedge \omega')] \vee (\omega \wedge \overline{\omega'}) = (\omega \wedge \emptyset) \vee (\omega \wedge \overline{\omega'}) = \emptyset \vee (\omega \wedge \overline{\omega'}) = \omega \wedge \overline{\omega'}$, cioè $\omega \to \overline{\omega'}$. Ne segue, dall'arbitrarietà di ω e ω', la tesi. □

Esempio 1.9.3 Con riferimento al gioco del lotto, consideriamo le partizioni $\mathcal{P}_{sec}$, $\mathcal{P}_{sem}$ (Esempio 1.8.1(iii)) e la partizione $\mathcal{P}_{pes}$ costituita dagli eventi E_n : "n è il primo numero estratto"($n = 1, \dots, 90$). Allora, $\mathcal{P}_{sec}$ è più fine delle rimanenti mentre non esiste alcun legame di finitezza tra $\mathcal{P}_{sem}$ e $\mathcal{P}_{pes}$. ♢

Con riferimento ad una famiglia non vuota di eventi $\mathcal{F} = (E_i)_{i \in I}$, il teorema seguente assicura che la famiglia di eventi $\mathcal{P}_{\mathrm{G}}(\mathcal{F}) = (\bigwedge_{i \in I} E_{if(i)})_{f \in \{0,1\}^I}$ - ove $E_{i0} = \overline{E_i}$ e $E_{i1} = E_i$ per ogni $i \in I$ - è una partizione dell'evento certo, che chiameremo **partizione generata** dalla famiglia $\mathcal{F}$.

Teorema 1.9.4 *Data una famiglia non vuota di eventi $\mathcal{F} = (E_i)_{i \in I}$, la famiglia $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$ è una partizione dell'evento certo.*

Dimostrazione Proviamo intanto che gli elementi di $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$ sono a due a due incompatibili. Sia $\bigwedge_{i \in I} E_{if(i)} \neq \bigwedge_{i \in I} E_{ig(i)}$. Esiste allora $i_0 \in I$ tale che $f(i_0) \neq g(i_0)$ e quindi $E_{i_0 f(i_0)} \wedge E_{i_0 g(i_0)} = \emptyset$. Ne segue, per il Teorema 1.6.2(xiii),(xv), $(\bigwedge_{i \in I} E_{if(i)}) \wedge (\bigwedge_{i \in I} E_{ig(i)}) = \bigwedge_{i \in I} (E_{if(i)} \wedge E_{ig(i)}) = \emptyset$.

Proviamo infine che $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$ è esaustiva. Tramite il Teorema 1.6.2(x),(xvi), otteniamo $\bigvee_{f \in \{0,1\}^I} (\bigwedge_{i \in I} E_{if(i)}) = \bigwedge_{i \in I} (E_{i0} \vee E_{i1}) = \bigwedge_{i \in I} (\overline{E}_i \vee E_i) = \Omega$. □

Nel prossimo esempio consideriamo alcune famiglie di eventi e forniamo i costituenti delle corrispondenti partizioni generate.

Esempio 1.9.5 (i) Con riferimento ad un evento E, i costituenti sono E e $\overline{E}$.

(ii) Dati gli eventi E e F, i costituenti sono $E \wedge F$, $\overline{E} \wedge F$, $E \wedge \overline{F}$ e $\overline{E} \wedge \overline{F}$.

(iii) Dati gli eventi E, F, G tali che $E \vee G = \Omega$ e $F \to E \wedge G$ i costituenti sono $F, E \wedge \overline{F} \wedge G, E \wedge \overline{G}$ e $\overline{E} \wedge G$.

(iv) Con riferimento alle estrazioni del gioco del lotto su una data ruota, consideriamo gli eventi:

A : viene estratto il numero 10
B : viene estratto l'ambo 20, 50
C : il terzo estratto è il numero 10.

Allora, i casi elementari della partizione generata sono:
- $A \wedge B \wedge C$: sono estratti i numeri 10, 20, 50 e 10 è il terzo estratto;
- $A \wedge B \wedge \overline{C}$: sono estratti i numeri 10, 20, 50 e 10 non è il terzo estratto;
- $A \wedge \overline{B} \wedge C$: 10 è il terzo estratto e non sono estratti il numero 20 o il numero 50;
- $A \wedge \overline{B} \wedge \overline{C}$: viene estratto il 10 che non è il terzo estratto e non sono estratti il numero 20 o il numero 50;
- $\overline{A} \wedge B \wedge \overline{C}$: sono estratti i numeri 20, 50 e non il 10;
- $\overline{A} \wedge \overline{B} \wedge \overline{C}$: non viene estratto il numero 10 e non viene estratto il numero 20 o il numero 50. $\lozenge$

Proviamo ora che la partizione generata è la partizione meno fine tra tutte quelle rispetto alle quali ogni elemento della famiglia è logicamente dipendente.

Teorema 1.9.6 *Data una famiglia non vuota di eventi $\mathcal{F} = (E_i)_{i\in I}$, sia $\mathcal{P}$ una partizione dell'evento certo tale che $E_i \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ per ogni $i \in I$. Allora, $\mathcal{P}$ è più fine di $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$.*

Dimostrazione Considerato il caso elementare $\omega \in \mathcal{P}$, esiste, per ogni $i \in I$, $f(i) \in \{0,1\}$ tale che $\omega \to E_{if(i)}$. Allora, per il Teorema 1.7.2(iv), $\omega \to \bigwedge_{i\in I} E_{if(i)}$. Dal Teorema 1.9.2 si ha allora la tesi, una volta osservato che $\bigwedge_{i\in I} E_{if(i)}$ è un caso elementare di $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$. $\square$

Concludiamo introducendo la nozione di eventi logicamente indipendenti. Data una famiglia di eventi $(E_i)_{i\in I}$ con almeno due elementi, diremo che gli eventi E_i sono **logicamente indipendenti** se E_i è logicamente indipendente dalla partizione $\mathcal{P}_{\mathrm{G}}((E_j)_{j\in I\setminus\{i\}})$ generata dai rimanenti eventi della famiglia.

Il prossimo teorema fornisce un criterio utile per verificare se gli eventi sono logicamente indipendenti. Ad esempio, consente di affermare che gli eventi A, B, C (Esempio 1.9.5(iv)) non sono logicamente indipendenti, in quanto il costituente $\overline{A} \wedge B \wedge C$ (come pure $\overline{A} \wedge \overline{B} \wedge C$) è impossibile.

Teorema 1.9.7 *Data la famiglia di eventi* $(E_i)_{i\in I}$ *con* $\#I \geq 2$ [12]*, sono equivalenti le proposizioni:*

(i) Gli eventi E_i $(i \in I)$ *sono logicamente indipendenti;*

(ii) Ogni costituente della partizione generata $\mathcal{P}_{\mathrm{G}}((E_i)_{i\in I})$ *è un caso elementare.*

DIMOSTRAZIONE (i) $\Rightarrow$ (ii) Considerati $i' \in I$ e $E = \bigwedge_{i\in I\setminus\{i'\}} E_{if(i)}$, risulta $E_{i'} \wedge E \neq \emptyset$, $\overline{E}_{i'} \wedge E \neq \emptyset$ e quindi, per il Teorema 1.6.2(xi), $\bigwedge_{i\in I} E_{if(i)} = (\bigwedge_{i\in\{i'\}} E_{if(i)}) \wedge (\bigwedge_{i\in I\setminus\{i'\}} E_{if(i)}) = E_{i'f(i')} \wedge E \neq \emptyset$ ogni valore di $f(i')$.

(ii) $\Rightarrow$ (i) Considerati $i' \in I$ e $E = \bigwedge_{i\in I\setminus\{i'\}} E_{if(i)}$, siano g, $h \in \{0,1\}^I$ tali che $g(i') = 1$, $h(i') = 0$ e $g(i) = f(i) = h(i)$ per ogni $i \neq i'$. Allora, $E_{i'} \wedge E = \bigwedge_{i\in I} E_{ig(i)} \neq \emptyset$, $\overline{E}_{i'} \wedge E = \bigwedge_{i\in I} E_{ih(i)} \neq \emptyset$ e quindi, data l'arbitrarietà del costituente $E \in \mathcal{P}_{\mathrm{G}}((E_i)_{i\in I\setminus\{i'\}})$, l'evento $E_{i'}$ è logicamente indipendente dalla partizione generata dagli altri eventi. □

Proviamo infine che l'indipendenza logica si conserva "in discesa".

Teorema 1.9.8 *Siano gli eventi* E_i $(i \in I)$ *logicamente indipendenti. Allora, dato un qualunque sottoinsieme* $J \subset I$ *con almeno due elementi, gli eventi della famiglia* E_j $(j \in J)$ *sono logicamente indipendenti.*

DIMOSTRAZIONE Sia $E = \bigwedge_{j\in J} E_{jf(j)}$ un costituente della partizione generata dagli eventi E_j $(j \in J)$. Considerata allora l'applicazione $g \in \{0,1\}^I$ tale che $g(i) = f(i)$, se $i \in J$, e $g(i) = 1$, se $i \notin J$, otteniamo, per il Teorema 1.9.7, $\bigwedge_{i\in I} E_{ig(i)} \neq \emptyset$ e quindi $E \neq \emptyset$, osservato che, per i teoremi 1.6.2(xi) e 1.7.2(vi), risulta $\bigwedge_{i\in I} E_{ig(i)} = (\bigwedge_{j\in J} E_{jf(j)}) \wedge (\bigwedge_{i\in I\setminus J} E_i) \to E$. Ne segue, dal Teorema 1.9.7, la tesi. □

1.10 Eventi condizionati

Considerato lo stato di conoscenza $\mathcal{C}$, sia $H = [h]_{\mathcal{C}}$ tale che $H \neq \emptyset$. Allora, nello stato di conoscenza $\mathcal{C}$ *non è deducibile* la falsità dell'enunciato h per cui l'assunzione che sia vero è compatibile con lo stato di conoscenza. L'assunzione (in via ipotetica o effettiva) che h sia un enunciato vero può

[12] Per ogni insieme A, con la notazione $\#A$ denotiamo la cardinalità di A.

quindi essere usata per considerare lo stato di conoscenza (non contraddittorio) $\mathcal{C} \cup \{h\}$, ottenuto a seguito dell'*incremento d'informazione* h. In questo modo, il passaggio dallo stato di conoscenza *iniziale* $\mathcal{C}$ a quello *finale* $\mathcal{C} \cup \{h\}$ permette di formalizzare, nell'ambito degli eventi, l'evoluzione della conoscenza a seguito dell'acquisizione di nuove informazioni (consentendo così un approccio dinamico ai problemi in condizioni d'incertezza)[13].

Le considerazioni appena svolte suggeriscono d'introdurre, per ogni evento $E = [p]_{\mathcal{C}}$, l'evento $E \,|\, H$ così definito:

$$E \,|\, H = [p]_{\mathcal{C}\cup\{h\}} \in \mathcal{E}_{\mathcal{C}\cup\{h\}},$$

che chiameremo **evento condizionato di** E **a** H, cioè l'evento descritto dall'enunciato p nello stato di conoscenza finale $\mathcal{C} \cup \{h\}$. In questo contesto, manterremo per E il nome di **evento** mentre per H useremo quello di **evento ipotesi**[14]. Che la definizione data sia ben fondata[15] deriva facilmente dalla proposizione seguente (di verifica immediata):

$$\mathcal{C} \cup \{h\} \models p \Leftrightarrow \mathcal{C} \models h \to p \tag{1.1}$$

qualunque siano gli enunciati h, p tali che $[h]_{\mathcal{C}} \neq \emptyset$.

Esempio 1.10.1 Con riferimento all'Esempio 1.5.2, siano $\mathcal{C}_1$ lo stato di conoscenza iniziale e $\mathcal{L}$ un qualsiasi insieme di enunciati chiuso per le solite operazioni logiche e includente gli enunciati (a)$\div$(l). Considerato allora come evento ipotesi l'evento $E_2 = [(e)]_{\mathcal{C}_1}$ (corrispondente all'incremento d'informazione: esce un numero pari) otteniamo $\mathcal{C}_1 \cup \{(e)\} = \mathcal{C}_2$ e quindi gli eventi condizionati a E_2 sono $\mathcal{C}_2$-eventi; in particolare, $F_1 = E_1 \,|\, E_2 = E_3 \,|\, E_2$, $\Omega_{\mathcal{C}_2} = E_2 \,|\, E_2 = \Omega_{\mathcal{C}_1} \,|\, E_2$, $F_2 = E_4 \,|\, E_2$ e $\emptyset_{\mathcal{C}_2} = E_5 \,|\, E_2 = \emptyset_{\mathcal{C}_1} \,|\, E_2$.

[13]Poichè nella definizione di evento intervengono sia l'insieme di enunciati $\mathcal{L}$ (che fornisce il *linguaggio* connesso con il problema aleatorio in esame) che lo stato di conoscenza $\mathcal{C}$ (che consente di identificare enunciati estensionalmente equivalenti), gli eventi possono essere modificati o tramite un ampliamento del linguaggio o per mezzo di un arricchimento della conoscenza. L'aspetto che qui consideriamo è dunque il secondo per cui l'incremento d'informazione dovuto all'enunciato h provocherà una modifica dell'intero quadro degli eventi rimanendo invariato il linguaggio sottostante.

[14]La nozione di evento condizionato viene dunque introdotta mediante una coppia di $\mathcal{C}$-eventi aventi però ruoli diversi. Il primo è usato ancora come evento - nel senso che l'enunciato p che lo rappresenta in $\mathcal{C}$ interviene a descrivere anche l'evento condizionato; il secondo invece come incremento d'informazione - nel senso che l'enunciato h che lo rappresenta in $\mathcal{C}$ contribuisce ad aumentare lo stato di conoscenza $\mathcal{C}$.

[15]Cioè che $E \,|\, H$ dipenda solamente dagli eventi E, H e non da come sono descritti.

Considerato ora come evento ipotesi l'evento $H = E_2 \wedge E_4 = [(e) \wedge (f)]_{\mathcal{C}_1}$ (corrispondente all'incremento d'informazione: esce il 2 oppure il 4) risulta $\mathcal{C}_1 \cup \{(e) \wedge (f)\} = \mathcal{C}_3$ e quindi gli eventi condizionati a H sono $\mathcal{C}_3$-eventi; in particolare, $G_1 = E_1 \,|\, H = E_3 \,|\, H$, $\Omega_{\mathcal{C}_3} = E_2 \,|\, H = \Omega_{\mathcal{C}_1} \,|\, H = E_4 \,|\, H$ e $\emptyset_{\mathcal{C}_3} = E_5 \,|\, H = \emptyset_{\mathcal{C}_1} \,|\, H$.

Scelto infine come evento ipotesi l'evento $K = E_2 \wedge \overline{E}_3 \wedge E_4 = [(e) \wedge \neg(b) \wedge (f)]_{\mathcal{C}_1}$ (relativo all'incremento d'informazione: esce il 4) riesce $\mathcal{C}_1 \cup \{(e) \wedge \neg(b) \wedge (f)\} = \mathcal{C}_5$ e quindi gli eventi condizionati a K sono $\mathcal{C}_5$-eventi, cioè l'evento certo $\Omega_{\mathcal{C}_5} = E_2 \,|\, K = E_4 \,|\, K = \Omega_{\mathcal{C}_1} \,|\, K$ e l'evento impossibile $\emptyset_{\mathcal{C}_5} = E_1 \,|\, K = E_3 \,|\, K = E_5 \,|\, K = \emptyset_{\mathcal{C}_1} \,|\, K$.$\diamondsuit$

Nel teorema seguente riportiamo alcune proprietà importanti degli eventi condizionati. La prima evidenzia che gli **eventi assoluti** (cioè quelli relativi allo stato di conoscenza iniziale) si possono considerare come caso particolare di quelli condizionati, scegliendo l'evento certo come evento ipotesi. La seconda assicura che l'uguaglianza di due eventi condizionati alla medesima ipotesi equivale all'uguaglianza delle congiunzioni dei rispettivi eventi con l'evento ipotesi; la terza evidenzia che l'evento ipotesi diviene il nuovo evento certo; la quarta che l'evento condizionato non cambia se sostituiamo il relativo evento con la sua congiunzione con l'evento ipotesi. L'ultima proprietà, chiamata *proprietà iterativa degli eventi condizionati*, mette in evidenza che l'acquisizione "passo per passo" delle informazioni equivale all'acquisizione "in blocco" delle stesse.

Teorema 1.10.2 *Siano E, H, K eventi assoluti tali che $H \neq \emptyset$ e $K \,|\, H \neq \emptyset \,|\, H$. Sussistono allora le proposizioni seguenti:*

(i) $E \,|\, \Omega = E$;

(ii) $E \,|\, H = F \,|\, H \Leftrightarrow E \wedge H = F \wedge H$;

(iii) $H \,|\, H = \Omega \,|\, H$;

(iv) $E \,|\, H = (E \wedge H) \,|\, H$;

(v) $E \,|\, H = \Omega \,|\, H$, *se* $H \to E$;

(vi) $E \,|\, H = \emptyset \,|\, H$, *se* $H \to \overline{E}$;

(vii) $E \,|\, H = \emptyset \,|\, H$, *se* $E \to \overline{H}$;

(viii) $(E \,|\, H) \,|\, (K \,|\, H) = E \,|\, (H \wedge K) = (E \,|\, K) \,|\, (H \,|\, K)$.

DIMOSTRAZIONE Poichè le proposizioni (iii)÷(vii) seguono facilmente da (ii), ci limitiamo a provare le rimanenti. A tal fine sia $E = [p]_{\mathcal{C}}$, $F = [q]_{\mathcal{C}}$, $H = [h]_{\mathcal{C}}$ e $K = [k]_{\mathcal{C}}$.

(i) Sia $\Omega = [v]_{\mathcal{C}}$. Allora $\mathcal{C}$ forza la verità di v e quindi, tramite (1.1), si ha $u \in [p]_{\mathcal{C}\cup\{v\}} \Leftrightarrow \mathcal{C} \cup \{v\} \models u \leftrightarrow p \Leftrightarrow \mathcal{C} \models v \to (u \leftrightarrow p) \Leftrightarrow \mathcal{C} \models u \leftrightarrow p \Leftrightarrow u \in [p]_{\mathcal{C}}$. Ne segue la tesi.

(ii) Sia intanto $E \mid H = F \mid H$. Allora, $\mathcal{C}\cup\{h\} \models p \leftrightarrow q$ e quindi, per (1.1), $\mathcal{C}$ forza la verità di $h \to (p \leftrightarrow q)$ e quindi, come facilmente si verifica, anche quella di $p\wedge h \leftrightarrow q\wedge h$. Conseguentemente, $E\wedge H = [p\wedge h]_{\mathcal{C}} = [q\wedge h]_{\mathcal{C}} = F\wedge H$.

Viceversa, sia $E \wedge H = F \wedge H$. Allora, nello stato di conoscenza $\mathcal{C}$, gli enunciati $p \wedge h$ e $q \wedge h$ hanno la medesima estensione; pertanto, assunto vero l'enunciato h, otteniamo che anche p e q hanno la medesima estensione. Conseguentemente, $\mathcal{C} \cup \{h\} \models p \leftrightarrow q$, cioè $E \mid H = F \mid H$.

(viii) Risulta intanto $H \wedge K \neq \emptyset$; infatti, in caso contrario, si avrebbe $K\wedge H = \emptyset\wedge H$ da cui, tramite (ii), seguirebbe la contraddizione $K \mid H = \emptyset \mid H$. Proviamo l'uguaglianza $(E \mid H) \mid (K \mid H) = E \mid (H \wedge K)$. Risulta $E \mid H = [p]_{\mathcal{C}\cup\{h\}}$, $K \mid H = [k]_{\mathcal{C}\cup\{h\}}$ da cui otteniamo $(E \mid H) \mid (K \mid H) = [p]_{(\mathcal{C}\cup\{h\})\cup\{k\}} = [p]_{\mathcal{C}\cup\{h\wedge k\}}$, osservato che $(\mathcal{C} \cup \{h\}) \cup \{k\}$ e $\mathcal{C} \cup \{h \wedge k\}$ sono lo stesso stato di conoscenza. Risulta allora $(E \mid H) \mid (K \mid H) = E \mid [h\wedge k]_{\mathcal{C}} = E \mid (H\wedge K)$. $\square$

Le definizioni delle operazioni di negazione, di congiunzione e disgiunzione (multiple o no) per gli eventi condizionati ad una medesima ipotesi $H = [h]_{\mathcal{C}}$, sono del tutto analoghe a quelle date per gli eventi assoluti; basta sostituire in esse lo stato di conoscenza iniziale $\mathcal{C}$ con quello finale $\mathcal{C} \cup \{h\}$. Rimangono pertanto valide tutte le proprietà riportate nel Teorema 1.6.2 con l'aggiunta beninteso del condizionamento ad H. Anche per le relazioni tra eventi condizionati (e relative proprietà) il discorso si riconduce a quello relativo agli eventi assoluti; basta precisare che le condizioni che definiscono le relazioni sono fatte nello stato di conoscenza finale e non in quello iniziale.

Ovviamente la medesima relazione comporta un diverso significato se considerata nell'ambito degli eventi assoluti o in quello degli eventi condizionati. Infatti, se gli eventi condizionati $E \mid H$ e $F \mid H$ sono incompatibili, questo significa che $(E \wedge F) \mid H = (E \mid H) \wedge (F \mid H) = \emptyset \mid H$ da cui risulta, per il Teorema 1.10.2(ii), $(E \wedge F) \wedge H = \emptyset \wedge H$ e quindi $(E \wedge H) \wedge (F \wedge H) = \emptyset$, cioè l'incompatibilità condizionata si traduce nell'incompatibilità degli eventi $E \wedge H$ e $F \wedge H$. Analogamente, se $E \mid H \to F \mid H$, allora $(E \wedge F) \mid H = (E \mid H) \wedge (F \mid H) = E \mid H$ e quindi, sempre per il Teorema 1.10.2(ii), l'implicazione condizionata significa che $E \wedge H \to F \wedge H$.

Poichè ogni asserzione fatta nello stato di conoscenza iniziale rimane, a fortiori, valida anche in quello finale, si conservano le relazioni tra gli eventi che sono state introdotte ricorrendo ai rappresentanti degli eventi coinvolti e facendo riferimento allo stato di conoscenza iniziale; pertanto, qualunque sia l'incremento d'informazione, l'implicazione sussistente tra due eventi si conserva, eventi incompatibili rimangono incompatibili e partizioni dell'evento certo rimangono partizioni dell'evento certo (con, eventualmente, meno casi elementari comportando così una riduzione del quadro delle possibilità, cioè dell'incertezza connessa con la descrizione scelta.

1.11 Numeri aleatori

Il numero che esce in un determinato lancio di uno specifico dado è - prima di effettuare il lancio o comunque non conoscendone l'esito - *non noto* e può essere uno qualsiasi dei numeri impressi; tuttavia esso è *ben definito*, poichè sono precisati sia il lancio che il dado considerati. Analogamente, lo è pure il primo numero estratto nella prossima estrazione sulla ruota di Venezia (che potrà essere uno qualsiasi dei primi 90 numeri naturali). In entrambi i casi siamo quindi in presenza di un *numero aleatorio*[16], cioè di un numero X *sconosciuto per mancanza di informazione, ma di valore ben determinato.*

Per passare ad una sua trattazione formale, conviene fare alcune considerazioni. Data la partizione dell'evento certo formata dai costituenti:

$$E_x^{(X)} : X \text{ assume il valore } x \qquad (x \in \mathbb{R}) \tag{1.2}$$

possiamo descrivere X mediante la funzione iniettiva che a ogni caso elementare $E_x^{(X)}$ associa il numero x. Questa però non è l'unica descrizione possibile; infatti, basta scegliere una partizione $\mathcal{P}$ più fine della partizione (1.2) e considerare la funzione che associa ad ogni caso elementare di $\mathcal{P}$ che implica $E_x^{(X)}$ il valore x[17]. Ad esempio, con riferimento al gioco del lotto (esempi 1.8.1(iii) e 1.9.3), nel caso del numero aleatorio "primo numero estratto", possiamo considerare per descriverlo, al posto della partizione $\mathcal{P}_{pes}$ (corrispondente alla partizione (1.2)), la partizione più fine $\mathcal{P}_{sec}$ e la funzione che associa ad ogni caso elementare $E_{(n_1,n_2,n_3,n_4,n_5)}$ il numero n_1 (funzione che, a differenza di quella associata alla $\mathcal{P}_{pes}$, non è iniettiva).

[16] Dal latino "àlea"(gioco di dadi).

[17] Funzione che risulta ben definita in forza del Teorema 1.9.2.

Questa molteplicità delle descrizioni è peraltro essenziale per consentire un trattazione agevole dei numeri aleatori. Per constatarlo, con riferimento al lancio simultaneo dei dadi A e B, siano X il numero che esce con il dado A, Y il numero che esce con il dado B e Z il numero aleatorio $X+Y$. Allora, volendo riferirci a partizioni del tipo (1.2), otteniamo che le associate descrizioni sono, rispettivamente, date dalle funzioni $f_X : E_i^{(X)} \to i$, $f_Y : E_j^{(Y)} \to j$ $(i,j = 1,\ldots,6)$ e $f_Z : E_h^{(Z)} \to h$ $(h = 2,\ldots,12)$. Ora, la descrizione f_Z può essere data *solamente dopo* aver eseguito effettivamente la somma $X+Y$; inoltre, *non indica esplicitamente* che il numero aleatorio Z è la somma dei numeri aleatori X e Y. Per farlo dovrebbe potersi esprimere come somma delle funzioni f_X e f_Y; richiesta che non può essere soddisfatta, poichè tali funzioni *non hanno il medesimo dominio* (condizione fondamentale per sommare due funzioni). Per ovviare a questo inconveniente, ricorriamo allora a una diversa descrizione dei tre numeri aleatori. Consideriamo come partizione di riferimento la partizione formata dai casi elementari $\omega_{ij} = E_i^{(X)} \wedge E_j^{(Y)}$ $(i,j = 1,\ldots,6)$. Ne segue che le descrizioni sono ora $g_X : \omega_{ij} \to i$, $g_Y : \omega_{ij} \to j$ e $g_Z : \omega_{ij} \to i+j$, esprimendo così anche al livello delle descrizioni che Z è somma di X e Y[18].

Non è quindi possibile identificare il numero aleatorio nè con la descrizione "naturale"relativa alla partizione (1.2), nè con la sua descrizione associata ad un'altra qualsiasi partizione più fine della (1.2).

D'altra parte, è evidente che ogni funzione reale di dominio una partizione dell'evento certo può essere scelta per introdurre un numero aleatorio nel senso sopra specificato. Inoltre, due funzioni reali f_1, f_2 - di dominio, rispettivo, le partizioni $\mathcal{P}_1, \mathcal{P}_2$ - definiranno il medesimo numero aleatorio se e solo se risulta $f_1(\omega_1) = f_2(\omega_2)$ per ogni $\omega_1 \in \mathcal{P}_1$ e $\omega_2 \in \mathcal{P}_2$ tali che $\omega_1 \wedge \omega_2 \neq \emptyset$; infatti, solo in questo caso il numero aleatorio descritto da f_1 sarà uguale a quello descritto da f_2, qualunque siano i casi elementari veri.

È proprio a quest'ultima osservazione che si rifà la definizione formale di numero aleatorio. Considerato l'insieme $\mathbb{F}$ delle funzioni di dominio partizioni dell'evento certo:

$$\mathbb{F} = \{f^{(\mathcal{P})} : f^{(\mathcal{P})} \in \mathbb{R}^{\mathcal{P}} \text{ con } \mathcal{P} \text{ partizione dell'evento certo}\},$$

[18]La situazione descritta è analoga a quella relativa alla somma di due numeri razionali. Infatti, per sommare il numero razionale rappresentato dalla frazione $\frac{1}{2}$ con quello rappresentato dalla frazione $\frac{1}{3}$, dobbiamo sostituire le due frazioni, rispettivamente, con le frazioni $\frac{3}{6}$,$\frac{2}{6}$ e poi rappresentare il numero razionale somma con la frazione $\frac{3}{6} + \frac{2}{6} = \frac{5}{6}$.

chiamiamo $f_1^{(\mathcal{P}_1)}, f_2^{(\mathcal{P}_2)} \in \mathbb{F}$ **equivalenti** (in simboli $f_1^{(\mathcal{P}_1)} \sim f_2^{(\mathcal{P}_2)}$) se $f_1^{(\mathcal{P}_1)}(\omega_1) = f_2^{(\mathcal{P}_2)}(\omega_2)$ per ogni $\omega_1 \in \mathcal{P}_1$ e $\omega_2 \in \mathcal{P}_2$ tali che $\omega_1 \wedge \omega_2 \neq \emptyset$. Osservato che tale relazione è una relazione di equivalenza sull'insieme $\mathbb{F}$, chiamiamo **numeri aleatori** le relative classi di equivalenza. Poichè le restrizioni di funzioni equivalenti sui casi elementari delle rispettive partizioni hanno medesimo insieme-imagine[19], chiamiamo **rango** (o insieme dei **valori possibili**) del numero aleatorio $X = [f^{(\mathcal{P})}]_\sim$ l'insieme-imagine $\mathrm{rg}(X)$ della *versione* $f^{(\mathcal{P})}$ considerata solo sui casi elementari dell'associata partizione $\mathcal{P}$. Chiamiamo infine **numeri certi** quei numeri aleatori X aventi rango formato da un solo elemento, detto **valore** di X [20].

Definiti i numeri aleatori come classi di equivalenza, procediamo introducendo per essi le usuali operazioni aritmetiche. A tal fine, dato il numero aleatorio X, siano $f_i = f^{(\mathcal{P}_i)}$ $(i = 1, 2)$ versioni di X. Considerata una funzione reale τ di dominio $\mathrm{rg}(X)$, risulta $(\tau \circ f_1)(\omega_1) = \tau(f_1(\omega_1)) = \tau(f_2(\omega_2)) = (\tau \circ f_2)(\omega_2)$ per ogni $\omega_1 \in \mathcal{P}_1$, $\omega_2 \in \mathcal{P}_2$ tali che $\omega_1 \wedge \omega_2 \neq \emptyset$. Allora, $\tau \circ f_1$ e $\tau \circ f_2$ sono funzioni equivalenti e quindi individuano lo stesso numero aleatorio che, dipendendo solo da X e non dalla particolare versione scelta, denotiamo con $\tau(X)$. Risulta pertanto:

- $\tau \circ f^{(\mathcal{P})}$ è una versione di $\tau(X)$, qualunque sia la versione $f^{(\mathcal{P})}$ di X.

Nel caso particolare che $\tau(x) = \alpha x$ per ogni reale x oppure $\tau(x) = \frac{1}{x}$ per ogni reale $x \neq 0$, otteniamo:

- $\alpha f^{(\mathcal{P})}$ è una versione di αX, qualunque sia la versione $f^{(\mathcal{P})}$ di X;
- $\frac{1}{f^{(\mathcal{P})}}$ è una versione di $\frac{1}{X}$, qualunque sia la versione $f^{(\mathcal{P})}$ di X, se $0 \notin \mathrm{rg}(X)$.

Passando alle operazioni di somma e prodotto di numeri aleatori, consideriamo anche il numero aleatorio Y e supponiamo che $g_1 = g^{(\overline{\mathcal{P}}_1)}$ sia una versione di Y. Introdotta la partizione:

$$\mathcal{P}^{(1)} = \{\omega^{(1)} = \omega_1 \wedge \overline{\omega}_1 \neq \emptyset : \omega_1 \in \mathcal{P}_1 \text{ e } \overline{\omega}_1 \in \overline{\mathcal{P}}_1\},$$

[19]Infatti, posto $f_i = f_i^{(\mathcal{P}_i)}$ e $A_i = f_i(\mathcal{P}_i)$ $(i = 1, 2)$, sia $f_1 \sim f_2$. Allora, dato $\omega_1 \in \mathcal{P}_1$, esiste, per il Lemma 1.9.1, $\omega_2 \in \mathcal{P}_2$ tale che $\omega_1 \wedge \omega_2 \neq \emptyset$. Allora, $f_1(\omega_1) = f_2(\omega_2) \in A_2$ da cui, data l'arbitrarietà di ω_1, si ha $A_1 \subset A_2$. In modo del tutto analogo si ottiene l'inclusione opposta.

[20]I numeri certi sono quindi quei particolari numeri aleatori che hanno come versioni funzioni costanti sui casi elementari della relativa partizione.

supponiamo che φ sia una funzione reale di dominio $D_1 = \{(f_1(\omega_1), g_1(\overline{\omega}_1)) : \omega^{(1)} \in \mathcal{P}^{(1)}\}$. Possiamo allora considerare la funzione $\varphi_1 = \varphi(f_1, g_1)^{(\mathcal{P}^{(1)})}$ così definita:

$$\varphi_1(\omega^{(1)}) = \varphi(f_1(\omega_1), g_1(\overline{\omega}_1))$$

per ogni $\omega^{(1)} \in \mathcal{P}^{(1)}$.
Sia ora $g_2 = g^{(\overline{\mathcal{P}}_2)}$ un'altra versione di Y. Posto, in analogia a quanto fatto sopra,

$$\mathcal{P}^{(2)} = \{\omega^{(2)} = \omega_2 \wedge \overline{\omega}_2 \neq \emptyset : \omega_2 \in \mathcal{P}_2 \text{ e } \overline{\omega}_2 \in \overline{\mathcal{P}}_2\}$$

e $D_2 = \{(f_2(\omega_2), g(\overline{\omega}_2)) : \omega^{(2)} \in \mathcal{P}^{(2)}\}$, risulta $D_1 = D_2$. Infatti, dato $\omega^{(1)}$, esiste, per il Lemma 1.9.1, $\omega^{(2)}$ tale che $\emptyset \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \omega_2) \wedge (\overline{\omega}_1 \wedge \overline{\omega}_2)$ e quindi $\omega_1 \wedge \omega_2 \neq \emptyset$ e $\overline{\omega}_1 \wedge \overline{\omega}_2 \neq \emptyset$; dunque $f_1(\omega_1) = f_2(\omega_2)$ ($f_1 \sim f_2$!) e $g_1(\overline{\omega}_1) = g_2(\overline{\omega}_2)$ ($g_1 \sim g_2$!), cioè $(f_1(\omega_1), g_1(\overline{\omega}_1)) \in D_2$. Ne segue $D_1 \subset D_2$. Per simmetria risulta anche $D_2 \subset D_1$ e quindi $D_1 = D_2$. Possiamo quindi considerare la funzione φ_2 di dominio $\mathcal{P}^{(2)}$ tale che:

$$\varphi_2(\omega^{(2)}) = \varphi(f_2(\omega_2), g_2(\overline{\omega}_2)).$$

Proviamo ora che φ_2 è equivalente a φ_1. A tal fine, sia $\emptyset \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \omega_2) \wedge (\overline{\omega}_1 \wedge \overline{\omega}_2)$ e quindi $f_1(\omega_1) = f_2(\omega_2)$ e $g_1(\overline{\omega}_1) = g_2(\overline{\omega}_2)$; ne segue, $\varphi_1(\omega^{(1)}) = \varphi(f_1(\omega_1), g_1(\overline{\omega}_1)) = \varphi(f_2(\omega_2), g_2(\overline{\omega}_2)) = \varphi_2(\omega^{(2)})$. Conseguentemente, φ_1 e φ_2 individuano lo stesso numero aleatorio che, dipendendo solo da X, Y e non dalle particolari versioni scelte, denotiamo con $\varphi(X, Y)$. Risulta pertanto:

- $\varphi(f^{(\mathcal{P}_1)}, g^{(\overline{\mathcal{P}}_1)})^{(\mathcal{P}^{(1)})}$ è una versione di $\varphi(X, Y)$, qualunque siano le versioni $f^{(\mathcal{P}_1)}$ di X e $g^{(\overline{\mathcal{P}}_1)}$ di Y.

Nel caso particolare che φ denoti la somma o il prodotto, otteniamo che sono versioni di $X + Y$ e di XY relative a $\mathcal{P}^{(1)}$, rispettivamente, le funzioni di dominio $\mathcal{P}^{(1)}$ tali che:

- $f^{(\mathcal{P}_1)}(\omega_1) + g^{(\overline{\mathcal{P}}_1)}(\overline{\omega}_1)$, qualunque siano le versioni $f^{(\mathcal{P}_1)}$ di X e $g^{(\overline{\mathcal{P}}_1)}$ di Y;
- $f^{(\mathcal{P}_1)}(\omega_1)\, g^{(\overline{\mathcal{P}}_1)}(\overline{\omega}_1)$, qualunque siano le versioni $f^{(\mathcal{P}_1)}$ di X e $g^{(\overline{\mathcal{P}}_1)}$ di Y.

È facile rendersi conto che, con queste definizioni, l'aritmetica dei numeri aleatori ha le medesime proprietà di quella dei numeri reali.

Passando infine alle relazioni d'ordine tra numeri aleatori, poniamo:

- $X \lhd Y$ se e solo se $f^{(\mathcal{P}_1)}(\omega_1) \lhd g^{(\overline{\mathcal{P}}_1)}(\overline{\omega}_1)$ per ogni $\omega_1 \wedge \overline{\omega}_1 \neq \emptyset$ $\quad (\lhd \in \{\leq, <\})$[21].

[21] Qualora risulti $\omega_1 \wedge \overline{\omega}_1 \neq \emptyset$ per ogni ω_1 e $\overline{\omega}_1$, la condizione $X \lhd Y$ assicura che i ranghi $\mathrm{rg}(X), \mathrm{rg}(Y)$ formano una coppia di classi separate.

Per verificare che la definizione data è ben fondata, supponiamo $f_1(\omega_1) \lhd g_1(\overline{\omega}_1)$ per ogni $\omega^{(1)}$ e consideriamo un generico $\omega^{(2)}$. Esiste allora, per il Lemma 1.9.1, $\omega^{(1)}$ tale che $\emptyset \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \omega_2) \wedge (\overline{\omega}_1 \wedge \overline{\omega}_2)$ da cui otteniamo $f_2(\omega_2) = f_1(\omega_1)$, $g_2(\overline{\omega}_2) = g_1(\overline{\omega}_1)$ e quindi $f_2(\omega_2) \lhd g_2(\overline{\omega}_2)$. Ovviamente, a differenza dell'usuale ordinamento per grandezza dei numeri reali, l'ordinamento $\leq$ che abbiamo considerato per i numeri aleatori non è totale (cioè due numeri aleatori non sono sempre confrontabili).

In questo ordine di idee, introduciamo ora l'evento:

- $\{X \trianglelefteq Y\} = \displaystyle\bigvee_{\omega_1 \wedge \overline{\omega}_1 \neq \emptyset \,:\, f^{(\mathcal{P}_1)}(\omega_1) \trianglelefteq g^{(\overline{\mathcal{P}}_1)}(\overline{\omega}_1)} (\omega_1 \wedge \overline{\omega}_1) \qquad (\trianglelefteq \in \{\leq, <, =\})$[22].

Per constatare che la definizione data è ben fondata, proviamo che $E_1 = E_2$, avendo posto $E_i = \bigvee_{\omega^{(i)} \in \mathcal{P}^{(i)} \,:\, f_i(\omega_i) \trianglelefteq g_i(\overline{\omega}_i)} \omega^{(i)}$ $(i = 1, 2)$. Dato $\omega^{(1)} \in \mathcal{P}^{(1)}$ tale che $f_1(\omega_1) \trianglelefteq g_1(\overline{\omega}_1)$, risulta

$$\begin{aligned}
\omega^{(1)} \wedge E_2 &= \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)} \,:\, f_2(\omega_2) \trianglelefteq g_2(\overline{\omega}_2)} (\omega^{(1)} \wedge \omega^{(2)}) \\
&= \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)} \,:\, \omega^{(2)} \wedge \omega^{(1)} \neq \emptyset \text{ e } f_2(\omega_2) \trianglelefteq g_2(\overline{\omega}_2)} (\omega^{(1)} \wedge \omega^{(2)}).
\end{aligned}$$

Osservato che $f_2(\omega_2) = f_1(\omega_1)$ e $g_2(\overline{\omega}_2) = g_1(\overline{\omega}_1)$ per ogni $\omega^{(1)} \wedge \omega^{(2)} \neq \emptyset$, otteniamo

$$\begin{aligned}
\omega^{(1)} \wedge E_2 &= \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)} \,:\, \omega^{(2)} \wedge \omega^{(1)} \neq \emptyset \text{ e } f_1(\omega_1) \trianglelefteq g_1(\overline{\omega}_1)} (\omega^{(1)} \wedge \omega^{(2)}) \\
&= \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)} \,:\, \omega^{(2)} \wedge \omega^{(1)} \neq \emptyset} (\omega^{(1)} \wedge \omega^{(2)}) = \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)}} (\omega^{(1)} \wedge \omega^{(2)}) \\
&= \omega^{(1)} \wedge \bigvee_{\omega^{(2)} \in \mathcal{P}^{(2)}} \omega^{(2)} = \omega^{(1)} \wedge \Omega = \omega^{(1)}.
\end{aligned}$$

Si ha quindi $E_1 \wedge E_2 = \bigvee_{\omega^{(1)} \in \mathcal{P}^{(1)} \,:\, f_1(\omega_1) \trianglelefteq g_1(\overline{\omega}_1)} \omega^{(1)} \wedge E_2 = E_1$, cioè $E_1 \to E_2$. Per simmetria risulta anche $E_2 \to E_1$ e quindi $E_1 = E_2$.

Dato un numero reale x qualsiasi, denotiamo con $X \lhd x$, $x \lhd X$ e $\{X \trianglelefteq x\}$, rispettivamente, le relazioni $X \lhd Y$, $Y \lhd X$ e l'evento $\{X \trianglelefteq Y\}$ relativi al numero aleatorio X e al numero certo Y di valore x. Si ha pertanto:

[22]Evento che, da un punto di vista interpretativo, si verifica qualora X assuma un valore non maggiore di Y, se $\trianglelefteq \equiv \leq$, e un valore minore di Y, se $\trianglelefteq \equiv <$ e un valore uguale a Y, se $\trianglelefteq \equiv =$.

- $\{X \trianglelefteq x\} = \displaystyle\bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, f^{(\mathcal{P}_1)}(\omega_1) \trianglelefteq x} \omega_1 \qquad (\trianglelefteq \in \{\leq, <, =\}).$

Di particolare interesse sono le relazioni seguenti che collegano gli eventi del tipo $\{X \leq x\}$ con quelli del tipo $\{X < x\}$. Per ogni x reale risulta:

$$\{X < x\} = \bigvee_{n \geq 1} \{X \lhd x - \frac{1}{n}\}, \qquad \{X \leq x\} = \bigwedge_{n \geq 1} \{X \lhd x + \frac{1}{n}\}, \tag{1.3}$$

ove $\lhd$ denota uno qualsiasi dei simboli $\leq, <$. Proviamo intanto la prima uguaglianza. Osservato che, per ogni ω_1, si ha $f_1(\omega_1) < x \Leftrightarrow f_1(\omega_1) \lhd x - \frac{1}{n}$ per qualche $n \geq 1$, otteniamo, tramite il Teorema 1.8.5(i),(iv),

$$\begin{aligned} \mathrm{set}(\{X < x\}) &= \{\omega_1 \in \mathcal{P}_1 : f_1(\omega_1) < x\} = \bigcup_{n \geq 1} \{\omega_1 \in \mathcal{P}_1 : f_1(\omega_1) \lhd x - \frac{1}{n}\} \\ &= \bigcup_{n \geq 1} \mathrm{set}(\{X \lhd x - \frac{1}{n}\}) = \mathrm{set}(\bigvee_{n \geq 1} \{X \lhd x - \frac{1}{n}\}) \end{aligned}$$

e quindi l'uguaglianza in esame. Per quanto riguarda l'altra uguaglianza, la dimostrazione è analoga, osservato che, per ogni ω_1, risulta $f_1(\omega_1) \leq x \Leftrightarrow f_1(\omega_1) \lhd x + \frac{1}{n}$ per ogni $n \geq 1$.

Concludiamo l'argomento osservando che, nel capitolo terzo, un ruolo centrale sarà svolto anche dagli eventi $\{X \neq x\} = \overline{\{X = x\}}$, $\{X \geq x\} = \overline{\{X < x\}}$, $\{X > x\} = \overline{\{X \leq x\}}$ e $\{a \lhd X \blacktriangleleft b\} = \{a \lhd X\} \wedge \{X \blacktriangleleft b\}$, ove $\lhd, \blacktriangleleft$ denotano uno qualunque dei simboli $<, \leq$. Con riferimento alla loro rappresentazione in termini delle versioni di X, tramite il Teorema 1.8.5(i),(v),(iii) risulta:

$$\{X = x\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, f^{(\mathcal{P}_1)}(\omega_1) = x} \omega, \quad \{X \geq x\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, f^{(\mathcal{P}_1)}(\omega_1) \geq x} ,$$

$$\{X > x\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, f^{(\mathcal{P}_1)}(\omega_1) > x} , \quad \{a \lhd X \blacktriangleleft b\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, \{a \lhd f^{(\mathcal{P}_1)}(\omega_1) \blacktriangleleft b\}} \omega.$$

Inoltre sussistono, come è facile verificare ricorrendo al solito teorema 1.8.5, le seguenti uguaglianze (ove si suppone $a < b$):

$$\begin{aligned} \{X < b\} &= \{X < a\} \vee \{a \leq X < b\} = \{X \leq a\} \vee \{a < X < b\}, \\ \{X = b\} &= \{X \leq b\} \wedge \{X \geq b\}, \\ \{X \leq b\} &= \{X < b\} \vee \{X = b\} \\ &= \{X < a\} \vee \{a \leq X \leq b\} = \{X \leq a\} \vee \{a < X \leq b\}. \end{aligned} \tag{1.4}$$

Dopo i numeri certi, i più semplici numeri aleatori sono quelli aventi due soli valori possibili. Tra di essi, di particolare importanza sono gli indicatori di evento che ora introduciamo. Chiamiamo **indicatore** dell'evento E il numero aleatorio $|E|$ avente come versione la funzione definita sulla partizione generata $\{E, \overline{E}\}$ che associa 1 al costituente E e 0 al costituente $\overline{E}$[23]. Ne segue che $|\Omega|$ e $|\emptyset|$ sono numeri certi di valore 1 e 0, rispettivamernte; inoltre, come è immediato constatare, $|\overline{E}|$ è il complemento a 1 di $|E|$, cioè $|\overline{E}| = 1 - |E|$.

Concludiamo elencando le relazioni esistenti tra gli indicatori della congiunzione e della disgiunzione con quelli delle loro componenti.

Teorema 1.11.1 *Sussistono le proposizioni seguenti:*

(i) $|E \wedge F| = |E|\,|F|$;

(ii) $|E \vee F| = |E| + |F| - |E|\,|F|$;

(iii) $|E \vee F| = |E| + |F|$, *se* $E \wedge F = \emptyset$;

(iv) $|E| \leq |F|$, *se* $E \to F$.

Dimostrazione Considerata come partizione di riferimento la partizione generata $\mathcal{P}_G(E, F)$, le proposizioni (i) ÷ (iii) si ottengono dalla tabella seguente dove vengono riportati, in funzione dei costituenti, i valori delle versioni degli indicatori $|E \wedge F|$, $|E \vee F|$ e della loro somma come pure i valori delle versioni degli indicatori $|E|$, $|F|$ e del loro prodotto e della loro somma.

	$\|E\|$	$\|F\|$	$\|E \wedge F\|$	$\|E\|\,\|F\|$	$\|E \vee F\|$	$\|E\| + \|F\|$	$\|E \vee F\| + \|E \wedge F\|$
$E \wedge F$	1	1	1	1	1	2	2
$E \wedge \overline{F}$	1	0	0	0	1	1	1
$\overline{E} \wedge F$	0	1	0	0	1	1	1
$\overline{E} \wedge \overline{F}$	0	0	0	0	0	0	0

(iv) Dalla $E \to F$, per il Teorema 1.7.2(i), risulta $E = E \wedge F$ e quindi, tramite (i), $|E| = |E \wedge F| = |E|\,|F| \leq |E|$. □

[23]Da un punto di vista interpretativo, l'indicatore di un evento può essere inteso come quella grandezza che assume valore 1, se l'evento si verifica, e valore 0, altrimenti.

Capitolo 2

Valutazione dell'incertezza

Affrontata, nel capitolo precedente, la descrizione dell'incertezza, tratteremo ora il problema della sua valutazione. Come la nozione di evento era la chiave di volta della descrizione, così la nozione di probabilità sarà quella della valutazione. Per introdurla seguiremo l'impostazione assiomatica (e quindi una trattazione astratta di natura ipotetico-deduttiva) ricorrendo, per giustificare gli assiomi, alla sua interpretazione in termini di quozienti di scommesse relative ad eventi (che riteniamo più consona, rispetto ad altre, agli studenti di materie attuariali e/o economiche).

2.1 Probabilità

Per **algebra di eventi** intendiamo una famiglia di eventi che contenga l'evento certo ed sia chiusa per negazione e disgiunzioni finite. Convenuto che, nel seguito, *$\mathcal{A}$ rappresenti un'algebra di eventi e che E, F (dotati o no di apici o pedici) denotino suoi elementi*, otteniamo (ricorrendo alle leggi di De Morgan) che sussistono le proprietà seguenti:

A1 $\emptyset, \Omega \in \mathcal{A}$.

A2 $\overline{E} \in \mathcal{A}$.

A3 $\bigvee_{i=1}^{n} E_i, \bigwedge_{i=1}^{n} E_i \in \mathcal{A}$.

Chiaramente $\{\emptyset, \Omega\}$ è la più "piccola" (nel senso dell'inclusione) algebra di eventi. Inoltre, è un'algebra la famiglia degli eventi logicamente dipendenti da una partizione dell'evento certo (Teorema 1.8.4).

Passiamo ora alla nozione astratta di probabilità. Con riferimento all'algebra $\mathcal{A}$, una funzione Pr di dominio $\mathcal{A}$ è una **probabilità su $\mathcal{A}$** se verifica i seguenti assiomi:

P1 $\Pr(\Omega) = 1$.

P2 $\Pr(E) \geq 0$.

P3 $\Pr(E \vee F) = \Pr(E) + \Pr(F)$, se $E \wedge F = \emptyset$.

Conveniamo che, nel seguito, Pr *denoti una probabilità sull'algebra $\mathcal{A}$.*

Osservazione 2.1.1 SCHEMA DELLA SCOMMESSA Per *scommessa relativa all'evento E* intendiamo un'operazione di scambio tra un importo certo p (la *puntata*) e un importo aleatorio di valore $v > 0$ (la *vincita*), se E si verifica, e zero, altrimenti. Si *scommette su E* se si *acquista* la scommessa (cioè si riceve $q|E|$ al prezzo p) e si *scommette contro E* se si *vende* la scommessa (cioè si vende $q|E|$ al prezzo p). A fronte delle due possibili *direzioni* della scommessa, "su"e "contro", i guadagni relativi sono $v|E| - p$ e $p - v|E|$. Introducendo il puntatore S che vale 1 o -1 a seconda che si acquisti o si venda la scommessa, otteniamo che il *guadagno* aleatorio relativo può esprimersi nel modo seguente:

$$G^{(E)}(p, v; S) = S(v|E| - p) = vS(|E| - \frac{p}{v}) = v\, S \cdot \begin{cases} 1 - q & \text{se } E \text{ è vero} \\ -q & \text{se } E \text{ è falso} \end{cases},$$

ove $q = \frac{p}{v}$ è detto *quoziente di scommessa.*

Assumiamo ora, per semplicità, che per Te sia indifferente "scommettere su E al quoziente q"o "scommettere contro E al quoziente q". Allora, viene naturale pensare che riterrai ammissibile la scommessa solamente se questa evita la perdita certa, cioè se $q \leq 1$; inoltre, che la quota aumenterà al crescere della tua fiducia sul verificarsi dell'evento, divenendo tanto più prossima a 1 quanto più ritieni che l'evento si verifichi e tanto più vicina a 0 quanto meno credi sul suo verificarsi. Conseguentemente, il quoziente di scommessa può essere inteso come una misura del *grado di fiducia* sul verificarsi dell'evento E e quindi in grado di esprimere una valutazione numerica dell'incertezza, cioè una probabilità[1].

[1] L'identificazione della probabilità con il grado di fiducia è sempre stata presente nello sviluppo del calcolo delle probabilità. Ad esempio, in *Recherches su la probabilité des jugements en matiére criminelle et en matiére civile* (1837), Siméon-Denis Poisson scrive: "La probabilità di un avvenimento è la ragione che abbiamo di credere che esso avrà luogo o abbia avuto luogo". Dunque, per Poisson la probabilità misura la ragione che abbiamo di credere che quell'avvenimento avrà o non avrà luogo; potrà quindi variare da persona

Passando ad esaminare un *sistema di scommesse relative ad eventi*, consideriamo gli eventi $E_1, \ldots E_n$ e mettiamo in evidenza, tramite il valori del puntatore S_i, se la relativa scommessa è su E_i oppure contro E_i $(1 \leq i \leq n)$. Indicati i quozienti di scommessa rispettivi con $q_1, \ldots q_n$ e le vincite con $v_1, \ldots v_n$, il guadagno aleatorio relativo sarà:

$$G^{(E_1,\ldots,E_n)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) = S_1 v_1(|E_1| - q_1) + \cdots + S_n v_n(|E_n| - q_n).$$

Ora, poichè per Te è indifferente scommettere su o contro l'evento E_i alla quota q_i $(i \leq n)$, viene naturale ritenere che considererai anche il sistema di scommesse "opposto"(ottenuto dal precedente scambiando ogni S_i con $-S_i$) e richiederai, al fine dell'ammissibilità, che i due sistemi non comportini una perdita certa, cioè che non risulti $\max G^{(E_1,\ldots,E_n)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) < 0$ e $\max G^{(E_1,\ldots,E_n)}(\mathbf{q}, \mathbf{v}; -\mathbf{S}) < 0$.

In questo ordine di idee le quote $q_1, \ldots, q_n$ si dicono *eque* se consentono di *evitare la perdita certa*, cioè se assicurano la validità delle disuguaglianze:

$$\min G^{(E_1,\ldots,E_n)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) \leq 0 \leq \max G^{(E_1,\ldots,E_n)}(\mathbf{q}, \mathbf{v}; \mathbf{S})$$

per ogni $\mathbf{v} = (v_1, \ldots, v_n) \in]0, +\infty[^n$ e $\mathbf{S} = (S_1, \ldots, S_n) \in \{-1, 1\}^n$.

Ciò precisato, assumiamo che d'ora in poi le quote siano eque. Ne segue che, per la scommessa relativa all'evento certo, risulta $G^{(\Omega)}(q, 1; 1) = (1 - q)$ e quindi,

---

a persona e da momento a momento perchè legata all'informazione soggettiva in un dato istante. Analogamente, nel suo *Formal Logic: or the Calculus of inference, necessary and Probable* (1847), Augustus De Morgan sottolinea che "per grado di probabilità noi intendiamo in realtà, o dovremmo intendere, il degree of belief (grado di convincimento o di fiducia)".

L'aspetto soggettivo della probabilità, che emerge dalla sua identificazione con il grado di fiducia, è uno dei due aspetti della natura della probabilità. Infatti, sin dalla metà del XVII secolo, l'idea di probabilità è stata una specie di Giano bifronte presentando una faccia *oggettiva* legata alle "frequenze"(e quindi di natura sperimentale-statistica) e una *soggettiva* legata ai "degrees of belief"(ognuna delle quali predominerà sull'altra in epoche diverse). Per comprendere la differenza sostanziale tra le due visioni consideriamo gli eventi descritti dai seguenti due enunciati:

- la cinquina secca 34, 12, 89, 45, 7 verrà estratta nella prossima estrazione del lotto sulla ruota di Venezia;

- il cavallo Furia arriverà primo nella prossima corsa all'ippodromo delle Capanelle di Roma.

Nel primo caso, ogni giocatore concorderà che la probabilità è $\frac{85!}{90!}$ - essendo precisate in modo inequivocabile le regole di estrazione (che non consentono di preferire l'estrazione di una pallina rispetto ad un'altra) - e quindi la sua valutazione avrà natura oggettiva. Nel secondo invece, due scommettitori potranno dare all'evento probabilità diverse - poichè potrebbero avere informazioni diverse (sulla salute del cavallo, sulla bravura del fantino, ...) - e quindi le loro valutazioni avranno natura soggettiva (rimanendo, in generale, quella di uno scommettitore sconosciuta all'altro).

essendo il guadagno un numero certo, deve essere $1-q=0$, cioè $q=1$. Dati infine due eventi incompatibili E e F, consideriamo il sistema di scommesse relative agli eventi E, F e $E\vee F$ avente vincite unitarie e, nell'ordine, quote q', q'', q e puntatori S', S'', S. Per il Teorema 1.11.1(iii) risulta allora,

$$\begin{aligned} G^{(E,F,E\vee F)}(\mathbf{q},\mathbf{1};\mathbf{S}) &= S'(|E|-q')+S''(|F|-q'')+S(|E\vee F|-q) \\ &= S'(|E|-q')+S''(|F|-q'')+S(|E|+|F|-q) \\ &= [(S'+S)|E|+(S''+S)|F|\,]-(S'q'+S''q''+Sq). \end{aligned}$$

Posto quindi $S'=S=1, S''=-1$ (in modo da annullare la parte aleatoria) otteniamo che il guadagno diviene il numero certo di valore $-(-q'-q''+q)$ e pertanto deve risultare, per evitare la perdita certa, $q=q'+q''$.

Le considerazioni fatte consentono dunque di giustificare gli assiomi della probabilità proposti, qualora si interpretino le probabilità come quote di scommessa eque nei sistemi di scommesse relative ad eventi[2]. $\triangle$

Nel teorema seguente riportiamo alcune importanti proprietà delle probabilità. In particolare, (iii), (v) e (vii) assicurano che la probabilità è una funzione d'evento additiva, monotona e subadditiva (rispetto alle disgiunzioni finite); la (viii) invece consente di calcolare la probabilità di una disgiunzione finita di eventi tramite i valori che la probabilità assume su tutte le loro possibili congiunzioni; la (ix) assicura che un **evento trascurabile** (cioè di probabilità nulla)[3] disgiunto ad un evento non ne altera la probabilità e che

[2]Il collegamento tra quota di scommessa e probabilità era presente sin dagli inizi del calcolo delle probabilità. Ad esempio, in *Essay Towards Solving a Problem in the Doctrine of Changes* (1763), Thomas Bayes afferma che:

> la probabilità di un evento è il rapporto tra il valore al quale un'aspettativa che dipende dall'accadere di quell'evento deve essere calcolata ed il valore che ciò che si intende assume una volta che l'evento si è verificato.

Riscrivendo questa definizione con la simbologia introdotta, otteniamo la formulazione:

> Sia E un evento. Considerato un oggetto B di valore $v>0$, sia p il valore dell'offerta "B se si verifica E; nulla altrimenti". Dicesi allora probabilità di E il rapporto $\frac{p}{v}$.

Sebbene Bayes non chiarisca cosa intenda con la parola "valore", appare verosimile ritenere che la identificasse con quella di "giusto prezzo"(espresso in termini monetari). Si verrebbe così a definire la probabilità dell'evento E come rapporto tra il giusto prezzo p della promessa di una vincita monetaria v, a condizione che si verifichi E, e v (inteso come giusto prezzo di B); cioè, nel linguaggio delle scommesse, come quota equa della scommessa relativa all'evento E.

[3]Anche chiamato **evento quasi impossibile**.

un **evento quasi certo** (cioè di probabilità uno) congiunto ad un evento non ne muta la probabilità; la (x) assicura che disgiunzioni finite di eventi quasi impossibili è quasi impossibile e che la congiunzione finita di eventi quasi certi è quasi certa; la (xi) fornisce infine due interessanti limitazioni della probabilità di un evento tramite le probabilità dei casi elementari di una partizione finita dell'evento certo.

Teorema 2.1.2 *Sussistono le proposizioni seguenti:*

(i) $\Pr(\overline{E}) = 1 - \Pr(E)$;

(ii) $\Pr(\emptyset) = 0 \leq \Pr(E) \leq 1 = \Pr(\Omega)$;

(iii) ADDITIVITÀ: $\Pr(\bigvee_{i=1}^n E_i) = \sum_{i=1}^n \Pr(E_i)$, *se* $E_i \wedge E_j = \emptyset$ $(i \neq j)$;

(iv) $\Pr(E \vee F) + \Pr(E \wedge F) = \Pr(E) + \Pr(F)$;

(v) MONOTONIA: $\Pr(E) \leq \Pr(F)$, *se* $E \to F$;

(vi) DISUGUAGLIANZA DI BONFERRONI: $\Pr\big(\bigwedge_{i=1}^n E_i\big) \geq \sum_{i=1}^n \Pr(E_i) + 1 - n$;

(vii) SUBADDITIVITÀ: $\Pr(\bigvee_{i=1}^n E_i) \leq \sum_{i=1}^n \Pr(E_i)$;

(viii) FORMULA D'INCLUSIONE-ESCLUSIONE:

$$\Pr\Big(\bigvee_{i=1}^n E_i\Big) = \sum_{\emptyset \neq J \subset \{1,\ldots,n\}} (-1)^{\#J-1} \Pr\Big(\bigwedge_{j \in J} E_j\Big);$$

(ix) *Se* $\Pr(E) = 0$, *allora* $\Pr(E \vee F) = \Pr(F)$. *Inoltre, se* $\Pr(E) = 1$, *allora* $\Pr(E \wedge F) = \Pr(F)$;

(x) *Se* $\Pr(E_i) = 0$ $(i = 1.\ldots, n)$, *allora* $\Pr(\bigvee_{i=1}^n E_i) = 0$. *Inoltre, se* $\Pr(E_i) = 1$ $(i = 1.\ldots, n)$, *allora* $\Pr(\bigwedge_{i=1}^n E_i) = 1$;

(xi) *Sia* $\mathcal{P} \subset \mathcal{A}$ *una partizione finita dell'evento certo. Riesce allora:*

$$\sum_{\omega \to E} \Pr(\omega) \leq \Pr(E) \leq 1 - \sum_{\omega \to \overline{E}} \Pr(\omega).$$

Dimostrazione (i)+(ii) Dalle $\Omega = E \vee \overline{E}$, P1 e P3, risulta $1 = \Pr(\Omega) = \Pr(E) + \Pr(\overline{E})$. Ne segue (i) e $\Pr(\emptyset) = 0$. Da qui, per P2, la tesi.

(iii) Per P3 la tesi sussiste per $n = 2$. Procedendo per induzione, assumiamo che la tesi sussista per n e proviamola per $n + 1$. Considerati allora gli eventi $E_1, \dots, E_{n+1}$ tali che $E_i \wedge E_j = \emptyset$ $(i \neq j)$, otteniamo

$$\begin{aligned}\Pr(\bigvee_{i=1}^{n+1} E_i) &= \Pr((\bigvee_{i=1}^{n} E_i) \vee E_{n+1}) = \Pr(\bigvee_{i=1}^{n} E_i) + \Pr(E_{n+1}) \\ &= \sum_{i=1}^{n} \Pr(E_i) + \Pr(E_{n+1}) = \sum_{i=1}^{n+1} \Pr(E_i).\end{aligned}$$

(iv) Poichè $E \vee F = E \vee (\overline{E} \wedge F)$ e $F = (\overline{E} \wedge F) \vee (E \wedge F)$, tramite (iii), risulta $\Pr(E \vee F) + \Pr(E \wedge F) = [\Pr(E) + \Pr(\overline{E} \wedge F)] + \Pr(E \wedge F) = \Pr(E) + [\Pr(\overline{E} \wedge F) + \Pr(E \wedge F)] = \Pr(E) + \Pr(F)$.

(v) Sia $E \to F$. Allora, per il Teorema 1.7.2(i), $F = E \vee F = E \vee (\overline{E} \wedge F)$ e quindi, per (iii), (ii), $\Pr(F) = \Pr(E) + \Pr(\overline{E} \wedge F) \geq \Pr(E)$.

(vi) Sia intanto $n = 2$. Allora, tramite (iv), (ii), risulta $\Pr(E_1 \wedge E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \vee E_2) \geq \Pr(E_1) + \Pr(E_2) - 1 = \Pr(E_1) + \Pr(E_2) + 1 - 2$. Poichè la tesi sussiste per $n = 2$, assumiamo che sussista per n e proviamola per $n + 1$. Considerati allora gli eventi $E_1, \dots, E_{n+1}$, otteniamo

$$\begin{aligned}\Pr(\bigwedge_{i=1}^{n+1} E_i) &= \Pr((\bigwedge_{i=1}^{n} E_i) \wedge E_{n+1}) \geq \Pr(\bigwedge_{i=1}^{n} E_i) + \Pr(E_{n+1}) - 1 \\ &\geq (\sum_{i=1}^{n} \Pr(E_i) + 1 - n) + \Pr(E_{n+1}) - 1 \\ &= \sum_{i=1}^{n+1} \Pr(E_i) + 1 - (n + 1).\end{aligned}$$

(vii) Per (iv), (ii), la tesi sussiste per $n = 2$. Assumiamo quindi che sussista per n e proviamola per $n+1$. Considerati allora gli eventi $E_1, \dots, E_{n+1}$, otteniamo

$$\begin{aligned}\Pr(\bigvee_{i=1}^{n+1} E_i) &= \Pr((\bigvee_{i=1}^{n} E_i) \vee E_{n+1}) \leq \Pr(\bigvee_{i=1}^{n} E_i) + \Pr(E_{n+1}) \\ &\leq \sum_{i=1}^{n} \Pr(E_i) + \Pr(E_{n+1}) = \sum_{i=1}^{n+1} \Pr(E_i).\end{aligned}$$

(viii) Poichè per $n = 2$ la formula si riduce all'uguaglianza considerata in (iv), assumiamo che sussista per n e proviamola per $n + 1$. Dati allora gli eventi $E_1, \dots, E_{n+1}$, da (iv) otteniamo

$$\begin{aligned}
\Pr\Big(\bigvee_{i=1}^{n+1} E_i\Big) &= \Pr\Big(E_{n+1} \vee \bigvee_{i=1}^{n} E_i\Big) = \Pr(E_{n+1}) + \Pr\Big(\bigvee_{i=1}^{n} E_i\Big) - \Pr\Big(\bigvee_{i=1}^{n}(E_i \wedge E_{n+1})\Big) \\
&= \Pr(E_{n+1}) + \sum_{\emptyset \neq J \subset \{1,\dots,n\}} (-1)^{\#J-1} \Pr\Big(\bigwedge_{j\in J} E_j\Big) \\
&\quad - \sum_{\emptyset \neq J \subset \{1,\dots,n\}} (-1)^{\#J-1} \Pr\Big(\bigwedge_{j\in J} (E_j \wedge E_{n+1})\Big) \\
&= \Pr(E_{n+1}) + \sum_{\emptyset \neq J \subset \{1,\dots,n\}} (-1)^{\#J-1} \Pr\Big(\bigwedge_{j\in J} E_j\Big) \\
&\quad + \sum_{\emptyset \neq J \subset \{1,\dots,n\}} (-1)^{\#[J\cup\{n+1\}]-1} \Pr\Big(\bigwedge_{j\in J} (E_j \wedge E_{n+1})\Big).
\end{aligned}$$

Osservato infine che, con riferimento all'insieme $I = \{1, \dots, n + 1\}$, il primo addendo della somma riguarda il sottoinsieme di I formato solo dall'elemento $n + 1$, il secondo i sottoinsiemi di I formati solo con elementi di $\{1, \dots, n\}$ e l'ultimo i sottoinsiemi di I contenenti l'elemento $n + 1$ e aventi almeno due elementi, otteniamo

$$\Pr\Big(\bigvee_{i=1}^{n+1} E_i\Big) = \sum_{\emptyset \neq J \subset \{1,\dots,n+1\}} (-1)^{\#J-1} \Pr\Big(\bigwedge_{j\in J} E_j\Big).$$

(ix) Dalla $\Pr(E) = 0$ si ha, per (v), $\Pr(E \wedge F) = 0$ ($E \wedge F \to E$!) e quindi, per (iv), la tesi. Sia ora $\Pr(E) = 1$. Da (i), otteniamo $\Pr(\overline{E}) = 0$ e $\Pr(E \wedge F) = 1 - \Pr(\overline{E \wedge F}) = 1 - \Pr(\overline{E} \vee \overline{F})$. Ne segue, per quanto appena provato, $\Pr(E \wedge F) = 1 - \Pr(\overline{F}) = \Pr(F)$.

(x) La prima parte della tesi segue immediatamente da (vii). Per quanto riguarda la seconda, da (i) e da quanto appena provato risulta $\Pr(\bigvee_{i=1}^{n} \overline{E}_i) = 0$. Ne segue, $\Pr(\bigwedge_{i=1}^{n} E_i) = 1 - \Pr(\overline{\bigwedge_{i=1}^{n} E_i}) = 1 - \Pr(\bigvee_{i=1}^{n} \overline{E}_i) = 1$.

(xi) Osservato che, per il Teorema 1.7.2(v), $\bigvee_{\omega\to E} \omega \to E$, $\bigvee_{\omega\to \overline{E}} \omega \to \overline{E}$, da (v), (i) otteniamo $\Pr(\bigvee_{\omega\to E} \omega) \leq \Pr(E)$, $\Pr(\bigvee_{\omega\to \overline{E}} \omega) \leq \Pr(\overline{E}) = 1 - \Pr(E)$ e quindi, tramite (iii), la tesi. $\square$

Il teorema seguente assicura che gli eventi di probabilità positiva di una qualsiasi famiglia di eventi a due a due incompatibili sono al più numerabili.

Pertanto, in ogni partizione dell'evento certo più che numerabile ci sono casi elementari di probabilità nulla. Conseguentemente, la nozione di evento trascurabile non coincide con quella di evento impossibile (che, per il Teorema 2.1.2(ii) è a sua volta trascurabile); analogamente, quella di evento quasi certo non coincide con quella di evento certo.

Teorema 2.1.3 *Sia $(E_i)_{i\in I}$ una famiglia di eventi a due a due incompatibili. Allora l'insieme $J = \{i : \Pr(E_i) > 0\}$ è al più numerabile.*

DIMOSTRAZIONE Basta osservare che $J = \bigcup_{n\geq 2}\{i : \Pr(E_i) > \frac{1}{n}\}$ e che l'insieme $\{i : \Pr(E_i) > \frac{1}{n}\}$ ha, per l'additività (Teorema 2.1.2(iii)), al più $n-1$ elementi. □

2.2 Probabilità numerabilmente additive

Nel Teorema 2.1.2 abbiamo considerato solamente congiunzioni e disgiunzioni finite di eventi di $\mathcal{A}$. Passando alle congiunzioni e disgiunzioni numerabili, bisogna richiedere, per poter calcolare la loro probabilità, che anch'esse appartengano ad $\mathcal{A}$. *Assumiamo quindi, in questa sezione, che l'algebra $\mathcal{A}$ sia una* σ**-algebra**, cioè sia chiusa anche per disgiunzioni e congiunzioni numerabili[4].

Ciò premesso, proviamo il teorema seguente che collega la probabilità della disgiunzione numerabile di eventi a due a due incompatibili con la somma della serie delle probabilità degli eventi della successione.

Teorema 2.2.1 *Per ogni successione $(E_n)_{n\geq 1}$ di eventi a due a due incompatibili risulta:*

$$\sum_{n\geq 1} \Pr(E_n) \leq \Pr(\bigvee_{n\geq 1} E_n).$$

DIMOSTRAZIONE Per l'additività e la monotonia (Teorema 2.1.2(iii),(v)) risulta $\sum_{i=1}^n \Pr(E_i) = \Pr(\bigvee_{i=1}^n E_i) \leq \Pr(\bigvee_{n\geq 1} E_n)$, osservato che, per il Teorema 1.7.2(vi),(v), $\bigvee_{i=1}^n E_i \to \bigvee_{n\geq 1} E_n$. Ne segue, passando al limite per $n \to +\infty$, la tesi. □

[4]Come lo è, ad esempio, l'algebra degli eventi logicamente dipendenti da una partizione dell'evento certo (Teorema 1.8.4).

Un ruolo chiave nello sviluppo del calcolo delle probabilità è svolto da quelle particolari probabilità su $\mathcal{A}$ - chiamate **probabilità numerabilmente additive** - che rendono la precedente disuguaglianza un'uguaglianza, cioè tali che soddisfino la proprietà di **additività numerabile**:

$$\Pr(\bigvee_{n\geq 1} E_n) = \sum_{n\geq 1} \Pr(E_n) \tag{2.1}$$

per ogni successione $(E_n)_{n\geq 1}$ di eventi a due a due incompatibili.

La loro importanza è dovuta essenzialmente al fatto che verificano, come ora proveremo, una particolare proprietà di continuità che riguarda le successioni $(E_n)_{n\geq 1}$ che sono **non decrescenti** (cioè tali che $E_n \to E_{n+1}$ per ogni $n \geq 1$) oppure **non crescenti** (cioè tali che $E_{n+1} \to E_n$ per ogni $n \geq 1$).

Teorema 2.2.2 *Sia* Pr *numerabilmente additiva. Allora, per ogni successione di eventi* $(E_n)_{n\geq 1}$, *sussistono le proposizioni seguenti:*

(i) CONTINUITÀ DAL BASSO: $\Pr(E_n) \uparrow \Pr(\bigvee_{n\geq 1} E_n)$, *se la successione è non decrescente;*

(ii) CONTINUITÀ DALL'ALTO: $\Pr(E_n) \downarrow \Pr(\bigwedge_{n\geq 1} E_n)$, *se la successione è non crescente;*

(iii) SUBADDITIVITÀ NUMERABILE: $\Pr(\bigvee_{n\geq 1} E_n) \leq \sum_{n\geq 1} \Pr(E_n)$*;*

(iv) Se $\Pr(E_n) = 0$ *per ogni* n*, allora* $\Pr\big(\bigvee_{n\geq 1} E_n\big) = 0$*. Inoltre, se* $\Pr(E_n) = 1$ *per ogni* n*, allora* $\Pr\big(\bigwedge_{n\geq 1} E_n\big) = 1$*;*

(v) Sia $\mathcal{P} \subset \mathcal{A}$ *una partizione al più numerabile dell'evento certo. Riesce allora:*

$$\sum_{\omega \to E} \Pr(\omega) \leq \Pr(E) \leq 1 - \sum_{\omega \to \overline{E}} \Pr(\omega).$$

DIMOSTRAZIONE (i) Sia la successione non decrescente. Risulta allora

$$E_m = E_1 \vee \bigvee_{i=2}^{m} (E_i \wedge \overline{E}_{i-1}) \qquad (m \geq 2).$$

Sia intanto $m = 2$. Dalla $E_1 \vee (E_2 \wedge \overline{E}_1) = (E_1 \vee E_2) \wedge (E_1 \vee \overline{E}_1) = (E_1 \vee E_2) \wedge \Omega = E_1 \vee E_2$ otteniamo, tramite il Teorema 1.7.2(i), $E_1 \vee (E_2 \wedge \overline{E}_1) = E_2$, ricordato che $E_1 \to E_2$. Poichè l'uguaglianza sussiste per $m = 2$, assumiamo che sussista per $m \geq 2$ e proviamola per $m + 1$. Risulta

$$\begin{aligned} E_1 \vee \bigvee_{i=2}^{m+1} (E_i \wedge \overline{E}_{i-1}) &= E_1 \vee (\bigvee_{i=2}^{m} (E_i \wedge \overline{E}_{i-1}) \vee (E_{m+1} \wedge \overline{E}_m)) \\ &= (E_1 \vee \bigvee_{i=2}^{m} (E_i \wedge \overline{E}_{i-1})) \vee (E_{m+1} \wedge \overline{E}_m) \\ &= E_m \vee (E_{m+1} \wedge \overline{E}_m) = E_{m+1}. \end{aligned}$$

Considerato ora $m \geq 2$, tramite i teoremi 1.7.2(v) e 1.6.2(xi), otteniamo

$$\bigvee_{n\geq 1} E_n = (E_1 \vee \bigvee_{i=2}^{m} (E_i \wedge \overline{E}_{i-1})) \vee (\bigvee_{n\geq m+1} (E_n \wedge \overline{E}_{n-1})) = E_m \vee \bigvee_{n\geq m+1} (E_n \wedge \overline{E}_{n-1})$$

e quindi

$$\Pr(\bigvee_{n\geq 1} E_n) = \Pr(E_m) + \sum_{n\geq m+1} \Pr(E_n \wedge \overline{E}_{n-1}),$$

tenuto conto dell'additività numerabile e osservato che la successione E_m, $E_{m+1} \wedge \overline{E}_m, E_{m+2} \wedge \overline{E}_{m+1}, \ldots$ è formata da eventi a due a due incompatibili.

A questo punto, notato che la successione $(\Pr(E_m))_{m\geq 1}$ è non decrescente (Teorema 2.1.2(v)), possiamo passare al limite per $m \to +\infty$ ottenendo

$$\Pr(\bigvee_{n\geq 1} E_n) = \lim_{m\to+\infty} \Pr(E_m) + \lim_{m\to+\infty} \sum_{n\geq m+1} \Pr(E_n \wedge \overline{E}_{n-1}) = \lim_{m\to+\infty} \Pr(E_m),$$

una volta osservato che la serie $\sum_{n\geq m+1} \Pr(E_n \wedge \overline{E}_{n-1})$ è il resto $(m+1)$-simo di una serie convergente.

(ii) Sia la successione non crescente. Osservato che la successione $(\overline{E}_n)_{n\geq 1}$ è, per il Teorema 1.7.2(i), non decrescente, dal Teorema 2.1.2(i), da (i) e dal Teorema 1.6.2(vi),(viii) otteniamo

$$\begin{aligned} \Pr\big(\bigwedge_{n\geq 1} E_n\big) &= 1 - \Pr\big(\overline{\bigwedge_{n\geq 1} E_n}\big) = 1 - \Pr\big(\bigvee_{n\geq 1} \overline{E}_n\big) = 1 - \lim_{n\to+\infty} \Pr(\overline{E}_n) \\ &= \lim_{n\to+\infty} (1 - \Pr(\overline{E}_n)) = \lim_{n\to+\infty} \Pr(\overline{\overline{E}}_n) = \lim_{n\to+\infty} \Pr(E_n). \end{aligned}$$

(iii) Usando la subadditività (Teorema 2.1.2(vii)) otteniamo $\Pr(\bigvee_{i=1}^n E_i) \leq \sum_{i=1}^n \Pr(E_i) \leq \sum_{n\geq 1} \Pr(E_n)$. Ne segue, tramite (i), la tesi, osservato che la successione $(\bigvee_{i=1}^n E_i)_{n\geq 1}$ è non decrescente.

(iv)+(v) Dimostrazioni analoghe a quelle delle proposizioni (x), (xi) del Teorema 2.1.2 (usando la subadditività numerabile al posto della subadditività). □

Osservazione 2.2.3 La continuità dal basso e dall'alto della probabilità, rispettivamente per successioni non decrescenti e non crescenti, non sono solo condizioni necessarie per l'additività numerabile ma anche sufficienti. Infatti, l'equivalenza tra i due tipi di continuità si ottiene con un ragionamento analogo a quello fatto per dimostrare la proposizione (ii) del Teorema 2.2.2. Che poi, data una successione $(E_n)_{n\geq 1}$ di eventi a due a due incompatibili, la continuità dal basso implichi l'additività numerabile segue immediatamente dall'osservazione che la successione $(\bigvee_{i=1}^{n} E_i)_{n\geq 1}$ è non decrescente e che, per l'additività, $\Pr(\bigvee_{i=1}^{n} E_i) = \sum_{i=1}^{n} \Pr(E_i)$. △

2.3 Probabilità condizionata

Dato un evento non trascurabile (cioè di probabilità positiva) $H \in \mathcal{A}$, possiamo considerare, per il Teorema 2.1.2(ii), l'evento condizionato $E \mid H$ per ogni evento $E \in \mathcal{A}$. Osservato che, per il Teorema 1.10.2(ii), gli eventi condizionati $E \mid H$ e $F \mid H$ sono uguali se e solo se $E \wedge H = F \wedge H$, introduciamo la probabilità dell'evento condizionato $E \mid H$, che chiameremo **probabilità condizionata di E a H**, ponendo:

$$\Pr(E \mid H) = \frac{\Pr(E \wedge H)}{\Pr(H)}.$$

Ne segue $\Pr(\Omega \mid H) = 1$, osservato che $\Omega \wedge H = H$. Inoltre, dati due eventi E, F incompatibili, dall'additività otteniamo

$$\Pr((E \vee F) \wedge H) = \Pr((E \wedge H) \vee (F \wedge H)) = \Pr(E \wedge H) + \Pr(F \wedge H)$$

e quindi

$$\begin{aligned} \Pr(E \vee F \mid H) &= \frac{\Pr((E \vee F) \wedge H)}{\Pr(H)} = \frac{\Pr(E \wedge H) + \Pr(F \wedge H)}{\Pr(H)} \\ &= \Pr(E \mid H) + \Pr(F \mid H). \end{aligned}$$

Infine, se la probabilità è numerabilmente additiva[5], risulta

$$\Pr(\bigvee_{n\geq 1} E_n \mid H) = \frac{\Pr((\bigvee_{n\geq 1} E_n) \wedge H)}{\Pr(H)} = \frac{\sum_{n\geq 1} \Pr(E_n \wedge H)}{\Pr(H)} = \sum_{n\geq 1} \Pr(E_n \mid H),$$

[5] In questo caso sarà sempre implicitamente supposto che $\mathcal{A}$ sia una σ-algebra.

qualunque sia la successione $(E_n)_{n\geq 1}$ di eventi a due a due incompatibili.

Conseguentemente, la probabilità condizionata $\Pr(\cdot \mid H)$ può essere intesa come la *probabilità finale* (numerabilmente additiva) su $\mathcal{A}$ che si ottiene aggiornando la *probabilità iniziale* (numerabilmente additiva) $\Pr(\cdot)$ a seguito dell'incremento d'informazione dovuto all'evento non trascurabile H.

Osservazione 2.3.1 CONTINUAZIONE DELL'OSSERVAZIONE 2.1.1. Per giustificare, mediante lo schema della scommessa, la definizione data, iniziamo col precisare cosa si debba intendere per *scommessa relativa all'evento condizionato* $E \mid H$. Poichè tale evento perde di significato qualora H risulti falso, la scommessa viene, in questo caso, annullata e quindi determina un guadagno nullo. Invece, nel caso che H risulti vero, la scommessa si comporta come una scommessa relativa all'evento E e quindi prevede uno scambio tra una puntata p e una vincita $v > 0$, se E è vero, e zero, altrimenti. Allora, il guadagno aleatorio relativo sarà:

$$G^{(E|H)}(p, v; S) = S[(v-p)|E\wedge H|+(-p)|\overline{E}\wedge H|\,] = S\cdot \begin{cases} v-p & \text{se } E,\ H \text{ veri} \\ -p & \text{se } E \text{ falso, } H \text{ vero} \\ 0 & \text{se } H \text{ falso} \end{cases} .$$

Introducendo la quota di scommessa q, dal Teorema 1.11.1(i) otteniamo

$$\begin{aligned} G^{(E|H)}(p, v; q) &= S\,v[(1-q)|E|\,|H| - q|\overline{E}|\,|H|\,] \\ &= S\,v\,([(1-q)|E| - q(1-|E|)])\,|H| = S\,v\,(|E| - q)|H|. \end{aligned}$$

Consideriamo ora un sistema di scommesse "misto" formato da scommesse relative agli eventi assoluti $E_1, \ldots, E_n$ e da scommesse relative agli eventi condizionati $E_{n+1} \mid H, \ldots, E_{n+m} \mid H$, con puntatori S_i, quote q_i e vincite v_i $(i = 1, \ldots, n+m)$. Allora, posto $\mathcal{E}' = \{E_1, \ldots, E_n\}$ e $\mathcal{E}''|H = \{E_{n+1}|H, \ldots, E_{n+m}|H\}$, il guadagno aleatorio relativo sarà:

$$G^{(\mathcal{E}', \mathcal{E}''|H)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) = \sum_{i=i}^{n} S_i v_i(|E_i| - q_i) + \sum_{j=n+1}^{n+m} S_j v_j(|E_j| - q_j)|H|.$$

Analogamente al caso delle scommesse relative a eventi assoluti, chiameremo le quote $q_1, \ldots, q_{n+m}$ *eque* se consentono di *evitare la perdita certa*, cioè se risulta

$$\min G^{(\mathcal{E}', \mathcal{E}''|H)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) \leq 0 \leq \max G^{(\mathcal{E}', \mathcal{E}''|H)}(\mathbf{q}, \mathbf{v}; \mathbf{S})$$

per ogni $\mathbf{v} = (v_1, \ldots, v_{n+m}) \in]0, +\infty[^{n+m}$ e $\mathbf{S} = (S_1, \ldots, S_{n+m}) \in \{-1, 1\}^{n+m}$.

Assunto che tutte le quote siano eque, dati gli eventi E e H tale che $\Pr(H) > 0$, consideriamo il sistema di scommesse relative agli eventi $E \wedge H$, H e $E \mid H$ con, rispettivamente, quote eque q', q'', q, vincite v', v'', v e puntatori S', S'', S. Riesce allora

$$\begin{aligned} G^{(E\wedge H, H, E|H)}(\mathbf{q}, \mathbf{v}; \mathbf{S}) &= S'v'(|E\wedge H| - q') + S''v''(|H| - q'') + Sv(|E| - q)|H| \\ &= S'v'|E||H| + S''v''|H| + Sv|E||H| - Sv\,q|H| \\ &\quad -(S'v'q' + S''q''v'') \\ &= [(S'v' + Sv)|E||H| + (S''v'' - Sv\,q)|H|\,] \\ &\quad -(S'v'q' + S''v''q'') \end{aligned}$$

e quindi, posto $S = S'' = 1, S' = -1$ e $v = v' = 1, v'' = q$ (in modo da annullare la parte aleatoria), otteniamo che il guadagno diviene il numero certo di valore $-(-p' + pp'')$. Ne segue, per evitare la perdita certa, che deve essere $-q' + qq'' = 0$, cioè $q' = qq''$.

Risulta così giustificata la definizione data per la probabilità condizionata, qualora si interpreti anche la probabilità di un evento condizionato come quota equa nella scommessa relativa a tale evento. △

Le proposizioni (ii), (iii) e (iv) del teorema seguente forniscono delle formule che svolgono un ruolo importante sia nello sviluppo del calcolo delle probabilità che nelle sue applicazioni. Le prime due consentono di calcolare delle probabilità assolute (cioè inerenti eventi assoluti) mediante probabilità condizionate (che spesso sono più facili da valutare nelle applicazioni); l'ultima invece permette di calcolare la probabilità condizionata di E a H tramite la probabilità condizionata di H a E e la probabilità di E[6]. La proposizione (vi) assicura che, nel caso di una partizione finita dell'evento certo formata da casi elementari di probabilità positiva, la probabilità di un qualsiasi evento appartiene al più piccolo intervallo contenente tutte le probabilità condizionate dell'evento ai casi elementari.

Teorema 2.3.2 *Sussistono le proposizioni seguenti:*

(i) Sia H un evento non trascurabile. Allora, $\Pr(E \wedge H \,|\, H) = \Pr(E \,|\, H)$. *Inoltre,* $\Pr(E \,|\, H) = 1$, *se* $H \to E$, *e* $\Pr(E \,|\, H) = 0$, *se* $H \to \overline{E}$. *Infine,* $\Pr(E \,|\, H) = \Pr(E)$, *se* $\Pr(E) = 1$ *o* $\Pr(H) = 1$, *e* $\Pr(E \,|\, H) = 0$, *se* $\Pr(E) = 0$;

[6]Coinvolgendo così tre stati di conoscenza: lo stato iniziale $\mathcal{C}$ e i due stati finali $\mathcal{C} \cup \{h\}$ e $\mathcal{C} \cup \{p\}$ determinati, rispettivamente, da una descrizione h dell'evento H e da una descrizione p dell'evento E.

(ii) *Sia $H_1, \dots, H_n \in \mathcal{A}$ una partizione finita dell'evento certo. Riesce allora:*

$$\Pr(E) = \sum_{i:\Pr(H_i)>0} \Pr(E \mid H_i) \Pr(H_i) \qquad \text{(formula di disintegrazione)}^7;$$

(iii) *Siano gli eventi $E_1, \dots, E_{n+1}$ tali che $\Pr(\bigwedge_{i=1}^n E_i) > 0$. Riesce allora:*

$$\Pr(\bigwedge_{i=1}^{n+1} E_i) = \Pr(E_1) \prod_{i=1}^{n} \Pr(E_{i+1} \mid \bigwedge_{j=1}^{i} E_j) \quad \text{(formula di fattorizzazione)};$$

(iv) *Siano E, H due eventi non trascurabili. Riesce allora:*

$$\Pr(E \mid H) = \frac{\Pr(H \mid E)}{\Pr(H)} \Pr(E) \qquad \text{(formula di Bayes)};$$

(v) *Siano $H, K \in \mathcal{A}$ tali che $\Pr(H \wedge K) > 0$. Riesce allora:*

$$\Pr(E \wedge H \mid K) = \Pr(H \mid K) \Pr(E \mid H \wedge K);$$

(vi) CONGLOMERABILITÀ: *Siano $H_1, \dots, H_n$ eventi non trascurabili a due a due incompatibili. Risulta allora:*

$$\min_{i=1,\dots n} \Pr(E \mid H_i) \leq \Pr(E \mid \bigvee_{i=1}^{n} H_i) \leq \max_{i=1,\dots n} \Pr(E \mid H_i).$$

DIMOSTRAZIONE (i) Segue dai teoremi 1.10.2(iv),(v),(vi) e 2.1.2(ii),(ix),(v).

(ii) Per l'additività (Teorema 2.1.2(iii)) si ha

$$\Pr(E) = \Pr(E \wedge \bigvee_{i=1}^{n} H_i) = \Pr(\bigvee_{i=1}^{n} (E \wedge H_i)) = \sum_{i=1}^{n} \Pr(E \wedge H_i).$$

Ora, poichè $E \wedge H_i \to H_i$ (Teorema 1.7.2(vi)), dalla monotonia (Teorema 2.1.2(v)) otteniamo $\Pr(E \wedge H_i) \leq \Pr(H_i)$ per ogni i. Risultano pertanto nulli

[7] Se la probabilità è numerabilmente additiva, possiamo considerare una partizione numerabile e sostituire la somma con la serie (la dimostrazione è del tutto analoga a quella relativa al caso finito).

tutti gli addendi della somma precedente relativi ad eventi H_i trascurabili. Allora

$$\Pr(E) = \sum_{i:\Pr(H_i)>0} \Pr(E \wedge H_i) = \sum_{i:\Pr(H_i)>0} \Pr(E \mid H_i) \Pr(H_i).$$

(iii) Osserviamo innanzitutto che dalla $\Pr(\bigwedge_{i=1}^{n} E_i) > 0$ segue, per monotonia, $\Pr(\bigwedge_{i=1}^{k} E_i) > 0$ per ogni $k \leq n$, notato che, per il Teorema 1.7.2(vi), (iv), $\bigwedge_{i=1}^{n} E_i \to \bigwedge_{i=1}^{k} E_i$. Ora, poichè per $n = 1$ la formula coincide con la definizione di probabilità condizionata, procediamo per induzione assumendo che sussista per n e provandola per $n + 1$. Risulta

$$\begin{aligned}
\Pr(\bigwedge_{i=1}^{n+1} E_i) &= \Pr(E_{n+1} \wedge \bigwedge_{i=1}^{n} E_i) = \Pr(E_{n+1} \mid \bigwedge_{i=1}^{n} E_i) \Pr(\bigwedge_{i=1}^{n} E_i) \\
&= \Pr(E_{n+1} \mid \bigwedge_{i=1}^{n} E_i) \left[\Pr(E_1) \prod_{i=1}^{n-1} \Pr(E_{i+1} \mid \bigwedge_{j=1}^{i} E_j) \right]
\end{aligned}$$

e quindi la tesi.

(iv) Segue dalla $\Pr(E \mid H) \Pr(H) = \Pr(E \wedge H) = \Pr(H \mid E) \Pr(E)$.

(v) Da (iii) risulta $\Pr(E \wedge H \wedge K) = \Pr(E \mid H \wedge K) \Pr(H \mid K) \Pr(K)$ e quindi

$$\Pr(E \wedge H \mid K) = \frac{\Pr(E \wedge H \wedge K)}{\Pr(K)} = \Pr(E \mid H \wedge K) \Pr(H \mid K).$$

(vi) Poichè gli eventi $H_1, \ldots, H_n$ hanno probabilità positiva, possiamo considerare, per la monotonia, la probabilità condizionata alla loro disgiunzione. Posto $m = \min_{i=1,\ldots n} \Pr(E \mid H_i)$ e $M = \max_{i=1,\ldots n} \Pr(E \mid H_i)$, otteniamo, tramite l'additività,

$$\begin{aligned}
\Pr(E \wedge \bigvee_{i=1}^{n} H_i) &= \Pr(\bigvee_{i=1}^{n} (E \wedge H_i)) = \sum_{i=1}^{n} \Pr(E \wedge H_i) \\
&= \sum_{i=1}^{n} \Pr(E \mid H_i) \Pr(H_i) \geq m \sum_{i=1}^{n} \Pr(H_i) = m \Pr(\bigvee_{i=1}^{n} H_i)
\end{aligned}$$

e, in modo analogo,

$$\Pr(E \wedge \bigvee_{i=1}^{n} H_i) \leq M \Pr(\bigvee_{i=1}^{n} H_i).$$

Ne segue la tesi. $\square$

2.4 Probabilità nel discreto

In questa sezione affrontiamo il problema della costruzione di probabilità sulla famiglia degli eventi logicamente dipendenti da una partizione **discreta** (cioè finita o numerabile) dell'evento certo. Il prossimo teorema assicura che per individuare una tale probabilità basta considerare una funzione di dominio la partizione e a valori non negativi di somma uno e prolungarla per additività numerabile sulla σ-algebra degli eventi logicamente dipendenti dalla partizione[8].

Teorema 2.4.1 *Sia* $\mathcal{P} = (\omega_i)_{i\in I}$ *una partizione discreta dell'evento certo. Sia inoltre la sequenza numerica* $(p_i)_{i\in I}$ *tale che* $p_i \geq 0$ *per ogni* $i \in I$ *e* $\sum_{i\in I} p_i = 1$[9]. *Sia infine*

$$\Pr(E) = \sum_{i\,:\,\omega_i \to E} p_i$$

per ogni $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. *Allora,* Pr *è una probabilità numerabilmente additiva sulla* σ*-algebra* $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$.

Dimostrazione Risulta $\Pr(\Omega) = \sum_{i\,:\,\omega_i\to\Omega} p_i = \sum_{i\in I} p_i = 1$. Sia ora $(E_n)_{n\geq 1}$ una successione in $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ formata da eventi a due a due incompatibili. Allora, per il teorema di caratterizzazione 1.8.3(i), $E_n = \bigvee_{i\,:\,\omega_i\to E_n} \omega_i$ per ogni n. Inoltre, se $n \neq m$, per il teorema di rappresentazione 1.8.5(iv), risulta

$$E_n \vee E_m = \bigvee_{i\,:\,\omega_i \to E_n \vee E_m} \omega_i = \bigvee_{i \in \{i:\omega_i\to E_n\}\cup\{i:\omega_i\to E_m\}} \omega_i$$

e, dalla $E_n \wedge E_m = \emptyset$ si ha, tramite il Teorema 1.8.5(iii),(ii),

$$\{i\,:\,\omega_i \to E_n\} \cap \{i\,:\,\omega_i \to E_m\} = \mathrm{set}(E_n) \cap \mathrm{set}(E_m) = \mathrm{set}(E_n \wedge E_m) = \emptyset.$$

Conseguentemente, sempre per il Teorema 1.8.5(iv), si ha[10]

$$\Pr(\bigvee_{n\geq 1} E_n) = \sum_{i\,:\,\omega_i \to \bigvee_{n\geq 1} E_n} p_i = \sum_{i \in \bigcup_{n\geq 1}\{i:\omega_i\to E_n\}} p_i = \sum_{n\geq 1} \sum_{i:\omega_i\to E_n} p_i$$

[8]Osserviamo che, nel caso finito, le nozioni di σ-algebra e di probabilità numerabilmente additiva coincidono, rispettivamente, con quelle di algebra e di probabilità. Rileviamo inoltre che i due teoremi che considereremo forniscono (relativamente al loro ambito) una caratterizzazione delle probabilità nel caso discreto (come facilmente si constata ricorrendo alla numerabile additività (per il primo) e alla formula di fattorizzazione (per il secondo)).

[9]Nel caso che I sia numerabile, la somma $\sum_{i\in I} p_i$ diviene una serie.

[10]Ricordando, nel caso numerabile, che le serie a termini positivi verificano le proprietà di associatività e permutabilità.

e quindi $\Pr(\bigvee_{n\geq 1} E_n) = \sum_{n\geq 1} \Pr(E_n)$. $\square$

Osservazione 2.4.2 (i) Ogni sequenza $(p_i)_{i\in I}$, che verifica le ipotesi del teorema precedente, verrà chiamata **distribuzione** (**di probabilità**) su $\mathcal{P}$ e, in particolare, distribuzione (di probabilità) **uniforme**, se è una sequenza costante.

(ii) Caso simmetrico Sia $\mathcal{P} = \{\omega_1, \dots, \omega_n\}$. Posto allora $p_i = p$ $(i = 1, \dots, n)$, otteniamo $1 = p_1 + \cdots + p_n = p\, n$ e quindi $p = \frac{1}{n}$. Ne segue,

$$\Pr(E) = \sum_{i\,:\,\omega_i \to E} p_i = \frac{\#\,\mathrm{set}(E)}{n} \tag{2.2}$$

e quindi, chiamato **caso favorevole a** E ogni caso elementare che lo implica, otteniamo la celeberrima formula:

$$\Pr(E) = \frac{\text{numero dei casi favorevoli a } E}{\text{numero dei casi possibili (giudicati ugualmente possibili)}}$$

che sta alla base dell'impostazione classica della probabilità[11].

(iii) Sia $\mathcal{P} = \{\omega_1, \omega_2, \dots\}$ numerabile. Allora, la convergenza della serie $\sum_{n\geq 1} p_n$ a uno obbliga ad attribuire *quasi tutta* la probabilità - a meno di un $\epsilon > 0$ arbitrario - a un *numero finito* di casi elementari e riservare, di conseguenza, probabilità *quasi nulla* - non più di ϵ - per gli *infiniti* casi elementari rimanenti. Pertanto, l'accettazione dell'additività numerabile comporta distribuzioni di probabilità decisamente sbilanciate e quindi l'impossibilità ad esprimere un giudizio di simmetria sui casi elementari[12]. $\triangle$

[11] In *Essai philosophique sur les probabilités* (1814), Pierre Simon de Laplace scrive (riguardo ai principi generali del calcolo delle probabilità):

> La teoria dei casi consiste nel ridurre tutti gli avvenimenti della stessa specie a un certo numero di casi ugualmente possibili, tali cioè da renderci ugualmente indecisi sulla loro esistenza, e nel determinare il numero di casi favorevoli all'avvenimento di cui si ricerca la probabilità. Il rapporto di tale numero con quello di tutti i casi possibili ci dà la misura di questa probabilità che non è altro che la frazione avente per numeratore il numero dei casi favorevoli e per denominatore il numero di tutti i casi possibili.

Qualora i casi elementari non siano ugualmente possibili, Laplace afferma (nel 2° Principio) che "si dovranno determinare prima le loro rispettive possibilità, la cui corretta valutazione è uno dei punti più delicati della teoria dei casi. Allora, la probabilità sarà la somma delle possibilità di ogni caso favorevole.".

[12] Giudizio che, in alcune applicazioni (ad esempio, nella teoria dei numeri, nella teoria delle decisioni, nella teoria dei giochi, ...) è talvolta naturale richiedere (abbandonando quindi l'additività numerabile e assumendo che tutti i casi elementari siano trascurabili).

Passando ad un'altra metodologia di costruzione di probabilità su partizioni finite dell'evento certo[13], introduciamo, date le partizioni finite $\mathcal{P}_1, \dots,$ $\mathcal{P}_{n+1}$, la **partizione prodotto** (delle $\mathcal{P}_1, \dots, \mathcal{P}_{n+1}$):

$$\mathcal{P}^{(h)} = \{\omega^{(h)} = \omega_1 \wedge \dots \wedge \omega_h \neq \emptyset : \omega_i \in \mathcal{P}_i \, (i = 1, \dots, h)\}$$

per ogni $h = 1, \dots, n+1$. Chiaramente, $\mathcal{P}^{(1)} = \mathcal{P}_1$ e $\omega^{(1)} = \omega_1$.

Consideriamo ora su $\mathcal{P}_1$ una distribuzione (di probabilità) $(p_{\omega_1})_{\omega_1 \in \mathcal{P}_1}$; su $\mathcal{P}_2$, per ogni $\omega_1 \in \mathcal{P}_1$, una distribuzione $p^{(\omega_1)}$ tale che $p^{(\omega_1)}_{\omega_2} = 0$, se $\omega_1 \wedge \omega_2 = \emptyset$; su $\mathcal{P}_3$, per ogni $\omega^{(2)} \in \mathcal{P}^{(2)}$, una distribuzione $p^{(\omega^{(2)})}$ tale che $p^{(\omega^{(2)})}_{\omega_3} = 0$, se $\omega^{(2)} \wedge \omega_3 = \emptyset$; ... ; su $\mathcal{P}_{n+1}$, per ogni $\omega^{(n)} \in \mathcal{P}^{(n)}$, una distribuzione $p^{(\omega^{(n)})}$ tale che $p^{(\omega^{(n)})}_{\omega_{n+1}} = 0$, se $\omega^{(n)} \wedge \omega_{n+1} = \emptyset$.

Prendendo spunto dalla formula di fattorizzazione delle probabilità condizionate (Teorema 2.3.2(iii)), definiamo:

$$p_{\omega^{(h+1)}} = p_{\omega^{(1)}} \prod_{i=1}^{h} p^{(\omega^{(i)})}_{\omega_{i+1}} = p_{\omega^{(h)}} \, p^{(\omega^{(h)})}_{\omega_{h+1}} \tag{2.3}$$

per ogni $\omega^{(h)} \in \mathcal{P}^{(h)}$ $(h = 1, \dots, n)$.

Il teorema seguente introduce una probabilità sull'algebra degli eventi logicamente dipendenti dalla partizione prodotto $\mathcal{P}^{(n+1)}$, consentendo di interpretare $p^{(\omega^{(h)})}$ come la distribuzione su $\mathcal{P}_{h+1}$ condizionata al caso elementare (non trascurabile) $\omega^{(h)}$.

Teorema 2.4.3 *Posto:*

$$\Pr(E) = \sum_{\omega^{(n+1)} \to E} p_{\omega^{(n+1)}}$$

per ogni $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P}^{(n+1)})$, la funzione Pr *è una probabilità su $\mathcal{E}_{\mathrm{L}}(\mathcal{P}^{(n+1)})$ tale che $p_{\omega^{(h)}} = \Pr(\omega^{(h)})$ $(h = 1, \dots, n+1)$ e $p^{(\omega^{(h)})}_{\omega_{h+1}} = \Pr(\omega_{h+1} \,|\, \omega^{(h)})$ per ogni $\omega^{(h)} \in \mathcal{P}^{(h)}$ non trascurabile $(h = 1, \dots, n)$.*

Dimostrazione Proviamo intanto che Pr è una probabilità. A tal fine, per il Teorema 2.4.1, basta verificare che

$$\sum_{\omega^{(n+1)} \in \mathcal{P}^{(n+1)}} p_{\omega^{(n+1)}} = 1.$$

[13]Metodologia che rimane valida anche nel caso di partizioni numerabili. Ci limitiamo qui al caso finito perchè interverrà nell'analisi dei modelli di estrazione da urne.

Per $n = 1$ otteniamo

$$\begin{aligned}\sum_{\omega^{(2)}\in\mathcal{P}^{(2)}} p_{\omega^{(2)}} &= \sum_{\omega_1\wedge\omega_2\in\mathcal{P}^{(2)}} p_{\omega_1}\, p_{\omega_2}^{(\omega_1)} = \sum_{\omega_1\in\mathcal{P}_1}\Bigg[p_{\omega_1} \sum_{\substack{\omega_2\in\mathcal{P}_2\\ \omega_1\wedge\omega_2\neq\emptyset}} p_{\omega_2}^{(\omega_1)}\Bigg]\\ &= \sum_{\omega_1\in\mathcal{P}_1}\Bigg[p_{\omega_1}\sum_{\omega_2\in\mathcal{P}_2} p_{\omega_2}^{(\omega_1)}\Bigg] = \sum_{\omega_1\in\mathcal{P}_1} p_{\omega_1} = 1.\end{aligned}$$

Assumiamo ora che la formula sussista per $n \geq 2$ e proviamola per $n+1$. Da (2.3) si ha

$$\begin{aligned}\sum_{\omega^{(n+1)}\in\mathcal{P}^{(n+1)}} p_{\omega^{(n+1)}} &= \sum_{\omega_1\wedge\cdots\wedge\,\omega_{n+1}\in\mathcal{P}^{(n+1)}} p_{\omega^{(n)}} p_{\omega_{n+1}}^{(\omega^{(n)})}\\ &= \sum_{\omega^{(n)}\in\mathcal{P}^{(n)}}\Bigg[p_{\omega^{(n)}} \sum_{\substack{\omega_{n+1}\in\mathcal{P}^{(n+1)}\\ \omega^{(n)}\wedge\omega_{n+1}\neq\emptyset}} p_{\omega_{n+1}}^{(\omega^{(n)})}\Bigg]\\ &= \sum_{\omega^{(n)}\in\mathcal{P}^{(n)}}\Bigg[p_{\omega^{(n)}} \sum_{\omega_{n+1}\in\mathcal{P}^{(n+1)}} p_{\omega_{n+1}}^{(\omega^{(n)})}\Bigg] = \sum_{\omega^{(n)}\in\mathcal{P}^{(n)}} p_{\omega^{(n)}} = 1.\end{aligned}$$

Passando alla seconda parte della tesi procediamo, notato che l'uguaglianza sussiste per $h = n + 1$, per induzione a ritroso supponendo che sussista per h e provandola per $h - 1$. Dato $\omega^{(h)}$, tramite (2.3), risulta

$$\begin{aligned}\Pr(\omega^{(h-1)}) &= \Pr(\omega^{(h-1)} \wedge \bigvee_{\omega_h\in\mathcal{P}_h}) = \sum_{\omega_h\in\mathcal{P}_h} \Pr(\omega^{(h-1)} \wedge \omega_h)\\ &= \sum_{\substack{\omega_h\in\mathcal{P}_h\\ \omega^{(h-1)}\wedge\,\omega_h\neq\emptyset}} \Pr(\omega^{(h-1)} \wedge \omega_h) = \sum_{\substack{\omega_h\in\mathcal{P}_h\\ \omega^{(h-1)}\wedge\,\omega_h\neq\emptyset}} p_{\omega^{(h-1)}\wedge\,\omega_h}\\ &= \sum_{\omega_h\in\mathcal{P}_h} p_{\omega^{(h-1)}} p_{\omega_h}^{(\omega^{(h-1)})} = p_{\omega^{(h-1)}} \sum_{\omega_h\in\mathcal{P}_h} p_{\omega_h}^{(\omega^{(h-1)})} = p_{\omega^{(h-1)}}.\end{aligned}$$

Passando alla parte conclusiva della tesi, sia $\omega^{(h+1)}$ tale che $\Pr(\omega^{(h)}) > 0$. Allora, per la seconda parte della tesi e (2.3), otteniamo

$$\Pr(\omega_{h+1} \wedge \omega^{(h)}) = \Pr(\omega^{(h+1)}) = p_{\omega^{(h+1)}} = p_{\omega^{(h)}} p_{\omega_{h+1}}^{(\omega^{(h)})} = \Pr(\omega^{(h)})\, p_{\omega_{h+1}}^{(\omega^{(h)})}$$

e quindi

$$p_{\omega_{h+1}}^{(\omega^{(h)})} = \frac{\Pr(\omega_{h+1} \wedge \omega^{(h)})}{\Pr(\omega^{(h)})} = \Pr(\omega_{h+1} \,|\, \omega^{(h)}).$$

La dimostrazione è così conclusa. $\square$

2.5 Correlazione e indipendenza stocastica

Per valutare quantitativamente l'influenza di un incremento d'informazione (reale o ipotetico) dovuto ad un evento non trascurabile H, viene naturale confrontare la probabilità iniziale $\Pr(\cdot)$ con la corrispondente probabilità finale $\Pr(\cdot \mid H)$. In questo ordine di idee, diremo che l'evento E:
- è **correlato negativamente con** H, se $\Pr(E \mid H) < \Pr(E)$;
- è **correlato positivamente con** H, se $\Pr(E \mid H) > \Pr(E)$;
- **non è correlato con** H, se $\Pr(E \mid H) = \Pr(E)$.

Quindi, da un punto di vista interpretativo, nel primo caso l'incremento di informazione fa diminuire la fiducia sul verificarsi dell'evento, nel secondo la fa aumentare mentre nel terzo la lascia inalterata.

Per il Teorema 2.3.2(i), gli eventi che sono quasi certi o trascurabili non sono correlati con ogni evento non trascurabile; inoltre, ogni evento non è correlato con gli eventi quasi certi. Passando all'evento negato, se E è correlato positivamente (negativamente) o non correlato con H, allora $\overline{E}$ è correlato negativamente (positivamente) o non correlato con H; risulta infatti

$$\Pr(E \mid H) \substack{>\\=\\<} \Pr(E) \Leftrightarrow 1 - \Pr(\overline{E} \mid H) \substack{>\\=\\<} 1 - \Pr(\overline{E}) \Leftrightarrow \Pr(\overline{E} \mid H) \substack{<\\=\\>} \Pr(\overline{E}).$$

Le nozioni di correlazione e di non correlazione sono evidentemente asimmetriche, in quanto viene stabilito quale, dei due eventi in esame, funge da evento e quale da incremento d'informazione. Però, se anche l'evento E è non trascurabile, allora divengono simmetriche. Infatti, in questo caso, per la formula di Bayes (Teorema 2.3.2(iv)), risulta

$$\frac{\Pr(E \mid H)}{\Pr(E)} = \frac{\Pr(H \mid E)}{\Pr(H)}$$

e quindi:
- se il rapporto è 1, allora ognuno degli eventi non è correlato con l'altro;
- se il rapporto è minore (maggiore) di 1, allora gli eventi sono reciprocamente correlati negativamente (positivamente) nella stessa misura (percentuale).

Conviene a questo punto introdurre la terminologia seguente: dati gli eventi $E_1, \ldots, E_n$, diremo che la probabilità **si fattorizza su** $E_1 \wedge \cdots \wedge E_n$ se la relativa probabilità è il prodotto delle probabilità delle componenti, cioè se $\Pr(E_1 \wedge \cdots \wedge E_n) = \Pr(E_1) \cdots \Pr(E_n)$.

Teorema 2.5.1 *Sia E non correlato con H. Allora, la probabilità si fattorizza sui costituenti della partizione generata dagli eventi E e H.*

DIMOSTRAZIONE Risulta $\Pr(E \wedge H) = \Pr(E \mid H)\Pr(H) = \Pr(E)\Pr(H)$. Dalla $E = (E \wedge H) \vee (E \wedge \overline{H})$ otteniamo allora, tramite l'additività,

$$\Pr(E) = \Pr(E \wedge H) + \Pr(E \wedge \overline{H}) = \Pr(E)\Pr(H) + \Pr(E \wedge \overline{H})$$

e quindi $\Pr(E \wedge \overline{H}) = \Pr(E)(1 - \Pr(H)) = \Pr(E)\Pr(\overline{H})$. A questo punto $\Pr(\overline{E} \wedge H) = \Pr(\overline{E})\Pr(H)$ si ottiene con lo stesso procedimento scambiando i ruoli di E e H. Infine, la fattorizzazione su $\overline{E} \wedge \overline{H}$ si consegue in modo analogo a partire da $\overline{E}$. □

Se gli eventi E, H non hanno *probabilità estreme* (cioè 0 o 1), possiamo considerare, per il Teorema 2.1.2(i), anche gli eventi condizionati $E \mid \overline{H}$ e $\overline{E} \mid \overline{H}$. Ora, per ogni costituente $E' \wedge H' = \emptyset$ della partizione generata $\mathcal{P}_{\mathrm{G}}(E, H)$, risulta $\Pr(E' \mid H') = \Pr(E' \wedge H' \mid H') = 0$ (Teorema 1.10.2(iv)) e quindi E è correlato negativamente (positivamente) con H', a seconda che $E' = E$ o $E' = \overline{E}$. Pertanto, la presenza di costituenti impossibili forza la probabilità di uno dei due eventi a mutare a fronte di un'informazione sulla estensione dell'altro. Dunque, se vogliamo introdurre una nozione di indipendenza tra gli eventi E, H che traduca l'idea che la probabilità di uno qualsiasi di essi non venga influenzata da una qualsiasi informazione sulla estensione dell'altro, dobbiamo *necessariamente* supporre che tutti i costituenti della partizione generata siano casi elementari, cioè che i due eventi siano logicamente indipendenti (Teorema 1.9.7).

Queste considerazioni suggeriscono quindi di chiamare gli eventi $E_1, \dots, E_n$ $(n \geq 2)$ **stocasticamente indipendenti**[14] se *sono logicamente indipendenti* e ciascuno di essi non è correlato con i casi elementari non trascurabili della partizione generata dagli altri, cioè se

$$\Pr(E_i \mid \bigwedge_{j \neq i} E'_j) = \Pr(E_i)$$

per ogni caso elementare non trascurabile $\bigwedge_{j \neq i} E'_j$ della partizione generata $\mathcal{P}_{\mathrm{G}}(E'_1, \dots, E'_{i-1}, E'_{i+1} \dots, E'_n)$ $(i = 1, \dots, n)$.

Ovviamente, per il Teorema 2.1.2(i), l'indipendenza stocastica degli eventi $E_1, \dots, E_n$ è equivalente all'indipendenza stocastica degli eventi $E_{1f(1)}, \dots, E_{nf(n)}$[15], osservato che $\mathcal{P}_{\mathrm{G}}(E_1, \dots, E_n) = \mathcal{P}_{\mathrm{G}}(E_{1f(1)}, \dots, E_{nf(n)})$.

[14] L'attributo stocastico, dal greco "stokastikòs" (congetturale), significa "nel senso del calcolo delle probabilità".

[15] Ove, come al solito, $f \in \{0, 1\}^I$ e $E_{i0} = \overline{E}_i$, $E_{i1} = E_i$ per ogni $i \in I$.

Il risultato seguente assicura la conservazione "in discesa" dell'indipendenza stocastica.

Teorema 2.5.2 *Siano gli eventi $E_1, \ldots, E_n$ stocasticamente indipendenti. Inoltre, sia non trascurabile ogni costituente della partizione generata $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$. Allora, gli eventi di un sottoinsieme di $\{E_1, \ldots, E_n\}$ con almeno due elementi sono stocasticamente indipendenti.*

DIMOSTRAZIONE (i) Sia $\mathcal{E} \subset \{E_1, \ldots, E_n\}$ con almeno due elementi. Dal Teorema 1.9.8 segue che gli eventi di $\mathcal{E}$ sono logicamente indipendenti. Scelto $E_{i'} \in \mathcal{E}$, denotiamo con $\mathcal{P}'$ la partizione generata dalla famiglia $\mathcal{E} \setminus \{E_{i'}\}$ e con $\mathcal{P}''$ la partizione generata dalla famiglia $\{E_1, \ldots, E_n\} \setminus (\mathcal{E} \cup \{E_{i'}\})$. Indicati infine i costituenti delle due partizioni, rispettivamente, con ω' e ω'', risulta

$$\omega' = \omega' \wedge \Omega = \omega' \wedge \bigvee_{\omega'' \in \mathcal{P}''} \omega'' = \bigvee_{\omega'' \in \mathcal{P}''} (\omega' \wedge \omega''). \tag{2.4}$$

Osservato che ω' e ogni $\omega' \wedge \omega''$ hanno, per monotonia, probabilità positiva (ogni costituente di $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$ non è trascurabile!) possiamo considerare la probabilità condizionata di $E_{i'}$ a ω' e a $\omega' \wedge \omega''$ per ogni $\omega'' \in \mathcal{P}''$. Risulta allora, per l'indipendenza stocastica, $\Pr(E_{i'} \,|\, \omega' \wedge \omega'') = \Pr(E_{i'})$ per ogni $\omega'' \in \mathcal{P}''$ (poichè $\omega' \wedge \omega'' \in \mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{i'-1}, E_{i'+1}, \ldots, E_n)$). Ne segue, per (2.4) e la conglomerabilità (Teorema 2.3.2(vi), $\Pr(E_{i'} \,|\, \omega') = \Pr(E_{i'})$. □

Il teorema seguente mette in luce il fondamentale legame esistente tra l'indipendenza stocastica e la fattorizzazione della probabilità sulla partizione generata.

Teorema 2.5.3 *Siano gli eventi $E_1, \ldots, E_{n+1}$ logicamente indipendenti. Sussitono allora le proposizioni seguenti:*

(i) Se gli eventi sono stocasticamente indipendenti e i costituenti della partizione generata $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{n+1})$ sono non trascurabili, allora la probabilità si fattorizza sui costituenti;

(ii) Se la probabilità si fattorizza sui costituenti della partizione generata, allora gli eventi sono stocasticamente indipendenti.

DIMOSTRAZIONE (i) Considerato il costituente $E'_1 \wedge \cdots \wedge E'_{n+1}$, dalla formula di fattorizzazione (Teorema 2.3.2(iii)) e dal Teorema 2.5.2 otteniamo

$$\Pr(\bigwedge_{i=1}^{n+1} E'_i) = \Pr(E'_1) \prod_{i=1}^{n} \Pr(E'_{i+1} \,|\, \bigwedge_{j=1}^{i} E'_j) = Pr(E'_1) \prod_{i=1}^{n} \Pr(E'_{i+1}) = \prod_{i=1}^{n+1} \Pr(E'_i).$$

(ii) Fissato i, sia $H = \bigwedge_{j\neq i} E'_j$ tale che $\Pr(H) > 0$. Risulta allora

$$\begin{aligned}
\Pr(H) &= \Pr(E_i \wedge H) + \Pr(\overline{E}_i \wedge H) \\
&= \Pr(E_i) \prod_{j\neq i} \Pr(E'_j) + \Pr(\overline{E}_i) \prod_{j\neq i} \Pr(E'_j) \\
&= [\Pr(E_i) + \Pr(\overline{E}_i)] \prod_{j\neq i} \Pr(E'_j) = \prod_{j\neq i} \Pr(E'_j)
\end{aligned}$$

e quindi

$$\Pr(E_i \,|\, H) = \frac{\Pr(E_i \wedge H)}{\Pr(H)} = \frac{\Pr(E_i) \prod_{j\neq i} \Pr(E'_i)}{\prod_{j\neq i} \Pr(E'_j)} = \Pr(E_i).$$

Osservato infine che gli eventi $E_1, \dots, E_{n+1}$ sono logicamente indipendenti (teoremi 1.9.6, 2.1.2(ii)), ne segue la tesi. □

Concludiamo mostrando, tramite un esempio famoso, che l'indipendenza stocastica non si conserva "in salita".

Esempio 2.5.4 Urna di Bernstein Un'urna contiene quattro palline numerate da 1 a 4. Con riferimento ad una estrazione dall'urna, consideriamo i seguenti tre eventi:

E_i : la pallina estratta ha impresso il numero i o il numero 4 $\quad (i = 1, 2, 3)$.

Poichè, come è facile constatare, i costituenti della relativa partizione generata:

ω_i : viene estratta la pallina con impresso il numero i $\quad (i = 1, \dots, 4)$

sono tutti possibili, possiamo concludere intanto che i tre eventi sono logicamente indipendenti. Osservato poi che $E_1 = \omega_1 \vee \omega_4$, $E_2 = \omega_2 \vee \omega_4$, $E_3 = \omega_3 \vee \omega_4$ e $E_i \wedge E_j = \omega_4$ $(i \neq j)$, otteniamo, per il Teorema 1.10.2(v), $E_3 \,|\, (E_1 \wedge E_2) = \Omega \,|\, (E_1 \wedge E_2)$.

Adottando ora, come appare naturale in questo contesto, lo schema simmetrico, otteniamo $\Pr(E_i) = \frac{1}{2}$ $(i = 1, 2, 3)$ e $\Pr(E_i \wedge E_j) = \frac{1}{4}$ $(i \neq j)$. Conseguentemente, $\Pr(E_3 \,|\, (E_1 \wedge E_2)) = 1 \neq \Pr(E_3)$ e quindi i tre eventi non sono stocasticamente indipendenti. Lo sono però a due a due. Infatti, gli eventi E_i, E_j $(i \neq j)$ sono, per il Teorema 1.9.8, logicamente indipendenti; inoltre, $\Pr(E_i \wedge E_j) = \Pr(\omega_4) = \frac{1}{4} = \Pr(\omega_i) = \Pr(E_i \wedge \overline{E}_j)$ e quindi $\Pr(E_i \,|\, E_j) = \frac{\Pr(E_i \wedge E_j)}{\Pr(E_j)} = \Pr(E_i) = \frac{\Pr(E_i \wedge \overline{E}_j)}{\Pr(\overline{E}_j)} = \Pr(E_i \,|\, \overline{E}_j)$. ♢

2.6 Tre modelli di estrazione

Molti problemi collegati a giochi d'azzardo (dadi, testa e croce, lotto, ...), come pure alcuni problemi relativi alla statistica (distribuzione di un carattere nella popolazione, ...) e all'epidemiologia (contagio tra individui di una popolazione colpita da una epidemia, ...) sono inquadrabili nei tre *modelli di estrazione* da un'urna con due alternative che considereremo in questa sezione: estrazioni senza rimessa, con rimessa e con contagio simmetrico.

Prima di iniziare il loro studio, conviene introdurre una proprietà fondamentale che è verificata da questi tre modelli di estrazione. Chiamiamo gli eventi $E_1, \ldots . E_n$ **scambiabili** se la probabilità di un qualunque costituente $E_{1f(1)}, \ldots, E_{nf(n)} \in \mathcal{P}_{\mathrm{G}}(E_1, \ldots . E_n)$ dipende *solamente* dal numero di eventi affermati, cioè:

$$\Pr(E_{1f(1)} \wedge \cdots \wedge E_{nf(n)}) = \Pr(E_{1g(1)} \wedge \cdots \wedge E_{ng(n)})$$

per ogni $f, g : \{1, \ldots, n\} \to \{0, 1\}$ tali che $\#\{i : g(i) = 1\} = \#\{i : f(i) = 1\}$.

Proviamo ora un importante risultato riguardante gli eventi scambiabili che fornisce, in un certo senso, un principio di riduzione all'origine: la probabilità di un costituente "parziale" formato da m eventi coincide con quella di un costituente "iniziale" formato con il primi m eventi e avente il medesimo numero k di elementi affermati ($0 \leq k \leq m$).

Teorema 2.6.1 *Siano gli eventi $E_1, \ldots . E_n$ scambiabili. Allora, qualunque sia il costituente $E'_{i_1} \wedge \cdots \wedge E'_{i_m} \in \mathcal{P}_{\mathrm{G}}(E_{i_1}, \ldots . E_{i_m})$ $(1 \leq i_1, \ldots, i_m \leq n)$, risulta*

$$\Pr(\underbrace{E'_{i_1} \wedge \cdots \wedge E'_{i_m}}_{k \text{ affermati}}) = \Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_m}_{k \text{ affermati}}).$$

DIMOSTRAZIONE Per ipotesi, tutti i costituenti $E'_1 \wedge \cdots \wedge E'_n$ con k affermazioni hanno la medesima probabilità, che indichiamo con p_k $(k = 0, \ldots, n)$.

Sia ora $E'_{i_1} \wedge \cdots \wedge E'_{i_m}$ un costituente con k affermazioni. Sia inoltre, $m < n$ (altrimenti la tesi è banale). Posto, $h = n - m$ e $\{j_1, \ldots, j_h\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_m\}$, consideriamo la partizione $\mathcal{P} = \mathcal{P}_{\mathrm{G}}(E_{j_1}, \ldots, E_{j_h})$ generata dagli eventi rimanenti. Risulta allora

$$\begin{aligned} E'_{i_1} \wedge \cdots \wedge E'_{i_m} &= (E'_{i_1} \wedge \cdots \wedge E'_{i_m}) \wedge \bigvee_{\omega \in \mathcal{P}} \omega = \bigvee_{\omega \in \mathcal{P}} (E'_{i_1} \wedge \cdots \wedge E'_{i_m} \wedge \omega) \\ &= \bigvee_{i=0}^{h} \left[\bigvee_{\omega \in A(i)} (E'_{i_1} \wedge \cdots \wedge E'_{i_m} \wedge \omega) \right], \end{aligned} \tag{2.5}$$

ove $A(i) = \{\omega \in \mathcal{P} : \omega$ ha i eventi affermati$\}$. Ora, l'insieme $A(i)$ ha $\binom{h}{i}$ costituenti (contando anche quelli eventualmente impossibili). Osservato che i costituenti che compaiono nella disgiunzione in parentesi quadra di (2.5) sono elementi di $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$ con $k + i$ elementi affermati, otteniamo

$$\Pr(E'_{i_1} \wedge \cdots \wedge E'_{i_m}) = \sum_{i=0}^{h} \binom{h}{i} p_{k+i}$$

e quindi la probabilità è indipendente da quali eventi tra gli $E_{i_1}, \ldots, E_{i_m}$ sono affermati. Per l'arbitrarietà del costituente considerato, si ha la tesi. $\square$

Passando ad esaminare i tre modelli di estrazione, supponiamo che da un'urna contenente inizialmente $s \geq 2$ palline colorate si eseguano in sequenza n estrazioni, di una pallina alla volta, con una delle seguenti modalità:
• senza remissione della pallina estratta nell'urna (*estrazioni senza rimessa*);
• con remissione della pallina estratta nell'urna (*estrazioni con rimessa*);
• con remissione della pallina estratta assieme ad un numero $a \geq 1$ di palline del medesimo colore di quella estratta (*estrazioni con contagio simmetrico*).
Supponiamo inoltre che le palline siano di due tipi: bianche con numerosità $b \geq 1$ e rosse con numerosità $r \geq 1$ $(b+r = s)$. Assumiamo infine (come è del tutto naturale in questo contesto) che le palline contenute nell'urna dopo k estrazioni abbiano la medesima probabilità di essere estratte alla $(h+1)$-sima estrazione $(h = 1, \ldots, n-1)$ (condizione di simmetria).

Introdotti gli eventi:

$$E_h : \text{nella } h\text{-sima estrazione viene estratta pallina bianca} \quad (h = 1, \ldots, n),$$

possiamo descrivere gli esiti relativi all'estrazione completa tramite la partizione $\mathcal{S}$ delle **storie** formata dai costituenti $E'_1 \wedge \cdots \wedge E'_n \neq \emptyset$ della partizione generata dagli eventi $E_1, \ldots, E_n$. Chiaramente, nelle estrazioni con rimessa o con contagio, tutti i costituenti sono delle storie; nel caso invece di estrazioni senza rimessa, un costituente avente $k \geq 0$ eventi affermati è una storia se e solo se $n \leq s$ e $k \leq b$, $n - k \leq r$, cioè $\max(0, n-r) \leq k \leq \min(b, n)$.

Tramite (2.2), introduciamo la seguente distribuzione di probabilità sulla partizione $\mathcal{P}_1 = \{E_1, \overline{E}_1\}$:

$$p = p_{E_1} = \frac{b}{s} > 0, \qquad q = p_{\overline{E}_1} = \frac{r}{s} > 0. \tag{2.6}$$

Scegliamo ora un numero relativo $a \geq -1$. Dato $E'_1 \wedge \cdots \wedge E'_h \in \mathcal{P}^{(h)} = \mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_h)^{(ce)}$ $(h < n)$ con k eventi affermati, consideriamo (ricorrendo ancora a (2.2)) la distribuzione di probabilità $p^{(E'_1 \wedge \cdots \wedge E'_h)}$ sulla partizione $\mathcal{P}_{h+1} = \{E_{h+1}, \overline{E}_{h+1}\}$:

$$\begin{aligned} p^{(E'_1 \wedge \cdots \wedge E'_h)}_{E_{h+1}} &= \begin{cases} \frac{b+ak}{s+ah} & \text{se } (E'_1 \wedge \cdots \wedge E'_h) \wedge E_{h+1} \neq \emptyset \\ 0 & \text{altrimenti} \end{cases} \\ p^{(E'_1 \wedge \cdots \wedge E'_h)}_{\overline{E}_{h+1}} &= \begin{cases} \frac{r+a(h-k)}{s+ah} & \text{se } (E'_1 \wedge \cdots \wedge E'_h) \wedge \overline{E}_{h+1} \neq \emptyset \\ 0 & \text{altrimenti} \end{cases} \end{aligned} \tag{2.7}$$

relativa all'estrazione $(h+1)$-sima seguente l'estrazione di k bianche e $h-k$ rosse in una sequenza di h estrazioni con rimessa, se $a = 0$; con contagio simmetrico, se $a \geq 1$; senza rimessa, se $a = -1$. Per quanto riguarda la condizione $(E'_1 \wedge \cdots \wedge E'_h) \wedge E'_{h+1} \neq \emptyset$, osserviamo che risulta sempre verificata nelle estrazioni con rimessa e con contagio simmetrico; nel caso invece di estrazioni senza rimessa, deve essere $h < s$ e $\max(0, h-r \leq k \leq \min(b,h)$ $(E'_1 \wedge \cdots \wedge E'_h \neq \emptyset!)$ e inoltre $k < b$, se $E'_{h+1} = E_{h+1}$, e $h-k < r$, se $E'_{h+1} = \overline{E}_{h+1}$.

Osservato a questo punto che la partizione $\mathcal{S}$ delle storie coincide con la partizione prodotto $\mathcal{P}^{(n)}$ delle partizioni $\mathcal{P}_1, \ldots, \mathcal{P}_n$ (considerata nella seconda parte della Sezione 2.4), possiamo introdurre (tramite la costruzione precedente) sugli eventi logicamente dipendenti dalle storie la probabilità Pr considerata nel Teorema 2.4.3. Il lemma seguente fornisce, in particolare, il valore che essa assume sulle storie aventi lo stesso numero di affermazioni.

Lemma 2.6.2 *Per ogni costituente $E'_1 \wedge \cdots \wedge E'_n$ risulta:*

$$\mathrm{Pr}(\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}}) = \frac{\displaystyle\prod_{u=0}^{k-1}(b+au) \prod_{v=0}^{n-k-1}(r+av)}{\displaystyle\prod_{t=0}^{n-1}(s+at)} > 0$$

per ogni $a \geq 0$ e, se $a = -1$, per ogni $n \leq s$ e k tale che $n - r \leq k \leq b$[16]. Pertanto, nei tre modelli di estrazione considerati, gli eventi $E_1, \ldots, E_n$ sono scambiabili in quanto la probabilità delle storie dipende solo dal numero delle affermazioni presenti e non da dove sono collocate.

[16]Ricordando che si pone $\prod_{i=h}^{m} c_i = 1$, se $h > m$.

DIMOSTRAZIONE Poichè l'uguaglianza sussiste per $n = 1$, assumiamo che sussista per n e proviamola per $n + 1$. Consideriamo dunque un costituente $E'_1 \wedge \cdots \wedge E'_{n+1}$ avente k affermazioni. Allora, $E'_1 \wedge \cdots \wedge E'_n$ avrà $k - 1 \geq 0$ affermazioni, se $E'_{n+1} = E_{n+1}$, e k affermazioni, altrimenti. Inoltre, le condizioni poste nel caso $a = -1$ assicurano che il costituente è una storia nelle tre modalità di estrazione considerate. Conseguentemente, $E'_1 \wedge \cdots \wedge E'_n \neq \emptyset$. Allora (ipotesi induttiva)

$$\Pr(E'_1 \wedge \cdots \wedge E'_n) = \begin{cases} \dfrac{\prod_{u=0}^{k-2}(b+au)\prod_{v=0}^{n-k}(r+av)}{\prod_{t=0}^{n-1}(s+at)} > 0 & \text{se } E'_{n+1} = E_{n+1} \\[2ex] \dfrac{\prod_{u=0}^{k-1}(b+au)\prod_{v=0}^{n-k-1}(r+av)}{\prod_{t=0}^{n-1}(s+at)} > 0 & \text{se } E'_{n+1} = \overline{E}_{n+1} \end{cases}$$

e, per (2.7) e il Teorema 2.4.3,

$$\Pr(E'_{n+1}|E'_1 \wedge \cdots \wedge E'_n) = p^{(E'_1\wedge\cdots\wedge E'_n)}_{E'_{n+1}} = \begin{cases} \frac{b+a(k-1)}{s+an} & \text{se } E'_{n+1} = E_{n+1} \\ \frac{r+a(n-k)}{s+an} & \text{se } E'_{n+1} = \overline{E}_{n+1} \end{cases}.$$

Ne segue la tesi, tenuto conto dell'uguaglianza $\Pr(E'_1\wedge\cdots\wedge E'_{n+1}) = \Pr(E'_{n+1}| E'_1 \wedge \cdots \wedge E'_n) \Pr(E'_1 \wedge \cdots \wedge E'_n)$. □

Chiamato infine **successo** l'estrazione di pallina bianca, rappresentiamo formalmente la **frequenza di successo** (numero globale di palline bianche estratte) tramite la versione del numero aleatorio $S_n = |E_1|+\cdots+|E_n|$ relativa alla partizione delle storie. Conseguentemente, considerato un qualsiasi valore possibile di S_n, l'evento $\{S_n = k\}$ risulta disgiunzione di $\binom{n}{k}$ storie (quelle con le k affermazioni dislocate in qualche modo) e quindi[17]

$$\Pr(S_n = k) = \binom{n}{k} \Pr(\underbrace{E'_1 \wedge \cdots\cdots \wedge E'_n}_{k \text{ affermati}}), \tag{2.8}$$

ove $E'_1 \wedge \cdots\cdots \wedge E'_n$ è una storia qualsiasi avente k eventi affermati.

Passiamo ora ad una analisi più dettagliata dei tre modelli di estrazione considerati.

[17]Adottando la notazione più snella $\Pr(S_n = k)$ al posto della $\Pr(\{S_n = k\})$.

2.6.1 Estrazioni senza rimessa

Deve intanto essere $n \leq s$. Inoltre, per il Lemma 2.6.2 (con $a = -1$), risulta

$$\Pr(\underbrace{E'_1 \wedge \cdots\cdots \wedge E'_n}_{\substack{k \text{ affermati} \\ \max(0,n-r)\leq k\leq\min(b,n)}}) = \frac{\left[\binom{b}{k}k!\right]\left[\binom{r}{n-k}(n-k)!\right]}{\binom{s}{n}n!} = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{s}{n}\binom{n}{k}}$$

da cui, per il Teorema 2.6.1, si ha

$$\Pr(\underbrace{E'_{i_1} \wedge \cdots\cdots \wedge E'_{i_m}}_{\substack{k \text{ affermati} \\ \max(0,m-r)\leq k\leq\min(b,m)}}) = \frac{\binom{b}{k}\binom{r}{m-k}}{\binom{s}{m}\binom{m}{k}} \qquad (1 \leq i_1, \ldots, i_m \leq n). \quad (2.9)$$

Ne segue, tramite (2.8), la **distribuzione ipergeometrica**:

$$\Pr(S_n = k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{s}{n}} \quad (\max(0, n-r) \leq k \leq \min(b, n)). \quad (2.10)$$

Osservato infine che, per (2.9), $\Pr(E_i) = p$ $(i = 1, \ldots n)$ e che, se $i \neq j$, si ha

$$\Pr(E_i \mid E_j) = \frac{\Pr(E_i \wedge E_j)}{\Pr(E_j)} = \frac{\binom{b}{2}}{\binom{s}{2}}\frac{1}{p} = \frac{b-1}{s-1} < p,$$

possiamo concludere che gli eventi $E_1, \ldots, E_n$ sono a due a due correlati negativamente[18].

Osservazione 2.6.3 Estrazioni in blocco Gli eventi $\{S_n = k\}$ descrivono il risultato della sequenza di n estrazioni di una pallina alla volta senza tener conto dell'ordine in cui appaiono le palline. Poichè si può giungere allo stesso risultato estraendo dall'urna in una sola volta le n palline, viene naturale confrontare gli eventi $\{S_n = k\}$ con i corrispondenti eventi:

$E_n^{(k)}$: su n palline estratte dall'urna *tutte in una volta* k sono bianche.

[18]Come suggerito anche dall'intuizione: l'uscita di pallina bianca diminuisce il grado di fiducia di vedere ancora estratta una bianca.

Si tratta evidentemente di eventi *diversi* in quanto sono diverse le modalità di estrazione e con esse i rispettivi stati di conoscenza. Però, se si numerano le palline da 1 a s e si suppone che i casi possibili:

$\omega_{\{i_1,\dots,i_n\}}$: vengono estratte dall'urna *tutte in una volta* le n palline $i_1,\dots,i_n$

abbiano la stessa probabilità di essere estratti (condizione di simmetria), allora i due procedimenti di estrazione comportano la medesima probabilità per gli eventi $\{S_n = k\}$ e $E_n^{(k)}$. Infatti, osservato (con riferimento a (2.10)), che il denominatore fornisce il numero dei casi posssibili e il numeratore quello dei casi favorevoli all'evento $E_n^{(k)}$, possiamo concludere, tramite (2.2), che *la probabilità di vedere k palline bianche in n estrazioni sequenziali coincide con la probabilità di vedere k palline bianche nell'estrazione in blocco di n palline* (in condizioni di simmetria).

2.6.2 Estrazioni con rimessa

Dal Lemma 2.6.2 (con $a = 0$) otteniamo la **distribuzione di Bernoulli**:

$$\Pr(\underbrace{E_1' \wedge \cdots \wedge E_n'}_{k \text{ affermati}}) = p^k q^{n-k}$$

e quindi, per il Teorema 2.6.1, si ha

$$\Pr(\underbrace{E_{i_1}' \wedge \cdots \wedge E_{i_m}'}_{k \text{ affermati}}) = p^k q^{m-k} \quad (1 \le i_1,\dots,i_m \le n). \tag{2.11}$$

Ne segue, tramite (2.8), la **distribuzione binomiale**:

$$\Pr(S_n = k) = \binom{n}{k} p^k q^{n-k} \quad (k = 0,1,\dots,n). \tag{2.12}$$

Osservato che $\Pr(E_1' \wedge \cdots \wedge E_n') = \prod_{i=1}^n \Pr(E_i')$ per ogni $E_1' \wedge \cdots \wedge E_n'$, possiamo concludere, tramite il Teorema 2.5.3, che gli eventi $E_1,\dots,E_n$ sono stocasticamente indipendenti, una volta notato che sono logicamente indipendenti.

Poichè le estrazioni avvengono con rimessa, possiamo considerare un numero arbitrario di estrazioni. Viene allora naturale introdurre il **tempo di attesa del primo successo** (numero di estrazioni che occorre fare per ottenere per la prima volta un successo), rappresentato formalmente dalla versione del numero aleatorio T relativa alla partizione numerabile $\mathcal{T}$:

$$\begin{aligned} \{T=1\} &= E_1 \\ \{T=n\} &= \overline{E}_1 \wedge \cdots \wedge \overline{E}_{n-1} \wedge E_n \quad (n \geq 2) \\ \{T=+\infty\} &= \bigwedge_{n\geq 1} \overline{E}_n. \end{aligned}$$

Ora, posto $p_n = p\,q^{n-1} = \Pr(T=n)$ (per (2.11)), otteniamo

$$\sum_{n\geq 1} pq^{n-1} = p\sum_{n\geq 1} q^{n-1} = p\,\frac{1}{1-q} = 1,$$

osservato che la serie in esame è una serie geometrica di ragione $q \in]0,1[$. Conseguentemente, posto $p_{+\infty} = 0$, possiamo introdurre, per il Teorema 2.4.1, una probabilità numerabilmente additiva su $\mathcal{E}_{\mathrm{L}}(\mathcal{T})$ tramite la **distribuzione geometrica**:

$$\Pr(T=t) = \begin{cases} p\,q^{t-1} & \text{se } t = 1, 2, \ldots \\ 0 & \text{se } t = +\infty \end{cases}. \tag{2.13}$$

Se infine si considera l'evento $\{T > n\} = \bigwedge_{i=1}^{n} \overline{E}_i$ (**ritardo di n estrazioni nell'apparizione del primo successo**), da (2.11) otteniamo

$$\Pr(T>n) = \Pr(\overline{E}_1)^n = \Pr(T>1)^n$$

e quindi

$$\begin{aligned} \Pr(T-n>m\,|\,T>n) &= \frac{\Pr(\{T>n+m\} \wedge \{T>n\})}{\Pr(T>n)} \\ &= \frac{\Pr(T>n+m)}{\Pr(T>n)} = \frac{\Pr(T>1)^{n+m}}{\Pr(T>1)^n} \\ &= \Pr(T>1)^m = \Pr(T>m). \end{aligned}$$

Dunque, la successione dei ritardi $(\{T>n\})_{n\geq 1}$ è una successione di eventi *priva di memoria*: la probabilità del ritardo di m estrazioni nell'apparizione del primo successo dopo un numero qualsiasi di insuccessi è la stessa di quella del ritardo di m estrazioni nell'apparizione del primo successo dopo un insuccesso iniziale.

Osservazione 2.6.4 (i) La mancanza di memoria insita nel modello probabilistico delle estrazioni con rimessa comporta che giocare sui numeri ritardatari nel gioco del lotto è privo di senso[19]: il fatto che un numero non sia uscito per molte estrazioni precedenti non aumenta in alcun modo la probabilità che venga estratto alla successiva estrazione, rimanendo questa probabilità sempre uguale a $\binom{89}{4}\binom{90}{5}^{-1} = \frac{1}{18}$.

(ii) Data una storia E con k eventi affermati, denotiamo con $\Pr^{(sr)}(E)$ e $\Pr^{(cr)}(E)$ le probabilità corrispondenti, rispettivamente, alle estrazioni senza rimessa e a quelle con rimessa. Per il Lemma 2.6.2 si ha

$$\Pr^{(sr)}(E) = \frac{\prod_{u=0}^{k-1}(b-u)\prod_{v=0}^{n-k+1}(r-v)}{\prod_{t=0}^{n-1}(s-t)} = \frac{\prod_{u=0}^{k-1}(p-\frac{u}{s})\prod_{v=0}^{n-k+1}[(1-p)-\frac{v}{s}]}{\prod_{t=0}^{n-1}(1-\frac{t}{s})}$$

ove l'ultima espressione si ottiene dalla precedente dividendo numeratore e denominatore per s^n. Passando quindi al limite per $s \to +\infty$ tenendo costante la composizione dell'urna (percentuale di palline bianche presenti), risulta $\Pr^{(sr)}(E) \to p^k q^{n-k} = \Pr^{(cr)}(E)$, tenuto conto di (2.11). La distribuzione di Bernoulli può dunque essere adoperata, se la numerosità dell'urna è elevata, per approssimare la distribuzione ipergeometrica[20].

2.6.3 Estrazioni con contagio simmetrico

Dal Lemma 2.6.2 (con $a \geq 1$) risulta

$$\Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}}) = \frac{\prod_{u=0}^{k-1}(b+au)\prod_{v=0}^{n-k+1}(r+av)}{\prod_{t=0}^{n-1}(s+at)}$$

[19]Supponendo, come è naturale tenuto conto del meccanismo di sorteggio, che estrazioni successive di cinquine semplici siano equiprobabili e stocasticamente indipendenti.

[20]Come suggerito dall'intuizione: se l'urna ha molte palline, immettere o non immettere la pallina estratta nell'urna non comporta alcuna differenza sostanziale.

da cui, dividendo numeratore e denominatore per $(-a)^n$, otteniamo[21]

$$\begin{aligned}
\Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}}) &= \frac{\prod_{u=0}^{k-1}(-\frac{b}{a}-u)\prod_{v=0}^{n-k+1}(-\frac{r}{a}-v)}{\prod_{t=0}^{n-1}(-\frac{s}{a}-t)} \\
&= \frac{\left[\binom{-\frac{b}{a}}{k}k!\right]\left[\binom{-\frac{r}{a}}{n-k}(n-k)!\right]}{\binom{-\frac{s}{a}}{n}n!} = \frac{\binom{-\frac{b}{a}}{k}\binom{-\frac{r}{a}}{n-k}}{\binom{-\frac{s}{a}}{n}\binom{n}{k}}
\end{aligned}$$

e quindi, per il Teorema 2.6.1, risulta

$$\Pr(\underbrace{E'_{i_1} \wedge \cdots \wedge E'_{i_m}}_{k \text{ affermati}}) = \frac{\binom{-\frac{b}{a}}{k}\binom{-\frac{r}{a}}{m-k}}{\binom{-\frac{s}{a}}{m}\binom{m}{k}} \qquad (1 \leq i_1, \ldots, i_m \leq n). \tag{2.14}$$

Ne segue, tramite (2.8), la **distribuzione di Polya**:

$$\Pr(S_n = k) = \frac{\binom{-\frac{b}{a}}{k}\binom{-\frac{r}{a}}{n-k}}{\binom{-\frac{s}{a}}{n}} \quad (k = 0, 1, \ldots, n). \tag{2.15}$$

Osservato infine che, per (2.14), $\Pr(E_i) = p$ $(i = 1, \ldots, n)$ e che, se $i \neq j$, si ha

$$\Pr(E_i \mid E_j) = \frac{\Pr(E_i \wedge E_j)}{\Pr(E_j)} = \frac{\binom{-\frac{b}{a}}{2}}{\binom{-\frac{s}{a}}{2}}\frac{1}{p} = \frac{b+a}{s+a} > p,$$

possiamo concludere che gli eventi $E_1, \ldots, E_n$ sono a due a due correlati positivamente[22].

[21] Convenendo, come è usuale, di estendere il coefficiente binomiale ad ogni numero reale x ponendo $\binom{x}{m} = \frac{\prod_{i=0}^{m-1}(x-i)}{m!}$.

[22] Come suggerito anche dall'intuizione: l'uscita di pallina bianca aumenta il grado di fiducia di vedere ancora estratta una bianca.

Capitolo 3

Valutazione di variabili aleatorie

Affrontate, nei capitoli precedenti, la descrizione e la valutazione dell'incertezza, tratteremo ora il problema della valutazione di quei particolari numeri aleatori che sono le variabili aleatorie. Come le nozioni di evento e di probabilità erano la chiave di volta, rispettivamente, della descrizione e della valutazione, così la nozione di speranza matematica sarà quella della valutazione delle variabili aleatorie. Per introdurla seguiremo l'impostazione assiomatica ricorrendo, per giustificare gli assiomi, alla sua interpretazione in termini di quozienti di scommesse (come peraltro fatto per la probabilità).

In questo capitolo, $\mathcal{A}$ *denoterà una* σ*-algebra di eventi e* Pr *una probabilità numerabilmente additiva su* $\mathcal{A}$.

3.1 Variabili aleatorie

Per **variabile aleatoria** (in breve v.a.) intendiamo ogni numero aleatorio X tale che $\{X \leq x\} \in \mathcal{A}$ per ogni x reale [1]. Il prossimo lemma (conseguenza im-

[1] Un'esempio semplice di un numero aleatorio che non è una v.a. può ottenersi facendo riferimento al gioco del lotto (Esempio 1.8.1(iii)). Sia $\mathcal{P}$ la partizione delle cinquine secche e X il numero aleatorio "primo numero estratto" che definiamo su $\mathcal{P}$ tramite la versione: $f : E_{(n_1,n_2,n_3,n_4,n_5)} \to n_1$. Supponiamo inoltre che $\mathcal{A}$ sia la famiglia degli eventi logicamente dipendenti dalla partizione $\mathcal{P}'$ delle cinquine semplici. Allora, ad esempio, l'evento $\{X \leq 10\} \notin \mathcal{A}$ in quanto è solo logicamente semidipendente da $\mathcal{P}'$. Chiaramente, la nozione di variabile aleatoria non è assoluta, ma è relativa alla σ-algebra presa in considerazione; infatti, nell'esempio, X sarebbe diventato una v.a. se avessimo preso come $\mathcal{A}$ la famiglia

mediata delle formule (1.3)) fornisce una caratterizzazione delle v.a. tramite gli eventi definiti dalla disuguaglianza stretta.

Lemma 3.1.1 *Il numero aleatorio X è una v.a. se e solo se $\{X < x\} \in \mathcal{A}$ per ogni x reale.*

Passiamo ora ad elencare alcune importanti proprietà delle variabili aleatorie che saranno adoperate nel seguito. Il teorema seguente, conseguenza immediata del Lemma 3.1.1 e di (1.4), assicura che gli eventi relativi a uguaglianze e disuguaglianze tra una v.a. e un numero certo appartengono ad $\mathcal{A}$.

Teorema 3.1.2 *Siano X una v.a. e a, b numeri reali. Allora, appartengono ad $\mathcal{A}$ gli eventi $\{X = a\}$, $\{X \geq a\}$, $\{X > a\}$ e $\{a \lhd X \blacktriangleleft b\}$, ove $\lhd, \blacktriangleleft$ denotano uno qualunque dei simboli $<$, $\leq$.*

Il teorema seguente assicura invece che l'insieme delle v.a. è chiuso sia rispetto alle usuali operazioni aritmetiche che a trasformazioni continue. Precisiamo che nel seguito $\mathbb{Q}$ denota l'insieme numerabile dei numeri razionali.

Teorema 3.1.3 *I numeri certi e gli indicatori degli eventi di $\mathcal{A}$ sono v.a.. Inoltre, se X, Y sono v.a., lo sono pure $X + Y$ e XY. Infine, se τ è una funzione definita e continua sul rango di X, la trasformata $\tau(X)$ è una v.a.; in particolare, sono v.a. X^+, X^-, $\frac{1}{X}$, se $0 \notin \mathrm{rg}(X)$, e αX per ogni α reale.*[2]

DIMOSTRAZIONE Se X è il numero certo di valore c, allora $\{X \leq x\} = \Omega$, se $x \geq c$, e $\{X \leq x\} = \emptyset$, se $x < c$. In ogni caso quindi $\{X \leq x\} \in \mathcal{A}$.

Dato $E \in \mathcal{A}$, sia $X = |E|$. Allora, $\{X \leq x\} = \emptyset$, se $x < 0$, e $\{X \leq x\} = \overline{E}$, se $0 \leq x < 1$, e $\{X \leq x\} = \Omega$, se $x \geq 1$. In ogni caso quindi $\{X \leq x\} \in \mathcal{A}$.

Date le v.a. X, Y, verifichiamo intanto che la somma $X+Y$ è una v.a.. A tal fine, siano f una versione di X relativa alla partizione $\mathcal{P}$, g una versione di Y rispetto alla partizione $\overline{\mathcal{P}}$ e $\mathcal{P}^{(1)} = \{\omega^{(1)} = \omega \wedge \overline{\omega} \neq \emptyset : \omega \in \mathcal{P} \text{ e } \overline{\omega} \in \overline{\mathcal{P}}\}$. Allora, come facilmente si verifica, le funzioni $f_1 : \omega^{(1)} \to f(\omega)$ e $g_1 : \omega^{(1)} \to g(\overline{\omega})$ sono, rispettivamente, versioni di X e Y relative alla partizione $\mathcal{P}^{(1)}$. Ne

---

degli eventi logicamente dipendenti da $\mathcal{P}$.

[2]Ove $X^+ = \max(0, X)$ è la *parte positiva* di X mentre $X^- = \max(0, -X)$ è la *parte negativa* di X. Chiaramente, $X = X^+ - X^-$ e $|X| = X^+ + X^-$.

segue che la funzione che associa ad ogni $\omega^{(1)}$ il valore $f_1(\omega^{(1)}) + g_1(\omega^{(1)}) = f(\omega) + g(\overline{\omega})$ è una versione di $X + Y$ relativa a $\mathcal{P}^{(1)}$. Risulta allora

$$\{X + Y < t\} = \bigvee_{\omega^{(1)} \in \mathcal{P}^{(1)} : f_1(\omega^{(1)}) + g_1(\omega^{(1)}) < t} \omega^{(1)}$$

e quindi, per il Teorema 1.8.5(i), $\mathrm{set}(\{X + Y < t\}) = \{\omega^{(1)} : f_1(\omega^{(1)}) + g_1(\omega^{(1)}) < t\}$. Osservato che, per ogni $\omega^{(1)}$ e per ogni t reale, risulta $f_1(\omega^{(1)}) + g_1(\omega^{(1)}) < t \Leftrightarrow f_1(\omega^{(1)}) < r < t - g_1(\omega^{(1)}) \Leftrightarrow f_1(\omega^{(1)}) < r$ e $g_1(\omega^{(1)}) < t - r$ per qualche razionale r, dal Teorema 1.8.5(i),(iii),(iv) si ha

$$\begin{aligned}
\mathrm{set}(\{X + Y < t\}) &= \bigcup_{r \in \mathcal{Q}} \{\omega^{(1)} : f_1(\omega^{(1)}) < r \text{ e } g_1(\omega^{(1)}) < t - r\} \\
&= \bigcup_{r \in \mathcal{Q}} [\{\omega^{(1)} : f_1(\omega^{(1)}) < r\} \cap \{\omega^{(1)} : g_1(\omega^{(1)}) < t - r\}] \\
&= \bigcup_{r \in \mathcal{Q}} [\mathrm{set}(\{X < r\}) \cap \mathrm{set}(\{Y < t - r\})] \\
&= \bigcup_{r \in \mathcal{Q}} \mathrm{set}(\{X < r\} \wedge \{Y < t - r\}) \\
&= \mathrm{set}(\bigvee_{r \in \mathcal{Q}} [\{X < r\} \wedge \{Y < t - r\}])
\end{aligned}$$

e quindi, per il Lemma 3.1.1, $\{X + Y < t\} = \bigvee_{r \in \mathcal{Q}} [\{X < r\} \wedge \{Y < t - r\}] \in \mathcal{A}$. Ne segue, data l'arbitrarietà di t, tramite il Lemma 3.1.1 che la somma è una v.a..

Prima di verificare che il prodotto è una v.a., proviamo l'ultima parte della tesi. Sia dunque τ una funzione definita e continua sul rango di X. Risulta, qualunque sia t reale,

$$\begin{aligned}
\mathrm{set}(\{\tau(X) < t\}) &= \{\omega^{(1)} : \tau(f_1(\omega^{(1)})) \in]-\infty, t[\,\} \\
&= \{\omega^{(1)} : f_1(\omega^{(1)}) \in \tau^{-1}(]-\infty, t[)\}.
\end{aligned}$$

Ora, essendo τ continua, $W = \tau^{-1}(]-\infty, t[)$ è un insieme aperto (in quanto controimagine di un semiretta aperta) e quindi, per ogni $x \in W$, esistono $r_y, r'_y \in \mathbb{Q}$ tali che $x \in [r_x, r'_x] \subseteq W$. Conseguentemente, $W = \bigcup_{x \in W} [r_x, r'_x]$

da cui otteniamo, tramite il Teorema 1.8.5(iv),

$$\begin{aligned}
\mathrm{set}(\{\tau(X) < t\}) &= \{\omega^{(1)} : f_1(\omega^{(1)}) \in \bigcup_{x\in W} [r_x, r'_x]\} \\
&= \bigcup_{x\in W} \{\omega^{(1)} : f_1(\omega^{(1)}) \in [r_x, r'_x]\} \\
&= \bigcup_{x\in W} \{\omega^{(1)} : r_x \leq f_1(\omega^{(1)}) \leq r'_x\} \\
&= \bigcup_{x\in W} \{r_x \leq X \leq r'_x\} = \mathrm{set}(\bigvee_{x\in W} \{r_x \leq X \leq r'_x\})
\end{aligned}$$

e quindi, per il Teorema 3.1.2, $\{\tau(X) < t\}) = \bigvee_{x\in W}\{r_x \leq X \leq r'_x\} \in \mathcal{A}$, essendo la disgiunzione numerabile. Ne segue, per il Lemma 3.1.1, che $\tau(X)$ è una v.a.. Per constatare che X^+, X^-, $\frac{1}{X}$ e αX sono delle v.a., basta a questo punto osservare che sono trasformate di X tramite, rispettivamente, le funzioni continue $\tau(x) = \max(0, x)$, $\tau(x) = \max(0, -x)$, $\tau(x) = \frac{1}{x}$ $(x \neq 0)$ e $\tau(x) = \alpha\, x$.

Infine, per provare che il prodotto XY è una v.a., basta osservare che $XY = \frac{1}{4}[(X+Y)^2 - (X-Y)^2]$ e che Z^2 è una v.a., qualunque sia la v.a. Z, in quanto trasformata di Z tramite la funzione continua $\tau(x) = x^2$. □

Osservazione 3.1.4 Date le v.a. X e Y, il teorema appena provato assicura che sono elementi di $\mathcal{A}$ gli eventi $\{X = Y\}$, $\{X < Y\}$ e $\{X \leq Y\}$. Infatti, basta notare che $\{X = Y\} = \{X - Y = 0\}$, $\{X < Y\} = \{X - Y < 0\}$ e $\{X \leq Y\} = \{X - Y \leq 0\}$ e tenere presente il Lemma 3.1.1 e il Teorema 3.1.2. △

3.2 Funzioni di ripartizione

Data un v.a. X, chiamiamo **funzione di ripartizione** (in breve f.r.) di X la funzione F_X così definita:

$$F_X(x) = \Pr(X \leq x)$$

per ogni x reale[3]. Inoltre, la v.a. X viene detta **continua** se la funzione di ripartizione F_X è continua.

[3]Come nella nota 17 del capitolo precedente, usiamo le notazioni più snelle $\Pr(X \leq x)$, $\Pr(X > x)$, $\Pr(a < X \leq b)$, etc. al posto delle $\Pr(\{X \leq x\})$, $\Pr(\{X > x\})$, $\Pr(\{a < X \leq b\})$, etc..

Il teorema seguente assicura che la funzione di ripartizione è una funzione non decrescente continua a destra e avente al più un numerabile di salti (discontinuità di 1° specie) corrispondenti ai valori x per i quali è positiva la probabilità che la v.a. assuma la determinazione x. Inoltre, mette in evidenza che la conoscenza della funzione di ripartizione è condizione necessaria e sufficiente per valutare le probabilità di eventi del tipo "X appartiene ad un intervallo"(limitato o no).

Teorema 3.2.1 *Sia X una v.a.. Sussistono allora le proposizioni seguenti:*

(i) $0 \leq F_X \leq 1$;

(ii) $F_X(x) \leq F_X(x')$, *se* $x \leq x'$;

(iii) $F_X(x) = 0$, *se* $x < \inf X$, *e* $F_X(x) = 1$, *se* $x \geq \sup X$ [4];

(iv) $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$;

(v) $F_X(+\infty) = \lim_{x \to +\infty} F_X(x) = 1$;

(vi) $F_X(x^+) = \lim_{t \to x^+} F_X(t) = F_X(x)$;

(vii) $F_X(x^-) = \lim_{t \to x^-} F_X(t) = \Pr(X < x)$;

(viii) $F_X(x^+) - F_X(x^-) = \Pr(X = x)$;

(ix) *Per ogni a, b tali che $a < b$, risulta:*
- $\Pr(X > a) = 1 - F_X(a)$ *e* $\Pr(X \geq a) = 1 - F_X(a^-)$;
- $\Pr(a \leq X < b) = F_X(b^-) - F_X(a^-)$;
- $\Pr(a \leq X \leq b) = F_X(b) - F_X(a^-)$;
- $\Pr(a < X \leq b) = F_X(b) - F_X(a)$;
- $\Pr(a < X < b) = F_X(b^-) - F_X(a)$.

Inoltre, F_X ha al più un numerabile di punti di discontinuità.

DIMOSTRAZIONE (ii) Segue dalla monotonia della probabilità, osservato che, per il Teorema 1.8.5(i),(vi), $\{X \leq x\} \to \{X \leq x'\}$ per ogni $x \leq x'$.

(iii) Risulta $\{X \leq x\} = \emptyset$, se $x < \inf X$, e $\{X \leq x\} = \Omega$, se $x \geq \sup X$.

[4] Ove, $\inf X$ e $\sup X$ sono, nell'ordine, l'estremo inferiore e superiore del rango di X.

Le proposizioni (iv) $\div$ (vii) seguono dai teoremi 1.8.5(i),(iii),(iv),(vi), 1.7.2(vi) e dalla continuità dal basso e dall'alto della probabilità (Teorema 2.2.2(i),(ii))[5]. Precisiamo che denotiamo con f una versione di X relativa alla partizione $\mathcal{P}$.

(iv) Considerata la successione non crescente $(\{X \leq -n\})_{n\geq 1}$, otteniamo

$$\lim_{n\to+\infty} F_X(-n) = \lim_{n\to+\infty} \Pr(X \leq -n) = \Pr(\bigwedge_{n\geq 1}\{X \leq -n\}) = \Pr(\emptyset) = 0,$$

osservato che $\mathrm{set}(\bigwedge_{n\geq 1}\{X \leq -n\}) = \bigcap_{n\geq 1}\{\omega : f(\omega) \leq -n\} = \emptyset$.

(v) Considerata la successione non decrescente $(\{X \leq n\})_{n\geq 1}$, otteniamo

$$\lim_{n\to+\infty} F_X(n) = \lim_{n\to+\infty} \Pr(X \leq n) = \Pr(\bigvee_{n\geq 1}\{X \leq n\}) = \Pr(\Omega) = 1,$$

osservato che $\mathrm{set}(\bigvee_{n\geq 1}\{X \leq n\}) = \bigcup_{n\geq 1}\{\omega : f(\omega) \leq n\} = \mathcal{P}$.

(vi) Considerata la successione non crescente $(\{X \leq x + \frac{1}{n}\})_{n\geq 1}$, da (1.3) otteniamo

$$\lim_{n\to+\infty} F_X(x+\frac{1}{n}) = \lim_{n\to+\infty} \Pr(X \leq x+\frac{1}{n}) = \Pr(\bigwedge_{n\geq 1}\{X \leq x+\frac{1}{n}\}) = \Pr(X \leq x).$$

(vii) Considerata la successione non decrescente $(\{X \leq x - \frac{1}{n}\})_{n\geq 1}$, da (1.3) risulta

$$\lim_{n\to+\infty} F_X(x-\frac{1}{n}) = \lim_{n\to+\infty} \Pr(X \leq x-\frac{1}{n}) = \Pr(\bigvee_{n\geq 1}\{X \leq x-\frac{1}{n}\}) = \Pr(X < x).$$

(viii) Conseguenza immediata di (vi), (vii), osservato che, per (1.4), $\Pr(X \leq x) = \Pr(X < x) + \Pr(X = x)$.

(ix) Conseguenza immediata del Teorema 3.1.2, di (vi), (vii), (viii) e di (1.4).

L'ultima parte della tesi, segue banalmente da (viii) e dal Teorema 2.1.3. □

Nell'osservazione seguente forniamo l'espressione della f.r. di v.a. aventi un numero discreto di valori possibili ordinati per grandezza. Inoltre, formalizziamo l'idea intuitiva di "scelta a caso di un numero reale di un intervallo" tramite la nozione di distribuzione uniforme in un intervallo.

[5] Ricordato che una funzione monotona ammette limite destro (sinistro) in ogni punto di accumulazione destro (sinistro) del dominio. Inoltre che, per il teorema del limite della restrizione, se $\lim_{x\to t} f(x) = \lambda$, allora $\lim_{n\to+\infty} f(x_n) = \lambda$, per ogni successione $x_1, x_2, \ldots$ convergente a t.

Osservazione 3.2.2 (i) Data una v.a. X **semplice** (cioè con rango finito), siano $x_1 < \cdots < x_n$ i suoi valori possibili. Supposto $p_i = \Pr(X = x_i) > 0$ per ogni $i \leq n$, si ha

$$F_X(x) = \begin{cases} 0 & \text{se } x < x_1 \\ \sum_{i=1}^{h} p_i & \text{se } x_h \leq x < x_{h+1}\, (h = 1, \dots, n-1) \\ 1 & \text{se } x \geq x_n \end{cases} .$$

In particolare, se X è un numero certo, risulta $n = 1$ e quindi $p_1 = \Pr(X = x_1) = 1$. Ne segue

$$F_X(x) = \begin{cases} 0 & \text{se } x < x_1 \\ 1 & \text{se } x \geq x_1 \end{cases} .$$

Se $X = |E|$ ($E \in \mathcal{A}$), si ha $x_1 = 0$, $x_2 = 1$ e quindi $p_1 = \Pr(\overline{E})$ e $p_2 = \Pr(E)$. Ne segue

$$F_{|E|}(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - \Pr(E) & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} .$$

Infine, se X ha un numerabile di valori possibili $x_1 < x_2 < \dots$ di probabilità $p_i = \Pr(X = x_i) > 0$ ($i \geq 1$), otteniamo

$$F_X(x) = \begin{cases} 0 & \text{se } x < x_1 \\ \sum_{i=1}^{h} p_i & \text{se } x_h \leq x < x_{h+1}\, (h = 1, 2, \dots) \\ 1 & \text{se } x \geq \sup\{x_1, x_2, \dots\} \end{cases} .$$

In ogni caso quindi la f.r. è una funzione a gradini con salti di ampiezza p_h in corrispondenza dei valori possibili x_h.

(ii) Distribuzione uniforme in $[a, b]$ Una v.a. X è **distribuita uniformemente in** $[a, b]$ ($a < b$) se la sua f.r. assume l'espressione:

$$F_X(x) = \begin{cases} 0 & \text{se } x < 0 \\ \dfrac{x-a}{b-a} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases} .$$

Che questa particolare f.r. sia una possibile formalizzazione dell'idea intuitiva di "scelta a caso di un numero reale nell'intervallo $[a, b]$", trova giustificazione nel risultato seguente (riferito, per semplicità, all'intervallo $[0, 1]$):

Proposizione *La v.a. X è distribuita uniformemente in $[0, 1]$ se e solo se sussistono le proposizioni seguenti:*

(i) $\Pr(0 \leq X \leq 1) = 1$;

(ii) $\Pr(X = x) = 0$, *per ogni* $0 \leq x \leq 1$;

(iii) Sia $0 = x_0 < x_1 < \cdots < x_{n-1} < x_n = 1$ *una arbitraria suddivisione di* $[0, 1]$ *in intervalli di uguale ampiezza. Posto* $E_1 = \{x_0 \leq X \leq x_1\}, E_2 = \{x_1 < X \leq x_2\}, \ldots, E_n = \{x_{n-1} < X \leq x_n\}$, *risulta* $\Pr(E_k) = \frac{1}{n}$ *per ogni* $k \leq n$.[6]

Posto $F = F_X$, sia intanto X distribuito uniformemente in $[0, 1]$. Allora, per il Teorema 3.2.1(ix), $\Pr(0 \leq X \leq 1) = F(1) - F(0^-) = 1$. Inoltre, per il Teorema 3.2.1(viii), $\Pr(X = x) = 0$ per ogni x dell'intervallo. Infine, ancora per il Teorema 3.2.1(ix), $\Pr(E_1) = F(x_1) - F(0^-) = x_1 = \frac{1}{n}$ e $\Pr(E_k) = F(x_k) - F(x_{k-1}) = x_k - x_{k-1} = \frac{1}{n}$ $(k = 2, \ldots, n)$.

Sussistano ora le proposizioni (i), (ii) e (iii). Osservato che $\{X < 0\} \to \{0 \leq X \leq 1\}$, da (i) otteniamo $\Pr(X < 0) = 0$. Sia intanto $x < 0$. Allora, dalla $\{X \leq x\} \to \{X < 0\}$ risulta $F(x) = 0$. Inoltre, se $x \geq 1$, otteniamo $\{0 \leq X \leq 1\} \to \{X \leq x\}$ e quindi, sempre per (i), $F(x) = 1$. Sia ora $x = 0$. Allora, per (1.4), $\{X \leq 0\} = \{X < 0\} \vee \{X = 0\}$ e quindi, tramite (ii), $F(0) = 0$. Sia infine $0 < x < 1$. Supponiamo intanto che x sia un numero razionale. Allora $x = \frac{m}{n}$ con $1 \leq m < n$. Ne segue (procedendo per induzione a partire da (1.4))

$$\{X \leq x\} = \{X < 0\} \vee \{0 \leq X \leq \frac{1}{n}\} \vee \bigvee_{k=2}^{m} \{\frac{k-1}{n} < X \leq \frac{k}{n}\}$$

e quindi, considerata la suddivisione $x_k = \frac{k}{n}$ $(k = 0, \ldots, n)$, da (iii) otteniamo

$$F(x) = \Pr(X < 0) + \Pr(E_1) + \sum_{k=2}^{m} \Pr(E_k) = \frac{1}{n} + \frac{m-1}{n} = \frac{m}{n} = x.$$

[6] Cioè, per ogni suddivisione dell'intervallo unitario in intervallini di uguale ampiezza, sono uguali le probabilità di scegliere X in uno di essi.

Sia infine x un numero irrazionale. Esistono allora una successione crescente di numeri razionali $(a_n)_{n\geq 1}$ e una successione decrescente di numeri razionali $(b_n)_{n\geq 1}$ in $[0,1]$ convergenti entrambe a x. Per ogni $n \geq 1$ otteniamo $\{X \leq a_n\} \to \{X \leq x\}$, $\{X \leq x\} \to \{X \leq b_n\}$ e quindi, per quanto visto sopra, $a_n = F(a_n) \leq F(x) \leq F(b_n) = b_n$ per ogni $n \geq 1$. Ne segue, passando al limite per $n \to +\infty$, $F(x) = x$. △

Concludiamo la sezione con un risultato che affronta, in un caso particolare ma importante per le applicazioni, il problema della determinazione delle funzioni di ripartizione di trasformate certe di variabili aleatorie (che sono v.a. in forza del Teorema 3.1.3).

Teorema 3.2.3 *Siano X una v.a. e τ una funzione strettamente monotona e continua definita su un intervallo contenente il rango di X. Allora, la trasformata $Y = \tau(X)$ è una v.a. tale che:*

$$F_Y(y) = \begin{cases} 0 & \textit{se } y < \inf Y \\ F_X(\inf X) & \textit{se } y = \inf Y \\ F_X(\tau^{-1}(y)) & \textit{se } \inf Y < y < \sup Y \\ 1 & \textit{se } y \geq \sup Y \end{cases} \qquad (\tau \text{ crescente});$$

$$F_Y(y) = \begin{cases} 0 & \textit{se } y < \inf Y \\ 1 - F_X((\sup X)^-) & \textit{se } y = \inf Y \\ 1 - F_X(\tau^{-1}(y)^-) & \textit{se } \inf Y < y < \sup Y \\ 1 & \textit{se } y \geq \sup Y \end{cases} \qquad (\tau \text{ decrescente}),$$

avendo posto $\inf Z = \inf \mathrm{rg}(Z)$ *e* $\sup Z = \sup \mathrm{rg}(Z)$*, per ogni v.a. Z.*

DIMOSTRAZIONE Posto $F = F_X$, $G = F_Y$ e $\alpha = \inf Y$, $\beta = \sup Y$, procediamo per casi denotando con f una versione di X relativa alla partizione $\mathcal{P}$.

Se $y < \alpha$, allora $\{\tau(X) \leq y\} = \emptyset$ e quindi $G(y) = 0$. Se $y \geq \beta$, allora $\{\tau(X) \leq y\} = \Omega$ e quindi $G(y) = 1$.

Sia ora $\alpha < y < \beta$. Esistono allora $\omega', \omega'' \in \mathcal{P}$ tali che $\tau(f(\omega')) < y < \tau(f(\omega''))$. Ne segue, per il teorema di connessione delle funzioni continue[7], $y = \tau(x)$ per qualche x nel dominio di τ. Dunque, $\{\tau(X) \leq y\} = \{\tau(X) \leq$

[7]Ricordiamo che per *teorema di connessione delle funzioni continue* si intende il teorema: una funzione continua definita su un intervallo ha come insieme-imagine un intervallo.

$\tau(x)\} = \{\tau(X) \leq \tau(\tau^{-1}(y))\}$ e quindi $\{\tau(X) \leq y\} = \{X \leq \tau^{-1}(y)\}$, se τ è crescente, e $\{\tau(X) \leq y\} = \{X \geq \tau^{-1}(y)\} = \overline{\{X < \tau^{-1}(y)\}}$, se τ è decrescente. Dunque, $G(y) = F(\tau^{-1}(y))$, se τ è crescente, e $G(y) = 1 - F(\tau^{-1}(y)^-)$, se τ è decrescente (Teorema 3.2.1(vii)).

Sia infine $y = \alpha$. Esiste allora una successione decrescente $(y_n = \tau(f(\omega_n))_{n\geq 1}$ convergente ad α. Supponiamo intanto che τ sia crescente. Ne segue che la successione $(x_n = f(\omega_n))_{n\geq 1}$ è decrescente e quindi ammette limite $x_0 \geq \inf X$ per $n \to +\infty$. Per provare che riesce $x_0 = \inf X$, ragioniamo per assurdo assumendo $x_0 > \inf X$. Ne segue che esistono ω, $\omega' \in \mathcal{P}$ tali che $x_n \geq x_0 > f(\omega) > f(\omega') > \inf X$ per ogni $n \geq 1$. Risulta pertanto $y_n > \tau(f(\omega)) > \tau(f(\omega'))$ per ogni $n \geq 1$ e quindi, passando al limite per $n \to +\infty$, si ha la contraddizione $\alpha \geq \tau(f(\omega)) > \tau(f(\omega')) \geq \alpha$. Dunque $x_0 = \inf X$. Allora

$$\{Y \leq \alpha\} = \bigwedge_{n\geq 1}\{Y \leq y_n\} = \bigwedge_{n\geq 1}\{Y \leq \tau(x_n)\} = \bigwedge_{n\geq 1}\{X \leq x_n\} = \{X \leq \inf X\}$$

e quindi $G(y) = F(\inf X)$.

Nel caso τ decrescente, si prova, in modo analogo, che $\{Y \leq \alpha\} = \{X \geq \sup X\}$ e quindi $G(\alpha) = 1 - F((\sup X)^-)$ (Teorema 3.2.1(vii)). $\square$

Osservazione 3.2.4 L'individuazione della funzione di ripartizione della v.a. trasformata $Y = \tau(X)$ può ottenersi, in qualche caso, anche quando la funzione τ non è monotona. Considerata infatti la funzione $\tau(x) = x^2$, otteniamo, tramite il Teorema 3.2.1(ix), $F_Y(y) = 0$, se $y < 0$, e $F_Y(y) = \Pr(X^2 \leq y) = \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X((-\sqrt{y})^-)$, se $y \geq 0$. $\triangle$

3.3 Funzioni di densità

Un ruolo centrale nel calcolo delle probabilità e nelle sue applicazioni è svolto da quelle particolari v.a. X che sono **assolutamente continue**, cioè tali che la funzione di ripartizione F_X può ottenersi tramite integrazione di una funzione non negativa, nel senso che esiste una funzione (detta **densità** (**di probabilità**) di X) $f \geq 0$ Riemann integrabile su ogni intervallo limitato tale che:

$$F_X(x) = \lim_{a\to-\infty}\int_a^x f(t)dt = \int_{-\infty}^x f(t)dt$$

per ogni x reale[8]. Risulta[9]

$$\begin{aligned} 0 \leq F_x(x) - F_X(x^-) &= \lim_{n\to+\infty}(F_X(x) - F_X(x - \frac{1}{n})) = \lim_{n\to+\infty}\int_{x-\frac{1}{n}}^{x} f(t)dt \\ &\leq \lim_{n\to+\infty}[(\sup_{x-1\leq t\leq x} f(t))\,\frac{1}{n}] = 0 \end{aligned}$$

e quindi $F_x(x) = F_X(x^-)$. Ne segue, tramite il Teorema 3.2.1(vi), la continuità della funzione di ripartizione, cioè che X è una v.a. continua.

Conseguentemente, per il Teorema 3.2.1(ix), si ha

$$\Pr(a \lhd X \blacktriangleleft b) = F_X(b) - F_X(a) = \int_a^b f(t)dt \tag{3.1}$$

per ogni a, b reali tali che $a < b$, avendo denotato con $\lhd, \blacktriangleleft$ uno qualunque dei simboli $<, \leq$.

Per comprendere l'appropriatezza dell'attributo "densità", sia x_0 un punto di continuità di f. Allora, dati $\delta, \delta' \geq 0$ tali che $\Delta = \delta + \delta' > 0$, da (3.1) e dal teorema della media integrale[10] otteniamo

$$\Pr(x_0 - \delta' < X < x_0 + \delta) = \int_{x_0-\delta'}^{x_0+\delta} f(t)dt = \lambda\,\Delta = [f(x_0) + (\lambda - f(x_0)]\Delta$$

per qualche $\lambda \in [\inf f([x_0 - \delta', x_0 + \delta]), \sup f([x_0 - \delta', x_0 + \delta])]$. Conseguentemente, $\lambda - f(x_0)$ è un infinitesimo per $\Delta \to 0$ (x_0 punto di continuità!) e quindi

$$\Pr(x_0 - \delta' < X < x_0 + \delta) = f(x_0)\Delta + o(\Delta), \tag{3.2}$$

ove, usando la notazione di Landau, $o(\Delta)$ denota un infinitesimo d'ordine superiore al primo rispetto a Δ. Dunque

$$\lim_{\Delta\to 0^+} \frac{\Pr(x_0 - \delta' < X < x_0 + \delta)}{\Delta} = f(x_0)$$

[8]L'esistenza del limite è assicurata dalla non decrescenza della funzione $a \to \int_a^x f(t)dt$.

[9]Osservato che, dati $y < a < b$, dall'additività dell'integrale di Riemann otteniamo $\int_y^b f(t)dt = \int_y^a f(t)dt + \int_a^b f(t)dt$ da cui, passando al limite per $y \to -\infty$, si ha $\int_{-\infty}^b f(t)dt = \int_{-\infty}^a f(t)dt + \int_a^b f(t)dt$, cioè $F_X(b) = F_X(a) + \int_a^b f(t)dt$.

[10]Ricordiamo che per *teorema della media integrale* si intende il teorema: se la funzione g è integrabile sull'intervallo $J = [a, b]$, risulta $\int_a^b g(x)dx = \lambda(b - a)$ per qualche $\lambda \in [\inf g(J), \sup g(J)]$.

e quindi, analogamente al concetto di densità di massa considerato nella Fisica, la densità di probabilità può essere interpretata come una misura di "addensamento" della probabilità intorno al punto di continuità x_0, in quanto limite del rapporto tra la probabilità che la v.a. assuma un valore appartenente ad un intorno fondamentale di x_0 e la misura dell'intorno stesso.

Sussistono inoltre le seguenti uguaglianze

$$\begin{aligned}\Pr(x_0 < X < x_0 + \Delta) &= f(x_0)\Delta + o(\Delta)\\ \Pr(x_0 - \Delta < X < x_0) &= f(x_0)\Delta + o(\Delta)\end{aligned}$$

che si ottengono ponendo in (3.2), rispettivamente, $\delta' = 0$ e $\delta = 0$. Allora, per (3.1),

$$\begin{aligned}\left(\frac{dF_X}{dx}\right)^+(x_0) &= \lim_{\Delta\to 0^+}\frac{F_X(x_0+\Delta) - F_X(x_0)}{\Delta} = \lim_{\Delta\to 0^+}\frac{\Pr(x_0 < X < x_0 + \Delta)}{\Delta}\\ &= f(x_0)\\ &= \lim_{\Delta\to 0^+}\frac{\Pr(x_0 - \Delta < X < x_0)}{\Delta} = \lim_{\Delta\to 0^+}\frac{F_X(x_0) - F_X(x_0-\Delta)}{\Delta}\\ &= \left(\frac{dF_X}{dx}\right)^-(x_0).\end{aligned}$$

Possiamo quindi concludere che la funzione di ripartizione è derivabile in ogni punto x_0 di continuità della densità e risulta:

$$\frac{dF_X}{dx}(x_0) = f(x_0). \tag{3.3}$$

Osservazione 3.3.1 La densità di una v.a. assolutamente continua non è unica (l'integrale di Riemann non cambia se cambiamo il valore di una funzione integrabile in un numero finito di punti!). Ad esempio, con riferimento a una v.a. X distribuita uniformemente in $[0,1]$ (Osservazione 3.2.2(ii)), sono densità di X le funzioni:

$$f_\alpha(x) = \begin{cases} 0 & \text{se } x < 0\\ \alpha & \text{se } x = 0\\ 1 & \text{se } 0 < x \le 1\\ 0 & \text{se } x > 1\end{cases}$$

per ogni reale $\alpha \ge 0$. △

Assumiamo ora che f sia una funzione continua e $F_X(x) < 1$ per ogni x reale. Allora, per il Teorema 3.2.1(ix), $\Pr(X > x) = 1 - F_X(x) > 0$ e quindi possiamo considerare, dato $\Delta > 0$, la probabilità condizionata

$$\begin{aligned}\Pr(x < X < x + \Delta \,|\, X > x) &= \frac{\Pr(\{x < X < x + \Delta\} \wedge \{X > x\})}{\Pr(X > x)} \\ &= \frac{\Pr(x < X < x + \Delta)}{\Pr(X > x)} = \frac{f(x)\Delta + o(\Delta)}{1 - F_X(x)} \\ &= \frac{f(x)}{1 - F_X(x)}\,\Delta + o(\Delta),\end{aligned}$$

ove la terza uguaglianza segue da (3.2) ponendo $\delta' = 0$. Introdotta infine la funzione di **intensità** (**di probabilità**) μ così definita:

$$\mu(x) = \frac{f(x)}{1 - F_X(x)}$$

per ogni x reale, risulta

$$\Pr(x < X < x + \Delta \,|\, X > x) = \mu(x)\Delta + o(\Delta). \tag{3.4}$$

Dunque, $\mu(x)\Delta$ fornisce un'approssimazione del primo ordine della probabilità che la v.a. X assuma valori minori di $x + \Delta$, sapendo (in via reale o ipotetica) che i suoi valori possibili sono maggiori di x [11].

Assumiamo infine che i valori possibili di X siano tutti non negativi (come avviene nelle situazioni descritte in nota). Tramite (3.3) si ha

$$\mu(x) = -\frac{d}{dx}\ln(1 - F_X(x))$$

e quindi, passando alla funzione integrale di $-\mu$, otteniamo $\int_0^x -\mu(t)dt = \ln(1 - F_X(x))$. Ne segue la relazione seguente che consente di ricavare la f.r. dalla funzione di intensità[12]:

$$F_X(x) = 1 - \exp\left[-\int_0^x \mu(t)dt\right]. \tag{3.5}$$

[11] L'intensità di probabilità è di particolare interesse nelle situazioni che riguardano la *durata* di un dato fenomeno (funzionamento di una macchinario, esposizione a radiazioni, vita di un individuo, disoccupazione, etc.). Nel caso particolare della "durata in vita di un individuo", $\mu(x)\Delta$ è (a meno di infinitesimi d'ordine superiore al primo) la probabilità che l'individuo, di età x, non raggiunga l'età $x + \Delta$ (in questo ambito, l'intensità viene chiamata, più propriamente, **intensità di mortalità**).

[12] Molto utile in quelle situazioni concrete in cui si possiedono maggiori informazioni sul comportamento dell'intensità rispettto alle conoscenze sulla densità.

Concludiamo la sezione con il teorema seguente (naturale completamento del teorema 3.2.3) che affronta il problema della determinazione delle funzioni di densità di trasformate certe di variabili aleatorie assolutamente continue.

Teorema 3.3.2 *Siano X una v.a. limitata*[13] *assolutamente continua con densità f continua su un intervallo chiuso I includente il rango di X e τ una funzione strettamente monotona derivabile con derivata continua mai nulla su I. Allora, la trasformata $Y = \tau(X)$ è una v.a. assolutamente continua con densità:*

$$f_Y(y) = \begin{cases} f(\tau^{-1}(y))\ |\frac{d\tau^{-1}}{dy}(y)| & \text{se } \inf Y < y < \sup Y \\ 0 & \text{altrimenti} \end{cases}.$$

Dimostrazione Posto $F = F_X$, $G = F_Y$, $a = \inf X$, $b = \sup X$, $\alpha = \inf Y$, $\beta = \sup Y$ e $\iota = \tau^{-1}$, assumiamo intanto che τ sia crescente. Ne segue $a = \iota(\alpha)$ e $b = \iota(\beta)$. Sia intanto $y \in]\alpha, \beta[$. Allora, tramite il Teorema 3.2.3, $G(y) = F(\iota(y))$. Notato che $a < \iota(y)$, dal Teorema 3.2.1(ix),(iii) otteniamo $F(\iota(y)) = F(a^-) + \Pr(a \le X < \iota(y)) = \Pr(a \le X < \iota(y))$ e quindi, per (3.1),

$$G(y) = \int_a^{\iota(y)} f(t)dt = \int_{\iota(\tau(a))}^{\iota(y)} f(t)dt.$$

Ora, per il teorema della derivata della funzione inversa[14], ι è derivabile con derivata prima continua. Effettuando nell'ultimo integrale il cambio di variabile $t \rightsquigarrow \iota(z)$ otteniamo, tramite il teorema d'integrazione per sostituzione[15],

$$G(y) = \int_{\tau(a)}^{y} f(\iota(z))\iota'(z)dz = \int_{\alpha}^{y} f(\iota(z))|\iota'(z)|dz = \int_{-\infty}^{y} f_Y(z)dz,$$

tenuto conto che la derivata ι' è positiva, essendo ι crescente.

Assumiamo infine che τ sia decrescente. Ne segue $a = \iota(\beta)$ e $b = \iota(\alpha)$. Sia intanto $y \in]\alpha, \beta[$. Allora, tramite il Teorema 3.2.3, $F(y) = 1 - F(\iota(y)^-) =$

[13] Un numero aleatorio di dice **limitato** se il suo rango è un insieme limitato.

[14] Per *teorema della derivata della funzione inversa* si intende il teorema: se g è una funzione strettamente monotona derivabile in t_0 con derivata diversa da zero, allora la funzione inversa è derivabile nel punto $z_0 = g(t_0)$ e risulta e $(g^{-1})'(z_0) = g'(t_0)^{-1}$.

[15] Per *teorema d'integrazione per sostituzione* si intende il teorema: sia g una funzione continua su un intervallo chiuso I e h una funzione derivabile con derivata continua su un intervallo J tale che $h(J) \subseteq I$. Allora, $\int_{h(a)}^{h(b)} g(t)dt = \int_a^b g(h(z))h'(z)dz$ per ogni $a, b \in J$ $(a < b)$.

$1-F(\iota(y))$, tenuto conto che F è continua. Poichè $a < \iota(y)$, risulta $F(\iota(y)) = \Pr(a \le X < \iota(y))$ e quindi

$$\begin{aligned} G(y) &= 1 - \int_a^{\iota(y)} f(t)dt = \int_a^b f(t)dt - \int_a^{\iota(y)} f(t)dt = \int_{\iota(y)}^b f(t)dt \\ &= \int_{\iota(y)}^{\iota(\tau(b))} f(t)dt, \end{aligned}$$

notato che, per (3.1) e il Teorema 3.2.1(iii), $\int_a^b f(t)dt = F(b)-F(a) = 1$. Ora, per il teorema della derivata della funzione inversa, ι è derivabile con derivata prima continua. Effettuando nell'ultimo integrale il cambio di variabile $t \rightsquigarrow \iota(z)$ otteniamo, tramite il teorema d'integrazione per sostituzione,

$$\begin{aligned} G(y) &= \int_y^{\tau(b)} f(\iota(z))\iota'(z)dz = \int_y^{\alpha} f(\iota(z))\iota'(z)dz = \int_\alpha^y f(\iota(z))(-\iota'(z))dz \\ &= \int_\alpha^y f(\iota(z))|\iota'(z)|dz = \int_{-\infty}^y f_Y(z)dz, \end{aligned}$$

tenuto conto che la derivata ι' è negativa, essendo ι decrescente.

Per completare la dimostrazione, rimane da provare che $\int_\alpha^\beta f_Y(z)dz = 1$. Risulta, per il teorema d'integrazione per sostituzione,

$$\begin{aligned} 1 &= \int_a^b f(t)dt = \begin{cases} \displaystyle\int_{\iota(\alpha)}^{\iota(\beta)} f(t)dt = \int_\alpha^\beta f(\iota(z))\iota'(z)dz & \text{se } \tau \text{ crescente} \\ \displaystyle -\int_{\iota(\beta)}^{\iota(\alpha)} f(t)dt = \int_\alpha^\beta f(\iota(z))\iota'(z)dz & \text{se } \tau \text{ decrescente} \end{cases} \\ &= \int_\alpha^\beta f(\iota(z))|\iota'(z)|dz = \int_\alpha^\beta f_Y(z)dz. \end{aligned}$$

La dimostrazione è così conclusa. □

3.4 Speranza matematica

Considerata la famiglia $\mathcal{V}_b$ delle v.a. limitate[16], si chiama **speranza matematica** (o **valor medio** o **valore atteso**) **su** $\mathcal{V}_b$ ogni applicazione $\mathrm{E} : \mathcal{V}_b \to \mathbb{R}$ che verifica gli assiomi seguenti:

[16] Che è chiusa per le operazioni di addizione, moltiplicazione e moltiplicazione scalare e include le parti positiva e negativa di ogni suo elemento come pure gli indicatori di tutti gli eventi di $\mathcal{A}$ (Teorema 3.1.3).

E1 $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$ (additività).

E2 $\mathrm{E}(X)\leq\mathrm{E}(Y)$, se $X\leq Y$ (monotonia).

E3 $\mathrm{E}(|E|)=\Pr(E)$ (compatibilità).[17]

Conveniamo che, come $\mathcal{A}$ rappresenta una σ-algebra di eventi, Pr una probabilità numerabilmente additiva su $\mathcal{A}$, così E denoti una speranza matematica su $\mathcal{V}_b$ e che X, Y, Z (dotati o no di apici o pedici) rappresentino sempre v.a. limitate.

Osservazione 3.4.1 Continuazione dell'Osservazione 2.1.1. Ricordiamo che il guadagno aleatorio associato alla scommessa relativa all'evento E di puntata p e vincita v è dato dalla:

$$G^{(E)}(p,v;S)=S(v|E|-p),$$

ove S è il puntatore che vale 1, se si scommette su E, e -1, se si scommette contro E. Ora, poichè nell'espressione del guadagno compare l'importo aleatorio $Y=v|E|$, possiamo rivedere tale scommessa come una operazione di scambio tra un importo certo p e l'importo aleatorio Y, nel senso che si acquista Y al prezzo p, se si scommette su Y, e si vende Y al prezzo p, se si scommette contro Y.

Alla luce di questa osservazione chiamiamo, dato un importo aleatorio X arbitrario, *scommessa relativa a* X una operazione di scambio tra un importo certo p (il *prezzo*) e X (la *vincita*). Introducendo il solito puntatore S che vale 1, se si scommette su X, e -1, se si scommette contro X, otteniamo che il guadagno aleatorio relativo è dato da $G^{(X)}(p;S)=S(X-p)$.

Passando ad esaminare un *sistema di scommesse relative a importi aleatori*, consideriamo gli importi aleatori $X_1,\dots X_n$ e mettiamo in evidenza, tramite il valori del puntatore S_i, se la relativa scommessa è su X_i oppure contro X_i ($1\leq i\leq n$). Indicati i prezzi rispettivi con $p_1,\dots p_n$, il guadagno aleatorio relativo assume l'espressione:

$$G^{(X_1,\dots,X_n)}(\mathbf{p};\mathbf{S})=S_1(X_1-p_1)+\cdots+S_n(X_n-p_n).$$

A questo punto, in analogia al caso delle scommesse relative a eventi, chiamiamo i prezzi $p_1,\dots,p_n$ *equi* se consentono di *evitare la perdita certa*, cioè se:

$$\inf G^{(X_1,\dots,X_n)}(\mathbf{p};\mathbf{S})\leq 0\leq\sup G^{(X_1,\dots,X_n)}(\mathbf{p};\mathbf{S})$$

per ogni $\mathbf{S}=(S_1,\dots,S_n)\in\{-1,1\}^n$.

[17]La cui esistenza verrà provata nel corso progredito di Calcolo delle Probabilità.

Ciò precisato, assumiamo che d'ora in poi i prezzi siano equi. Ora, dati gli importi aleatori X e Y, consideriamo il sistema di scommesse relative agli importi X, Y e $X+Y$ avente, nell'ordine, prezzi p', p'', p e puntatori S', S'', S. Risulta allora

$$\begin{aligned} G^{(X,Y,X+Y)}(\mathbf{p};\mathbf{S}) &= S'(X-p') + S''(Y-p'') + S(X+Y-p) \\ &= [(S'+S)X + (S''+S)Y] - (S'p' + S''p'' + Sp). \end{aligned}$$

Posto quindi $S = 1, S' = S'' = -1$ (in modo da annullare la parte aleatoria) otteniamo che il guadagno diviene il numero certo di valore $-(-p' - p'' + p)$ e pertanto deve risultare $p = p' + p''$. Dunque, il prezzo equo relativo alla somma di due importi aleatori è uguale alla somma dei prezzi equi degli addendi.

Supponiamo ora $X \leq Y$. Risulta allora

$$G^{(X,Y)}(\mathbf{p};\mathbf{S}) = S'(X-p') + S''(Y-p'') = (S'X + S''Y) - (S'p' + S''p'')$$

e quindi, posto $S' = -1$ e $S'' = 1$, otteniamo

$$G^{(X,Y)}(\mathbf{p};\mathbf{S}) = (Y - X) + (p' - p'').$$

Ne segue, $0 \geq \inf G^{(X,Y)}(\mathbf{p};\mathbf{S}) = \inf(Y-X) + (p'-p'')$ e quindi $p' \leq p''$, osservato che $\inf(Y-X) \geq 0$. Dunque, se $X \leq Y$, il prezzo equo relativo all'importo X è minore o uguale a quello relativo all'importo Y.

Infine, se $X = |\Omega|$, otteniamo, come già visto, $p = 1$. Conseguentemente, la funzione che ad ogni evento E di $\mathcal{A}$ associa il prezzo equo della scommessa relativa all'indicatore $|E|$, è una probabilità sulla σ-algebra $\mathcal{A}$. Infatti, indicato con $P(E)$ tale prezzo equo, otteniamo banalmente gli assiomi P1 e P2; inoltre, l'assioma P3 deriva dall'additività dei prezzi equi e dalla relazione $|E \vee F| = |E| + |F|$ valida ogniqualvolta i due eventi sono incompatibili (Teorema 1.11.1(iii)).

Le considerazioni fatte consentono dunque di giustificare gli assiomi E1, E2 della speranza matematica, qualora si interpretino le speranze matematiche come prezzi equi nei sistemi di scommesse relative a v.a. (considerate come importi aleatori)[18]; inoltre, l'assioma E3 serve per assicurare che la probabilità $P(\cdot)$ generata dai prezzi equi degli indicatori coincida con la probabilità Pr prefissata sulla σ-algebra $\mathcal{A}$.

[18]Il collegamento tra i prezzi equi e la speranza matematica era già presente agli albori del calcolo delle probabilità. Ad esempio, nel primo trattato di calcolo delle probabilità *De Ratiociniis in Ludo Aleæ* (1656), Christiaan Huygens pone a fondamento del calcolo delle probabilità la nozione (senza darne una definizione esplicita) di speranza matematica (da lui chiamata *expectatio* (attesa)), prescindendo dal concetto di probabilità, e ne desume l'espressione analitica, in alcuni casi particolari, identificandola con i prezzi di scommesse eque (non definite formalmente ma identificate con quelle situazioni aleatorie che non comportano "alcun danno alla persona").

Nel prossimo teorema riportiamo le principali proprietà della speranza matematica. In particolare, (i), (v) e l'assioma E2 assicurano che la speranza matematica è un funzionale lineare monotono su $\mathcal{V}_b$ che associa al numero certo di valore 1 il valore 1 (proprietà di normalizzazione). Inoltre, nel caso di v.a. semplici, da (vi) otteniamo

$$\mathrm{E}(X) = \frac{\sum_{i=1}^n \Pr(X = x_i)x_i}{\sum_{i=1}^n \Pr(X = x_i)}$$

e quindi la speranza matematica è la media ponderata dei valori possibili pesati con le probabilità che la v.a. assuma quei valori. La disuguaglianza di Markov fornisce una limitazione superiore della coda destra della distribuzione della v.a. e mette in evidenza che la speranza matematica è un *indice di posizione*; quella di Kolmogorov invece fornisce una limitazione inferiore, tramite il momento secondo, delle code della distribuzione della v.a.. Infine, (ix) assicura che una v.a. non negativa con speranza matematica nulla assume quasi certamente il valore zero.

Teorema 3.4.2 *Sussistono le proposizioni seguenti:*

(i) LINEARITÀ: $\mathrm{E}(\sum_{i=1}^n \alpha_i X_i) = \sum_{i=1}^n \alpha_i \mathrm{E}(X_i)$*;*

(ii) INTERNALITÀ: $\inf X \leq \mathrm{E}(X) \leq \sup X$*;*

(iii) POSITIVITÀ: $\mathrm{E}(X) \geq 0$*, se* $X \geq 0$*;*

(iv) $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$*;*

(v) Se X *è il numero certo di valore* α*, allora* $\mathrm{E}(X) = \alpha$*;*

(vi) Se X *è una v.a. semplice con valori possibili* $x_1, \ldots, x_n$*, allora*

$$\mathrm{E}(X) = \sum_{i=1}^n x_i \Pr(X = x_i);$$

(vii) Se X *è una v.a. con un numerabile di valori possibili* $x_1, x_2, \ldots$*, allora la serie* $\sum_{n\geq 1} x_n \Pr(X = x_n)$ *è assolutamente convergente e riesce:*

$$\mathrm{E}(X) = \sum_{n\geq 1} x_n \Pr(X = x_n);$$

(viii) Se $X \geq 0$ *e* $a > 0$*, allora*

$$\Pr(X \geq a) \leq \frac{\mathrm{E}(X)}{a} \qquad \text{(disuguaglianza di Markov);}$$

(ix) Se $X \neq 0$ *e* $a > 0$*, allora*

$$\Pr(|X| > a) \geq \frac{\mathrm{E}(X^2) - a^2}{\sup(X^2)} \qquad \text{(disuguaglianza di Kolmogorov);}$$

(x) Se $X \geq 0$ *e* $\mathrm{E}(X) = 0$*, allora* $\Pr(X = 0) = 1$.

Dimostrazione Le proposizioni (iii), (v) derivano immediatamente da (ii) mentre la proposizione (iv) da E2, (i) (osservato che $-|X| \leq X \leq |X|$).

(i) Da E1 risulta, tramite induzione,

$$\mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{E}(X_i) \tag{3.6}$$

per ogni $X_1, \ldots, X_n$. Basta dunque provare che $\mathrm{E}(\alpha X) = \alpha \mathrm{E}(X)$ per ogni α reale. Se $\alpha = 0$, da E3 si ha $\mathrm{E}(0 \cdot X) = \mathrm{E}(|\emptyset|) = \Pr(\emptyset) = 0$. Se $\alpha = -1$, da E1, E3 otteniamo $\mathrm{E}(X) + \mathrm{E}(-X) = \mathrm{E}(X + (-X)) = \mathrm{E}(|\emptyset|) = 0$ e quindi

$$\mathrm{E}(-X) = -\mathrm{E}(X). \tag{3.7}$$

Supponiamo ora $\alpha > 0$ e procediamo per casi. Sia intanto $\alpha = n$. Allora, da (3.6) si ha

$$\mathrm{E}(nX) = \mathrm{E}(\overbrace{X + \cdots + X}^{n\text{volte}}) = \overbrace{\mathrm{E}(X) + \cdots + \mathrm{E}(X)}^{n\text{volte}} = n\mathrm{E}(X).$$

Sia ora $\alpha = \frac{1}{n}$. Allora, $\mathrm{E}(X) = \mathrm{E}(n(\frac{1}{n}X)) = n\mathrm{E}(\frac{1}{n}X)$ e quindi $\mathrm{E}(\frac{1}{n}X) = \frac{1}{n}\mathrm{E}(X)$.
Sia ora $\alpha = \frac{m}{n}$. Allora, $\mathrm{E}(\frac{m}{n}X) = \mathrm{E}(m(\frac{1}{n}X)) = m\mathrm{E}(\frac{1}{n}X) = \frac{m}{n}\mathrm{E}(X)$.
Sia ora α un numero irrazionale. Esistono allora due successioni $(r_n)_{n\geq 1}$, $(q_n)_{n\geq 1}$ di numeri razionali positivi tali che $r_n \uparrow \alpha$ e $q_n \downarrow \alpha$. Sia intanto $X \geq 0$. Ne segue $r_n X \leq \alpha X \leq q_n X$ da cui, per E2, $r_n \mathrm{E}(X) = \mathrm{E}(r_n X) \leq \mathrm{E}(\alpha X) \leq \mathrm{E}(q_n X) = q_n \mathrm{E}(X)$ e quindi, passando al limite per $n \to +\infty$, si ha $\mathrm{E}(\alpha X) = \alpha \mathrm{E}(X)$. Sia infine X arbitrario. Considerate allora le parti positiva e negativa di X, otteniamo, tramite (3.6) e (3.7), $\mathrm{E}(\alpha X) = \mathrm{E}(\alpha(X^+ - X^-)) = \mathrm{E}(\alpha X^+ +$

$(-\alpha X^-)) = \mathrm{E}(\alpha X^+) + \mathrm{E}(-(\alpha X^-)) = \mathrm{E}(\alpha X^+) - \mathrm{E}(\alpha X^-) = \alpha\mathrm{E}(X^+) - \alpha\mathrm{E}(X^-) = \alpha[\mathrm{E}(X^+) - \mathrm{E}(X^-)] = \alpha\mathrm{E}(X^+ - X^-) = \alpha\mathrm{E}(X)$.

Supponiamo infine $\alpha < 0$. Allora, per (3.7), $\mathrm{E}(\alpha\, X) = \mathrm{E}(-|\alpha|\, X) = -\mathrm{E}(|\alpha|\, X) = -|\alpha|\,\mathrm{E}(X) = \alpha\mathrm{E}(X)$.

(ii) Posto $\alpha = \inf X$, $\beta = \sup X$, tramite E3, (i) e E2, risulta $\alpha = \alpha\,\mathrm{E}(|\Omega|) = \mathrm{E}(\alpha\,|\Omega|) \leq \mathrm{E}(X) \leq \mathrm{E}(\beta\,|\Omega|) = \beta\,\mathrm{E}(|\Omega|) = \beta$.

(vi) Per il Teorema 3.1.2, $\{X = x_i\} \in \mathcal{A}$ e quindi $|\{X = x_i\}| \in \mathcal{V}_b$ per ogni $i \leq n$. Osservato che $X = \sum_{i=1}^n x_i\,|\{X = x_i\}|$, da (i) e E3 otteniamo $\mathrm{E}(X) = \sum_{i=1}^n x_i\,\mathrm{E}(|\{X = x_i\}|) = \sum_{i=1}^n x_i \Pr(X = x_i)$.

(vii) Risulta $X = \sum_{n\geq 1} x_n\,|\{X = x_n\}|$ e, per quanto constatato nella dimostrazione precedente, $|\{X = x_n\}| \in \mathcal{V}_b$ per ogni $n \geq 1$. Assunto $x_n \in\,]a, b[$ per ogni $n \geq 1$ (X v.a. limitata!), la convergenza assoluta della serie discende immediatamente dalla:

$$\begin{aligned}\sum_{n\geq 1} |x_n| \Pr(X = x_n) &\leq \sum_{n\geq 1} \max(|a|, |b|) \Pr(X = x_n)\\ &= \max(|a|, |b|) \sum_{n\geq 1} \Pr(X = x_n) = \max(|a|, |b|).\end{aligned}$$

Per provare la seconda parte della tesi, poniamo $E_n = \bigvee_{i=1}^n \{X = x_i\}$ e consideriamo le v.a. semplici:

$$\underline{X}_n = \sum_{i=1}^n x_i\,|\{X = x_i\}| + a\,|\overline{E}_n|, \qquad \overline{X}_n = \sum_{i=1}^n x_i\,|\{X = x_i\}| + b\,|\overline{E}_n|$$

per ogni $n \geq 1$. Risulta allora $\underline{X}_n \leq X \leq \overline{X}_n$ da cui, per E2, otteniamo $\mathrm{E}(\underline{X}_n) \leq \mathrm{E}(X) \leq \mathrm{E}(\overline{X}_n)$ e quindi, per (vi) e E3,

$$\sum_{i=1}^n x_i \Pr(X = x_i) + a \Pr(\overline{E}_n) \leq \mathrm{E}(X) \leq \sum_{i=1}^n x_i \Pr(X = x_i) + b \Pr(\overline{E}_n). \quad (3.8)$$

Ora, poichè $\bigvee_{n\geq 1} E_n = \bigvee_{n\geq 1} \{X = x_n\} = \Omega$ [19] e la successione $(E_n)_{n\geq 1}$ è non decrescente, si ha, per la continuità dal basso della probabilità (Teorema 2.2.2(i)), $\Pr(E_n) \uparrow \Pr(\Omega) = 1$ e quindi $\Pr(\overline{E}_n) = 1 - \Pr(E_n) \downarrow 0$. Conseguentemente, passando al limite per $n \to +\infty$ (la serie è convergente in

[19] Notato che, come facilmente si verifica, $\bigvee_{i\in I}(\bigvee_{j\in J_i} E_j) = \bigvee_{j\in\bigcup_{i\in I} J_i} E_j$, qualunque siano gli eventi E_j.

quanto assolutamente convergente!), da (3.8) otteniamo

$$\sum_{n\geq 1} x_n \Pr(X = x_n) \leq \mathrm{E}(X) \leq \sum_{n\geq 1} x_n \Pr(X = x_n),$$

cioè $\mathrm{E}(X) = \sum_{n\geq 1} x_n \Pr(X = x_n)$.

(viii) Osservato che $a\,|\{X \geq a\}| \leq X$, da E3, (i) e E2 si ha $a \Pr(X \geq a) = a\,\mathrm{E}(|\{X \geq a\}|) = \mathrm{E}(a|\{X \geq a\}|) \leq \mathrm{E}(X)$ e quindi la tesi ($a > 0$!).

(ix) Osservato che

$$\begin{aligned} X^2 &= X^2(|\{X \leq a\}| + |\{X > a\}|) = X^2|\{X \leq a\}| + X^2|\{X > a\}| \\ &\leq a^2 + \sup(X^2)|\{X > a\}| \end{aligned}$$

da E2, (i) e E3 otteniamo

$$\mathrm{E}(X^2) \leq a^2 + \sup(X^2)\mathrm{E}(|\{X > a\}|) = a^2 + \sup(X^2) \Pr(X > a)$$

e quindi la tesi ($\sup(X^2) > 0$!).

(x) Considerata la successione non crescente di eventi $(\{X < \frac{1}{n}\})_{n\geq 1}$, tramite (1.3) e la continuità dall'alto della probabilità (Teorema 2.2.2(ii)) risulta

$$\begin{aligned} \Pr(X = 0) &= \Pr(X \leq 0) = \Pr(\bigwedge_{n\geq 1} \{X < \frac{1}{n}\}) \\ &= \lim_{n\to+\infty} \Pr(X < \frac{1}{n}) = 1 - \lim_{n\to+\infty} \Pr(X \geq \frac{1}{n}) = 1, \end{aligned}$$

osservato che, per (viii), $\Pr(X \geq \frac{1}{n}) \leq \mathrm{E}(X)\,(\frac{1}{n})^{-1} = 0$. □

Poichè la serie considerata nella proposizione (vii) del teorema precedente è permutabile (essendo assolutamente convergente), la sua somma non dipende dal modo in cui vengono numerati gli elementi del rango $\mathrm{rg}(X)$ di X. Conviene allora indicare tale valore con la notazione $\sum_{x\in \mathrm{rg}(X)} x \Pr(X = x)$ che consente di non esplicitare l'ordinamento scelto per determinarlo. Tale notazione verrà anche adottata nel caso che il rango $\mathrm{rg}(X)$ sia finito, come supposto nella proposizione (vi) del teorema precedente.

I due teoremi seguenti forniscono l'espressione della speranza matematica di trasformate certe di v.a. con rango discreto e di v.a. assolutamente continue.

Teorema 3.4.3 *Siano X una v.a. con rango discreto e τ una funzione con dominio includente* $\mathrm{rg}(X)$ *tale che $\tau(X) \in \mathcal{V}_b$. Risulta allora:*

$$\mathrm{E}(\tau(X)) = \sum_{x \in \mathrm{rg}(X)} \tau(x) \Pr(X = x).$$

Dimostrazione Dal Teorema 3.4.2(vi),(vii) risulta

$$\mathrm{E}(\tau(X)) = \sum_{y \in \mathrm{rg}(\tau(X))} y \Pr(\tau(X) = y).$$

Ora, denotato con f una versione di X relativa alla partizione $\mathcal{P}$ e posto $I = \{x \in \mathrm{rg}(X) : \tau(x) = y\}$, $J_x = \{\omega : f(\omega) = x\}$, tramite la nota 19 di pagina 86 otteniamo

$$\bigvee_{x \in I} \{X = x\} = \bigvee_{x \in I} (\bigvee_{\omega \in J_x} \omega) = \bigvee_{\omega \in \bigcup_{x \in I} J_x} \omega = \bigvee_{\omega : \tau(f(\omega)) = y} \omega = \{\tau(X) = y\}$$

e quindi

$$\begin{aligned}
\mathrm{E}(\tau(X)) &= \sum_{y \in \mathrm{rg}(\tau(X))} y \Pr(\bigvee_{x \in I} \{X = x\}) = \sum_{y \in \mathrm{rg}(\tau(X))} y \sum_{x \in I} \Pr(X = x) \\
&= \sum_{x \in \mathrm{rg}(X)} \tau(x) \Pr(X = x),
\end{aligned}$$

notato che l'ultima sommatoria è una permutazione della precedente. □

Teorema 3.4.4 *Sia X una v.a. assolutamente continua con valori nell'intervallo chiuso $[a, b]$ e densità f continua su $[a, b]$. Sia inoltre $\tau : [a, b] \to \mathbb{R}$ una funzione continua. Risulta allora:*

$$\mathrm{E}(\tau(X)) = \int_a^b \tau(x) f(x) dx.$$

In particolare, posto τ la funzione identica, si ha $\mathrm{E}(X) = \int_a^b x f(x) dx$.

Dimostrazione Per il Teorema 3.1.3 e il teorema di Weierstrass[20], $\tau(X)$ è una v.a. limitata e quindi appartenente a $\mathcal{V}_b$.

[20] Per *teorema di Weierstrass* si intende il teorema: Ogni funzione continua definita su un insieme compatto è limitata.

Per il teorema da Heine[21], f e τf sono uniformemente continue e quindi, dato $\epsilon > 0$, esiste una suddivisione $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$ di $[a, b]$ tale che $|f(x') - f(x'')| < \epsilon$, $|\tau(x')f(x') - \tau(x'')f(x'')| < \epsilon$ per ogni $x', x'' \in J_i = [x_{i-1}, x_i]$ $(i = 1, \ldots, n)$. Indicati con α_i, β_i. rispettivamente, il minimo e il massimo valore di τf nell'intervallo J_i e posto $\Delta_i = x_i - x_{i-1}$ $(i = 1, \ldots, n)$, si ha, per la definizione dell'integrale di Riemann,

$$\sum_{i=1}^{n} \alpha_i \Delta_i \leq \int_a^b \tau(x) f(x) dx \leq \sum_{i=1}^{n} \beta_i \Delta_i,$$

osservato che le sommatorie sono, rispettivamente, una somma inferiore e una somma superiore relative alla suddivisione considerata. Poichè

$$\sum_{i=1}^{n} (\beta_i - \alpha_i) \Delta_i < \epsilon \sum_{i=1}^{n} \Delta_i = (b - a)\, \epsilon \tag{3.9}$$

otteniamo

$$\sum_{i=1}^{n} \alpha_i \Delta_i \leq \int_a^b \tau(x) f(x) dx < \sum_{i=1}^{n} \alpha_i \Delta_i + (b - a)\, \epsilon. \tag{3.10}$$

Considerata infine la partizione dell'evento certo $E_1 = \{x_0 \leq X \leq x_1\}$, $E_i = \{x_{i-1} < X \leq x_i\}$ $(i = 2, \ldots, n)$ e indicati con $\underline{x}_i$, $\overline{x}_i$, rispettivamente, i punti di minimo e di massimo di τ nell'intervallo J_i $(i = 1, \ldots, n)$, risulta $\sum_{i=1}^{n} \tau(\underline{x}_i)|E_i| \leq \tau(X) \leq \sum_{i=1}^{n} \tau(\overline{x}_i)|E_i|$ e quindi, per la monotonia E2 e il Teorema 3.4.2(vi), $\sum_{i=1}^{n} \tau(\underline{x}_i) \Pr(E_i) = \mathrm{E}(\sum_{i=1}^{n} \tau(\underline{x}_i)|E_i|) \leq \mathrm{E}(\tau(X)) \leq \mathrm{E}(\sum_{i=1}^{n} \tau(\overline{x}_i)|E_i|) = \sum_{i=1}^{n} \tau(\overline{x}_i) \Pr(E_i)$. Osservato che F_X è continua, dal Teorema 3.2.1(ix), si ha $\Pr(E_i) = F_X(x_i) - F_X(x_{i-1})$ $(i = 1, \ldots, n)$ e quindi

$$\sum_{i=1}^{n} \tau(\underline{x}_i)[F_X(x_i) - F_X(x_{i-1})] \leq \mathrm{E}(\tau(X)) \leq \sum_{i=1}^{n} \tau(\overline{x}_i)[F_X(x_i) - F_X(x_{i-1})].$$

Poichè la densità è continua sull'intervallo $[a, b]$, la funzione di ripartizione è derivabile, per (3.3), in ogni punto dell'intervallo e quindi, per il teorema di

[21] Per *teorema di Heine* si intende il teorema: Ogni funzione continua definita su un insieme compatto è uniformemente continua.

Lagrange[22], $F_X(x_i) - F_X(x_{i-1}) = f(\xi_i)\Delta_i$ con $\xi_i \in]x_{i-1}, x_i[$. Ne segue

$$\begin{aligned}\sum_{i=1}^{n} \tau(\underline{x}_i)[f(\underline{x}_i) + (f(\xi_i) - f(\underline{x}_i))]\Delta_i &\leq \mathrm{E}(\tau(X)) \\ &\leq \sum_{i=1}^{n} \tau(\overline{x}_i)[f(\overline{x}_i) + (f(\xi_i) - f(\overline{x}_i))]\Delta_i\end{aligned}$$

da cui otteniamo

$$\begin{aligned}\sum_{i=1}^{n} \alpha_i\Delta_i &+ \sum_{i=1}^{n} \tau(\underline{x}_i)(f(\xi_i) - f(\underline{x}_i))\Delta_i \\ &\leq \sum_{i=1}^{n} \tau(\underline{x}_i)f(\underline{x}_i)\Delta_i + \sum_{i=1}^{n} \tau(\underline{x}_i)(f(\xi_i) - f(\underline{x}_i))\Delta_i \\ &= \sum_{i=1}^{n} \tau(\underline{x}_i)[f(\underline{x}_i) + (f(\xi_i) - f(\underline{x}_i))]\Delta_i \leq \mathrm{E}(\tau(X)) \\ \sum_{i=1}^{n} \beta_i\Delta_i &+ \sum_{i=1}^{n} \tau(\overline{x}_i)(f(\xi_i) - f(\overline{x}_i))\Delta_i \\ &\geq \sum_{i=1}^{n} \tau(\overline{x}_i)f(\overline{x}_i)\Delta_i + \sum_{i=1}^{n} \tau(\overline{x}_i)(f(\xi_i) - f(\overline{x}_i))\Delta_i \\ &= \sum_{i=1}^{n} \tau(\overline{x}_i)[f(\overline{x}_i) + (f(\xi_i) - f(\underline{x}_i))]\Delta_i \geq \mathrm{E}(\tau(X)).\end{aligned}$$

Dunque

$$\begin{aligned}\sum_{i=1}^{n} \alpha_i\Delta_i &+ \sum_{i=1}^{n} \tau(\underline{x}_i)(f(\xi_i) - f(\underline{x}_i))\Delta_i \\ &\leq \mathrm{E}(\tau(X)) \leq \sum_{i=1}^{n} \beta_i\Delta_i + \sum_{i=1}^{n} \tau(\overline{x}_i)(f(\xi_i) - f(\overline{x}_i))\Delta_i. \qquad (3.11)\end{aligned}$$

[22]Per *teorema di Lagrange* si intende il teorema: se la funzione f è continua nell'intervallo chiuso $[a, b]$ ed è derivabile in tutti i punti interni, allora esiste un punto ξ interno all'intervallo tale che $f(b) - f(a) = f'(\xi)(b - a)$.

Ora, tramite (3.9), si ha

$$\begin{aligned}
&\sum_{i=1}^{n}\beta_i\Delta_i+\sum_{i=1}^{n}\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))\Delta_i \\
&\qquad -\Big[\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i\Big] \\
&=\sum_{i=1}^{n}(\beta_i-\alpha_i)\Delta_i+\sum_{i=1}^{n}[\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))-\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))]\Delta_i \\
&<(b-a)\,\epsilon+\sum_{i=1}^{n}\left[\,|\tau(\overline{x}_i)|\,|f(\xi_i)-f(\overline{x}_i)|+|\tau(\underline{x}_i)|\,|f(\xi_i)-f(\underline{x}_i)|\,\right]\Delta_i \\
&<(b-a)\,\epsilon+2(\max|\tau|)\,\epsilon\sum_{i=1}^{n}\Delta_i=(1+2\max|\tau|)(b-a)\,\epsilon
\end{aligned}$$

e quindi

$$\begin{aligned}
&\sum_{i=1}^{n}\beta_i\Delta_i+\sum_{i=1}^{n}\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))\Delta_i \\
&<\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i+(1+2\max|\tau|)(b-a)\,\epsilon \\
&<\sum_{i=1}^{n}\alpha_i\Delta_i+(1+3\max|\tau|)(b-a)\,\epsilon,
\end{aligned}$$

osservato che anche $\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i<(\max|\tau|)(b-a)\,\epsilon$. Allora, per (3.11),

$$\begin{aligned}
\sum_{i=1}^{n}\alpha_i\Delta_i \quad &+ \quad \sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i \\
&\leq \quad \mathrm{E}(\tau(X)) \\
&< \quad \sum_{i=1}^{n}\alpha_i\Delta_i+(1+3\max|\tau|)(b-a)\,\epsilon
\end{aligned}$$

e quindi, per (3.10),

$$\left|\mathrm{E}(\tau(X))-\int_a^b\tau(x)f(x)dx\right|<(1+3\max|\tau|)(b-a)\,\epsilon.$$

Ne segue, per l'arbitrarietà di ϵ, la tesi. $\square$

Osservazione 3.4.5 (i) La proposizione (vii) del Teorema 3.4.2 fornisce una espressione della speranza matematica di una v.a. avente rango numerabile tramite la serie assolutamente convergente di termine generale il prodotto di un valore possibile per la probabilità che la v.a. assuma quel valore. Alla luce di questo risultato, viene allora naturale estendere la nozione di speranza matematica a quelle v.a. *non limitate* U con un numerabile di valori possibili $u_1, u_2, \ldots$ tali che la serie $\sum_{n\geq 1} u_n \Pr(U = u_n)$ sia assolutamente convergente (e quindi convergente e permutabile) ponendo:

$$\mathrm{E}(U) = \sum_{n\geq 1} u_n \Pr(U = u_n).$$

In questo modo è possibile, con riferimento alle estrazioni con rimessa da un'urna e alla σ-algebra $\mathcal{E}_L(\mathcal{T})$ (Sezione 2.6.2), considerare la speranza matematica della v.a. T tempo di attesa del primo successo. Osservato che

$$nq^{n-1} = \sum_{\substack{m+m'=n-1 \\ m,m'\geq 0}} q^m q^{m'}$$

per ogni $n \geq 1$, otteniamo che la serie $\sum_{n\geq 1} nq^{n-1}$ è il prodotto alla Cauchy di due serie geometriche di ragione $q > 0$ e quindi, per il teorema di Mertens[23],

$$\sum_{n\geq 1} nq^{n-1} = [\sum_{m\geq 0} q^m]\,[\sum_{m'\geq 1} q^{m'}] = \left(\frac{1}{p}\right)^2.$$

Conseguentemente, per (2.13),

$$\mathrm{E}(T) = \sum_{n\geq 1} n \Pr(T = n) = \sum_{n\geq 1} npq^{n-1} = p \sum_{n\geq 1} nq^{n-1} = \frac{1}{p}$$

Dunque il tempo medio di attesa del primo successo è inversamente proporzionale alla probabilità di avere un successo. Inoltre, per valori di p del tipo $\frac{1}{m}$ $(m > 1)$, l'uguaglianza ottenuta diviene estremamente significativa

[23]Ricordiamo che il *prodotto alla Cauchy* delle serie $\sum_{m\geq 0} a_m$ e $\sum_{m\geq 0} b_m$ è la serie di termine generale $c_m = a_0 b_m + a_1 b_{m-1} + \cdots + a_{m-1} b_1 + a_m b_0$. Per *teorema di Mertens* intendiamo il teorema: se la serie $\sum_{m\geq 0} a_m$ converge assolutamente ad a e se la serie $\sum_{m\geq 0} b_m$ converge assolutamente a b, allora la serie prodotto alla Cauchy $\sum_{m\geq 0} c_m$ converge ad ab.

in quanto assicura che si devono effettuare *in media* m estrazioni per veder estratta per la prima volta una pallina bianca; se poi si considerano più di m estrazioni, ci si attenderà, essendo il processo privo di memoria, di estrarre *in media* una pallina bianca ogni qualvolta si fanno m estrazioni.

È importante osservare che questa interpretazione del tempo medio di attesa del primo successo è strettamente connessa con le estrazioni con rimessa. Infatti, se consideriamo, ad esempio, estrazioni con contagio unitario ($a = 1$) da un'urna di due palline, otteniamo $p = \frac{1}{2}$ e, per (2.14),

$$p_n = \Pr(T = n) = \frac{\binom{-1}{1}\binom{-1}{n-1}}{\binom{-2}{n}\binom{n}{1}} = \frac{n!}{n(n+1)!} = \frac{1}{n(n+1)}$$

per ogni $n \geq 1$. Osservato che

$$\begin{aligned}\sum_{n\geq 1} p_n &= \sum_{n\geq 1} \frac{1}{n(n+1)} = \sum_{n\geq 1}(\frac{1}{n} - \frac{1}{n+1}) = \lim_{n\to+\infty} \sum_{i=1}^{n}(\frac{1}{n} - \frac{1}{n+1}) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n}(1 - \frac{1}{n+1}) = 1,\end{aligned}$$

possiamo considerare, tramite il Teorema 2.4.1, una probailità numerabilmente additiva sulla σ-algebra $\mathcal{E}_{\mathrm{L}}(\mathcal{T})$ tramite la distribuzione:

$$\Pr(T = t) = \begin{cases} \frac{1}{t(t+1)} & \text{se } t = 1, 2, \ldots \\ 0 & \text{se } t = +\infty \end{cases}.$$

Ne segue che, in questo contesto, il tempo medio di attesa del primo successo non è 2 ($= p^{-1}$) ma infinito in quanto

$$\mathrm{E}(T) = \sum_{n\geq 1} n \Pr(T = n) = \sum_{n\geq 1} = \frac{1}{n+1} = +\infty,$$

notato che la serie in esame è il resto 2-simo della serie armonica.

(ii) Alla luce del Teorema 3.4.4, viene anche naturale estendere la nozione di speranza matematica alle v.a. *non limitate* U assolutamente continue con densità f tali che l'integrale improprio $\int_{-\infty}^{+\infty} xf(x)dx$ esista finito, ponendo:

$$\mathrm{E}(U) = \int_{-\infty}^{+\infty} xf(x)dx.^{24}$$

In questo modo è possibile considerare, ad esempio, la speranza matematica di una v.a. U distribuita secondo la normale di densità $f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})$. Infatti, considerati a, b tali che $a < b$, tramite il cambio di variabile $t \rightsquigarrow \frac{x-\mu}{\sqrt{2\sigma^2}}$ e posto $c = \sqrt{2\sigma^2}$, dal teorema d'integrazione per sostituzione otteniamo

$$\begin{aligned}
\int_a^b x\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx &= c\int_{\frac{a-\mu}{c}}^{\frac{b-\mu}{c}} (c\,t+\mu)e^{-t^2}dt \\
&= c\left[c\int_{\frac{a-\mu}{c}}^{\frac{b-\mu}{c}} te^{-t^2}dt + \mu\int_{\frac{a-\mu}{c}}^{\frac{b-\mu}{c}} e^{-t^2}dt\right] \\
&= c^2\left[-\frac{1}{2}e^{-t^2}\right]_{\frac{a-\mu}{c}}^{\frac{b-\mu}{c}} + c\,\mu\int_{\frac{a-\mu}{c}}^{\frac{b-\mu}{c}} e^{-t^2}dt
\end{aligned}$$

e quindi, passando al limite per $(a,b) \to (-\infty,+\infty)$, risulta

$$\int_{-\infty}^{+\infty} x\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx = c\,\mu\int_{-\infty}^{+\infty} e^{-t^2}dt = c\,\mu\sqrt{\pi} = \sqrt{2\pi\sigma^2}\mu,$$

ricordata l'uguaglianza fondamentale $\int_{-\infty}^{+\infty} e^{-t^2}dt = \sqrt{\pi}$. Conseguentemente, poichè l'integrale è finito, si ha $\mathrm{E}(U) = \int_{-\infty}^{+\infty} xf(x;\mu,\sigma^2)dx = \mu$. $\triangle$

3.5 Varianza e covarianza

Data una v.a. limitata X, dal Teorema 3.1.3 otteniamo $(X - \mathrm{E}(X))^2 \in \mathcal{V}_b$ e quindi possiamo considerare la speranza matematica:

$$\mathrm{Var}(X) = \mathrm{E}\big((X - \mathrm{E}(X))^2\big),$$

[24] Ricordiamo che, data una funzione g integrabile su ogni intervallo limitato, l'integrale improprio $\int_{-\infty}^{+\infty} g(x)dx$ è così definito: $\int_{-\infty}^{+\infty} g(x)dx = \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b g(x)dx$.

chiamata **varianza** (o **momento centrale secondo**) di X.

Nel teorema seguente riportiamo alcune proprietà basilari della varianza. In particolare, (iii) assicura che le v.a. con varianza nulla assumono quasi certamente il valore della loro speranza matematica; (v) che la varianza di un indicatore d'evento è il prodotto delle probabilità dell'evento e della sua negazione.

Teorema 3.5.1 *Sussistono le proposizioni seguenti:*

(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* α, β *reali;*

(iii) Se $\mathrm{Var}(X) = 0$*, allora* $\Pr(X = \mathrm{E}(X)) = 1$*;*

(iv) Se $\epsilon > 0$*, allora*

$$\Pr(|X - \mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2} \text{ (disuguaglianza di Bienaymé-Čebičev);}$$

(v) $\mathrm{Var}(|E|) = \Pr(E) \Pr(\overline{E})$*;*

(vi) Se X *è una v.a. con rango discreto, allora:*

$$\mathrm{Var}(X) = \sum_{x \in \mathrm{rg}(X)} (x - \mathrm{E}(X))^2 \Pr(X = x);$$

(vii) Se X *è una v.a. assolutamente continua con valori nell'intervallo chiuso* $[a, b]$ *e densità* f *continua su* $[a, b]$*, allora:*

$$\mathrm{Var}(X) = \int_a^b (x - \mathrm{E}(X))^2 f(x) dx.$$

DIMOSTRAZIONE Le proposizioni (vi), (vii) seguono dai teoremi 3.4.3 e 3.4.4 (ponendo $\tau(x) = (x - \mathrm{E}(X))^2$).

(i) Dal Teorema 3.4.2(i)) risulta $\mathrm{Var}(X) = \mathrm{E}\big(X^2 - 2\mathrm{E}(X)X + X^2\big) = \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

(ii) Per il Teorema 3.4.2(i),(v) si ha $\mathrm{Var}(\alpha X + \beta) = \mathrm{E}\big([(\alpha X + \beta) - \mathrm{E}(\alpha X + \beta)]^2\big) = \mathrm{E}\big([\alpha X + \beta - (\alpha \, \mathrm{E}(X) + \beta)]^2\big) = \mathrm{E}\big(\alpha^2 \, (X - \mathrm{E}(X))^2\big) = \alpha^2 \, \mathrm{Var}(X)$.

(iii) Conseguenza immediata del Teorema 3.4.2(x).

(iv) Dalla disuguaglianza di Markov (Teorema 3.4.2(viii)) risulta

$$\Pr(|X-\mathrm{E}(X)| \geq \epsilon) = \Pr((X-\mathrm{E}(X))^2 \geq \epsilon^2) \leq \frac{\mathrm{E}((X-\mathrm{E}(X))^2)}{\epsilon^2} = \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

(v) Per (i), E3 si ha $\mathrm{Var}(|E|) = \mathrm{E}(|E|^2) - \mathrm{E}(|E|)^2 = \mathrm{E}(|E|) - \mathrm{E}(|E|)^2 = \mathrm{E}(|E|)(1-\mathrm{E}(|E|)) = \Pr(E)\Pr(\overline{E})$. □

Osservazione 3.5.2 (i) La varianza può essere intesa come un *indice di dispersione*. Infatti, fissato un valore $\epsilon > 0$, la disuguaglianza di Bienaymé-Čebičev fornisce una limitazione superiore della probabilità che la v.a. assuma un valore che disti dalla speranza matematica per almeno ϵ (che diviene significativa se $\mathrm{Var}(X) < \epsilon^2$); inoltre, tramite al disuguaglianza di Kolmogorov (Teorema 3.4.2(ix)) otteniamo la limitazione inferiore:

$$\Pr(|X-\mathrm{E}(X)| > \epsilon) \geq \frac{\mathrm{Var}(X) - \epsilon^2}{\sup[(X-\mathrm{E}(X))^2]}$$

che diviene significativa se $\mathrm{Var}(X) > \epsilon^2$. Quindi la conoscenza della varianza fornisce un'informazione di quanto i valori possibili di una v.a. siano più o meno concentrati attorno alla corrispondente speranza matematica (che, come sappiamo, è invece un indice di posizione).

(ii) Chiamata **deviazione standard** di X la quantità $\mathrm{Ds}(X) = \sqrt{\mathrm{Var}(X)}$, poniamo $n\sigma_X$ al posto di ϵ nella disuguaglianza di Bienaymé-Čebičev. Otteniamo allora la disuguaglianza:

$$\Pr(|X-\mathrm{E}(X)| < n\mathrm{Ds}(X)) \geq 1 - \frac{1}{n^2}$$

che fornisce una limitazione inferiore (di calcolo immediato) della probabilità che la v.a. differisca dalla sua speranza matematica per meno di multipli della deviazione standard. Ad esempio, per $n = 2, 3, 4$ e 5 si hanno, nell'ordine, le seguenti limitazioni inferiori $0.75, 0.\overline{8}, 0.9375$ e 0.96. △

Osservazione 3.5.3 Urna di composizione incognita. La nozione di varianza consente un'analisi interessante del modello di estrazioni con rimessa da un'urna di *composizione incognita*. Da un'urna formata da s palline sia bianche che rosse, di cui non si conosca la composizione (percentuale di palline bianche presenti nell'urna), si effettuano n estrazioni con rimessa. Sia allora $\mathcal{P}_1$ la partizione formata dai casi elementari:

$$H_j : \text{la composizione dell'urna è } \theta_j = \frac{j}{s} \qquad (j = 0, \ldots, s).$$

Introdotti inoltre gli eventi:

E_h : nella h-sima estrazione viene estratta pallina bianca $\quad (h = 1, \dots, n)$,

denotiamo, come al solito, con $E_1' \wedge \cdots \wedge E_n'$ l'elemento generico della partizione $\mathcal{S}$ delle storie.

Al fine d'introdurre sulla σ-algebra $\mathcal{A} = \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ degli eventi logicamente dipendenti dalla partizione prodotto:

$$\mathcal{P} = \{H_j \wedge E_1' \wedge \cdots \wedge E_n' \neq \emptyset : H_j \in \mathcal{P}_1 \,\mathrm{e}\, E_1' \wedge \cdots \wedge E_n' \in \mathcal{S}\}$$

una probabilità "naturale", consideriamo su $\mathcal{P}_1$ una distribuzione (di probabilità) $(p_j)_{j \in \{1,\dots,s\}}$ e, tenuto conto che le estrazioni avvengono con rimessa, su $\mathcal{S}$ la distribuzione:

$$p^{(j)}(\underbrace{E_1' \wedge \cdots \wedge E_n'}_{k \text{ affermati}}) = \begin{cases} \theta_j^k (1-\theta_j)^{n-k} & \text{se } H_j \wedge E_1' \wedge \cdots \wedge E_n' \neq \emptyset \\ 0 & \text{altrimenti} \end{cases},$$

per ogni $j = 0, \dots, s$. Siamo dunque nello schema considerato nella seconda parte della Sezione 2.4 e quindi, per il Teorema 2.4.3, esiste una probabilità Pr su $\mathcal{A}$ tale che $p_j = \Pr(H_j)$ $(j = 0. \dots, s)$ e

$$p^{(j)}(\underbrace{E_1' \wedge \cdots \wedge E_n'}_{k \text{ affermati}} \mid H_j) = \theta_j^k (1-\theta_j)^{n-k} \qquad (\Pr(H_j) > 0).$$

Conseguentemente, gli eventi $E_1, \dots, E_n$ sono scambiabili con riferimento alla probabilità condizionata $\mathrm{P}(\cdot \mid H_j)$ e quindi, per il Teorema 2.6.1,

$$\Pr(\underbrace{E_{i_1}' \wedge \cdots \wedge E_{i_m}'}_{k \text{ affermati}} \mid H_j) = \theta_j^k (1-\theta_j)^{m-k} \quad (\Pr(H_j) > 0; 1 \leq m \leq n) \tag{3.12}$$

qualunque sia $j = 1, \dots, s$.

Considerata infine la funzione $f : H_j \wedge E_1' \wedge \cdots \wedge E_n' \to \theta_j$ di dominio $\mathcal{P}$, introduciamo il numero aleatorio $Z = [f]_\sim$ avente θ_j $(j = 0, 1, \dots, s)$ valori possibili, che rappresenta, da un punto di vista interpretativo, la sconosciuta "composizione dell'urna". Osservato che $\{Z = \theta_j\} = H_j \in \mathcal{A}$, si ha $\Pr(Z = \theta_j) = \Pr(H_j)$ $(j = 0, \dots, s)$. Dalla formula di disintegrazione (Teorema 2.3.2 (ii)) e (3.12) risulta allora

$$\begin{aligned} \Pr(E_i) &= \sum_{j:\Pr(H_j)>0} \Pr(E_i \mid H_j) \Pr(H_j) \\ &= \sum_{j:\Pr(H_j)>0} \theta_j \Pr(H_j) = \sum_{j=0}^{s} \theta_j \Pr(Z = \theta_j) \end{aligned}$$

e quindi, per il Teorema 3.4.2(vi),

$$\Pr(E_i) = \mathrm{E}(Z) \qquad (i = 1, \ldots, n). \tag{3.13}$$

Dati ora gli eventi E_m, E_r ($m \neq r$), da (3.12) e dal Teorema 3.4.3 (con $\tau(x) = x^2$), otteniamo

$$\begin{aligned}
\Pr(E_m \wedge E_r) &= \sum_{j:\Pr(Z=\theta_j)>0} \Pr(E_m \wedge E_r \,|\, H_j) \Pr(Z = \theta_j) \\
&= \sum_{j:\Pr(Z=\theta_j)>0} \theta_j^2 \Pr(Z = \theta_j) = \sum_{j=0}^{s} \theta_j^2 \Pr(Z = \theta_j) = \mathrm{E}(Z^2)
\end{aligned}$$

e quindi, per (3.13) e il Teorema 3.5.1(i),

$$\begin{aligned}
\Pr(E_m \wedge E_r) - \Pr(E_m)\Pr(E_r) &= \mathrm{E}(Z^2) - \Pr(E_m)\Pr(E_r) \\
&= \mathrm{E}(Z^2) - \mathrm{E}(Z)^2 = \mathrm{Var}(Z).
\end{aligned}$$

Conseguentemente, se $\mathrm{Var}(Z) > 0$, si ha $\Pr(E_m \wedge E_r) > 0$ da cui otteniamo $\Pr(E_m) > 0$, $\Pr(E_r) > 0$ e $\Pr(E_m \,|\, E_r) = \frac{\Pr(E_m \wedge E_r)}{\Pr(E_r)} > \Pr(E_m)$, cioè i due eventi sono correlati positivamente; se $\mathrm{Var}(Z) = 0$ (cioè se è quasi certa la composizione dell'urna (Teorema 3.5.1(iii)) e $\mathrm{E}(Z) > 0$, allora, per (3.13), $\Pr(E_m) > 0$, $\Pr(E_r) > 0$ e $\Pr(E_m \,|\, E_r) = \Pr(E_m)$, cioè i due eventi sono non correlati.

Supponiamo ora di effettuare le n estrazioni realizzando la storia $E_1' \wedge \cdots \wedge E_n'$ avente k affermazioni. Viene allora naturale chiedersi quale sia la composizione dell'urna più probabile vista tale evidenza, cioè per quale valore di j è massima la probabilità condizionata

$$\begin{aligned}
\Pr(H_j \,|\, \underbrace{E_1' \wedge \cdots \wedge E_n'}_{k \text{ affermati}}) &= \frac{\Pr(E_1' \wedge \cdots \wedge E_n' \,|\, H_j)}{\Pr(E_1' \wedge \cdots \wedge E_n')} \Pr(H_j) \\
&= \frac{\theta_j^k (1-\theta_j)^{n-k}}{\Pr(E_1' \wedge \cdots \wedge E_n')} \Pr(H_j),
\end{aligned}$$

ove le due uguaglianza sono ottenute, nell'ordine, ricorrendo alla formula di Bayes (Teorema 2.3.2(iv)) e a (3.12). Ovviamente, la soluzione del problema dipende da come è distribuita la probabilità nella partizione $\mathcal{P}_1$. Al fine di evitare complicazioni, ci limitiamo a determinarla nel caso simmetrico. Il problema diviene allora equivalente al seguente problema di massimo:

$$\max_{j=0,\ldots,s} \theta_j^k (1-\theta_j)^{n-k}.$$

Considerata ora la funzione $g(x) = x^k(1-x)^{n-k}$ su $[0,1]$, risulta, via derivazione, che assume il valore massimo in corrispondenza della *frequenza relativa di successo osservata* $x_0 = \frac{k}{n}$ che, in generale, non è detto sia una delle composizioni possibili dell'urna. In questo caso, tenuto conto che la funzione g è strettamente crescente prima di x_0 e strettamente decrescente dopo, si procede a confrontare tra di loro i valori di g corrispondenti alla massima composizione che precede x_0 e alla minima composizione che lo segue e scegliere quella di valore massimo. Comunque, nei casi reali, la numerosità dell'urna è di norma elevata per cui la frequenza relativa di successo osservata fornisce, nel caso simmetrico, una buona approssimazione della composizione più probabile vista l'evidenza. $\triangle$

Considerate ora due v.a. limitate X, Y, risulta $(X - \mathrm{E}(X))(Y - \mathrm{E}(Y)) \in \mathcal{V}_b$ e quindi possiamo considerare la speranza matematica:

$$\mathrm{Cov}(X,Y) = \mathrm{E}\big((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\big)$$

detta **covarianza** di X e Y. Riesce quindi $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ e, in particolare, $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

Il teorema seguente fornisce alcune proprietà basilari della covarianza. In particolare, (ii), (iii) mostrano che la covarianza è un funzionale bilineare su ${\mathcal{V}_b}^2$; (iv) assicura che la varianza della somma di v.a. a due a due **non correlate** (cioè di covarianza nulla) è uguale alla somma delle rispettive varianze; (v) invece che, se due eventi sono non correlati, lo sono pure i relativi indicatori (tenendo presente il Teorema 2.5.1))[25]; infine, la disuguaglianza di Cauchy-Schwarz fornisce un importante collegamento tra la covarianza di due v.a. e le rispettive varianze.

Teorema 3.5.4 *Sussistono le proposizioni seguenti:*

(i) $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$*;*

(ii) $\mathrm{Cov}(\alpha X, Y) = \alpha\,\mathrm{Cov}(X,Y)$ *per ogni* α *reale;*

(iii) $\mathrm{Cov}(\sum_{i=1}^n X_i, Y) = \sum_{i=1}^n \mathrm{Cov}(X_i, Y)$*;*

(iv) $\mathrm{Var}\big(\sum_{i=1}^n X_i\big) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i} \mathrm{Cov}(X_i, X_j)$*;*

[25]Evidentemente non vale il viceversa, in quanto gli indicatori sono non correlati anche quando gli eventi sono entrambi trascurabili.

(v) $\mathrm{Cov}(|E|,|F|) = \Pr(E \wedge F) - \Pr(E)\Pr(F)$*;*

(vi) Risulta:

$$\mathrm{Cov}(X,Y)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y) \quad \text{(Disuguaglianza di Cauchy-Schwarz).}$$

Inoltre, nel caso di uguaglianza e $\mathrm{Cov}(X,Y) \neq 0$*, ogni v.a. è quasi certamente una trasformata affine dell'altra di tipo crescente, se* $\mathrm{Cov}(X,Y) > 0$*, e decrescente, se* $\mathrm{Cov}(X,Y) < 0$*.*

Dimostrazione Poniamo, per semplicità, $\mu = \mathrm{E}(X)$ e $\nu = \mathrm{E}(Y)$.

(i) Dal Teorema 3.4.2(i),(v) risulta $\mathrm{E}\big((X-\mu)(Y-\nu)\big) = \mathrm{E}\big(XY - \nu X - \mu Y + \mu\nu\big) = \mathrm{E}(XY) - \nu\mu - \mu\nu + \mu\nu$ e quindi la tesi.

(ii) Per (i) e la linearità della speranza matematica si ha $\mathrm{Cov}(\alpha X, Y) = \mathrm{E}(\alpha XY) - \mathrm{E}(\alpha X)\nu = \alpha[\mathrm{E}(XY) - \mu\nu] = \alpha\,\mathrm{Cov}(X,Y)$.

(iii) Sia intanto $n = 2$. Posto $\mathrm{E}(Z) = \xi$, per la linearità della speranza matematica e (i), risulta $\mathrm{Cov}(X+Z,Y) = \mathrm{E}\big[(X+Z)Y\big] - \mathrm{E}(X+Z)\,\nu = \mathrm{E}(XY) + \mathrm{E}(ZY) - \mu\nu - \xi\nu = \big[\mathrm{E}(XY) - \mu\nu\big] + \big[\mathrm{E}(YZ) - \nu\xi\big] = \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,Z)$. Procedendo per induzione, supponiamo che l'uguaglianza sussista per $n \geq 2$ e proviamola per $n+1$. Risulta

$$\begin{aligned}
\mathrm{Cov}\Big(\sum_{i=1}^{n+1} X_i, Y\Big) &= \mathrm{Cov}\Big(\sum_{i=1}^{n} X_i + X_{n+1}, Y\Big) = \mathrm{Cov}\Big(\sum_{i=1}^{n} X_i, Y\Big) + \mathrm{Cov}(X_{n+1}, Y) \\
&= \sum_{i=1}^{n} \mathrm{Cov}(X_i, Y) + \mathrm{Cov}(X_{n+1}, Y) = \sum_{i=1}^{n+1} \mathrm{Cov}(X_i, Y).
\end{aligned}$$

(iv) Sia intanto $n = 2$. Posto $\mathrm{E}(X_i) = \mu_i$ $(i = 1, 2)$, dalla linearità della speranza matematica risulta

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= \mathrm{E}\big([(X_1+X_2) - \mathrm{E}(X_1+X_2)]^2\big) = \mathrm{E}\Big(\Big[\sum_{i=1}^{2}(X_i - \mu_i)\Big]^2\Big) \\
&= \mathrm{E}\Big(\sum_{i=1}^{2}(X_i - \mu_i)^2 + 2(X_1-\mu_1)(X_2-\mu_2)\Big) \\
&= \sum_{i=1}^{2} \mathrm{E}\big((X_i-\mu_i)^2\big) + 2\,\mathrm{E}\big((X_1-\mu_1)(X_2-\mu_2)\big) \\
&= \sum_{i=1}^{2} \mathrm{Var}(X_i) + 2\,\mathrm{Cov}(X_1, X_2).
\end{aligned}$$

Procedendo per induzione, supponiamo che l'uguaglianza sussista per $n \geq 2$ e proviamola per $n+1$. Posto $Y = \sum_{i=1}^{n} X_i$, da (iii) si ha

$$\begin{aligned} \mathrm{Var}\Big(\sum_{i=1}^{n+1} X_i\Big) &= \mathrm{Var}\big(Y + X_{n+1}\big) = \mathrm{Var}\big(Y\big) + \mathrm{Var}(X_{n+1}) + 2\,\mathrm{Cov}\big(Y, X_{n+1}\big) \\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i} \mathrm{Cov}(X_i, X_j) \\ &\qquad + \mathrm{Var}(X_{n+1}) + 2\sum_{i=1}^{n} \mathrm{Cov}(X_i, X_{n+1}) \end{aligned}$$

e quindi la tesi.

(v) Per (i), Teorema 1.11.1(i) e E3 si ha $\mathrm{Cov}(|E|, |F|) = \mathrm{E}(|E|\,|F|) - \mathrm{E}(|E|)\,\mathrm{E}(|F|) = \mathrm{E}(|E \wedge F|) - \mathrm{E}(|E|)\,\mathrm{E}(|F|) = \mathrm{Pr}(E \wedge F) - \mathrm{Pr}(E)\,\mathrm{Pr}(F)$.

(vi) Posto $U = X - \mu$ e $V = Y - \nu$, dobbiamo provare che riesce

$$\mathrm{E}(UV)^2 \leq \mathrm{E}(U^2)\mathrm{E}(V^2). \tag{3.14}$$

A tal fine, dato un numero reale t, consideriamo la v.a. $tU + V$. Allora, dalla solita linearità della speranza matematica risulta

$$\mathrm{E}\big((tU+V)^2\big) = \mathrm{E}(t^2U^2 + 2t\,UV + V^2) = \mathrm{E}(U^2)\,t^2 + 2\,\mathrm{E}(UV)\,t + \mathrm{E}(V^2). \tag{3.15}$$

Ne segue, per la positività (Teorema 3.4.2(iii)), che la disequazione a coefficienti reali

$$\mathrm{E}(U^2)\,t^2 + 2\,\mathrm{E}(UV)\,t + \mathrm{E}(V^2) \geq 0$$

ha come soluzione, data l'arbitrarietà di t, *ogni* t reale. Ora, se $\mathrm{E}(U^2) = 0$, deve essere $\mathrm{E}(UV) = 0$ [26] e quindi (3.14) è verificata. Se invece $\mathrm{E}(U^2) \neq 0$, l'equazione di secondo grado

$$\mathrm{E}(U^2)\,t^2 + 2\,\mathrm{E}(UV)\,t + \mathrm{E}(V^2) = 0 \tag{3.16}$$

può ammettere al più una sola soluzione. Dunque, il suo discriminante non può essere maggiore di zero, cioè $4\,(\mathrm{E}(UV)^2 - \mathrm{E}(U^2)\,\mathrm{E}(V^2)) \leq 0$ da cui otteniamo (3.14).

[26] In caso contrario le soluzioni della disequazione risulterebbero maggiori o uguali di $t' = -\frac{\mathrm{E}(V^2)}{2\mathrm{E}(UV)}$, se $\mathrm{E}(UV) > 0$, e minori o uguali di t', se $\mathrm{E}(UV) < 0$.

Assumiamo infine $\mathrm{Cov}(X,Y) \neq 0$ e che (3.14) sia un'uguaglianza. Allora, l'equazione (3.16) ammette l'unica soluzione:

$$t^* = -\frac{\mathrm{E}(UV)}{\mathrm{E}(U^2)}.$$

Ne segue, per (3.15), $\mathrm{E}\big((t^*U+V)^2\big) = 0$ e quindi, per il Teorema 3.4.2(x), $1 = \Pr(V = -t^*U) = \Pr(Y-\nu = -t^*(X-\mu)) = \Pr(Y = -t^*X + (t^*\mu+\nu))$, cioè Y è una trasformata affine di X crescente, se $\mathrm{E}(UV) > 0$, e decrescente, se $\mathrm{E}(UV) < 0$. Per simmetria, otteniamo allora che anche X è una trasformata affine di Y crescente, se $\mathrm{E}(UV) > 0$, e decrescente, se $\mathrm{E}(UV) < 0$. □

Osservazione 3.5.5 (i) Forniamo l'espressione della speranza matematica e della varianza della frequenza di successo S_n nei tre modelli di estrazione considerati nella Sezione 2.6. Per quanto riguarda la speranza matematica risulta, per linearità (Teorema 3.4.2(i)) e E3, $\mathrm{E}(S_n) = \sum_{i=1}^n \mathrm{E}(|E_i|) = \sum_{i=1}^n \Pr(E_i) = n\Pr(E_1) = np$. Passando a considerare la varianza, dai teoremi 3.5.4(iv),(v) e 3.5.1(v), risulta

$$\begin{aligned}
\mathrm{Var}(S_n) &= \sum_{i=1}^n \mathrm{Var}(|E_i|) + 2\sum_{i=1}^{n-1}\sum_{j>i}\mathrm{Cov}(|E_i|,|E_j|)\\
&= \sum_{i=1}^n \Pr(E_i)\Pr(\overline{E}_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}[\Pr(E_i\wedge E_j) - \Pr(E_i)\Pr(E_j)]\\
&= npq + 2\sum_{i=1}^{n-1}\sum_{j>i}[\Pr(E_i\wedge E_j) - p^2]\\
&= npq + 2\binom{n}{2}[\Pr(E_1\wedge E_2) - p^2]
\end{aligned}$$

e quindi

$$\mathrm{Var}(S_n) = npq + n(n-1)[\Pr(E_1\wedge E_2) - p^2].$$

• Estrazioni senza rimessa Da (2.9) (con $b \geq 2$, $2 \leq n \leq s$) otteniamo

$$\begin{aligned}
\Pr(E_1\wedge E_2) - p^2 &= \frac{\binom{b}{2}}{\binom{s}{2}} - p^2 = p\,\frac{b-1}{s-1} - p^2 = p\left[\frac{b-1}{s-1} - \frac{b}{s}\right]\\
&= p\,\frac{b-s}{s(s-1)} = p\,\frac{p-1}{s-1} = -\frac{pq}{s-1}
\end{aligned}$$

e quindi

$$\mathrm{Var}(S_n) = npq + n(n-1)\left[-\frac{pq}{s-1}\right] = np\,q\left[1 - \frac{n-1}{s-1}\right] = npq\,\frac{s-n}{s-1}.$$

• Estrazioni con rimessa Da (2.11) risulta $\Pr(E_1 \wedge E_2) - p^2 = 0$ e quindi

$$\mathrm{Var}(S_n) = npq.$$

• Estrazioni con contagio simmetrico Da (2.14) otteniamo

$$\begin{aligned}\Pr(E_1 \wedge E_2) - p^2 &= \frac{\binom{-\frac{b}{a}}{2}}{\binom{-\frac{s}{a}}{2}} - p^2 = p\,\frac{\frac{b}{a}+1}{\frac{s}{a}+1} - p^2 = p\left[\frac{b+a}{s+a} - \frac{b}{s}\right] \\ &= p\,a\,\frac{s-b}{s(s+a)} = a\,\frac{pq}{s+a}\end{aligned}$$

e quindi

$$\mathrm{Var}(S_n) = npq + n(n-1)\left[a\,\frac{pq}{s+a}\right] = npq\left[1 + \frac{(n-1)a}{s+a}\right] = npq\,\frac{s+na}{s+a}.$$

(ii) Sia X distribuita uniformemente nell'intervallo $[a, b]$ (Osservazione 3.2.2(ii)). Allora, X è assolutamente continua con densità $f(x) = (b-a)^{-1}$, se $a \leq x \leq b$, e $f(x) = 0$, altrimenti. Ne segue, per il Teorema 3.4.4,

$$\mathrm{E}(X) = \int_a^b x f(x) dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{1}{b-a}\,\frac{b^2-a^2}{2} = \frac{a+b}{2}$$

e quindi la speranza matematica è il punto medio dell'intervallo. Per calcolare la varianza ricorriamo al Teorema 3.5.1(vii) ottenendo

$$\mathrm{Var}(X) = \int_a^b (x - \mathrm{E}(X))^2 f(x) dx = \frac{1}{b-a}\int_a^b \left(x - \frac{a+b}{2}\right)^2 dx = \frac{(b-a)^2}{12}.$$

(iii) Sia X una v.a. semplice distribuita uniformemente sui primi n numeri naturali. Dal Teorema 3.4.2(vi) otteniamo allora

$$\mathrm{E}(X) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\,\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Inoltre, per il Teorema 3.5.1(vi), si ha

$$\begin{aligned}\mathrm{Var}(X) &= \frac{1}{n}\sum_{i=1}^{n}(i - \mathrm{E}(X))^2 = \frac{1}{n}[\sum_{i=1}^{n} i^2 - 2\mathrm{E}(X)\sum_{i=1}^{n} i + n\mathrm{E}(X)^2] \\ &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{2} + \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.\end{aligned}$$

Nel caso particolare che X sia il numero che esce nel lancio di un dado equilibrato, otteniamo $\mathrm{E}(X) = \frac{7}{2} = 3.5$ e $\mathrm{Var}(X) = \frac{35}{12} = 29.1\overline{6}$. △

3.6 Il Teorema di Bernoulli

A partire dall'opera di Gerolamo Cardano *De ludo aleæ* (1526), la probabilità è sempre stata collegata alla frequenza relativa osservata (rapporto tra il numero di eventi che si verificano tra gli n considerati e n); ci si attendeva infatti che, a fronte di una successione di prove, a lungo andare la frequenza relativa osservata fosse all'incirca uguale alla probabilità. Questa opinione, peraltro molto popolare, venne espressa con la cosiddetta *legge empirica del caso*:

> In una serie di prove ripetute nelle stesse condizioni, la frequenza relativa di un evento si avvicina alla probabilità dell'evento stesso, e l'approssimazione tende a migliorare con l'aumentare del numero delle prove.

che mette in relazione la frequenza relativa rilevata sperimentalmente con la probabilità definita aprioristicamente.

Nell'opera *Ars conjectandi* (pubblicata postuma nel 1713), Jakob Bernoulli prova il teorema seguente che fornisce una prima versione, in termini matematici precisi, della formulazione piuttosto vaga di questa legge e che è divenuto, da subito, uno dei risultati più celebri del calcolo delle probabilità[27].

Teorema 3.6.1 (di Bernoulli) *La successione $(E_n)_{n\geq 1}$ sia costituita da eventi equiprobabili e tali che la probabilità si fattorizzi sulla partizione generata $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$ per ogni $n \geq 2$ [28]. Allora, posto $S_n = \sum_{i=1}^{n} |E_i|$ $(n \geq 1)$ e*

[27] Nella lettera del 3 ottobre 1703 indirizzata a Gottfrieds Wilhelm Leibniz, Bernoulli scrive: "Anche la più stupida delle persone sa - per non so quale istinto di natura e senza nessun precedente ammaestramento - che, più cresce il numero delle osservazioni, minore è il pericolo di allontanarsi dal vero; tuttavia darne accurata dimostrazione matematica è indagine tutt'altro che spregevole.

Mi sono proposto inoltre di ricercare se, col crescere delle osservazioni, la probabilità di non scostarsi dal vero rapporto tenda asintoticamente ad un limite inferiore all'unità - nel qual caso vano sarebbe lo sforzo di desumere empiricamente la probabilità dal numero di casi osservati - o invece tenda alla certezza.

Ora ho trovato che le cose stanno in questo secondo modo. Sono in grado così di determinare qual'è il numero delle osservazioni da fare perchè risulti 100, 1000, 10000 volte più verosimile (e quindi moralmente certo) che il rapporto tra il numero dei casi così determinato corrisponda, anzi non corrisponda, alla probabilità vera - ciò che nell'uso della vita civile è sufficiente affinchè le nostre previsioni vengano dirette in ogni contingenza non meno scientificamente che nei giochi d'azzardo."

[28] Gli eventi $E_1, \ldots, E_n$ sono quindi, per il Teorema 2.5.3(ii), stocasticamente indipendenti per ogni $n \geq 2$.

indicato con p la probabilità comune, risulta:

$$\lim_{n\to+\infty} \Pr\left(|\, S_n - np\,| \le \epsilon\right) = 0$$
$$\lim_{n\to+\infty} \Pr\left(\left|\, \frac{S_n}{n} - p\,\right| < \epsilon\right) = 1,$$

qualunque sia $\epsilon > 0$.

Dimostrazione Per provare il primo limite, iniziamo con l'osservare che, per l'ipotesi di fattorizzabilità, riesce

$$\begin{aligned}\Pr(S_n = h) > Pr(S_n = h-1) \quad &\Leftrightarrow \quad \binom{n}{h} p^h q^{n-h} > \binom{n}{h} p^{h-1} q^{n-h+1} \\ &\Leftrightarrow \quad h < (n+1)p.\end{aligned}$$

Conseguentemente, il valore massimo della probabilità $p_n(\cdot) = \Pr(S_n = \cdot)$ si ottiene in corrispondenza ad un valore h_n^* tale che $p_n(h_n^* - 1) \le p_n(h_n^*)$ e $p_n(h_n^* + 1) \le p_n(h_n^*)$, cioè $np - q \le h_n^* \le (n+1)p$. Posto allora $d_n = h_n^* - np$ otteniamo $-q \le d_n \le p$ e quindi $|d_n| \le 1$; risulta inoltre

$$\Pr(S_n = h_n^*) = \binom{n}{np + d_n} p^{np+d_n} q^{nq-d_n}.$$

Ciò osservato, ricordando la formula di Stirling:

$$n! = \sqrt{2\pi n}\, n^n e^{-n} \left(1 + \frac{\theta_n}{12n}\right) \qquad (0 < \theta_n < 1)$$

e la relazione $\binom{n}{h} = \frac{n!}{h!\,(n-h)!}$, otteniamo

$$\begin{aligned}\binom{n}{np+d_n} \quad = \quad &\sqrt{\frac{n}{2\pi(np+d_n)(nq-d_n)}} \left(\frac{n}{np+d_n}\right)^{np+d_n} \left(\frac{n}{nq-d_n}\right)^{nq-d_n} \\ &\times \frac{1 + \frac{\theta_n}{12n}}{(1 + \frac{\theta_{np+d_n}}{12(np+d_n)})(1 + \frac{\theta_{nq-d_n}}{12(nq-d_n)})}\end{aligned}$$

da cui, posto

$$a_n = \sqrt{\frac{n}{2\pi(np+d_n)(nq-d_n)}}, \qquad b_n = \frac{1 + \frac{\theta_n}{12n}}{(1 + \frac{\theta_{np+d_n}}{12(np+d_n)})(1 + \frac{\theta_{nq-d_n}}{12(nq-d_n)})},$$

risulta

$$\Pr(S_n = h_n^*) = a_n \left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} b_n. \tag{3.17}$$

Ricordato pure la disuguaglianza $e^x \geq 1 + x$ per ogni x reale, si ha

$$\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n}$$
$$= \left(1 - \frac{d_n}{np+d_n}\right)^{np+d_n} \left(1 + \frac{d_n}{nq-d_n}\right)^{nq-d_n} \leq e^{-d_n} e^{d_n} = 1$$

$$\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} = \left(1 + \frac{d_n}{np}\right)^{-(np+d_n)} \left(1 - \frac{d_n}{nq}\right)^{-(nq-d_n)}$$
$$\geq \exp\left(-\frac{d_n}{np}(np+d_n)\right) \exp\left(\frac{d_n}{nq}(nq-d_n)\right) = \exp\left(-\frac{d_n^2}{npq}\right)$$

e quindi

$$\exp\left(-\frac{d_n^2}{npq}\right) \leq \left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} \leq 1.$$

Ne segue, passando al limite, per $n \to +\infty$, e tenendo conto che $|d_n| \leq 1$,

$$\lim_{n\to+\infty} \left[\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n}\right] = 1.$$

Osservato infine che $b_n \to 1$, per $n \to +\infty$, e che

$$\lim_{n\to+\infty} \sqrt{2\pi npq}\; a_n = \lim_{n\to+\infty} \sqrt{\frac{pq}{(p+\frac{d_n}{n})(q-\frac{d_n}{n})}} = 1,$$

da (3.17) otteniamo

$$\lim_{n\to+\infty} \sqrt{2\pi npq}\,\Pr(S_n = h_n^*) = 1. \tag{3.18}$$

Siamo ora finalmente in grado di provare il limite in esame. A tal fine, sia $\epsilon > 0$. Notato che nell'intervallo $[np-\epsilon, np+\epsilon]$ ci possono essere al più $2\epsilon+1$ valori di S_n, risulta

$$\Pr(|S_n - np| \leq \epsilon) = \sum_{h\in[np-\epsilon,np+\epsilon]} \Pr(S_n = h) \leq (2\epsilon+1)\Pr(S_n = h_n^*)$$

e quindi, per (3.18),

$$\Pr(|S_n - np| \leq \epsilon) \leq \frac{(2\epsilon+1)}{\sqrt{2\pi npq}} \left(\sqrt{2\pi npq}\ \Pr(S_n = h_n^*)\right) \to 0, \text{ per } n \to +\infty,$$

come volevasi.

Passiamo ora a considerare il secondo limite. Per l'ipotesi di fattorizzabilità, gli eventi $E_1, \ldots, E_{n+1}$ sono scambiabili; dunque, per il Teorema 2.6.1, $\Pr(E_i \wedge E_j) = \Pr(E_1 \wedge E_2) = p^2$ $(i \neq j)$ e quindi, per i teoremi 3.5.4(v),(iv) e 3.5.1(v), $\mathrm{Var}(S_n) = \sum_{i=1}^{n} \mathrm{Var}(|E_i|) = npq$; inoltre, per la linearità della speranza matematica (Teorema 3.4.2(i)), si ha $\mathrm{E}(\frac{S_n}{n}) = p$. Ne segue, tramite la disuguaglianza di Bienaymé-Čebičev (Teorema 3.5.4(vi)) e il Teorema 3.5.1(ii),

$$\Pr\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}(\frac{S_n}{n})}{\epsilon^2} = \frac{\frac{1}{n^2}\,npq}{\epsilon^2} = \frac{pq}{n\epsilon^2}. \tag{3.19}$$

Passando infine al limite, per $n \to +\infty$, otteniamo la seconda parte della tesi. La dimostrazione è così conclusa. □

Osservazione 3.6.2 (i) È di fondamentale importanza tenere presente che il Teorema di Bernoulli non risolve il problema di risalire dalla frequenza relativa osservata (*probabilità empirica*) alla probabilità (supposta ignota) degli eventi della successione - che corrisponde a un problema di *probabilità inversa* (ricavare la probabilità data la frequenza relativa) - bensì risolve un problema di *probabilità diretta* (prevedere le frequenze relative data la probabilità), insegnando a desumere (con alta probabilità) dalla probabilità degli eventi *nota a priori* (*fatta nell'ambito di un modello teorico*) quella empirica con un'approssimazione grande a piacere, qualora, da un punto di vista applicativo, le ipotesi di equiprobabilità e fattorizzabilità siano ragionevoli.[29]

[29]L'indebito passaggio dalla probabilità diretta alla probabilità inversa ha comportato, tra l'altro, l'utilizzazione del Teorema di Bernoulli come test per confermare o rigettare l'ipotesi iniziale (ad esempio, "moneta perfetta" con riferimento ad una successione di lanci di una data moneta), confermandola, se la probabilità empirica risulta molto vicina a quella prevista (nell'esempio $\frac{1}{2}$), e rigettarla, se ne è lontana. A tale proposito, in *Sul ruolo concettuale delle "considerazioni asintotiche" nella teoria delle probabilità* (1969), Bruno de Finetti scrive: "Questo ragionamento è assurdo (oppure è una formulazione falsata di un ragionamento corretto ma diverso). Infatti, se l'opinione iniziale consisteva effettivamente nel giudicare ugualmente probabili e stocasticamente indipendenti le successive prove,

(ii) Da (3.19) otteniamo la disuguaglianza

$$\Pr\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \geq 1 - \frac{pq}{n\epsilon^2}$$

che consente di determinare, nelle ipotesi del Teorema di Bernoulli, una valutazione n_0 del numero di osservazioni da fare affinchè la frequenza relativa differisca dalla probabilità comune per non più della soglia ϵ con probabilità almeno π. Basta infatti porre $1 - \frac{pq}{n\epsilon^2} \geq \pi$ per ottenere $n_0 \geq \frac{pq}{\epsilon^2(1-\pi)}$. △

ciò significava esattamente questo: che tale opinione non si doveva modificare *qualunque risultato si fosse osservato*; si trattava di una valutazione definitiva e impegnativa, non modificabile senza contraddirla; non si può dire che una modificazione è giustificata per il verificarsi di *un caso imprevisto* perchè tutti i casi, tutti i possibili risultati sulle diverse prove, erano previsti come possibili e l'indipendenza stocastica ammessa significava che nessun risultato avrebbe giustificato ripensamenti. Comportarsi diversamente sarebbe come non mantenere la promessa di aiutare un amico in caso di malattia ritenendosi giustificati dicendo che al momento della promessa non era prevedibile che subentrasse una malattia.".