ISBN 978-88-8303-924-9 (online)

EUT Edizioni Università di Trieste
via Weiss 21, 34128 Trieste
http://eut.units.it
https://www.facebook.com/EUTEdizioniUniversitaTrieste

Probabilità elementare

Silvano Holzer

EUT EDIZIONI UNIVERSITÀ DI TRIESTE

The subject of Probabilities belongs equally to
the science of Number and to that of Logic
George Boole, *An Investigation of the Laws of Thought*, 1853

Indice

Premessa

Il testo è una esposizione degli argomenti trattati nelle lezioni di natura teorica del corso di Calcolo delle Probabilità del 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 primo anno del corso magistrale in Scienze Statistiche e Attuariali.

Per rendere sufficientemente completa l'esposizione, sono state incluse delle parti opzionali (di norma scritte in carattere più piccolo), di carattere sia dimostrativo che concettuale, la cui lettura potrà essere omessa (e rimandata eventualmente ad un secondo momento) senza pregiudicare in alcun modo la comprensione della maggior parte del materiale esposto.

Il lettore che desiderasse approfondire e ampliare gli argomenti trattati potrà consultare le opere riportate nella bibliografia.

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

Silvano Holzer

Trieste, ottobre 2016

Capitolo 1

Descrizione dell'incertezza

Impostando il problema della descrizione dell'incertezza in termini logici, si introducono, tramite lo *stato di conoscenza*, le nozioni chiave di evento assoluto e condizionato come classi di equivalenza di enunciati. Analogamente, introdotta la nozione di partizione dell'evento certo, si definiscono le nozioni basilari di numero aleatorio e di vettore aleatorio, tramite classi di equivalenza costituite da funzioni di dominio partizioni dell'evento certo.

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);
- 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. $\sqrt{2}$ è un numero razionale;
2. Otello è geloso di Desdemona;
3. $7 + 3 = 10$;
4. L'ambo "7,45" uscirà nella prossima estrazione del lotto sulla ruota di Venezia;

5. Nei due lanci consecutivi di un dado[1] il 5 esce 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 dato individuo; 2, 5 enunciano il fatto che una certa relazione binaria sussiste per una data coppia di individui; 3, 6 enunciano il fatto che una certa relazione ternaria sussiste per una data 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:
- x è un numero razionale;
- x è geloso di y;
- x + y = z
- L'ambo "x,y" uscirà nella prossima estrazione del lotto sulla ruota di Venezia;
- nei due lanci consecutivi di un dado il numero x esce 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*)[2].

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

[1]Cubo con le facce punzonate da 1 a 6.

[2]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).

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*[3]). Indicandoli, rispettivamente, 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;
- il condizionale $p \rightarrow q$ 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

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

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 $(\sqrt{2})^2$ 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 e bicondizionali apparentemente paradossali come, ad esempio, "se $3 > 7$, allora Socrate era il maestro di Platone" e "10 è un numero pari se solo se Socrate era il maestro di Platone".

Osserviamo infine che solo l'implicazione materiale assicura sia la validità dell'usuale *procedura di andata e ritorno*[4] 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

[4] Per dimostrare che due enunciati hanno medesima estensione verificare che il primo implica il secondo e che il secondo implica il primo.

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 $\forall x \in U(P(x))$. 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 $\exists x \in U(P(x))$.

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:

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

Conseguentemente:
- la quantificazione universale $\forall x \in U(P(x))$ è vera se e solo se l'enunciato $P(u)$ è vero qualunque sia l'individuo u in U;
- la quantificazione esistenziale $\exists x \in U(P(x))$ 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 $\forall x \in U(P(x))$ e $\exists x \in U(P(x))$, 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 $P = (p_i)_{i \in I}$ con $I \neq \emptyset$, la:
- **congiunzione multipla** di P (in simboli $\bigwedge_{i \in I} p_i$), cioè l'enunciato "per ogni $i \in I$ vale p_i";
- **disgiunzione multipla** di 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 P è vera se e solo se ogni enunciato della famiglia P è vero;
- la disgiunzione multipla di P è vera se e solo se qualche enunciato della famiglia 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[5].

Data una famiglia bivariata di enunciati $P = (p_{ij})_{(i,j)\in I\times J}$ possiamo considerare la famiglia di enunciati $(\wedge_{j\in J} p_{ij})_{i\in I}$ e poi l'enunciato $\vee_{i\in I}(\wedge_{j\in J} p_{ij})$, oppure la famiglia di enunciati $(\vee_{i\in I} p_{ij})_{j\in J}$ e poi l'enunciato $\wedge_{j\in J}(\vee_{i\in I} p_{ij})$. È bene tenere presente che i due procedimenti conducono, in generale, a enunciati di estensione diversa. Infatti, posto $I = J =]0, +\infty[$ e considerata la famiglia di enunciati $(i < j)_{(i,j)\in I^2}$, otteniamo che gli enunciati $\vee_{i\in I}(\wedge_{j\in J}(i < j))$ e $\wedge_{j\in J}(\vee_{i\in I}(i < j))$ sono, rispettivamente, falso e vero.

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 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.

L1 $\bigwedge_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I} p_{\pi(i)}, \quad \bigvee_{i\in I} p_i \leftrightarrow \bigvee_{i\in I} p_{\pi(i)}$ (legge commutativa)

L2 $(p \wedge q) \wedge r \leftrightarrow p \wedge (q \wedge r), \quad (p \vee q) \vee r \leftrightarrow p \vee (q \vee r)$ (l. associativa)

L3 $p \wedge \bigvee_{i\in I} p_i \leftrightarrow \bigvee_{i\in I}(p \wedge p_i), \quad p \vee \bigwedge_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I}(p \vee p_i)$ (l. distributiva)

[5]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 risulterebbe $I \neq \emptyset$.

L4 $p \leftrightarrow \bigwedge_{i\in I} p_i, \quad p \leftrightarrow \bigvee_{i\in I} p_i$, se $p_i \leftrightarrow p$ per ogni $i \in I$ (l. di idempotenza)

L5 $p \wedge (p \vee q) \leftrightarrow p, \quad p \vee (p \wedge q) \leftrightarrow p$ (l. di assorbimento)

L6 $\neg \bigwedge_{i\in I} p_i \leftrightarrow \bigvee_{i\in I} \neg p_i, \quad \neg \bigvee_{i\in I} p_i \leftrightarrow \bigwedge_{i\in I} \neg p_i$ (leggi di De Morgan)

L7 $\neg(\neg p) \leftrightarrow p$ (l. della doppia negazione)

L8 $\neg p \vee p$ (l. del terzo escluso)

L9 $\neg(\neg p \wedge p)$ (l. di non contraddizione)

L10 $(p \to q) \leftrightarrow \neg p \vee q$ (l. di Filone di Megara)

L11 $\bigwedge_{i\in I}(p_i \wedge q_i) \leftrightarrow (\bigwedge_{i\in I} p_i) \wedge (\bigwedge_{i\in I} q_i), \quad \bigvee_{i\in I}(p_i \vee q_i) \leftrightarrow (\bigvee_{i\in I} p_i) \vee (\bigvee_{i\in I} q_i)$

L12 $\bigwedge_{i\in I_1\cup I_2} p_i \leftrightarrow (\bigwedge_{i\in I_1} p_i) \wedge (\bigwedge_{i\in I_2} p_i), \quad \bigvee_{i\in I_1\cup I_2} p_i \leftrightarrow (\bigvee_{i\in I_1} p_i) \vee (\bigvee_{i\in I_2} p_i)$

1.5 Eventi

Considerato uno **stato di conoscenza** $\mathcal{C}$ (cioè una famiglia di enunciati ritenuti veri (in via ipotetica o effettiva)), denotiamo, qualsiasi sia l'enunciato p, con $\mathcal{C} \models p$ la proposizione "lo stato di conoscenza $\mathcal{C}$ *costringe* l'enunciato p *ad essere vero*" (in breve *"p è vero sub $\mathcal{C}$"*)[6]. Sussistono allora le proposizioni:

C1 $\mathcal{C} \models p$, se $p \in \mathcal{C}$;

C2 $\mathcal{C} \models \neg p$ se e solo se $\mathcal{C}$ costringe p ad essere falso;

C3 $\mathcal{C} \models (p \wedge q)$ se e solo se $\mathcal{C} \models p$ e $\mathcal{C} \models q$;

C4 $\mathcal{C} \models (p \vee q)$ se almeno uno tra p e q è vero sub $\mathcal{C}$;

[6] 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.

C5 Se $\mathcal{C} \models (p \to q)$, allora $\mathcal{C} \models p$ implica $\mathcal{C} \models q$;

C6 $\mathcal{C} \models (p \leftrightarrow q)$ se solo se p, q hanno la medesima estensione sub $\mathcal{C}$.

Considerato inoltre un **linguaggio** $\mathcal{L}$ (cioè un insieme non vuoto di enunciati)[7], 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 ogni $p \in \mathcal{L}$, 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*. Poiché 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 di E coincida con quella di p.[8]

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 $\varnothing_{\mathcal{C}} = [p'']_{\mathcal{C}}$; 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 $\varnothing_{\mathcal{C}}$.

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 $\varnothing$. Inoltre, denoteremo sempre gli eventi con lettere latine maiuscole. △

[7] Da un punto di vista interpretativo, lo stato di conoscenza rappresenta l'insieme delle conoscenze di un soggetto interessato ad una data situazione incerta (ad esempio, il lancio di un dado), mentre il linguaggio raccoglie quegli enunciati relativi al problema in esame che sono di interesse per l'individuo (ad esempio, "esce un numero pari ", "esce un numero primo", "esce un numero maggiore di 3", ...).

[8] L'idea di associare l'evento ad uno stato di conoscenza, 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 stabilire 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".

Esempio 1.5.2. Sia $\mathcal{L}$ costituito dagli enunciati relativi al lancio di un dado:

a. esce il numero 2;	f. esce un numero maggiore di 4;
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 10;	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.

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

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

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: $\Omega_{\mathcal{C}_3} = \{\text{d,e,h,i}\}$, $\varnothing_{\mathcal{C}_3} = \{\text{c,f,l}\}$ e $G_1 = \{\text{a,b,g}\}$.

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: $\Omega_{\mathcal{C}_4} = \{\text{b,d,i}\}$, $\varnothing_{\mathcal{C}_4} = \{\text{c,f}\}$, $H_1 = \{\text{a,g}\}$, $H_2 = \{\text{e,h}\}$ e $H_3 = \{\text{l}\}$.

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 $\Omega_{\mathcal{C}_5} = \{\text{d,e,h,i}\}$ e quello impossibile $\varnothing_{\mathcal{C}_5} = \{\text{a,b,c,f,g,l}\}$. $\diamondsuit$

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 IL LINGUAGGIO $\mathcal{L}$ 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 L8 e di non contraddizione L9, l'evento certo $\Omega = [\,\neg p \vee p\,]_{\mathcal{C}}$ e quello impossibile $\varnothing = [\,\neg p \wedge p\,]_{\mathcal{C}}$ 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 del lemma seguente che assicura la compatibilità dell'equivalenza $\sim_{\mathcal{C}}$ sia con la negazione che con la congiunzione e disgiunzione (multiple o no). Precisiamo che, al fine di snellire l'esposizione, useremo le abbreviazioni metalinguistiche: "$\Rightarrow$" per "se $\ldots$, allora" e "$\Leftrightarrow$" per "se e solo se"; inoltre assumeremo implicitamente che le famiglie di enunciati che considereremo siano in $\mathcal{L}$.

Lemma 1.6.1. *Sussistono le proposizioni:*

(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, $\bigwedge_{i \in \emptyset} E_i = \Omega$, $\bigvee_{i \in \emptyset} E_i = \varnothing$ 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, tramite le proposizioni (i),(iii),(vii) e (ix), 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) $E \wedge \overline{E} = \varnothing, \quad E \vee \overline{E} = \Omega;$

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

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

(x) $E = (E \wedge F) \vee (E \wedge \overline{F});$

(xi) $E \vee F = E \vee (\overline{E} \wedge F) = F \vee (\overline{F} \wedge E);$

(xii) $\bigwedge_{i\in I}(E_i \wedge F_i) = (\bigwedge_{i\in I} E_i) \wedge (\bigwedge_{i\in I} F_i);$

(xiii) $\bigwedge_{i\in I_1\cup I_2} E_i = (\bigwedge_{i\in I_1} E_i) \wedge (\bigwedge_{i\in I_2} E_i);$

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

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

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

Sussistono inoltre anche le proprietà che si ottengono dalle (xii), (xiii) sostituendo la congiunzione multipla con la disgiunzione multipla.

Dimostrazione. Le proposizioni (i)÷(vi) seguono banalmente dalle leggi logiche omonime; la (vii) dalle L8, L9, mentre le (xii), (xiii) dalle L11, L12, rispettivamente.

(viii) La prima uguaglianza si ottiene dalla legge della doppia negazione L7. Per quanto riguarda la seconda, da quanto appena provato e da (vii), (vi) si ha $\overline{\Omega} = \overline{\Omega \vee \overline{\Omega}} = \overline{\Omega} \wedge \overline{\overline{\Omega}} = \overline{\Omega} \wedge \Omega = \varnothing$. Ne segue, passando alla terza uguaglianza, $\Omega = \overline{(\overline{\Omega})} = \overline{\varnothing}$.

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

(x) Da (ix),(vii),(iii) risulta $E = E \wedge \Omega = E \wedge (F \vee \overline{F}) = (E \wedge F) \vee (E \wedge \overline{F})$.

(xi) Da (x),(ii),(i),(v), si ha $E \vee F = E \vee [(F \wedge E) \vee (F \wedge \overline{E})] = [E \vee (F \wedge E)] \vee (F \wedge \overline{E}) = [E \vee (E \wedge F)] \vee (\overline{E} \wedge F) = E \vee (\overline{E} \wedge F)$ Per simmetria, si ottiene l'altra uguaglianza.

(xiv) Posto $I_1 = \{i \in I : E_i \neq \varnothing\}$ e $I_2 = \{i \in I : E_i \neq \Omega\}$, da (xiii), (iv), (ix) si ha $\bigvee_{i\in I} E_i = \bigvee_{i\in I_1\cup I_1^c} E_i = \bigvee_{i\in I_1} E_i \vee \bigvee_{i\in I_1^c} E_i = \bigvee_{i\in I_1} E_i \vee \varnothing =$

$\bigvee_{i \in I_1} E_i$ e $\bigwedge_{i \in I} E_i = \bigwedge_{i \in I_2 \cup I_2^c} E_i = \bigwedge_{i \in I_2} E_i \wedge \bigwedge_{i \in I_2^c} E_i = \bigwedge_{i \in I_2} E_i \wedge \Omega = \bigwedge_{i \in I_1} E_i$.

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

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

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 $E_1 \to E_3$; 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. ♢.

Il prossimo teorema assicura che l'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:*

(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* $\varnothing$ *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 seguenti cinque passi.

1. $E \to F \Leftrightarrow \overline{E} \vee F = \Omega$. Posto $E = [p]_{\mathcal{C}}$ e $F = [q]_{\mathcal{C}}$, dalla legge di Filone di Megara L10 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 $E \to F \Leftrightarrow \mathcal{C} \models p \to q \Leftrightarrow \Omega = [p \to q]_{\mathcal{C}} \Leftrightarrow \Omega = \overline{E} \vee F$.

2. $E \to F \Leftrightarrow \overline{F} \to \overline{E}$ discende dal passo precedente e dalla $\overline{E} \vee F = \overline{(\overline{F})} \vee \overline{E}$ (conseguenza del Teorema 1.6.2(i),(viii)).

3. $\overline{E} \vee F = \Omega \Rightarrow E = E \wedge F$. Supposto $\overline{E} \vee F = \Omega$, tramite il Teorema 1.6.2(ix),(iii),(vii), otteniamo $E = E \wedge \Omega = E \wedge (\overline{E} \vee F) = (E \wedge \overline{E}) \vee (E \wedge F) = \varnothing \vee (E \wedge F) = E \wedge F$.

4. $E = E \wedge F \Rightarrow F = E \vee F$. Supposto $E = E \wedge F$, tramite il Teorema 1.6.2(i),(v), risulta $E \vee F = (E \wedge F) \vee F = F \vee (F \wedge E) = F$.

5. $F = E \vee F \Rightarrow \overline{E} \vee F = \Omega$. Supposto $F = E \vee F$, per il Teorema 1.6.2(ii), (vii),(xvi), riesce $\overline{E} \vee F = \overline{E} \vee (E \vee F) = (\overline{E} \vee E) \vee F = \Omega \vee F = \Omega$.

(ii) Dal Teorema 1.6.2(iv) 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 il Teorema 1.6.2(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, per il Teorema 1.6.2(ii), $E = E \wedge (F \wedge G) = (E \wedge F) \wedge G = E \wedge G$ e quindi, per (i), $E \to G$. Dunque $\to$ è un ordinamento. Che $\varnothing, \Omega$ siano, rispettivamente, il minimo e il massimo deriva, tramite (i), dalle $\varnothing = \varnothing \wedge E$ 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(xii), $(\bigwedge_{i\in I} E_i) \wedge (\bigwedge_{i\in I} F_i) = \bigwedge_{i\in I}(E_i \wedge F_i) = \bigwedge_{i\in I} E_i$ 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 dal Teorema 1.6.2(iv), 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(xiv), 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 $E_{i_0} = \bigvee_{i\in\{i_0\}} H_i = \bigvee_{i\in I} H_i \to \bigvee_{i\in I} E_i$. $\square$

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 termini (detti **costituenti**) sono a due a due incompatibili (cioè $E_i \wedge E_j = \varnothing$ per ogni $i \neq j$). Nel seguito, data una partizione $\mathcal{P}$ dell'evento certo, denoteremo con $\mathcal{P}^*$ la famiglia dei **casi elementari** di $\mathcal{P}$, cioè dei costituenti di $\mathcal{P}$ che sono diversi dall'evento impossibile.

Esempio 1.8.1. (i) La più "piccola" partizione è la famiglia formata solo 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(xiv), anche la famiglia $\mathcal{P}^*$.

(ii) Con riferimento al lancio di un dado, consideriamo gli eventi:[9]

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

ove m è un fissato elemento di $\{1, \ldots, 6\}$. Allora, sono partizioni le famiglie $(E_{<m}, E_{=m}, E_{>m})$ e $(E_i)_{i=1,\ldots,6}$.

(iii) Con riferimento al gioco del lotto[10], consideriamo - relativamente alle estrazioni su una data ruota - gli eventi:

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

[9]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).

[10]Che consiste nell'estrazione sequenziale, senza rimessa, da un'urna (detta *ruota*) di 90 palline (numerate da 1 a 90) di cinque palline.

Allora, le famiglie:

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

sono partizioni costituite, rispettivamente da $\binom{90}{5}5! = \frac{90!}{85!} = 5.273.912.160$ e da $\binom{90}{5} = 43.949.268$ casi elementari[11].

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

$$\begin{aligned} E_{vi} &: A \text{ vince}, \quad E_{pe} : A \text{ perde}, \quad E_{pa} : A \text{ pareggia}\\ E_{ij} &: A \text{ segna } i \text{ reti e } B \text{ segna } j \text{ reti} \quad (i,j \ge 0).\end{aligned}$$

Se si assume che la partita si svolga regolarmente, sono partizioni le famiglie (E_{vi}, E_{pe}, E_{pa}) e $(E_{ij})_{i,j\ge 0}$. $\Diamond$

Dati un evento E e un caso elementare ω della partizione $\mathcal{P}$, dal Teorema 1.6.2(x) otteniamo $\varnothing \ne \omega = (\omega \wedge E) \vee (\omega \wedge \overline{E})$ 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:

- $\omega = \omega \wedge E$, cioè $\omega \to E$ (Teorema 1.7.2(i)), quando $\omega \wedge \overline{E} = \varnothing$;
- $\omega = \omega \wedge \overline{E}$, cioè $\omega \to \overline{E}$, quando $\omega \wedge E = \varnothing$;
- $\omega \wedge E \ne \varnothing$ e $\omega \wedge \overline{E} \ne \varnothing$.

Conseguentemente, 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).

Diremo che l'evento E è:

- **logicamente dipendente** da $\mathcal{P}$ se ogni caso elementare di $\mathcal{P}$ implica E oppure implica $\overline{E}$;
- **logicamente semidipendente** da $\mathcal{P}$ se esiste un caso elementare ω' di $\mathcal{P}$ che implica E oppure implica $\overline{E}$ ed esiste un caso elementare ω'' di $\mathcal{P}$ tale che $\omega'' \wedge E \ne \varnothing$ e $\omega'' \wedge \overline{E} \ne \varnothing$;
- **logicamente indipendente** da $\mathcal{P}$ se, per ogni caso elementare ω di $\mathcal{P}$, risulta $\omega \wedge E \ne \varnothing$ e $\omega \wedge \overline{E} \ne \varnothing$.[12]

---

[11] 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*.

[12] La risoluzione dell'incertezza descritta da $\mathcal{P}$ consente di risolvere l'incertezza anche per l'evento E 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.

Esempio 1.8.2. (i) Per i teoremi 1.6.2(viii) e 1.7.2(ii), gli eventi impossibile e certo sono logicamente dipendenti da ogni partizione dell'evento certo.

(ii) Con riferimento al gioco del lotto (Esempio 1.8.1(iii)), consideriamo gli eventi:

$$\begin{aligned} A &: \text{il primo numero estratto ha una sola cifra} \\ B &: \text{esce l'ambo 20, 50} \\ C &: \text{i numeri sono estratti in modo crescente.} \end{aligned}$$

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.

(iii) Con riferimento alla partita di calcio (Esempio 1.8.1(iv)), consideriamo gli eventi:

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

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}). $\Diamond$

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; inoltre non richiameremo più esplicitamente le proposizioni (i)$\div$(xii), (xv) e (xvi) del Teorema 1.6.2 come pure le proposizioni (i), (ii) del Teorema 1.7.2.

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

(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. (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

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

Osservato che, per ogni $\omega \to \overline{E}$, si ha $\omega = \omega \wedge \overline{E}$ e quindi $E \wedge \omega = \omega \wedge E = (\omega \wedge \overline{E}) \wedge E = \omega \wedge (\overline{E} \wedge E) = \omega \wedge \varnothing = \varnothing$, risulta anche

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

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

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

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

(ii) Considerato $I \subset \mathcal{P}^*$, poniamo $E = \bigvee_{\omega \in I} \omega$. Ora, se $I = \emptyset$, si ha $E = \varnothing \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ (Esempio 1.8.2(i)). Sia quindi $I \neq \emptyset$. Dato un arbitrario $\omega_0 \in \mathcal{P}$, sia intanto $\omega_0 \in I$. Allora, per il Teorema 1.6.2(xiii), risulta

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

e quindi $\omega_0 \to E$. Sia ora $\omega_0 \notin I$. Allora, $\omega_0 \wedge E = \bigvee_{\omega \in I} (\omega_0 \wedge \omega) = \varnothing$ e quindi $\omega_0 = (\omega_0 \wedge E) \vee (\omega_0 \wedge \overline{E}) = \omega_0 \wedge \overline{E}$, cioè $\omega_0 \to \overline{E}$. Ne segue $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. $\square$

Proviamo ora che l'insieme degli eventi logicamente dipendenti da una partizione include la partizione ed è chiuso per negazione e per congiunzioni e disgiunzioni (multiple o no).

Teorema 1.8.4. *Risulta $\mathcal{P} \subset \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. Inoltre, se E e E_i ($i \in I$) sono 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. L'inclusione $\mathcal{P} \subset \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ segue banalmente dal Teorema 1.8.3(ii). Dalla logica dipendenza di E otteniamo che ogni caso elementare implica $E = \overline{(\overline{E})}$ o $\overline{E}$, cioè $\overline{E} \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. Dalla logica dipendenza di E_i per ogni i, si ha, per il Teorema 1.8.3(i), $\bigvee_{i \in I} E_i = \bigvee_{i \in I} (\bigvee_{\omega \to E_i} \omega) = \bigvee_{\omega : \exists i \in I (\omega \to E_i)} \omega$ e quindi, per il Teorema 1.8.3(ii), $\bigvee_{i \in I} E_i \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$. Per quanto riguarda la congiunzione, basta osservare che $\bigwedge_{i \in I} E_i = \overline{\bigvee_{i \in I} \overline{E_i}}$ e tenere presente quanto appena provato. $\square$

Consideriamo infine l'applicazione set : $\mathcal{E}_{\mathrm{L}}(\mathcal{P}) \to 2^{\mathcal{P}^*}$ che associa ad ogni evento E logicamente dipendente da $\mathcal{P}$ l'insieme $\mathrm{set}(E)$ 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 $\mathcal{P}^*$ che muta la relazione di implicazione (tra eventi) in quella di inclusione (tra insiemi).[13]

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* $I \subset \mathcal{P}^*$;

(ii) $\mathrm{set}(\Omega) = \mathcal{P}^*$ *e* $\mathrm{set}(\varnothing) = \emptyset$;

(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. Che gli eventi considerati nelle sei proposizioni siano eventi appartenenti a $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ è assicurato dal Teorema 1.8.4.

Passando alla biunivocità, proviamo, osservato che la suriettività è conseguenza immediata di (i), 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 $I \subset \mathcal{P}^*$ e $E = \bigvee_{\omega \in I} \omega$. Allora, per il Teorema 1.8.3, $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$ e quindi, sempre per lo stesso 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 $\varnothing \neq \omega_0 = \omega_0 \wedge E = \omega_0 \wedge \bigvee_{\omega \in I} \omega = \bigvee_{\omega \in I} (\omega_0 \wedge \omega) = \varnothing$.

(ii) Segue da (i), osservato che $\varnothing = \bigvee_{\omega \in \emptyset} \omega$ e $\Omega = \bigvee_{\omega \in \mathcal{P}^*} \omega$.

(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 $\omega \in \mathrm{set}(\bigwedge_{i \in I} E_i)$.

[13] 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).

(iv) Sia intanto $\omega_0 \in \mathrm{set}(\bigvee_{i\in I} E_i)$. Allora $\varnothing \neq \omega_0 = \omega_0 \wedge \bigvee_{i\in I} E_i = \bigvee_{i\in I}(\omega_0 \wedge E_i)$ e quindi $\omega_0 \wedge E_{i_0} \neq \varnothing$ per qualche $i_0 \in I$. Notato che, per il Teorema 1.8.3(i), $E_{i_0} = \bigvee_{\omega\to E_{i_0}} \omega$ ($E_{i_0} \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$!), si ha $\varnothing \neq \omega_0 \wedge \bigvee_{\omega\to E_{i_0}} \omega = \bigvee_{\omega\to E_{i_0}}(\omega_0 \wedge \omega)$ e quindi esiste $\omega \to E_{i_0}$ tale che $\omega = \omega_0$, cioè $\omega_0 \in \mathrm{set}(E_{i_0}) \subset \bigcup_{i\in I} \mathrm{set}(E_i)$. 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) Risulta, per (iii) e (ii), $\mathrm{set}(\overline{E}) \cap \mathrm{set}(E) = \mathrm{set}(\overline{E} \wedge E) = \mathrm{set}(\varnothing) = \emptyset$ e, per (iv) e (ii), $\mathrm{set}(\overline{E}) \cup \mathrm{set}(E) = \mathrm{set}(E \vee \overline{E}) = \mathrm{set}(\Omega) = \mathcal{P}^*$. Da qui 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(xiii), $F = \bigvee_{\omega\in\mathrm{set}(F)} \omega = \bigvee_{\omega\in\mathrm{set}(E)} \omega \vee G = E \vee G$ e quindi $E \vee F = E \vee (E \vee G) = E \vee G = F$, cioè $E \to F$. L'implicazione rimanente segue dalla proprietà transitiva dell'implicazione tra eventi. □

1.9 Partizioni notevoli

Date le partizioni dell'evento certo $\mathcal{P}$ e $\mathcal{P}'$, diremo $\mathcal{P}$ **più fine di** $\mathcal{P}'$ (oppure $\mathcal{P}'$ **meno fine di** $\mathcal{P}$) se ogni caso elementare di $\mathcal{P}'$ è logicamente dipendente da $\mathcal{P}$. Conseguentemente, la frase "$\mathcal{P}$ 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 $\mathcal{P}$ esibisce, alla luce del Teorema di caratterizzazione 1.8.3(i), 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. Per snellire l'esposizione, data una partizione $\mathcal{P}$ e un evento E, usiamo la notazione $E \in \mathcal{P}$ per indicare che E è un costituente di $\mathcal{P}$.

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

Dimostrazione. Dato $\omega \in \mathcal{P}^*$, si ha $\omega = \omega \wedge \Omega = \omega \wedge \bigvee_{\omega'\in\mathcal{P}'} \omega' = \bigvee_{\omega'\in\mathcal{P}'}(\omega \wedge \omega')$ e quindi esiste $\omega' \in \mathcal{P}'$ tale che $\omega \wedge \omega' \neq \varnothing$[14]. □

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

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

[14] In caso contrario, per il Teorema 1.7.2(v),(ii), si avrebbe la contraddizione $\varnothing \neq \omega \to \varnothing$.

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

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

(ii) $\Rightarrow$ (i). Sia ω' un caso elementare di $\mathcal{P}'$. Considerato allora il caso elementare $\omega \in \mathcal{P}$, esiste un caso elementare $\omega'_0 \in \mathcal{P}'$ tale che $\omega \rightarrow \omega'_0$, cioè $\omega = \omega \wedge \omega'_0$. Quindi, se $\omega' = \omega'_0$, risulta $\omega \rightarrow \omega'$; se invece $\omega' \neq \omega'_0$, riesce $\omega = (\omega \wedge \omega') \vee (\omega \wedge \overline{\omega'}) = [(\omega \wedge \omega'_0) \wedge \omega'] \vee (\omega \wedge \overline{\omega'}) = [\omega \wedge (\omega'_0 \wedge \omega')] \vee (\omega \wedge \overline{\omega'}) = (\omega \wedge \emptyset) \vee (\omega \wedge \overline{\omega'}) = \varnothing \vee (\omega \wedge \overline{\omega'}) = \omega \wedge \overline{\omega'}$, cioè $\omega \rightarrow \overline{\omega'}$. Dunque, $\omega \rightarrow \omega'$ o $\omega \rightarrow \overline{\omega'}$. 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" ($n = 1, \ldots, 90$). Allora, $\mathcal{P}_{sec}$ è più fine delle rimanenti mentre non esiste alcun legame di finitezza tra $\mathcal{P}_{sem}$ e $\mathcal{P}_{pes}$. ◊

1.9.1 Partizione generata

Con riferimento all'evento generico E_i della famiglia $\mathcal{F} = (E_i)_{i \in I}$ e ad una qualsiasi applicazione f di dominio I e a valori in $\{0, 1\}$, poniamo $E_{i0} = \overline{E_i}$ e $E_{i1} = E_i$. Allora, il teorema seguente assicura che la famiglia di eventi $\mathcal{P}_{\mathrm{G}}(\mathcal{F}) = (\bigwedge_{i \in I} E_{if(i)})_{f \in \{0,1\}^I}$ è una partizione dell'evento certo, che chiameremo **partizione generata** da $\mathcal{F}$[15].

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 $E_{i_0 f(i_0)} \wedge E_{i_0 g(i_0)} = \varnothing$. Ne segue, per il Teorema 1.6.2(xv), $(\bigwedge_{i \in I} E_{if(i)}) \wedge (\bigwedge_{i \in I} E_{ig(i)}) = \bigwedge_{i \in I} (E_{if(i)} \wedge E_{ig(i)}) = \varnothing$.

Proviamo ora che $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$ è esaustiva. Risulta, per il Teorema 1.7.2(iv),(ii), $\bigvee_{f \in \{0,1\}^I} (\bigwedge_{i \in I} E_{if(i)}) = \bigwedge_{i \in I} (E_{i0} \vee E_{i1}) = \bigwedge_{i \in I} (\overline{E}_i \vee E_i) = \Omega$. □

[15] Ricordiamo che $\{0, 1\}^I$ denota l'insieme delle applicazioni di I in $\{0, 1\}$.

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 casi elementari sono $F = E \wedge F \wedge G, E \wedge \overline{F} \wedge G, E \wedge \overline{G} = E \wedge \overline{F} \wedge \overline{G}$ e $\overline{E} \wedge G = \overline{E} \wedge \overline{F} \wedge G$.

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

$$\begin{aligned} A &: \text{ viene estratto il numero 10} \\ B &: \text{ viene estratto l'ambo 20, 50} \\ C &: \text{ il terzo estratto è il numero 10.} \end{aligned}$$

Allora, i casi elementari sono $A \wedge B \wedge C$, $A \wedge B \wedge \overline{C}$, $A \wedge \overline{B} \wedge C$, $A \wedge \overline{B} \wedge \overline{C}$, $\overline{A} \wedge B \wedge \overline{C}$ e $\overline{A} \wedge \overline{B} \wedge \overline{C}$. ◊

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 $\mathcal{F} = (E_i)_{i \in I}$, sussistono le proposizioni:*

(i) $E_i \in \mathcal{E}_{\mathrm{L}}(\mathcal{P}_{\mathrm{G}}(\mathcal{F}))$ per ogni $i \in I$;

(ii) Sia $\mathcal{P}$ è 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. (i) Dato $i_0 \in I$, tramite il Teorema 1.7.2(vi) otteniamo $\bigwedge_{i \in I} E_{if(i)} \to E_{i_0}$, se $f(i_0) = 1$, e $\bigwedge_{i \in I} E_{if(i)} \to \overline{E}_{i_0}$, se $f(i_0) = 0$.

(ii) Dato $\omega \in \mathcal{P}^*$, 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, osservato che $\bigwedge_{i \in I} E_{if(i)}$ è un caso elementare di $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$. □

Il prossimo teorema, di interesse pratico, assicura che la presenza di eventi logicamente dipendenti in una data famiglia di eventi può essere eliminata ai fini dell'individuazione dei casi elementari della partizione generata.

Teorema 1.9.7. *Dato $I = I_1 \cup I_2$, sia la famiglia di eventi $\mathcal{F} = (E_i)_{i \in I}$ tale che E_i risulti logicamente dipendente dalla partizione generata dalla famiglia $\mathcal{F}_1 = (E_i)_{i \in I_1}$ per ogni $i \in I_2$. Allora, le partizioni generate $\mathcal{P}_{\mathrm{G}}(\mathcal{F})$ e $\mathcal{P}_{\mathrm{G}}(\mathcal{F}_1)$ hanno i medesimi casi elementari.*

Dimostrazione. Posto $\mathcal{P} = \mathcal{P}_{\mathrm{G}}(\mathcal{F})$ e $\mathcal{P}_1 = \mathcal{P}_{\mathrm{G}}(\mathcal{F}_1)$ proviamo intanto che $\mathcal{P}_1^* \subset \mathcal{P}^*$. Sia $\omega' = \bigwedge_{i\in I_1} E_{if(i)}$ ($f \in \{0,1\}^{I_1}$) un caso elementare di $\mathcal{P}_1$. Ora, per ogni $i \in I_2$, l'evento E_i è logicamente dipendente da $\mathcal{P}_1$ e quindi esiste $g(i) \in \{0,1\}$ tale che $\omega' \to E_{ig(i)}$. Allora, per il Teorema 1.7.2(iv), $\omega' \to \bigwedge_{i\in I_2} E_{ig(i)}$ e quindi $\omega' = \omega' \wedge \bigwedge_{i\in I_2} E_{ig(i)} = (\bigwedge_{i\in I_1} E_{if(i)}) \wedge (\bigwedge_{i\in I_2} E_{ig(i)})$, cioè ω' è un caso elementare di $\mathcal{P}$.

Proviamo ora che $\mathcal{P}^* \subset \mathcal{P}_1^*$. Poiché gli eventi della famiglia $\mathcal{F}$ sono tutti logicamente dipendenti da $\mathcal{P}_1$, la partizione $\mathcal{P}$ è, per il Teorema 1.9.6(ii), più fine di $\mathcal{P}_1$. Ciò osservato, sia ω un caso elementare di $\mathcal{P}$. Allora, per il Teorema 1.9.2, esiste $\omega' \in \mathcal{P}_1^* \subset \mathcal{P}^*$ tale che $\omega \to \omega'$. Ne segue $\varnothing \neq \omega = \omega \wedge \omega'$ da cui otteniamo $\omega = \omega'$ e quindi ω è un caso elementare di $\mathcal{P}_1$. □

Esempio 1.9.8. Con riferimento all'Esempio 1.9.5(iii), consideriamo gli eventi $D = \overline{E \wedge G} \vee F$ e $H = \overline{F} \wedge G$. Osservato che, per il Teorema 1.8.4, questi eventi sono logicamente dipendenti dalla partizione generata dagli eventi E, F, G, possiamo concludere, per il teorema precedente, che la partizione generate $\mathcal{P}_{\mathrm{G}}(E,F,G,D,H)$ ha come casi elementari quelli riportati nell'esempio in questione. ◊

1.9.2 Partizione prodotto

Con riferimento a due partizioni $\mathcal{P}$ e $\mathcal{P}'$ dell'evento certo, il teorema seguente assicura che la famiglia $\mathcal{P}\wedge\mathcal{P}' = (\omega\wedge\omega')_{(\omega,\omega')\in\mathcal{P}\times\mathcal{P}'}$ è una partizione dell'evento certo che chiameremo **partizione prodotto** di $\mathcal{P}$ e $\mathcal{P}'$.

Teorema 1.9.9. *Date le partizioni $\mathcal{P}$ e $\mathcal{P}'$, la famiglia $\mathcal{P}\wedge\mathcal{P}'$ è una partizione dell'evento certo.*

Dimostrazione. Proviamo intanto che gli elementi di $\mathcal{P}\wedge\mathcal{P}'$ sono a due a due incompatibili. Sia $\omega \wedge \omega' \neq \widetilde{\omega} \wedge \widetilde{\omega}'$. Allora, $\omega \neq \widetilde{\omega}$ oppure $\omega' \neq \widetilde{\omega}'$. Supposto, ad esempio, $\omega \neq \widetilde{\omega}$, risulta $(\omega \wedge \omega') \wedge (\widetilde{\omega} \wedge \widetilde{\omega}') = (\omega \wedge \widetilde{\omega}) \wedge (\omega' \wedge \widetilde{\omega}') = \varnothing$.

Per quanto riguarda l'esaustività, osserviamo che $\bigvee_{(\omega,\omega')\in\mathcal{P}\times\mathcal{P}'}(w \wedge \omega') = \bigvee_{\omega\in\mathcal{P}}[\bigvee_{\omega'\in\mathcal{P}'}(\omega \wedge \omega')] = \bigvee_{\omega\in\mathcal{P}}(\omega \wedge \bigvee_{\omega'\in\mathcal{P}'} \omega') = \bigvee_{\omega\in\mathcal{P}}(\omega \wedge \Omega) = \Omega$. □

Proviamo ora che la partizione prodotto è la partizione meno fine tra tutte quelle che sono più fini di $\mathcal{P}$ e $\mathcal{P}'$.

Teorema 1.9.10. *Date le partizioni $\mathcal{P}$ e $\mathcal{P}'$, la partizione prodotto $\mathcal{P} \wedge \mathcal{P}'$ è più fine di $\mathcal{P}$ e $\mathcal{P}'$. Inoltre, se $\mathcal{P}''$ è una partizione più fine di $\mathcal{P}$ e $\mathcal{P}'$, allora $\mathcal{P}''$ è più fine di $\mathcal{P} \wedge \mathcal{P}'$.*

Dimostrazione. Che $\mathcal{P} \wedge \mathcal{P}'$ sia più fine di $\mathcal{P}$ e $\mathcal{P}'$ segue banalmente dal Teorema 1.9.2. Considerato poi il caso elementare $\omega'' \in \mathcal{P}''$, esistono, sempre per lo stesso teorema, $\omega \in \mathcal{P}$ e $\omega' \in \mathcal{P}'$ tali che $\omega'' \rightarrow \omega$ e $\omega'' \rightarrow \omega'$. Allora, per il Teorema 1.7.2(iv), $\omega'' \rightarrow \omega \wedge \omega'$. Ne segue, per il Teorema 1.9.2, la tesi, una volta osservato che $\omega \wedge \omega'$ è un caso elementare di $\mathcal{P} \wedge \mathcal{P}'$. □

La nozione di partizione prodotto può essere introdotta anche per più di due partizioni. Considerate infatti le partizioni dell'evento certo $\mathcal{P}_i$ $(i \in I)$, la relativa partizione prodotto $\bigwedge_{i \in I} \mathcal{P}_i$ è costituita dagli eventi $\bigwedge_{i \in I} \omega_i$ tali che $\omega_i \in \mathcal{P}_i$ per ogni $i \in I$. Anche in questo caso (come si verifica adattando le dimostrazioni precedenti) la partizione prodotto è la partizione meno fine tra tutte quelle che sono più fini di $\mathcal{P}_i$, qualunque sia $i \in I$.

Esempio 1.9.11. (i) Il gioco del totocalcio richiede che chi vi partecipa pronostichi i risultati di 13 partite. Se A_i è la squadra che gioca in casa nell'i-sima partita, la relativa situazione d'incertezza può essere descritta dalla partizione costituita dagli eventi:

$$E_i^{(vi)} : A_i \text{ vince}, \quad E_i^{(pe)} : A_i \text{ perde}, \quad E_i^{(pa)} : A_i \text{ pareggia}.$$

La situazione d'incertezza complessiva relativa alle 13 partite, viene allora descritta dalla partizione prodotto $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_{13}$ avente $3^{13} = 1.594.323$ casi elementari.

(ii) Con riferimento a n lanci consecutivi di un dado, sia $\mathcal{P}_i$ (i = 1, . . . , n) la partizione costituita dagli eventi:

$$E_j^{(i)} : \text{esce il numero } j \text{ nel lancio } i\text{-simo} \qquad (j = 1, \ldots, 6).$$

La situazione d'incertezza complessiva relativa agli n lanci, viene allora descritta dalla partizione prodotto $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_n$ avente 6^n casi elementari. ♢

1.10 Eventi logicamente indipendenti

Tramite la nozione di partizione prodotto introduciamo ora quella di eventi logicamente indipendenti. Data una famiglia di eventi $(E_i)_{i \in I}$ con almeno due termini, 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.

Teorema 1.10.1. *Data la famiglia di eventi* $(E_i)_{i \in I}$ *con almeno due termini, 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 $E_{i'} \wedge E \neq \varnothing$, $\overline{E}_{i'} \wedge E \neq \varnothing$ e quindi, per il Teorema 1.6.2(xiii), $\bigwedge_{i \in I} E_{if(i)} = (\bigwedge_{i \in \{i'\}} E_{if(i)}) \wedge (\bigwedge_{i \in I \setminus \{i'\}} E_{if(i)}) = E_{i'f(i')} \wedge E \neq \varnothing$ per 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, $E_{i'} \wedge E = \bigwedge_{i \in I} E_{ig(i)} \neq \varnothing$, $\overline{E}_{i'} \wedge E = \bigwedge_{i \in I} E_{ih(i)} \neq \varnothing$ 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. □

Esempio 1.10.2. (i) Gli eventi A, B e C, considerati:
- nell'Esempio 1.8.2(ii), sono logicamente indipendenti, in quanto ogni costituente della partizione generata è un caso elementare;
- nell'Esempio 1.9.5(iv), non sono logicamente indipendenti, in quanto il costituente $\overline{A} \wedge B \wedge C$ (come pure il costituente $\overline{A} \wedge \overline{B} \wedge C$) è impossibile. ◊

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

Teorema 1.10.3. *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 1.10.1, $\bigwedge_{i \in I} E_{ig(i)} \neq \varnothing$ e quindi $E \neq \varnothing$, osservato che, per i teoremi 1.6.2(xiii) 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 1.10.1, la tesi. □

1.11 Eventi condizionati

Considerato lo stato di conoscenza $\mathcal{C}$, sia $H = [h]_{\mathcal{C}}$ tale che $H \neq \varnothing$. 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 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)[16].

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**[17]. Che la definizione data sia ben fondata[18] deriva facilmente dalla proposizione:

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

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

Esempio 1.11.1. 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 tutti gli enunciati ivi considerati. Posto allora $H = E_2 = [e]_{\mathcal{C}_1}$ (corrispondente all'incremento d'informazione: esce un numero pari) otteniamo $\mathcal{C}_1 \cup \{e\} = \mathcal{C}_2$ e quindi gli eventi condizionati a H sono $\mathcal{C}_2$-eventi; in particolare, $F_1 = E_1 \,|\, H = E_3 \,|\, H$, $\Omega_{\mathcal{C}_2} = E_2 \,|\, H = \Omega_{\mathcal{C}_1} \,|\, H$, $F_2 = E_4 \,|\, H$ e $\varnothing_{\mathcal{C}_2} = E_5 \,|\, H = \varnothing_{\mathcal{C}_1} \,|\, H$.

[16]Poiché 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.

[17]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 $\mathcal{C}$.

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

Posto ora $H = E_2 \wedge \overline{E}_4 = [e \wedge \neg f]_{\mathcal{C}_1}$ (corrispondente all'incremento d'informazione: esce il 2 oppure il 4) risulta $\mathcal{C}_1 \cup \{e \wedge \neg f\} = \mathcal{C}_3$ 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 $\varnothing_{\mathcal{C}_3} = E_4 \,|\, H = E_5 \,|\, H = \varnothing_{\mathcal{C}_1} \,|\, H$.

Scelto infine $H = E_2 \wedge \overline{E}_3 \wedge \overline{E}_4 = [e \wedge \neg b \wedge \neg f]_{\mathcal{C}_1}$ (relativo all'incremento d'informazione: esce il 4) riesce $\mathcal{C}_1 \cup \{e \wedge \neg b \wedge \neg f\} = \mathcal{C}_5$ e quindi gli eventi condizionati a H sono $\mathcal{C}_5$-eventi, cioè l'evento certo $\Omega_{\mathcal{C}_5} = E_2 \,|\, H = \Omega_{\mathcal{C}_1} \,|\, H$ e l'evento impossibile $\varnothing_{\mathcal{C}_5} = E_1 \,|\, H = E_3 \,|\, H = E_4 \,|\, H = E_5 \,|\, H = \varnothing_{\mathcal{C}_1} \,|\, H$. $\diamondsuit$

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 1.11.2. *Siano E, H, K eventi assoluti tali che $H \neq \varnothing$ e $K \,|\, H \neq \varnothing \,|\, H$. Sussistono allora le proposizioni:*

(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* $H \rightarrow E$;

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

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

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

DIMOSTRAZIONE. Poiché 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 $H \wedge K \neq \varnothing$; infatti, in caso contrario, si avrebbe $K \wedge H = \varnothing \wedge H$ da cui, tramite (ii), la contraddizione $K \mid H = \varnothing \mid H$.

Proviamo ora 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

$$(E \mid H) \mid (K \mid H) = [p]_{(\mathcal{C}\cup\{h\})\cup\{k\}} = [p]_{\mathcal{C}\cup\{h\wedge k\}},$$

una volta osservato che $(\mathcal{C} \cup \{h\}) \cup \{k\}$ e $\mathcal{C} \cup \{h \wedge k\}$ sono lo stesso stato di conoscenza. Ne segue $(E \mid H) \mid (K \mid H) = E \mid [h \wedge k]_{\mathcal{C}} = E \mid (H \wedge K)$. □

Osservazione 1.11.3. 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 $(E \wedge F) \mid H = (E \mid H) \wedge (F \mid H) = \varnothing \mid H$ da cui risulta, per il Teorema 1.11.2(ii), $(E \wedge F) \wedge H = \varnothing \wedge H$ e quindi $(E \wedge H) \wedge (F \wedge H) = \varnothing$, cioè l'incompatibilità condizionata si traduce nell'incompatibilità degli eventi $E \wedge H$ e $F \wedge H$. Analogamente, se $E \,|\, H \to F \,|\, H$, allora $(E \wedge F) \,|\, H = (E \,|\, H) \wedge (F \,|\, H) = E \,|\, H$ e quindi, sempre per il Teorema 1.11.2(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. △

1.12 Numeri e vettori aleatori

1.12.1 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*, poiché sono precisati sia il lancio che il dado considerati. Analogamente, lo è pure il primo numero estratto nella prossima estrazione del lotto su una data ruota (che potrà essere uno qualsiasi dei primi 90 numeri naturali). In entrambi i casi siamo quindi in presenza di un *numero aleatorio*[19], cioè di un numero X *sconosciuto per mancanza d'informazione, ma di valore ben determinato.*

Per passare ad una sua trattazione formale, conviene fare alcune considerazioni preliminari. Considerata la partizione $\mathcal{P}_{\mathrm{C}}(X)$ (chiamata **partizione canonica di** X) formata dai costituenti:

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

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 canonica e considerare la funzione che associa ad ogni caso elementare di $\mathcal{P}$ che implica $E_x^{(X)}$ il valore x[20]. Ad esempio, con riferimento al gioco del lotto (esempi

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

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

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 canonica), 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).

Questa molteplicità delle descrizioni è peraltro essenziale per consentire un trattazione agevole dei numeri aleatori. Per constatarlo, con riferimento al lancio simultaneo di due dadi distinguibili 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 $X + Y$. Allora, volendo riferirci a partizioni canoniche, 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 *sommato effettivamente* i possibili valori di X con quelli di Y; inoltre, *non indica esplicitamente* che il numero aleatorio Z è in effetti 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, poiché 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 prodotto $\mathcal{P}_{\mathrm{C}}(X) \wedge \mathcal{P}_{\mathrm{C}}(Y)$ e poniamo $\omega_{ij} = E_i^{(X)} \wedge E_j^{(Y)}$ $(i, j = 1, \ldots, 6)$. 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[21].

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

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 caso elementare $\omega_1 \wedge \omega_2 \in \mathcal{P}_1 \wedge \mathcal{P}_2$; 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 ω_1' e ω_2'.

[21] 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 $\frac{3}{6} + \frac{2}{6} = \frac{5}{6}$.

È proprio a quest'ultima osservazione che si rifà la definizione formale di numero aleatorio. Considerato l'insieme delle funzioni reali di dominio partizioni dell'evento certo costituite solo da casi elementari:

$$\mathbb{F}_1 = \{f^{(\mathcal{P})} : f^{(\mathcal{P})} \in \mathbb{R}^{\mathcal{P}} \,\text{con}\, \mathcal{P} = \mathcal{P}^* \,\text{partizione dell'evento certo}\},$$

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 caso elementare $\omega_1 \wedge \omega_2 \in \mathcal{P}_1 \wedge \mathcal{P}_2$. Osservato che tale relazione è una relazione di equivalenza sull'insieme $\mathbb{F}_1$, chiamiamo **numeri aleatori** le relative classi di equivalenza. Poiché funzioni equivalenti hanno medesimo insieme-imagine[22], chiamiamo **rango** (o insieme dei **valori possibili**) del numero aleatorio $X = [f^{(\mathcal{P})}]_\sim$ l'insieme-imagine $\mathrm{rg}(X) = f^{(\mathcal{P})}(\mathcal{P})$ della *versione* $f^{(\mathcal{P})}$. Chiamiamo infine **numeri certi** quei numeri aleatori X aventi rango formato da un solo elemento, detto **valore** di X[23].

Operazioni con i numeri aleatori

Definiti i numeri aleatori come classi di equivalenza, procediamo introducendo per essi le usuali operazioni aritmetiche. A tal fine, siano $f_i = f^{(\mathcal{P}_i)}$ versioni del numero aleatorio X e $g_i = g^{(\widetilde{\mathcal{P}}_i)}$ versioni del numero aleatorio Y $(i = 1, 2)$.

¶ Considerata una funzione reale τ di dominio $\mathrm{rg}(X)$, poniamo

$$\tau(X) = [\tau \circ f_1]_\sim.$$

La definizione è ben fondata; infatti, $\tau \circ f_1 \sim \tau \circ f_2$ in quanto $(\tau \circ f_1)(\omega_1) = \tau(f_1(\omega_1)) = \tau(f_2(\omega_2)) = (\tau \circ f_2)(\omega_2)$ per ogni caso elementare $\omega_1 \wedge \omega_2 \in \mathcal{P}_1 \wedge \mathcal{P}_2$.

In particolare, per $\tau(x) = \alpha x$, $\tau(x) = x^{-1}$ $(x \neq 0)$ e $\tau(x) = x^\alpha$ $(x > 0)$:

- αf_1 è la versione di αX relativa a $\mathcal{P}_1$;
- $\frac{1}{f_1}$ è la versione di $\frac{1}{X}$ relativa a $\mathcal{P}_1$, se $0 \notin \mathrm{rg}(X)$.
- f_1^α è la versione di X^α relativa a $\mathcal{P}_1$, se $\mathrm{rg}(X) \subset]0, +\infty[$.

[22]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 $\omega_1 \wedge \omega_2 \neq \varnothing$. 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.

[23]I numeri certi sono quindi quei particolari numeri aleatori che hanno come versioni funzioni costanti di medesimo valore.

¶ Passando alle operazioni di somma e prodotto di numeri aleatori, introduciamo la partizione $\mathcal{P}^{(1)}$ costituita dai casi elementari $\omega^{(1)} = \omega_1 \wedge \widetilde{\omega}_1 \in \mathcal{P}_1 \wedge \widetilde{\mathcal{P}}_1$ e supponiamo che φ sia una funzione reale di dominio $D_1 = \{(f_1(\omega_1), g_1(\widetilde{\omega}_1)) : \omega^{(1)} \in \mathcal{P}^{(1)}\}$. Considerata allora la funzione $\varphi_1 = \varphi(f_1, g_1)^{(\mathcal{P}^{(1)})}$ così definita:

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

per ogni $\omega^{(1)} \in \mathcal{P}^{(1)}$, poniamo

$$\varphi(X, Y) = [\varphi(f_1, g_1)^{(\mathcal{P}^{(1)})}]_\sim.$$

Per constatare che la definizione è ben fondata, introduciamo la partizione $\mathcal{P}^{(2)}$ costituita dai casi elementari $\omega^{(2)} = \omega_2 \wedge \widetilde{\omega}_2 \in \mathcal{P}_2 \wedge \widetilde{\mathcal{P}}_2$ e poniamo $D_2 = \{(f_2(\omega_2), g_2(\widetilde{\omega}_2)) : \omega^{(2)} \in \mathcal{P}^{(2)}\}$. Risulta allora $D_1 = D_2$. Infatti, dato $\omega^{(1)}$, esiste, per il Lemma 1.9.1, $\omega^{(2)}$ tale che $\varnothing \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \omega_2) \wedge (\widetilde{\omega}_1 \wedge \widetilde{\omega}_2)$ e quindi $\omega_1 \wedge \omega_2 \neq \varnothing$ e $\widetilde{\omega}_1 \wedge \widetilde{\omega}_2 \neq \varnothing$; dunque $f_1(\omega_1) = f_2(\omega_2)$ ($f_1 \sim f_2$!) e $g_1(\widetilde{\omega}_1) = g_2(\widetilde{\omega}_2)$ ($g_1 \sim g_2$!), cioè $(f_1(\omega_1), f_1(\widetilde{\omega}_1)) \in D_2$. Ne segue $D_1 \subset D_2$. Per simmetria risulta anche $D_2 \subset D_1$ e quindi $D_1 = D_2$. Possiamo allora considerare la funzione φ_2 di dominio $\mathcal{P}^{(2)}$ tale che:

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

A questo punto basta provare che φ_2 è equivalente a φ_1. A tal fine, sia $\varnothing \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \omega_2) \wedge (\widetilde{\omega}_1 \wedge \widetilde{\omega}_2)$. Allora, $f_1(\omega_1) = f_2(\omega_2)$ e $g_1(\widetilde{\omega}_1) = g_2(\widetilde{\omega}_2)$ e quindi $\varphi_1(\omega^{(1)}) = \varphi(f_1(\omega_1), g_1(\widetilde{\omega}_1)) = \varphi(f_2(\omega_2), g_2(\widetilde{\omega}_2)) = \varphi_2(\omega^{(2)})$.

In particolare, se φ denota la somma o il prodotto:

- $f_1(\omega_1) + g_1(\widetilde{\omega}_1)$ è la versione di $X + Y$ relativa a $\mathcal{P}^{(1)}$;
- $f_1(\omega_1)\, g_1(\widetilde{\omega}_1)$ è la versione di XY relativa a $\mathcal{P}^{(1)}$.

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

Relazioni d'ordine fra numeri aleatori

Adottando le medesime notazioni della sezione precedente, poniamo:

$$X \lhd Y \Leftrightarrow f_1(\omega_1) \lhd g_1(\widetilde{\omega}_1) \text{ per ogni } \omega_1 \wedge \widetilde{\omega}_1 \in \mathcal{P}^{(1)}$$

ove $\lhd$ denota uno qualunque dei simboli $<$, $\leq$[24].

[24] Qualora risulti $\omega_1 \wedge \widetilde{\omega}_1 \neq \varnothing$ per ogni ω_1 e $\widetilde{\omega}_1$, l'essere $X \lhd Y$ assicura che i ranghi $\mathrm{rg}(X)$ e $\mathrm{rg}(Y)$ formano una coppia di classi separate.

Per verificare che la definizione data è ben fondata, supponiamo $f_1(\omega_1) \lhd g_1(\widetilde{\omega}_1)$ per ogni $\omega^{(1)}$ e consideriamo un generico $\omega^{(2)}$. Esiste allora, per il Lemma 1.9.1, $\omega^{(1)}$ tale che $\varnothing \neq \omega^{(1)} \wedge \omega^{(2)} = (\omega_1 \wedge \widetilde{\omega}_1) \wedge (\omega_2 \wedge \widetilde{\omega}_2) = (\omega_1 \wedge \omega_2) \wedge (\widetilde{\omega}_1 \wedge \widetilde{\omega}_2)$ da cui otteniamo $f_2(\omega_2) = f_1(\omega_1)$, $g_2(\widetilde{\omega}_2) = g_1(\widetilde{\omega}_1)$ e quindi $f_2(\omega_2) \lhd g_2(\widetilde{\omega}_2)$.

Ovviamente, a differenza dell'usuale ordinamento per grandezza dei numeri reali, l'ordinamento $\leq$ che abbiamo ora introdotto non è totale (cioè due numeri aleatori non sono sempre confrontabili).

Eventi notevoli

Usando sempre le stesse notazioni delle ultime due sezioni, consideriamo, per ogni $A \subset \mathbb{R}$, l'evento:

$$\{X \in A\} = \bigvee_{\omega_1 \in \mathcal{P}_1 :\, f_1(\omega_1) \in A} \omega_1$$

che, da un punto di vista interpretativo, si verifica qualora il numero aleatorio X assume un valore appartenente all'insieme A.

Per constatare che la definizione data è ben fondata, proviamo che $E_1 = E_2$, avendo posto $E_i = \bigvee_{\omega_i \in \mathcal{P}_i :\, f_i(\omega_i) \in A} \omega_i$ $(i = 1, 2)$. Dato ω_1 tale che $f_1(\omega_1) \in A$, risulta

$$\omega_1 \wedge E_2 = \bigvee_{\omega_2 \in \mathcal{P}_2 :\, f_1(\omega_2) \in A} (\omega_1 \wedge \omega_2) = \bigvee_{\omega_2 \in \mathcal{P}_2 :\, \omega_1 \wedge \omega_2 \neq \varnothing \text{ e } f_2(\omega_2) \in A} (\omega_1 \wedge \omega_2).$$

Osservato che $f_2(\omega_2) = f_1(\omega_1)$ per ogni $\omega_1 \wedge \omega_2 \neq \varnothing$, otteniamo

$$\begin{aligned} \omega_1 \wedge E_2 &= \bigvee_{\omega_2 \in \mathcal{P}_2 :\, \omega_1 \wedge \omega_2 \neq \varnothing \text{ e } f_1(\omega_1) \in A} (\omega_1 \wedge \omega_2) = \bigvee_{\omega_2 \in \mathcal{P}_2 :\, \omega_1 \wedge \omega_2 \neq \varnothing} (\omega_1 \wedge \omega_2) \\ &= \bigvee_{\omega_2 \in \mathcal{P}_2} (\omega_1 \wedge \omega_2) = \omega_1 \wedge \bigvee_{\omega_2 \in \mathcal{P}_2} \omega_2 = \omega_1 \wedge \Omega = \omega_1. \end{aligned}$$

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

Osservato che, per il Teorema di rappresentazione 1.8.5, $\operatorname{set}(\{X \in A\}) = f_1^{-1}(A)$, proviamo, ricorrendo ancora allo stesso teorema, che le uguaglianze seguenti sussistono qualunque siano i sottoinsiemi A, B della retta reale.

$$\begin{aligned} \{X \in A\} \vee \{X \in B\} &= \{X \in A \cup B\} \\ \{X \in A\} \wedge \{X \in B\} &= \{X \in A \cap B\} \\ \overline{\{X \in A\}} &= \{X \in A^c\} \end{aligned} \tag{1.2}$$

Posto $E = \{X \in A\}$ e $F = \{X \in B\}$, proviamo la prima uguaglianza. Risulta $\mathrm{set}(\{X \in A \cup B\}) = f_1^{-1}(A \cup B) = f_1^{-1}(A) \cup f_1^{-1}(B) = \mathrm{set}(E) \cup \mathrm{set}(F) = \mathrm{set}(E \vee F)$. La seconda uguaglianza si prova in modo analogo, ricordato che $f_1^{-1}(A \cap B) = f_1^{-1}(A) \cap f_1^{-1}(B)$. Passando all'ultima uguaglianza, riesce $\mathrm{set}(\{X \in A^c\}) = f_1^{-1}(A^c) = (f_1^{-1}(A))^c = \mathrm{set}(\overline{E})$.

Un ruolo centrale sarà svolto nel capitolo terzo dagli eventi:

$$\begin{aligned} \{X \trianglelefteq \alpha\} &= \{X \in \{t \in \mathbb{R} : t \trianglelefteq \alpha\}\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, f_1(\omega_1) \trianglelefteq \alpha} \omega_1 \\ \{\alpha \trianglelefteq X\} &= \{X \in \{t \in \mathbb{R} : \alpha \trianglelefteq t\}\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, \alpha \trianglelefteq f_1(\omega_1)} \omega_1 \end{aligned}$$

ove $\alpha \in \mathbb{R} \cup \{-\infty + \infty\}$ e $\trianglelefteq$ denota uno qualsiasi dei simboli $\leq, <, =$[25]. Tramite questi eventi è allora possibile introdurre un ulteriore evento notevole (di chiaro significato intuitivo):

$$\{a \lhd X \prec b\} = \{a \lhd X\} \wedge \{X \prec b\} = \bigvee_{\omega_1 \in \mathcal{P}_1 \,:\, a \lhd f_1(\omega_1) \prec b} \omega_1$$

ove a, b ($a < b$) sono due numeri reali arbitrari e $\prec$ denota (come $\lhd$) uno qualunque dei simboli $<, \leq$[26].

Ricorrendo a (1.2) è ora agevole verificare che sussistono le uguaglianze seguenti (ove si suppone $a < c < b$):

$$\begin{aligned} \{X \lhd b\} &= \{X < a\} \vee \{a \leq X \lhd b\} = \{X \leq a\} \vee \{a < X \lhd b\} \\ \{X \leq b\} &= \{X < b\} \vee \{X = b\}, \quad \{b \leq X\} = \{b < X\} \vee \{X = b\} \\ \{X = b\} &= \{X \leq b\} \wedge \{b \leq X\} \\ \{x \leq X\} &= \overline{\{X < x\}}, \quad \{x < X\} = \overline{\{X \leq x\}} \\ \{a \leq X \lhd b\} &= \{a < X \lhd b\} \vee \{X = a\} \\ \{a \lhd X \leq b\} &= \{a \lhd X < b\} \vee \{X = b\} \\ \{a \lhd X \prec b\} &= \{a \lhd X \leq c\} \vee \{c < X \prec b\} = \{a \lhd X < c\} \vee \{c \leq X \prec b\} \end{aligned} \tag{1.3}$$

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

$$\{X < x\} = \bigvee_{n \geq 1} \{X \lhd a_n\}, \qquad \{X \leq x\} = \bigwedge_{n \geq 1} \{X \lhd b_n\} \tag{1.4}$$

[25] Eventi che, da un punto di vista interpretativo, si verificano qualora X assuma, rispettivamente, un valore non maggiore (non minore) di x, se $\trianglelefteq$ è $\leq$, e un valore minore (maggiore) di x, se $\trianglelefteq$ è $<$, e un valore uguale a x, se $\trianglelefteq$ è $=$.

[26] L'ultima uguaglianza si ottiene facilmente tramite (1.2).

con $(a_n)_{n\geq 1}$, $(b_n)_{n\geq 1}$ successioni reali tali che $a_n \uparrow x$ e $b_n \downarrow x$[27]:

Ricorrendo a (1.2), basta osservare, per provare la prima uguaglianza, che $f_1(\omega_1) < x \Leftrightarrow f_1(\omega_1) \lhd a_n$ per qualche $n \geq 1$ e, per provare la seconda, che $f_1(\omega_1) \leq x \Leftrightarrow f_1(\omega_1) \lhd b_n$ per ogni $n \geq 1$.

1.12.2 Indicatore di evento

Oltre ai numeri certi, 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 $\overline{E}$.[28] Ne segue che $|\Omega|$ e $|\varnothing|$ sono numeri certi di valore 1 e 0, rispettivamente; 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 1.12.1. *Sussistono le proposizioni:*

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

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

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

(iv) $|E| \leq |F|$, *se* $E \rightarrow F$;

(v) $|\bigvee_{n\geq 1} E_n| = \sum_{n\geq 1} |E_n|$, *se* $E_i \wedge E_j = \varnothing\ (i \neq j)$.

Dimostrazione. Considerata come partizione di riferimento la partizione generata $\mathcal{P}_{\mathrm{G}}(E, F)$, le proposizioni (i) $\div$ (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

[27]La notazione $x_n \uparrow x$ ($x_n \downarrow x$) significa che $(x_n)_{n\geq 1}$ è una successione reale crescente (decrescente) convergente a x.

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

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 $E \to F$ si ha $E = E \wedge F$ e quindi, per (i), $|E| = |E \wedge F| = |E|\,|F| \leq |E|$.

(v) Si ottiene facilmente ricorrendo alle versioni, notato che i casi elementari della partizione generata $\mathcal{P}_{\mathrm{G}}((E_n)_{n \geq 1})$ sono tra i costituenti $E_n \wedge \bigwedge_{i \neq n} \overline{E}_i$ $(n \geq 1)$ e $\bigwedge_{n \geq 1} \overline{E}_n = \overline{\bigvee_{n \geq 1} E_n}$. □

1.12.3 Vettori aleatori

Passando dal caso unidimensionale a quello multidimensionale, consideriamo l'insieme delle applicazioni definite su partizioni formate solo da casi elementari e a valori nello spazio $\mathbb{R}^{m+1}$ delle $(m+1)$-uple reali $(m \geq 1)$:

$$\mathbb{F}_m = \{f^{(\mathcal{P})} : f^{(\mathcal{P})} \in (\mathbb{R}^{m+1})^{\mathcal{P}} \text{ con } \mathcal{P} = \mathcal{P}^* \text{ partizione dell'evento certo}\}$$

e chiamiamo $f_1^{(\mathcal{P}_1)}, f_2^{(\mathcal{P}_2)} \in \mathbb{F}_m$ **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 caso elementare $\omega_1 \wedge \omega_2 \in \mathcal{P}_1 \wedge \mathcal{P}_2$. Osservato che tale relazione è una relazione di equivalenza sull'insieme $\mathbb{F}_m$, chiamiamo **vettori aleatori** (**m-dimensionali**) le relative classi di equivalenza. Notato infine che funzioni equivalenti hanno medesimo insieme-imagine, chiamiamo **rango** (o insieme dei **valori possibili**) del vettore aleatorio $\mathbf{X} = [f^{(\mathcal{P})}]_\sim$ l'insieme-imagine $\mathrm{rg}(\mathbf{X}) = f^{(\mathcal{P})}(\mathcal{P})$ della *versione* $f^{(\mathcal{P})}$.

Dato un sottoinsieme qualsiasi A di $\mathbb{R}^{m+1}$, introduciamo l'evento:

$$\{\mathbf{X} \in A\} = \bigvee_{\omega \in \mathcal{P}:\, f^{(\mathcal{P})}(\omega) \in A} \omega.$$

che risulta ben definito. Per questi eventi sussistono le uguaglianze:

$$\begin{aligned} \{\mathbf{X} \in A\} \vee \{\mathbf{X} \in B\} &= \{\mathbf{X} \in A \cup B\} \\ \{\mathbf{X} \in A\} \wedge \{\mathbf{X} \in B\} &= \{\mathbf{X} \in A \cap B\} \\ \overline{\{\mathbf{X} \in A\}} &= \{\mathbf{X} \in A^c\} \end{aligned} \tag{1.5}$$

qualunque siano i sottoinsiemi A, B dello spazio delle m-ple reali[29].

Dati infine $m+1$ numeri aleatori $X_1 = [f_1^{(\mathcal{P}_1)}]_\sim, \ldots, X_{m+1} = [f_{m+1}^{(\mathcal{P}_{m+1})}]_\sim$, consideriamo la funzione f - di dominio la partizione $\mathcal{P}^{(m)}$ dei casi elementari $\omega^{(m)} = \omega_1 \wedge \cdots \wedge \omega_{m+1}$ della partizione prodotto $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_{m+1}$ - così definita:

$$f(\omega^{(m)}) = (f_1(\omega_1), \ldots, f_m(\omega_{m+1}))$$

e chiamiamo $(m+1)$**-upla aleatoria** il vettore aleatorio:

$$\underline{X} = (X_1, \ldots, X_{m+1}) = [f^{(\mathcal{P}^{(m)})}]_\sim.$$

Per constatare che la definizione è ben fondata, poniamo $X_i = [\widetilde{f}_i^{(\widetilde{\mathcal{P}}_i)}]_\sim$ $(i = 1, \ldots, m+1)$ e $\widetilde{f}(\widetilde{\omega}^{(m)}) = (\widetilde{f}_1(\widetilde{\omega}_1), \ldots, \widetilde{f}_m(\widetilde{\omega}_{m+1}))$ per ogni caso elementare $\widetilde{\omega}^{(m)} = \widetilde{\omega}_1 \wedge \cdots \wedge \widetilde{\omega}_{m+1}$ della partizione prodotto $\widetilde{\mathcal{P}}_1 \wedge \cdots \wedge \widetilde{\mathcal{P}}_{m+1}$. Al fine di verificare che $f \sim \widetilde{f}$, sia $\varnothing \neq \omega^{(m)} \wedge \widetilde{\omega}^{(m)} = \bigwedge_{i=1}^m (\omega_i \wedge \widetilde{\omega}_i)$. Allora, $\omega_i \wedge \widetilde{\omega}_i \neq \varnothing$ da cui otteniamo $f_i(\omega_i) = \widetilde{f}_i(\widetilde{\omega}_i)$ $(i = 1, \ldots, m+1)$ e quindi $f(\omega^{(m)}) = \widetilde{f}(\widetilde{\omega}^{(m)})$.

Concludiamo la sezione provando che sussiste l'uguaglianza:

$$\{X_1 \in A_1\} \wedge \cdots \wedge \{X_{m+1} \in A_{m+1}\} = \{\underline{X} \in A_1 \times \cdots \times A_{m+1}\} \tag{1.6}$$

qualunque siano i sottoinsiemi $A_1, \ldots, A_{m+1}$ della retta reale.

Poiché la dimostrazione per $m > 1$ è analoga a quella relativa a $m = 1$, ci limitiamo a considerare il caso bidimensionale. Osservato che la funzione f_i di dominio $\mathcal{P}^{(1)}$ tale che $f_i(\omega^{(1)}) = f_i^{(\mathcal{P}_i)}(\omega_i)$ per ogni $\omega^{(1)}$ è equivalente a $f_i^{(\mathcal{P}_i)}(i = 1, 2)$, risulta $\{X_1 \in A_1\}, \{X_2 \in A_2\} \in \mathcal{E}_\mathrm{L}(\mathcal{P}^{(1)})$ da cui, tramite il Teorema di caratterizzazione 1.8.5(iii), otteniamo

$$\begin{aligned}
\mathrm{set}(\{X_1 \in A_1\} \wedge \{X_2 \in A_2\}) &= \mathrm{set}(\{X_1 \in A_1\}) \cap \mathrm{set}(\{X_2 \in A_2\}) \\
&= \{\omega^{(1)} : f_1(\omega^{(1)}) \in A_1\} \cap \{\omega^{(1)} : f_2(\omega^{(1)}) \in A_2\} \\
&= \{\omega^{(1)} : (f_1(\omega^{(1)}), f_2(\omega^{(1)})) \in A_1 \times A_2\} \\
&= \{\omega^{(1)} : (f_1^{(\mathcal{P}_1)}(\omega_1), f_2^{(\mathcal{P}_2)}(\omega_2)) \in A_1 \times A_2\} \\
&= \{\omega^{(1)} : f(\omega^{(1)}) \in A_1 \times A_2\} = \mathrm{set}(\{\underline{X} \in A_1 \times A_2\}).
\end{aligned}$$

Osservazione 1.12.2. Oltre ai numeri e vettori aleatori, compaiono nelle applicazioni anche enti aleatori che assumono come valori possibili $\pm\infty$. Per passare a una loro formalizzazione basta considerare, al posto di $\mathbb{F}_m$ $(m \geq 1)$, l'insieme delle funzioni di dominio partizioni e a valori nella *retta reale*

[29] Le prove che la definizione è ben fondata e delle uguaglianze (1.5) sono del tutto analoghe a quelle fatte nel caso unidimensionale.

ampliata $\mathbb{R}\cup\{-\infty,+\infty\}$ e riproporre le definizioni introdotte nel caso reale. Sussistono allora, nel nuovo contesto dei numeri e vettori aleatori **estesi**, tutti i risultati ottenuti per i numeri e vettori aleatori, con la solita precauzione di evitare espressioni aritmetiche illecite (somma di infiniti di segno opposto e rapporti di infiniti). $\triangle$

1.13 Esercizi

1. Provare le leggi logiche L1÷L12, le proposizioni C1÷C6, il Lemma 1.6.1 e le uguaglianze considerate in (1.3).

2. Esprimere in modo più semplice l'evento: $*\,\overline{(E\vee\overline{F})\wedge F}\vee[\overline{\overline{G}\vee F}\wedge(E\wedge F)]$; $*\,E\vee\overline{E\vee\overline{\overline{F\vee\overline{\overline{G}}}}}$; $*\,(\overline{E}\wedge\overline{\overline{F}\vee G})\vee F$; $*\,\overline{E\wedge\overline{\overline{E\vee\overline{\overline{F\wedge\overline{\overline{G}}}}}}})$.

3. Provare che sussistono le uguaglianze ($I,J\neq\emptyset$):

$$(\bigvee_{i\in I}E_i)\wedge(\bigvee_{j\in J}F_j)=\bigvee_{(i,j)\in I\times J}(E_i\wedge F_j),\quad(\bigwedge_{i\in I}E_i)\vee(\bigwedge_{j\in J}F_j)=\bigwedge_{(i,j)\in I\times J}(E_i\vee F_j),$$

$$(\bigvee_{i\in I}E_i)\vee(\bigvee_{j\in J}F_j)=\bigvee_{(i,j)\in I\times J}(E_i\vee F_j),\quad(\bigwedge_{i\in I}E_i)\wedge(\bigwedge_{j\in J}F_j)=\bigwedge_{(i,j)\in I\times J}(E_i\wedge F_j).$$

4. Posto $E\Delta F=(E\wedge\overline{F})\vee(\overline{E}\wedge F)$, provare che: $*\,E\Delta F\to(E\Delta G)\vee(G\Delta F)$; $*\,E\Delta F=\overline{E}\Delta\overline{F}$.

5. Come devono essere gli eventi E,F affinchè: $*\,\overline{\overline{\overline{E}}\vee\overline{F}}\to\varnothing$; $*\,E\vee F\to E\wedge F$?

6. Date le partizioni $\mathcal{P}$, $\mathcal{P}'$ provare che: $*$ se $\mathcal{P}$ è più fine di $\mathcal{P}'$ e $\mathcal{P}'$ è più fine di $\mathcal{P}$, allora le due partizioni hanno gli stessi casi elementari; $*$ se $\mathcal{P}$ è più fine di $\mathcal{P}'$, allora $\mathcal{E}_{\mathrm{L}}(\mathcal{P}')\subseteq\mathcal{E}_{\mathrm{L}}(\mathcal{P})$.

7. Provare che la partizione generata da una partizione ha i medesimi casi elementari della partizione stessa.

8. Provare che, se $E\wedge F=E\wedge G$ e $E\vee F=E\vee G$, allora $F=G$ (suggerimento: ricorrere alla partizione generata).

9. Dati gli eventi incompatibili E e F, individuare gli eventi G tali che $G\to E\vee F$ e $F\to\overline{F}\vee\overline{G}$ (suggerimento: ricorrere alla partizione generata).

10. Determinare la partizione generata dagli eventi E,F,G, $\overline{E\wedge G}\vee F$ e $F\wedge\overline{G}$, sapendo che E,F sono esaustivi e $F\wedge G\to\overline{E}$.

11. Determinare la partizione generata dagli eventi E, F, G, H tali che: $*$ E incompatibile sia con F che con G e $E \vee F \to H$; $*$ G, H incompatibili, $H \to \overline{F}$ e $G \vee H \to E$; $*$ E, F incompatibili, G, H incompatibili, E, F, H esaustivi e $G \to E$; $*$ E, F esaustivi, $G \to \overline{F}$ e $H \to \overline{G}$; $*$ $E \to F \to G \to H$.

12. Con riferimento ad una partita di calcio tra le squadre A e B determinare la partizione generata dagli eventi:
- E : B non perde o pareggia, F : A perde,
- G : B vince uno a zero, H : B fa meno di 3 reti.

È l'evento K : "B segna due reti" logicamente dipendente da tali eventi?

13. Con riferimento ad una gara ciclistica a cui partecipano 30 ciclisti numerati da 1 a 30, individuare la partizione generata dagli eventi:
- E : il primo arrivato ha un numero dispari,
- F : il primo arrivato ha un numero minore di 20,
- G : il primo arrivato ha un numero non minore di 10.

14. Con riferimento al lancio simultaneo di due dadi distinguibili, individuare la partizione generata dagli eventi:
- E : la somma dei numeri usciti non supera 3,
- F : la somma dei numeri usciti è minore di 8,
- G : la somma dei numeri usciti è uguale a 5 o 6,
- H : la somma dei numeri usciti è maggiore di 8.

15. Siano E, F eventi possibili con F logicamente semidipendente dalla partizione generata da E, Cosa si può dire del legame logico sussistente tra E e la partizione generata da F? E se F fosse logicamente dipendente?

16. Rappresentare $|E_1 \vee E_2 \vee E_3 \vee E_4|$ solamente con gli indicatori $|E_i|$ $(i = 1, \ldots, 4)$.

17. Provare che riesce $|\bigwedge_{i \in I} E_i| = \min_{i \in I} |E_i|$ e $|\bigvee_{i \in I} E_i| = \max_{i \in I} |E_i|$.

18. Dati gli eventi E, F, G, fornire una versione del numero aleatorio $Y = \max(|\overline{E}|, |G| + |E \wedge F|)$.

19. Dati gli eventi incompatibili E, F e i numeri aleatori $X = |E|^3 + 1$ e $Y = |E| + 2|F|$, fornire una versione della coppia aleatoria (X, Y).

Capitolo 2

Valutazione dell'incertezza I

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 e 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:

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

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

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

Chiaramente $\{\varnothing, \Omega\}$ 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).

Con riferimento all'algebra $\mathcal{A}$, una funzione reale Pr di dominio $\mathcal{A}$ è una **probabilità su $\mathcal{A}$** se verifica gli assiomi:

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

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

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

Conveniamo che, nel seguito, Pr DENOTI UNA PROBABILITÀ SU $\mathcal{A}$.

Nell'osservazione seguente forniamo una giustificazione degli assiomi proposti ricorrendo allo *schema della scommessa*.

Osservazione 2.1.1. SCHEMA DELLA SCOMMESSA Per *scommessa relativa all'evento E* intendiamo un'operazione di scambio tra un importo certo $p \geq 0$ (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 $v|E|$ al prezzo p) e si *scommette contro E* se si *vende* la scommessa (cioè si vende $v|E|$ 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 $q = \frac{p}{v} \geq 0$ 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 $q \leq 1$[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[2].

[1]Infatti, per perdere certamente (sia che si scommetta su o contro) deve risultare $G^{(E)}(p, v; S) = vS(|E| - q) < 0$ $(S = 1, 2)$, cioè $1 - q < 0$.

[2]L'identificazione della probabilità con il grado di fiducia è sempre stata presente nello sviluppo del calcolo delle probabilità. Ad esempio, in *Ars conjectandi* (1713), Jakob Bernoulli scrive: "La certezza di qualunque cosa si può trattare o oggettivamente, cioè in

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:

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

sè - e in questo caso non mostra altro che la reale, presente o futura, presenza di quella cosa - o soggettivamente, cioè in rapporto a noi - e in questo caso consiste nella misura della nostra conoscenza di quella realtà.*[omissis]* La certezza soggettiva delle cose - quella relativa a noi - non è in tutti uguale, ma varia di molto in più e in meno. Le cose di cui siamo certi per rivelazione, riflessione, percezione sensoriale, esperienza, introspezione o per altri modi, cioè le cose per le quali non dovremmo avere dubbi sulla loro esistenza presente o futura, possiedono per noi la certezza massima e assoluta. Tutte le altre cose contengono, a seconda della nostra conoscenza, una misura incompleta di certezza, che può essere più grande o più piccola, a seconda che esistano maggiori o minori probabilità che una certa cosa sia stata, sia o sarà. La probabilità è, infatti, un grado di certezza e si differenzia dalla certezza come la parte dal tutto". Dunque, per Bernoulli, la probabilità misura il grado di certezza che abbiamo nel credere che un avvenimento avrà o non avrà luogo; potrà quindi variare da persona a persona e da momento a momento perchè legata all'informazione soggettiva in un dato istante. Analogamente, 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". Infine, in *Calcolo delle probabilità* (1950), Bruno de Finetti scrive, con riferimento all'esito di un evento: "possiamo attenderlo con un grado minore o maggiore di fiducia, attribuirgli cioè un grado più o meno alto di probabilità".

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, rispettivamente, dagli 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 - poiché 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).

Ora, poiché 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 comportino 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, 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 1.12.1(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[3]. $\triangle$

[3]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 di ciò che s'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 esplichi il significato della parola "valore", è verosimile ritenere che lo identificasse con quello di "giusto prezzo" (espresso in termini monetari). Si verrebbe così

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 la probabilità di un evento non si altera qualora si consideri la sua disgiunzione con un **evento trascurabile** (cioè di probabilità nulla)[4] oppure la sua congiunzione con un **evento quasi certo** (cioè di probabilità uno); la (x) assicura che disgiunzioni finite di eventi quasi impossibili sono trascurabili e di eventi quasi certi sono quasi certe; la (xi) fornisce infine due 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:*

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

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

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

(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) $\Pr\big(\bigwedge_{i=1}^n E_i\big) \geq \sum_{i=1}^n \Pr(E_i) + 1 - n$ (disuguaglianza di Bonferroni);

(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* $\Pr(E \wedge F) = \Pr(F)$;

---

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.

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

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

(xi) Sia $\mathcal{P} \subset \mathcal{A}$ *una partizione finita. 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, P3 e P2, si ha $1 = \Pr(\Omega) = \Pr(E) + \Pr(\overline{E}) \geq \Pr(E)$. Ne segue (i) e $\Pr(\varnothing) = 0$. Da qui (ii), tramite P2.

(iii) Poiché la tesi sussiste per $n = 1$, procediamo per induzione, assumendo che sussista per $n \geq 1$ e provandola per $n + 1$. Considerati gli eventi $E_1, \dots, E_{n+1}$ tali che $E_i \wedge E_j = \varnothing$ $(i \neq j)$, da P3 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) Poiché, per il Teorema 1.6.2(xi),(x), $E \vee F = E \vee (\overline{E} \wedge F)$ 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, $F = E \vee F = E \vee (\overline{E} \wedge F)$ e quindi, per P3 e P2, $\Pr(F) = \Pr(E) + \Pr(\overline{E} \wedge F) \geq \Pr(E)$.

(vi) Poiché la tesi sussiste per $n = 1$, assumiamo che sussista per $n \geq 1$ e proviamola per $n + 1$. Considerati gli eventi $E_1, \dots, E_{n+1}$, da (iv), (ii) si ha

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

(vii) Si procede per induzione in modo analogo a quanto fatto nella dimostrazione di (iii) (sostituendo la seconda e terza uguaglianza con $\leq$).

(viii) Poiché la tesi sussiste per $n = 1$, assumiamo che sussista per $n \geq 1$ e proviamola per $n+1$. Considerati gli eventi $E_1, \ldots, E_{n+1}$, da (iv) otteniamo

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

Osservato infine che, con riferimento all'insieme $I = \{1, \ldots, n+1\}$, 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 $\{1, \ldots, n\}$ e l'ultimo i sottoinsiemi di I contenenti l'elemento $n+1$ e aventi almeno due elementi, otteniamo

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

(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. □

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 $J = \{i \in I : \Pr(E_i) > 0\}$ è al più numerabile.*

Dimostrazione. Basta osservare che $J = \bigcup_{n\geq 2}\{i \in I : \Pr(E_i) > \frac{1}{n}\}$ e che l'insieme $\{i \in I : \Pr(E_i) > \frac{1}{n}\}$ 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 poterne calcolare la 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[5].

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 loro probabilità.

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 i teoremi 1.7.2(vi) e 1.6.2(xiii), $\bigvee_{i=1}^{n} E_i \to (\bigvee_{i=1}^{n} E_i) \vee (\bigvee_{i>n} E_i) = \bigvee_{n\geq 1} E_n$. Ne segue, passando al limite per $n \to +\infty$, la tesi. □

[5] 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 verificano 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[6].

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:*

(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. Allora:*

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

---

[6]La grande messe di risultati significativi che si ottengono (come vedremo nei capitoli seguenti) nell'ambito delle probabilità numerabilmente additive, ha indotto ad assumere "for mathematical convenience" la numerabile additività (nata nell'ambito della teoria della misura) come proprietà caratterizzante la nozione di probabilità (rafforzando così l'assiome A3). È bene tenere presente però che questa impostazione è del tutto arbitraria, come dichiarato da Andrej N. Kolmogorov in *Grundbegriffe der Wahrscheinlichkeitrechnung* (1933): "Noi ci limitiamo, arbitrariamente, a considerare quei modelli che soddisfano l'Assioma VI" (che è una formulazione equivalente della numerabile additività).

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 $m = 2$. Poichè $E_1 \to E_2$, risulta $E_1 \vee (E_2 \wedge \overline{E}_1) = (E_1 \vee E_2) \wedge (E_1 \vee \overline{E}_1) = (E_1 \vee E_2) \wedge \Omega = E_1 \vee E_2 = E_2$. Assumiamo ora che l'uguaglianza sussista per $m \geq 2$ e proviamola per $m+1$. Poichè $E_m \to E_{m+1}$, si ha

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

Considerato ora $m \geq 2$, tramite il Teorema 1.6.2(xiii), 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 non decrescente, dal Teorema 2.1.2(i) e da (i) 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)) risulta $\Pr(\bigvee_{i=1}^n E_i) \leq \sum_{i=1}^n \Pr(E_i) \leq \sum_{n\geq 1} \Pr(E_n)$. Ne segue, tramite (i), la tesi, osservato che la successione $(\bigvee_{i=1}^n E_i)_{n\geq 1}$ è non decrescente e che $\bigvee_{n\geq 1} E_n = \bigvee_{n\geq 1}(\bigvee_{i=1}^n E_i)$.

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

Osservazione 2.2.3. La continuità dal basso e dall'alto della probabilità 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, $\Pr(\bigvee_{i=1}^n E_i) = \sum_{i=1}^n \Pr(E_i)$. $\triangle$

2.3 Probabilità condizionata

Dato un evento non trascurabile (cioè di probabilità positiva) H, 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 1.11.2(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 P3 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

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

Dunque $\Pr(\cdot \mid H)$ è una probabilità su $\mathcal{A}$ che, qualora la probabilità Pr sia numerabilmente additiva[7], risulta numerabilmente additiva. Risulta infatti

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

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** su $\mathcal{A}$ che si ottiene aggiornando la **probabilità iniziale** $\Pr(\cdot)$ a seguito dell'informazione che l'evento non trascurabile H è vero, *e null'altro in più*.[8]

Nell'osservazione seguente forniamo una giustificazione della definizione di probabilità condizionata data mediante lo schema della scommessa.

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* $E \mid H$. Poiché 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 $p \geq 0$ 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 1.12.1(i) otteniamo

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

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

[8] A tale proposito, in *Teoria delle probabilità* (vol. II, 1970), Bruno de Finetti scrive: "Perciò è errato dire o pensare semplicisticamente (senza almeno sottintendere simili precauzioni) che $\Pr(E \mid H)$ è la probabilità di E dopo conosciuto H. Al momento in cui verremo a sapere che si è verificato H avremo in genere appreso altre circostanze suscettibili di influire sul nostro giudizio, e comunque già l'informazione sul verificarsi di H giunge quasi sempre inevitabilmente, in modo più o meno esplicito, arricchita di particolari e circostanze che alterano la conoscenza finale e quindi, verosimilmente, il giudizio di probabilità in quella situazione".

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 $E_{n+1} \,|\, H, \ldots, E_{n+m} \,|\, H$, 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 $E \,|\, H$ con, rispettivamente, quote eque q', q'', q, vincite v', v'', v e puntatori S', S'', S. Riesce allora

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

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 $-(-q' + qq'')$. 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. $\triangle$

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 a 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 [9]. 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:*

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

(ii) *Sia $H_1, \ldots, H_n \in \mathcal{A}$ una partizione finita. Riesce allora:*

$$\Pr(E) = \sum_{i:\Pr(H_i)>0} \Pr(E \mid H_i) \Pr(H_i) \qquad \text{(formula di disintegrazione)}^{10};$$

(iii) *Siano gli eventi $E_1, \ldots, E_{n+1}$ 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:*

$$\Pr(E \mid H) = \frac{\Pr(H \mid E)}{\Pr(H)} \Pr(E) \qquad \text{(formula di Bayes)};^{11}$$

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

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

[9] Coinvolgendo così tre stati di conoscenza: lo stato iniziale $\mathcal{C}$ e gli 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.

[10] 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).

[11] Il fattore $\Pr(H \mid E)$ viene denominato **fattore di verosimiglianza** di H per E.

(vi) CONGLOMERABILITÀ: *Siano* $H_1, \ldots, H_n$ *eventi a due a due incompatibili tali che* $\Pr(\bigvee_{i=1}^n H_i) > 0$. *Posto allora* $I = \{i : \Pr(H_i) > 0\}$, *risulta:*

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

DIMOSTRAZIONE. (i) Segue dai teoremi 1.11.2(iv),(v),(vi) e 2.1.2(ix),(v),(ii).

(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, poiché $E \wedge H_i \to H_i$, dalla monotonia (Teorema 2.1.2(v)) si ha $\Pr(E \wedge H_i) \leq \Pr(H_i)$ per ogni i. Sono pertanto nulli tutti gli addendi della somma precedente relativi ad eventi H_i trascurabili. Allora

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

(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 $\bigwedge_{i=1}^n E_i = (\bigwedge_{i=1}^k E_i) \wedge (\bigwedge_{i=k+1}^n E_i) \to \bigwedge_{i=1}^k E_i$. Ora, poiché 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

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

e quindi la tesi.

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

(v) Da (iii) si ha $\Pr(E \wedge H \wedge K) = \Pr(E \,|\, H \wedge K) \Pr(H \,|\, K) \Pr(K)$ e quindi

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

(vi) Posto $m = \min_{i\in I} \Pr(E \,|\, H_i)$ e $M = \max_{i\in I} \Pr(E \,|\, H_i)$, otteniamo, per additività e monotonia (Teorema 2.1.2(iii),(v)),

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

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. $\square$

Supposto che gli eventi $E \wedge H$, $H \wedge K$ non siano trascurabili, risulta

$$\Pr(E \,|\, H \wedge K) = \frac{\Pr((E \wedge K) \wedge H))}{\Pr(K \wedge H)} = \frac{\Pr(E \wedge K \,|\, H))}{\Pr(K \,|\, H)}$$

da cui, tramite la proposizione (iv) del teorema precedente, otteniamo la seguente generalizzazione della formula di Bayes:

$$\Pr(E \,|\, H \wedge K) = \frac{\Pr(K \,|\, E \wedge H)}{\Pr(K \,|\, H)} \Pr(E \,|\, H) \tag{2.2}$$

chiamata **proprietà iterativa della formula di Bayes**[12].

[12]Per giustificare questa denominazione, facciamo alcune considerazioni. Per l'Osservazione 1.11.3, la famiglia $\mathcal{A} \,|\, H = \{E \,|\, H : E \in \mathcal{A}\}$ degli eventi di $\mathcal{A}$ condizionati ad H è un'algebra e la funzione P di dominio $\mathcal{A} \,|\, H$ tale che $\mathrm{P}(E \,|\, H) = \Pr(E \,|\, H)$ è una probabilità su tale algebra. Dato allora l'evento condizionato $K \,|\, H$ tale che $\mathrm{P}(K \,|\, H) > 0$, possiamo applicare la regola di Bayes ottenendo

$$\mathrm{P}((E \,|\, H) \,|\, (K \,|\, H)) = \frac{\mathrm{P}((K \,|\, H) \,|\, (E \,|\, H))}{\mathrm{P}(K \,|\, H)} \mathrm{P}(E \,|\, H)$$

da cui, tramite il Teorema 1.11.2(viii), risulta (2.2). Dunque, tale formula può essere interpretata come il risultato della procedura iterativa seguente: dato l'evento non trascurabile H di $\mathcal{A}$, applicare la formula di Bayes alla probabilità Pr definita sugli eventi assoluti di $\mathcal{A}$ per ottenere la probabilità P definita sugli eventi condizionati di $\mathcal{A} \,|\, H$; dato poi l'evento condizionato non trascurabile (rispetto P) $K \,|\, H$ di $\mathcal{A} \,|\, H$, riapplicare la formula di Bayes per ottenere la probabilità $\mathrm{P}((\cdot \,|\, H) \,|\, (K \,|\, H))$ sull'algebra $(\mathcal{A} \,|\, H) \,|\, K \,|\, H$ degli eventi condizionati di $\mathcal{A} \,|\, H$ all'evento ipotesi $K \,|\, H$.

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 reale 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[13].

Teorema 2.4.1. *Sia $\mathcal{P} = (\omega_i)_{i\in I}$ una partizione discreta. Sia inoltre la sequenza numerica $(p_i)_{i\in I}$ tale che $p_i \geq 0$ per ogni $i \in I$ e $\sum_{i\in I} p_i = 1$*[14]*. Sia infine*

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

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. Riesce $\Pr(\Omega) = \sum_{i\,:\,\omega_i\to\Omega} p_i = \sum_{i\in I} p_i = 1$. Sia ora $(E_n)_{n\geq 1}$ una successione in $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ formata da eventi a due a due incompatibili e $E = \bigvee_{n\geq 1} E_n$. Allora, per il Teorema di caratterizzazione 1.8.3(i), $E_n = \bigvee_{i\,:\,\omega_i\in\,\mathrm{set}(E_n)} \omega_i$ per ogni n e $E = \bigvee_{i\,:\,\omega_i\in\,\mathrm{set}(E)} \omega_i$. Osservato che, per il Teorema di rappresentazione 1.8.5(iv),(iii),(ii), si ha $\mathrm{set}(E) = \bigcup_{n\geq 1} \mathrm{set}(E_n)$ e $\mathrm{set}(E_n) \cap \mathrm{set}(E_m) = \mathrm{set}(E_n \wedge E_m) = \mathrm{set}(\varnothing) = \emptyset$, se $n \neq m$, otteniamo

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

e quindi $\Pr(E) = \sum_{n\geq 1} \Pr(E_n)$[15]. □

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

[13] 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 il teorema fornise una caratterizzazione delle probabilità nel caso discreto (come facilmente si constata ricorrendo alla numerabile additività).

[14] Nel caso numerabile, la somma $\sum_{i\in I} p_i$ diviene la serie $\sum_{n\geq 1} p_n$.

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

(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,

$$\Pr(E) = \sum_{i\,:\,\omega_i \to E} p_i = \frac{\#\,\mathrm{set}(E)}{n} \tag{2.3}$$

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[16].

(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 di esprimere un giudizio di simmetria sui casi elementari[17]. △

Il prossimo teorema riguarda invece la costruzione di una probabilità sull'algebra degli eventi logicamente dipendenti dalla partizione prodotto di un numero finito di partizioni, che supponiamo tutte finite. Date quindi le

[16]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.".

[17]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).

partizioni finite $\mathcal{P}_1, \ldots, \mathcal{P}_{n+1}$ e denotato con ω_i il generico caso elementare di $\mathcal{P}_i$ $(i = 1, \ldots, n+1)$, sia $\mathcal{P}^{(h)}$ la partizione costituita dai casi elementari $\omega^{(h)} = \omega_1 \wedge \cdots \wedge \omega_h$ della partizione prodotto $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_h$ $(h = 1, \ldots, n+1)$. Chiaramente, $\mathcal{P}^{(1)} = \mathcal{P}_1$ e $\omega^{(1)} = \omega_1$.

Consideriamo ora su $\mathcal{P}_1$ una distribuzione $(p_{\omega_1})_{\omega_1 \in \mathcal{P}_1}$; su $\mathcal{P}_2$, per ogni ω_1, una distribuzione $p^{(\omega_1)}$ tale che $p^{(\omega_1)}_{\omega_2} = 0$, se $\omega_1 \wedge \omega_2 = \varnothing$; su $\mathcal{P}_3$, per ogni $\omega^{(2)}$, una distribuzione $p^{(\omega^{(2)})}$ tale che $p^{(\omega^{(2)})}_{\omega_3} = 0$, se $\omega^{(2)} \wedge \omega_3 = \varnothing$; ... ; su $\mathcal{P}_{n+1}$, per ogni $\omega^{(n)}$, una distribuzione $p^{(\omega^{(n)})}$ tale che $p^{(\omega^{(n)})}_{\omega_{n+1}} = 0$, se $\omega^{(n)} \wedge \omega_{n+1} = \varnothing$.

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

$$p_{\omega^{(h+1)}} = p_{\omega^{(1)}} \prod_{i=1}^{h} p^{(\omega^{(i)})}_{\omega_{i+1}} = p_{\omega^{(h)}}\, p^{(\omega^{(h)})}_{\omega_{h+1}} \tag{2.4}$$

per ogni $\omega^{(h)} \in \mathcal{P}^{(h)}$ $(h = 1, \ldots, n)$[18].

Il prossimo teorema introduce una probabilità sull'algebra degli eventi logicamente dipendenti dalla partizione $\mathcal{P}^{(n+1)}$ che consente 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. *Per ogni evento $E \in \mathcal{E}_{\mathrm{L}}(\mathcal{P}^{(n+1)})$, poniamo:*

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

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

Dimostrazione. Proviamo intanto che Pr è una probabilità. A tal fine, per il Teorema 2.4.1, basta verificare che $\sum_{\omega^{(n+1)}} p_{\omega^{(n+1)}} = 1$. Da (2.4) otteniamo

$$\begin{aligned} \sum_{\omega^{(n+1)}} p_{\omega^{(n+1)}} &= \sum_{\omega^{(n)} \wedge \omega_{n+1} \neq \varnothing} p_{\omega^{(n)}} p^{(\omega^{(n)})}_{\omega_{n+1}} = \sum_{\omega^{(n)}} \left[p_{\omega^{(n)}} \sum_{\omega_{n+1} \,:\, \omega^{(n)} \wedge \omega_{n+1} \neq \varnothing} p^{(\omega^{(n)})}_{\omega_{n+1}} \right] \\ &= \sum_{\omega^{(n)}} \left[p_{\omega^{(n)}} \sum_{\omega_{n+1}} p^{(\omega^{(n)})}_{\omega_{n+1}} \right] = \sum_{\omega^{(n)}} p_{\omega^{(n)}} = 1. \end{aligned}$$

[18] La seconda uguaglianza si ottiene facilmente procedendo per induzione su h.

Passando alla seconda parte della tesi, osservato che l'uguaglianza sussiste per $h = n+1$, procediamo per induzione a ritroso supponendo che sussista per h e provandola per $h-1$. Dato $\omega^{(h)}$, tramite (2.4), si ha

$$\begin{aligned}
\Pr(\omega^{(h-1)}) &= \Pr(\omega^{(h-1)} \wedge \bigvee_{\omega_h} \omega_h) = \sum_{\omega_h} \Pr(\omega^{(h-1)} \wedge \omega_h) \\
&= \sum_{\omega_h \,:\, \omega^{(h-1)} \wedge \omega_h \neq \varnothing} \Pr(\omega^{(h-1)} \wedge \omega_h) = \sum_{\omega_h \,:\, \omega^{(h-1)} \wedge \omega_h \neq \varnothing} p_{\omega^{(h-1)} \wedge \omega_h} \\
&= \sum_{\omega_h \,:\, \omega^{(h-1)} \wedge \omega_h \neq \varnothing} p_{\omega^{(h-1)}} p_{\omega_h}^{(\omega^{(h-1)})} = p_{\omega^{(h-1)}} \sum_{\omega_h} p_{\omega_h}^{(\omega^{(h-1)})} = p_{\omega^{(h-1)}},
\end{aligned}$$

notato che $\omega^{(h-1)} \wedge \omega_h \in \mathcal{P}^{(h)}$, ogniqualvolta $\omega^{(h-1)} \wedge \omega_h \neq \varnothing$.

Passando alla parte conclusiva della tesi, assumiamo $\Pr(\omega^{(h)}) > 0$. Sia intanto $\omega_{h+1} \wedge \omega^{(h)} \neq \varnothing$. Allora, per la seconda parte della tesi e (2.4), risulta

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

e quindi

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

Sia infine $\omega_{h+1} \wedge \omega^{(h)} = \varnothing$. Allora, per il Teorema 2.3.2(i), $\Pr(\omega_{h+1} \,|\, \omega^{(h)}) = \Pr(\omega_{h+1} \wedge \omega^{(h)} \,|\, \omega^{(h)}) = 0 = p_{\omega_{h+1}}^{(\omega^{(h)})}$. □

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 $\Pr(\cdot \,|\, H)$. In questo ordine di idee, diremo che l'evento E è:
- **correlato negativamente con** H, se $\Pr(E \,|\, H) < \Pr(E)$;
- **correlato positivamente con** H, se $\Pr(E \,|\, H) > \Pr(E)$;
- **non correlato con** H, se $\Pr(E \,|\, H) = \Pr(E)$.

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 sono non correlati con ogni evento non trascurabile; inoltre, ogni evento è non 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; infatti, per il Teorema 2.1.2(i), risulta

$$\Pr(E \mid H) \gtreqless \Pr(E) \Leftrightarrow 1 - \Pr(\overline{E} \mid H) \gtreqless 1 - \Pr(\overline{E}) \Leftrightarrow \Pr(\overline{E} \mid H) \lesseqgtr \Pr(\overline{E}).$$

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 è non 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 anche le probabilità condizionate $\Pr(E \mid \overline{H})$ e $\Pr(\overline{E} \mid \overline{H})$. Ora, per ogni costituente $E' \wedge H' = \varnothing$ della partizione generata $\mathcal{P}_{\mathrm{G}}(E, H)$, risulta $\Pr(E' \mid H') = \Pr(E' \wedge H' \mid H') = 0$ (Teorema 2.3.2(i)) e quindi E è correlato negativamente (positivamente) con H', a seconda che $E' = E$ o $E' = \overline{E}$. Dunque, la presenza di costituenti impossibili forza una correlazione positiva o negativa tra i due eventi. Allora, 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 sull'estensione dell'altro, dobbiamo *necessariamente* supporre che i costituenti della partizione generata siano casi elementari, cioè che i due eventi siano logicamente indipendenti (Teorema 1.10.1).

Queste considerazioni suggeriscono quindi di chiamare gli eventi $E_1, \ldots, E_{n+1}$ **stocasticamente indipendenti**[19] se *sono logicamente indipendenti* e ciascuno di essi è non 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 $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{i-1}, E_{i+1}, \ldots, E_{n+1})$ $(i = 1, \ldots, n+1)$[20].

Ovviamente, l'indipendenza stocastica degli eventi $E_1, \ldots, E_{n+1}$ è equivalente all'indipendenza stocastica degli eventi $E'_1, \ldots, E'_{n+1}$, notato che $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{n+1}) = \mathcal{P}_{\mathrm{G}}(E'_1, \ldots, E'_{n+1})$.

Proviamo ora che, analogamente a quanto avviene per l'indipendenza logica, anche l'indipendenza stocastica si conserva "in discesa".

Teorema 2.5.2. *Siano gli eventi $E_1, \ldots, E_{n+1}$ stocasticamente indipendenti. Allora, dati $j_1, \ldots, j_k \in \{1, \ldots, n+1\}$ $(k \geq 2)$, gli eventi $E_{j_1}, \ldots, E_{j_k}$ sono stocasticamente indipendenti.*

Dimostrazione. Per il Teorema 1.10.3, gli eventi $E_{j_1}, \ldots, E_{j_k}$ sono logicamente indipendenti. Dato E_{j_h}, denotiamo con $\mathcal{P}'$ la partizione generata dagli eventi $E_{j_1}, \ldots, E_{j_k}$ diversi da E_{j_h} e con $\mathcal{P}''$ la partizione generata dagli eventi $E_1, \ldots, E_{n+1}$ diversi da $E_{j_1}, \ldots, E_{j_k}$. Indicati infine i costituenti delle due partizioni, rispettivamente, con ω' e ω'', risulta

$$\omega' = \omega' \wedge \Omega = \omega' \wedge \bigvee_{\omega'' \in \mathcal{P}''} \omega'' = \bigvee_{\omega'' \in \mathcal{P}''} (\omega' \wedge \omega'') \tag{2.5}$$

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

[20] Qui e nel seguito, per snellire l'esposizione, usiamo, con riferimento allla partizione generata $\mathcal{P}_{\mathrm{G}}((E_i)_{i \in I})$, la notazione E'_i al posto della $E_{if(i)}$ $(f \in \{0,1\}^I)$. Conseguentemente, E'_i rappresenta sia l'evento E_i che il suo negato $\overline{E}_i$.

per ogni $\omega' \in \mathcal{P}'$. Sia ora ω' non trascurabile. Considerato un qualsiasi ω'' tale che $\omega' \wedge \omega''$ non è trascurabile, dall'indipendenza stocastica degli eventi $E_1, \ldots, E_{n+1}$ otteniamo $\Pr(E_{j_h} \,|\, \omega' \wedge \omega'') = \Pr(E_{j_h})$, poiché $\omega' \wedge \omega'' \in \mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{j_h-1}, E_{j_h+1}, \ldots, E_n)$. Ne segue, considerati gli eventi $\omega' \wedge \omega''$ per ogni $\omega'' \in \mathcal{P}''$, tramite (2.5) e la conglomerabilità (Teorema 2.3.2(vi), $\Pr(E_{j_h} \,|\, \omega') = \Pr(E_{j_h})$. □

Il teorema seguente fornisce una caratterizzazione dell'indipendenza stocastica tramite la fattorizzabilità della probabilità sulla partizione generata.

Teorema 2.5.3 (di fattorizzazione)**.** *Dati gli eventi $E_1, \ldots, E_{n+1}$, siano non trascurabili i costituenti della relativa partizione generata. Allora, sono equivalenti le proposizioni:*

(i) Gli eventi sono stocasticamente indipendenti;

(ii) La probabilità si fattorizza sui costituenti della partizione generata.

Dimostrazione. (i) $\Rightarrow$ (ii) Osservato che, per $n = 1$, (ii) segue dal Teorema 2.5.1, assumiamo che sussista per n e proviamola per $n + 1$. Siano dunque gli eventi $E_1, \ldots, E_{n+2}$ stocasticamente indipendenti e i costituenti della partizione $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{n+2})$ non trascurabili. Allora, per la monotonia (Teorema 2.1.2(v)) i costituenti della partizione $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_{n+1})$ sono non trascurabili e gli eventi $E_1, \ldots, E_{n+1}$ sono, per il Teorema 2.5.2, stocasticamente indipendenti. Conseguentemente, dato il costituente $H = E'_1 \wedge \cdots \wedge E'_{n+2}$, otteniamo

$$\Pr\Big(\bigwedge_{i=1}^{n+2} E'_i\Big) = \Pr(E'_{n+2} \,|\, H) \Pr(H) = \Pr(E'_{n+2}) \prod_{i=1}^{n+1} \Pr(E'_i) = \prod_{i=1}^{n+2} \Pr(E'_i).$$

(ii) $\Rightarrow$ (i) 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

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

Ricordato infine che i costituenti non sono trascurabili, dal Teorema 1.10.1 otteniamo che gli eventi $E_1, \ldots, E_{n+1}$ sono logicamente indipendenti. La proposizione (i) è così dimostrata. $\square$

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 gli eventi:

$$\omega_i : \text{viene estratta la pallina con impresso il numero i} \quad (i = 1, \ldots, 4)$$

e poniamo, adottando lo schema simmetrico (Osservazione 2.4.2(ii)), $\Pr(\omega_i) = \frac{1}{4}$ $(i = 1, \ldots, 4)$.

Introdotti gli eventi:

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

e osservato, ad esempio, che il costituente $\overline{E}_1 \wedge \overline{E}_2 \wedge \overline{E}_3$ è impossibile, possiamo concludere che i tre eventi non sono stocasticamente indipendenti in quanto, per il Teorema 1.10.1, non lo sono logicamente.

Però, lo sono a due a due. Infatti, notato che $E_i = \omega_i \vee \omega_4$, $(i = 1, 2, 3)$ e, se $i \neq j$, $E_i \wedge E_j = \omega_4$, $E_i \wedge \overline{E}_j = \omega_i$, $\overline{E}_i \wedge \overline{E}_j = \omega_h$ $(h \neq i, j, 4)$, otteniamo $\Pr(E_i) = \frac{1}{2}$ $(i = 1, 2, 3)$ e $\Pr(E_i \wedge E_j) = \Pr(E_i \wedge \overline{E}_j) = \Pr(\overline{E}_i \wedge \overline{E}_j) = \frac{1}{4}$ $(i \neq j)$. Conseguentemente, gli eventi E_i, E_j $(i \neq j)$ sono, sempre per lo stesso teorema, logicamente indipendenti; inoltre, $\Pr(E_i \wedge E_j) = \Pr(\omega_4) = \Pr(\omega_i) = \Pr(E_i \wedge \overline{E}_j)$ e quindi $\Pr(E_i \,|\, E_j) = \Pr(E_i) = \Pr(E_i \,|\, \overline{E}_j)$, cioè E_i non è correlato con i casi elementari della partizione generata da E_j. Pertanto, gli eventi E_i, E_j sono stocasticamente indipendenti[21]. $\Diamond$

2.6 Tre modelli di estrazione

Molti problemi relativi ai giochi d'azzardo (dadi, testa o croce, lotto, ...), come pure alcuni riguardanti la statistica (distribuzione di un carattere nella popolazione, ...) e l'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.

[21] Osserviamo che ognuno dei tre eventi è correlato positivamente con l'evento congiunzione degli altri due.

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 $f, g \in \{0,1\}^{\{1,\ldots,n\}}$ tali che $\#\{i : g(i) = 1\} = \#\{i : f(i) = 1\}$[22].

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 di elementi affermati.

Teorema 2.6.1. *Siano gli eventi $E_1, \ldots E_n$ scambiabili. Allora, qualunque sia il costituente $E'_{i_1} \wedge \cdots \wedge E'_{i_m} \in \mathcal{P}_{\mathrm{G}}(E_{i_1}, \ldots E_{i_m})$ $(i_h \neq i_l (h \neq l); m \leq n)$, 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 $\mathcal{P} = \mathcal{P}_{\mathrm{G}}(E_{j_1}, \ldots, E_{j_h})$. Risulta allora

$$\begin{aligned} E'_{i_1} \wedge \cdots \wedge E'_{i_m} &= (E'_{i_1} \wedge \cdots \wedge E'_{i_m}) \wedge \bigvee_{\omega \in \mathcal{P}} \omega = \bigvee_{\omega \in \mathcal{P}} (E'_{i_1} \wedge \cdots \wedge E'_{i_m} \wedge \omega) \\ &= \bigvee_{i=0}^{h} \Big[\bigvee_{\omega \in A(i)} (E'_{i_1} \wedge \cdots \wedge E'_{i_m} \wedge \omega) \Big], \end{aligned} \tag{2.6}$$

ove $A(i) = \{\omega \in \mathcal{P} : \omega \text{ ha } i \text{ eventi affermati}\}$. 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 (2.6) sono elementi di $\mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$ con $k + i$ elementi affermati, otteniamo

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

[22]In altre parole, la probabilità che si verifichino esattamente k eventi tra gli $E_1, \ldots E_n$ dipende solo da k e non da quali eventi, tra di essi, si verificano.

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

Passando ai 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 modalità seguenti:

- senza remissione nell'urna della pallina estratta (*estrazioni senza rimessa*);
- con remissione nell'urna della pallina estratta (*estrazioni con rimessa*);
- con remissione nell'urna della pallina estratta assieme ad un numero prefissato $a \geq 1$ di palline del medesimo colore di quella estratta (*estrazioni con contagio simmetrico*)[23].

Supponiamo che le palline siano di due tipi: bianche con numerosità $b \geq 1$ e rosse con numerosità $r \geq 1$ $(b+r = s)$. Assumiamo inoltre (come è del tutto naturale in questo contesto) che le palline presenti nell'urna dopo h estrazioni abbiano la medesima probabilità di essere estratte alla $(h+1)$-sima estrazione $(h = 1, \dots, n-1)$ (condizione di simmetria).

Introdotti, per $h = 1, \dots, n$, l'evento:

E_h : si effettua l'h-sima estrazione e viene estratta pallina bianca[24]

e la relativa partizione generata $\mathcal{P}_h = \{E_h, \overline{E}_h\}$, possiamo descrivere gli esiti relativi alle prime h estrazioni tramite la partizione $\mathcal{S}^{(h)}$ dei casi elementari (chiamati **storie iniziali**) della partizione prodotto $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_h$. Chiaramente, nelle estrazioni con rimessa o con contagio, tutti i costituenti (di $\mathcal{P}_1 \wedge \cdots \wedge \mathcal{P}_h$) sono storie iniziali; nel caso invece di estrazioni senza rimessa, un costituente con $k \geq 0$ eventi affermati è una storia iniziale se e solo se $h \leq s$ e $k \leq b$, $h - k \leq r$, cioè $\max(0, h-r) \leq k \leq \min(b, h)$.

Ciò precisato, ricorrendo alla supposta simmetria introduciamo, tramite (2.3), la seguente distribuzione di probabilità sulla partizione generata $\mathcal{P}_1$:

$$p = p_{E_1} = \frac{b}{s} > 0, \qquad q = p_{\overline{E}_1} = \frac{r}{s} > 0. \tag{2.7}$$

[23]Modello introdotto da György Polya e F. Eggenberger nel 1923 per rappresentare fenomeni aventi caratteristiche simili a quelle del contagio delle malattie.

[24]Poiché l'evento E_h fa riferimento (dal punto di vista descrittivo) allo stato di conoscenza $\mathcal{C}$, è sottinteso che $\mathcal{C}$ contenga le informazioni che riguardano sia la composizione dell'urna che la modalità di estrazione. Chiaramente, se non è possibile effettuare l'estrazione, tale evento risulta impossibile. Situazione che non si verifica mai se le estrazioni avvengono con rimessa o con contagio; si verifica invece nel caso di estrazioni senza rimessa quando l'indice h supera il numero s di palline presenti nell'urna. Concludiamo osservando che $\overline{E}_h$ significa che non viene effettuata l'h-sima estrazione oppure viene effettuata e esce pallina rossa.

Scegliamo ora un numero relativo $a \geq -1$. Data la storia iniziale $E'_1 \wedge \cdots \wedge E'_h$ ($h < n$) con k eventi affermati, consideriamo (ricorrendo ancora a (2.3)) la distribuzione di probabilità $p^{(E'_1 \wedge \cdots \wedge E'_h)}$ sulla partizione $\mathcal{P}_{h+1}$:

$$\begin{aligned} p^{(E'_1 \wedge \cdots \wedge E'_h)}_{E_{h+1}} &= \begin{cases} \dfrac{b+ak}{s+ah}, & \text{se } E'_1 \wedge \cdots \wedge E'_h \wedge E_{h+1} \in \mathcal{S}^{(h+1)} \\ 0, & \text{altrimenti} \end{cases} \\ p^{(E'_1 \wedge \cdots \wedge E'_h)}_{\overline{E}_{h+1}} &= \begin{cases} \dfrac{r+a(h-k)}{s+ah}, & \text{se } E'_1 \wedge \cdots \wedge E'_h \wedge \overline{E}_{h+1} \in \mathcal{S}^{(h+1)} \\ 0, & \text{altrimenti} \end{cases} \end{aligned} \tag{2.8}$$

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 $E'_1 \wedge \cdots \wedge E'_h \wedge E'_{h+1} \in \mathcal{S}^{(h+1)}$, 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 $\max(0, h-r \leq k \leq \min(b,h)$ (poiché $E'_1 \wedge \cdots \wedge E'_h \neq \varnothing$) 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, con riferimento alla partizione $\mathcal{S} = \mathcal{S}^{(n)}$ delle **storie**, siamo nella situazione descritta nella seconda parte della Sezione 2.4, possiamo introdurre 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 iniziali aventi lo stesso numero di affermazioni.

Lemma 2.6.2. *Per ogni storia iniziale $E'_1 \wedge \cdots \wedge E'_h \in \mathcal{S}^{(h)}$ risulta:*

$$\Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_h}_{k \text{ affermati}}) = \frac{[b(b+a)\cdots(b+a(k-1))]\,[r(r+a)\cdots(r+a(h-k-1))]}{s(s+a)\cdots(s+a(h-1))}$$

qualora si assuma, per quanto concerne i due fattori al numeratore in parentesi quadra, che il primo valga 1, se $k = 0$, mentre il secondo valga 1, se $k = h$.

Dimostrazione. Dal Teorema 2.4.3 otteniamo $\Pr(E'_1) = p_{E'_1}$ e quindi, per (2.7), l'uguaglianza sussiste per $h = 1$. Assumiamo allora che sussista per $h < n$ e proviamola per $h+1$. Sia dunque $E'_1 \wedge \cdots \wedge E'_{h+1}$ una storia iniziale di $\mathcal{S}^{(h+1)}$ con k affermazioni. Allora, $\omega^{(h)} = E'_1 \wedge \cdots \wedge E'_h$ è una storia iniziale di $\mathcal{S}^{(h)}$ avente $k-1 \geq 0$ affermazioni, se

$E'_{h+1} = E_{h+1}$, e k affermazioni, altrimenti. Allora, per (2.8) e il Teorema 2.4.3, si ha

$$\Pr(E'_{h+1}|\omega^{(h)}) = p_{E'_{h+1}}^{(\omega^{(h)})} = \begin{cases} \dfrac{b+a(k-1)}{s+ah}, & \text{se } E'_{h+1} = E_{h+1} \\ \dfrac{r+a(h-k)}{s+ah}, & \text{se } E'_{h+1} = \overline{E}_{h+1} \end{cases},$$

notato che $\Pr(\omega^{(h)}) > 0$ in quanto, tramite l'ipotesi induttiva, risulta

$$\Pr(\omega^{(h)}) = \begin{cases} \dfrac{\prod_{u=0}^{k-2}(b+au)\prod_{v=0}^{h-k}(r+av)}{\prod_{t=0}^{h-1}(s+at)} > 0, & \text{se } E'_{h+1} = E_{h+1} \\ \dfrac{\prod_{u=0}^{k-1}(b+au)\prod_{v=0}^{h-k-1}(r+av)}{\prod_{t=0}^{h-1}(s+at)} > 0, & \text{se } E'_{h+1} = \overline{E}_{h+1} \end{cases}.$$

Ne segue la tesi, ricordato che $\Pr(E'_1 \wedge \cdots \wedge E'_{h+1}) = \Pr(E'_{h+1} \,|\, \omega^{(h)}) \Pr(\omega^{(h)})$. □

Dunque, nei tre modelli di estrazione considerati, gli eventi $E_1, \ldots, E_n$ sono scambiabili, in quanto, per il lemma appena provato, la probabilità di una qualsiasi storia $E'_1 \wedge \cdots \wedge E'_n$ dipende solo dal numero delle affermazioni presenti e non da dove sono collocate.

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 $\mathcal{S}$ delle storie. Conseguentemente, considerato un qualsiasi valore possibile $k \in \{0, \ldots, n\}$ 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[25]

$$\Pr(S_n = k) = \binom{n}{k} \Pr(\underbrace{E'_1 \wedge \cdots\cdots \wedge E'_n}_{k \text{ affermati}}), \tag{2.9}$$

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

[25] 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

$$\Pr(\underbrace{E_1' \wedge \cdots\cdots \wedge E_n'}_{\substack{k \text{ affermati} \\ \max(0,n-r)\leq k\leq \min(b,n)}}) = \frac{[b(b-1)\cdots(b-k+1)]\,[r(r-1)\cdots(r-(n-k)+1)]}{s(s-1)\cdots(s-n+1)}$$

da cui otteniamo

$$\Pr(\underbrace{E_1' \wedge \cdots\cdots \wedge E_n'}_{\substack{k \text{ affermati} \\ \max(0,n-r)\leq k\leq \min(b,n)}}) = \frac{\left[\binom{b}{k}k!\right]\left[\binom{r}{n-k}(n-k)!\right]}{\binom{s}{n}n!} = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{s}{n}\binom{n}{k}}$$

e quindi, tramite (2.9), la **distribuzione ipergeometrica**:

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

Inoltre, per il Teorema 2.6.1 e il lemma sopra citato, sussiste l'uguaglianza:

$$\Pr(\underbrace{E_{i_1}' \wedge \cdots\cdots \wedge E_{i_h}'}_{\substack{k \text{ affermati} \\ \max(0,h-r)\leq k\leq \min(b,h)}}) = \frac{\binom{b}{k}\binom{r}{h-k}}{\binom{s}{h}\binom{h}{k}} \quad (i_m \neq i_l (m \neq l); h \leq n) \tag{2.11}$$

e quindi, dipendendo tale probabilità solo dal numero di affermazioni e non da quali eventi sono affermati, risultano equiprobabili tutte le storie parziali aventi lo stesso numero di eventi (qualunque essi siano) e di affermazioni.

Infine, per (2.11), $\Pr(E_i) = p$ $(i = 1, \ldots n)$ e, se $b, n \geq 2$ e $i \neq j$, $\Pr(E_i \wedge E_j) = \binom{b}{2}\binom{s}{2}^{-1} = \frac{b(b-1)}{s(s-1)}$; riesce allora (tramite (2.7))

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

Conseguentemente, gli eventi $E_1, \ldots, E_n$ sono, qualora $b, n \geq 2$, a due a due correlati negativamente[26].

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

Osservazione 2.6.3. Estrazioni in blocco L'evento $\{S_n = k\}$ rappresenta l'estrazione di k palline bianche senza tener conto dell'ordine in cui appaiono in una sequenza di n estrazioni. Poiché 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:

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

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 possibili e il numeratore quello dei casi favorevoli all'evento $E_n^{(k)}$, possiamo concludere, tramite (2.3), 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, tramite (2.9), la **distribuzione binomiale**:

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

Inoltre, per il Teorema 2.6.1 e il lemma sopra citato, sussiste l'uguaglianza:

$$\Pr(\underbrace{E_{i_1}' \wedge \cdots \wedge E_{i_h}'}_{k \text{ affermati}}) = p^k q^{h-k} \quad (i_m \neq i_l (m \neq l); h \leq n) \tag{2.13}$$

e quindi, come nel caso delle estrazioni senza rimessa, risultano equiprobabili tutte le storie parziali aventi lo stesso numero di eventi (qualunque essi siano) e di affermazioni.

Osservato infine che, per (2.7), $\Pr(E_1' \wedge \cdots \wedge E_n') = \prod_{i=1}^n \Pr(E_i') > 0$ per ogni $E_1' \wedge \cdots \wedge E_n'$, possiamo concludere, tramite il Teorema 2.5.3, che gli eventi $E_1, \ldots, E_n$ sono stocasticamente 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 esteso (Osservazione 1.12.2) T relativa alla partizione canonica $\mathcal{T}$ costituita dai casi elementari:

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

Posto $p_n = p\, q^{n-1} = \Pr(T = n)$ (per (2.13)), otteniamo

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

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**:

$$\Pr(T = t) = \begin{cases} p\,q^{t-1}, & \text{se } t = 1, 2, \ldots \\ 0, & \text{se } t = +\infty \end{cases} . \tag{2.14}$$

Se ora si considera l'evento $\{T > n\} = \bigwedge_{i=1}^n \overline{E}_i$ (**ritardo di n estrazioni nell'apparizione del primo successo**), da (2.13) otteniamo

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

e quindi

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

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.

Consideriamo infine, dato $k \geq 2$, il **tempo di attesa del k-simo successo** (numero di estrazioni che occorre fare per ottenere per la prima volta k successi), rappresentato formalmente dalla versione del numero aleatorio esteso $T^{(k)}$ relativa alla partizione canonica $\mathcal{T}^{(k)}$ formata dai casi elementari:

$$\begin{aligned}
\{T^{(k)} = k\} &= E_1 \wedge \cdots \wedge E_k \\
\{T^{(k)} = n\} &= \bigvee_{\underbrace{E'_1 \wedge \cdots \wedge E'_{n-1}}_{k-1 \text{ affermati}}} (E'_1 \wedge \cdots \wedge E'_{n-1} \wedge E_n) \quad (n > k) \\
\{T^{(k)} = +\infty\} &= \bigwedge_{n\geq k} \overline{\{T^{(k)} = n\}}.
\end{aligned}$$

Posto $p_{h+k} = \binom{h+k-1}{k-1} p^k\, q^h = \Pr(T^{(k)} = h + k)$ $(h \geq 0)$ (per 2.12)), risulta

$$\sum_{h\geq 0} p_{h+k} = \sum_{h\geq 0} \binom{h+k-1}{k-1} p^k\, q^h = 1.$$

Infatti, dalla

$$\binom{h+k-1}{k-1} = \binom{h+k-1}{(h+k-1)-(k-1)} = \binom{h+k-1}{h}$$

otteniamo

$$\begin{aligned}
\binom{h+k-1}{k-1} &= \frac{(h+k-1)(h+k-2)\cdots(k+1)k}{h!} \\
&= (-1)^h\, \frac{[(-k)-h+1][(-k)-h+2]\cdots[(-k)-1](-k)}{h!} \\
&= (-1)^h \binom{-k}{h}
\end{aligned}$$

e quindi

$$\sum_{h\geq 0} p_{h+k} = \sum_{h\geq 0} (-1)^h \binom{-k}{h} p^k\, q^h = p^k \sum_{h\geq 0} (-1)^h \binom{-k}{h} q^h = \frac{p^k}{(1-q)^k} = 1,$$

ricordato che $\sum_{h\geq 0} (-1)^h \binom{-k}{h} x^h = (1-x)^{-k}$ per ogni x tale che $|x| < 1$.

Conseguentemente, posto $p_{+\infty} = 0$, possiamo introdurre, per il Teorema 2.4.1, una probabilità numerabilmente additiva su $\mathcal{E}_{\mathrm{L}}(\mathcal{T}^{(k)})$ tramite la **distribuzione di Pascal**:

$$\Pr(T^{(k)} = t) = \begin{cases} \binom{t-1}{k-1} p^k\, q^{t-k}, & \text{se } t = k, k+1, \ldots \\ 0, & \text{se } t = +\infty \end{cases} . \tag{2.15}$$

Nell'esempio seguente applichiamo i due modelli di estrazione sinora considerati al gioco del lotto e quello del "testa o croce".

Esempio 2.6.4. (i) GIOCO DEL LOTTO Le scommesse più usuali sono quelle relative agli eventi del tipo: la cinquina estratta includerà un numero prefissato (*ambata*); due numeri prefissati (*ambo*); tre numeri prefissati (*terno*); quattro numeri prefissati (*quaterna*); cinque numeri prefissati (*cinquina semplice*).

Assumendo, come è del tutto naturale, la condizione di simmetria, possiamo usare i risultati di questa sezione per calcolare le relative probabilità, specificando il numero di palline (a partire da 1 fino a 5) da identificare con le palline bianche. Otteniamo allora che le probabilità sono, rispettivamente,

$$\frac{\binom{89}{4}}{\binom{90}{5}} = \frac{1}{18}, \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{2}{801}, \frac{\binom{87}{2}}{\binom{90}{5}} = \frac{1}{11.748}, \frac{\binom{86}{1}}{\binom{90}{5}} = \frac{1}{511.038}, \frac{\binom{85}{0}}{\binom{90}{5}} = \frac{1}{43.949.268}[27].$$

Passando infine a considerare estrazioni successive di cinquine semplici da una medesima ruota, possiamo ritenere, poichè ogni pallina ha la stessa probabilità di essere estratta (pari a $\frac{1}{18}$), di essere in presenza di estrazioni con rimessa da un'urna costituita da una pallina bianca e da 17 rosse. Conseguentemente, la mancanza di memoria insita nel modello probabilistico delle estrazioni con rimessa, comporta che scommettere sui *numeri ritardatari* è privo di senso: il fatto che un numero non sia uscito per molte estrazioni non aumenta in alcun modo la probabilità che esca nell'estrazione successiva, rimanendo questa probabilità sempre uguale a $\frac{1}{18}$. Inoltre, la presenza di numeri ritardatari non deve sorprendere in quanto, considerati gli eventi:

$$R_i : \text{il numero } j \text{ non esce per } n \text{ estrazioni consecutive},$$

[27] Osservato che il gestore adotta come coefficienti (da moltiplicare per la puntata per ottenere la vincita lorda corrispondente) rispettivamente i valori 11,232, 250, 4.500, 120.000 e 6.000.000, possiamo asserire che il lotto non è un gioco equo (qualora si identifichino, come fatto nell'Osservazione 2.1.1, i quozienti di scommessa con le probabilità); notato poi che i rapporti tra la vincita equa e la vincita reale diventano, rispettivamente, 1,6, 1,6, 2,61, 4,26 e 7,32, possiamo concludere che l'iniquità a favore del gestore cresce con l'aumentare dell'improbabilità dell'evento relativo alle scommesse considerate.

la probabilità che almeno un numero non esca nelle n estrazioni consecutive è data, per la formula d'inclusione-esclusione (Teorema 2.1.2(viii)) dalla:

$$\Pr(\bigvee_{i=1}^{90} R_i) = \sum_{\emptyset \neq J \subset \{1,\ldots,90\}} (-1)^{\#J-1} \Pr(\bigwedge_{j\in J} R_j) = \sum_{m=1}^{85} \binom{90}{m} \left[\frac{\binom{90-m}{5}}{\binom{90}{5}} \right]^n ,$$

che, nel caso dei "numeri centenari" ($n = 100$) è pari a 0,26 (mentre la probabilità che un prefissato numero sia centenario è pari a 0,0033).

(ii) Gioco del testa o croce Il gioco consiste nel lancio (anche ripetuto) di una moneta equilibrata, le cui facce sono chiamate, rispettivamente, *Testa* e *Croce*. Ricorrendo al modello delle estrazioni da un'urna, formalizziamo il gioco mediante le estrazioni con rimessa da un'urna formata da due palline, una bianca (Testa) e una rossa (Croce). Ciò precisato, procediamo calcolando $\Pr(E)$, ove E rappresenta uno qualsiasi dei due eventi seguenti:

• *Esce testa per la prima volta in un lancio $n \geq k > 1$ tale che la divisione di n per k ha resto $0 \leq r < k$.* Allora, $E = \bigvee_{m\geq 1}\{T = km + r\}$ e quindi

$$\begin{aligned} \Pr(E) &= \sum_{m\geq 1} \Pr(T = km + r) = \sum_{m\geq 1} \frac{1}{2^{km+r}} = \frac{1}{2^r} \sum_{m\geq 1} \left(\frac{1}{2^k}\right)^m \\ &= \frac{1}{2^r} \left[\sum_{m\geq 0} \left(\frac{1}{2^k}\right)^m - 1 \right] = \frac{1}{2^r(2^k-1)} , \end{aligned}$$

osservato che la serie è geometrica di ragione $\frac{1}{2^k}$. Quindi, la probabilità che esca testa per la prima volta in un lancio multiplo di k è $\frac{1}{2^k-1}$ ($r = 0$). Allora, la probabilità che esca testa per la prima volta in un lancio pari è $\frac{1}{3}$ e quindi la probabilità che esca testa per la prima volta in un lancio dispari è $1 - \frac{1}{3} = \frac{2}{3}$.

• *Esce un numero pari di teste nei primi n lanci.* Allora, $E = \bigvee_{i=1}^{\lfloor \frac{n}{2} \rfloor}\{S_n = 2\,i\}$[28] e quindi, per (2.12),

$$\Pr(E) = \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \Pr(S_n = 2i) = \frac{1}{2^n} \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2i} = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Andiamo ora a valutare la probabilità che *il primo successo avvenga al lancio m sapendo che il secondo avviene al lancio $n > 1$.* Considerato l'evento

$$H : \text{il secondo successo avviene al lancio } n$$

[28] Ove, per ogni x reale, $\lfloor x \rfloor$ denota la *parte intera* di x, cioè il massimo numero relativo minore o uguale a x.

e posto $H_k = E_k \wedge E_n \wedge \bigwedge_{i\in\{1,\dots,n-1\}\setminus\{k\}} \overline{E}_i$ $(k = 1,\dots,n-1)$, otteniamo $H = \bigvee_{k=1}^{n-1} H_k$ e $\Pr(H_k) = \frac{1}{2^n}$. Ne segue, $\Pr(H) = (n-1)\frac{1}{2^n}$ e quindi

$$\Pr(T = m \mid H) = \frac{\Pr(\{T = m\} \wedge H)}{\Pr(H)} = \frac{\Pr(H_m)}{\Pr(H)} = \frac{1}{n-1}.$$

Dunque, gli eventi $\{T = 1\}, \dots, \{T = n-1\}$ sono tutti equiprobabili a fronte dell'informazione che esce testa per la seconda volta al lancio n-simo.

Determiniamo infine il minimo numero $n^\star$ di lanci che bisogna effettuare affinchè esca *Testa con probabilità almeno pari a* $\theta \in]0,1[$. Osservato che la probabilità che esca Testa in n lanci è, per (2.12), pari a $1 - \Pr(S_n = 0) = 1 - q^n$, basta risolvere la disequazione $1 - q^n \geq \theta$. Risulta, $q^n \leq 1 - \theta$, cioè

$$n \geq \frac{\ln(1-\theta)}{\ln q} = \frac{\ln(1-\theta)}{\ln\frac{1}{2}} = -\frac{\ln(1-\theta)}{\ln 2} = x^\star$$

e quindi $n^\star = x^\star$, se $x^\star$ è un numero naturale, e $n^\star = \lfloor x^\star \rfloor + 1$, altrimenti. $\Diamond$

2.6.3 Estrazioni con contagio simmetrico

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

$$\Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}}) = \frac{[b(b+a)\cdots(b+a(k-1))]\,[r(r+a)\cdots(r+a(n-k-1))]}{s(s+a)\cdots(s+a(n-1))}$$

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

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

e quindi, tramite (2.9), la **distribuzione di Polya**:

$$\Pr(S_n = k) = \frac{\binom{-\frac{b}{a}}{k}\binom{-\frac{r}{a}}{n-k}}{\binom{-\frac{s}{a}}{n}} \quad (k = 0, 1, \dots, n). \tag{2.16}$$

Inoltre, per il Teorema 2.6.1 e il lemma sopra citato, sussiste l'uguaglianza:

$$\Pr(\underbrace{E'_{i_1} \wedge \cdots \wedge E'_{i_h}}_{k \text{ affermati}}) = \frac{\binom{-\frac{b}{a}}{k}\binom{-\frac{r}{a}}{h-k}}{\binom{-\frac{s}{a}}{h}\binom{h}{k}} \qquad (i_m \neq i_l (m \neq l); h \leq n) \quad (2.17)$$

e quindi, anche per questa modalità, sono equiprobabili tutte le storie parziali aventi lo stesso numero di eventi (qualunque essi siano) e di affermazioni.

Infine, per (2.17), $\Pr(E_i) = p$ $(i = 1, \dots, n)$ e, se $b, n \geq 2$ e $i \neq j$, $\Pr(E_i \wedge E_j) = \binom{-\frac{b}{a}}{2}\binom{-\frac{s}{a}}{2}^{-1} = \frac{b(b+a)}{s(s+a)}$; riesce allora (tramite (2.7)),

$$\Pr(E_i \,|\, E_j) = \frac{\Pr(E_i \wedge E_j)}{\Pr(E_j)} = \frac{b+a}{s+a} > p.$$

Conseguentemente, gli eventi $E_1, \dots, E_n$ sono, se $b, n \geq 2$, a due a due correlati positivamente[29].

2.7 Esercizi

1. Se $\Pr(E) = \frac{1}{3}$ e $\Pr(\overline{F}) = \frac{1}{4}$, possono essere incompatibili gli eventi E, F?

2. Siano gli eventi E_1, E_2, E_3 stocasticamente indipendenti tali che $\Pr(E_i) = \frac{1}{1+i}$ $(i = 1, 2, 3)$. Considerato il numero aleatorio $S = |E_1| + |E_2| + |E_3|$, si determini la probabilità che S assuma valore 1.

3. Con riferimento al lancio simultaneo di due dadi A, B distinguibili, siano X, Y i numeri che escono, rispettivamente, con il dado A e con il dado B. Considerati gli eventi E : "X è pari" e F: "$X + Y$ è pari", dire se gli eventi E, F sono stocasticamente indipendenti.

4. Individuare la partizione generata dagli eventi E_i $(i = 1, 2, 3, 4)$ sapendo che E_1, E_2 sono incompatibili, $E_3 \to E_1$ e $E_3 \to \overline{E}_4$. Considerata poi su tale partizione la distribuzione uniforme: $*$ calcolare $\Pr(E_3 \vee (E_1 \wedge \overline{E}_4))$ e $\Pr(E_1 \vee E_2 \,|\, E_3 \vee E_4)$; $*$ cosa si può dire della probabilità dell'evento E tale che $E_3 \vee (E_1 \wedge E_4) \to E$?

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

5. Con riferimento al lancio simultaneo di otto dadi equilibrati distinguibili, calcolare la probabilità che escano: $*$ due numeri uguali; $*$ due coppie distinte di numeri uguali; $*$ almeno due facce dispari; $*$ più facce pari che dispari.

6. Un'urna contiene 200 dadi di cui la metà sono truccati in modo tale che il numero 6 esca con probabbilità $\frac{1}{4}$ e gli altri con probabilità $\frac{3}{20}$. Estratto a caso[30] un dado si effettuano tre lanci ottenendo i numeri 1, 2 e 3. Calcolare la probabilità che il dado estratto sia truccato.

7. Provare che gli eventi E, F di probabilità non estrema sono non correlati se $\Pr(E \mid F) + \Pr(\overline{E} \mid \overline{F}) = 1$.

8. Dato l'evento H di probabilità non estrema, individuare gli eventi E tali che $\Pr(E) = \Pr(E \mid H) + \Pr(E \mid \overline{H})$.

9. Con riferimento a 10 lanci di una moneta, la probabilità di ottenere 4 teste è uguale a quella di ottenere 5 teste. È la moneta equilibrata?

10. Dati gli eventi E, H tali che $\Pr(E) = \frac{1}{3}$, $\Pr(E \mid H) = \frac{1}{4}$ e $\Pr(H \mid E) = \frac{1}{3}$, calcolare $\Pr(\overline{H} \mid \overline{E})$.

11. Con riferimento al numero X che esce nel lancio di un dado equilibrato, risolvere l'equazione (in t) $\Pr(X^2 \leq t) = \Pr(X^2 \geq t)$.

12. La successione $a_n = \frac{2}{3^{n+1}}$ $(n \geq 1)$ è una distribuzione di probabilità sui numeri naturali?

Nei prossimi esercizi, una qualsiasi urna U, avente b palline bianche e r palline rosse, la indicheremo con la notazione $U(b, r)$; inoltre E_n denoterà sempre l'evento descritto dall'enunciato "viene effettuata l'estrazione n-sima e viene estratta pallina bianca".

13. Con riferimento all'urna $U(4, 12)$ calcolare le probabilità degli eventi descritti dai seguenti enunciati nelle tre modalità di estrazione considerate: $*$ la prima pallina estratta è rossa sapendo che la terza estratta è bianca; $*$ la quarta pallina estratta è bianca sapendo che nelle prime tre estrazioni sono uscite solo due palline dello stesso colore; $*$ la terza pallina estratta è rossa sapendo che la prima e la quarta estratta sono di colore diverso.

14. Dall'urna $U(b, r)$ si effettuano n estrazioni con la modalità seguente: ad ogni estrazione si rimette la pallina estratta nell'urna solamente se è bianca. Dire se gli eventi $E_1, \ldots, E_n$ sono scambiabili.

[30] Con la frase "a caso" intendiamo far riferimento al caso simmetrico.

15. Dall'urna $U(4,6)$ si effettuano tre estrazioni con la seguente modalià: si rimette la pallina estratta nell'urna, se esce bianca, e si mettono nell'urna due palline rosse, altrimenti. Calcolare: $*$ $\Pr(E_3 \mid E_1 \vee E_2)$; $*$ la probabilità che ci siano sette palline rosse nell'urna alla fine delle tre estrazioni.
16. Date le urne $A(4,6)$ e $B(1,1)$, si tolgono da A due palline e le si immettono *senza guardarne il colore* in B. Si procede poi a quattro estrazioni con contagio unitario (cioè $a = 1$) dall'urna B. Calcolare la probabilità che siano state inserite in B due palline di colore diverso, sapendo che sono state estratte tre palline di medesimo colore.
17. Date le urne $A(4,2)$ e $B(3,3)$, si eseguono estrazioni senza rimessa con la seguente modalità: prima di ogni estrazione si lancia una moneta equilibrata e si estrae la pallina da A, se esce testa, e da B altrimenti. Dati gli eventi H_i: "esce testa al lancio i-simo", calcolare: $*$ $\Pr(E_2 \mid E_1 \wedge H_2)$; $*$ la probabilità che esca testa nel secondo lancio sapendo che nelle prime due estrazioni sono uscite due palline bianche.
18. Data l'urna $U(5,5)$, si lancia un dado equilibrato e si effettua un numero di estrazioni pari al numero uscito con la seguente modalità: senza rimessa, se il numero uscito è maggiore di tre, e con rimessa altrimenti. Calcolare la probabilità che tutte le palline estratte siano rosse.
19. Data l'urna $U(6,4)$ si lanciano due dadi equilibrati e si procede poi a estrazioni senza rimessa, se la somma dei numeri usciti è maggiore di 6, e a estrazioni con contagio unitario altrimenti. Calcolare: $*$ la probabilità che la seconda pallina estratta sia rossa sapendo che nella prima e terza estrazione sono uscite palline di colore diverso; $*$ la distribuzione di probabilità sulle possibili composizioni dell'urna alla fine della terza estrazione sapendo che nelle prime tre estrazioni sono state estratte più palline rosse che bianche; $*$ la correlazione tra gli eventi E: "la somma dei numeri usciti è maggiore di cinque" e H: "nelle prime tre estrazioni sono uscite due palline bianche".

Capitolo 3

Valutazione dell'incertezza II

Affrontate, nei capitoli precedenti, la descrizione e la valutazione della 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 limitate. Per introdurla seguiremo l'impostazione assiomatica ricorrendo, per giustificare gli assiomi, alla sua interpretazione in termini di prezzi equi di scommesse relative a importi aleatori (come peraltro fatto, con riferimento alle quote eque di scommesse relative a eventi, per la probabilità).

In questo capitolo, $\mathcal{A}$ denota una σ-algebra di eventi e Pr una probabilità numerabilmente additiva su $\mathcal{A}$. Precisiamo inoltre che non richiameremo più esplicitamente i teoremi 2.1.2, 2.2.2 e 2.3.2.

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

[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)). Siano X il numero aleatorio "primo numero estratto", che definiamo su $\mathcal{P}_{sec}$ tramite la versione: $f : E_{(n_1,n_2,n_3,n_4,n_5)} \to n_1$, e $\mathcal{A} = \mathcal{E}_{\mathrm{L}}(\mathcal{P}_{sem})$. Allora, ad esempio, l'evento $\{X \leq 10\} \notin \mathcal{A}$ in quanto è solo logicamente semidipendente da $\mathcal{P}_{sem}$. Chiaramente, la nozione di variabile aleatoria non è assoluta, ma

immediata di (1.4)) 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.*

Nei due teoremi seguenti elenchiamo alcune proprietà delle variabili aleatorie che saranno adoperate nel seguito. Il primo, conseguenza immediata del lemma precedente e (1.3), assicura che gli eventi costruiti tramite uguaglianze e/o disuguaglianze tra v.a. e numeri certi appartengono ad $\mathcal{A}$. Il secondo invece rileva 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 3.1.2. *Siano X una v.a. e a, b numeri reali. Allora, appartengono ad $\mathcal{A}$ gli eventi $\{a \leq X\}$, $\{a < X\}$, $\{X = a\}$ e $\{a \lhd X \prec b\}$, ove $\lhd, \prec$ denotano uno qualunque dei simboli $<$, $\leq$.*

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^-, $\frac{1}{X}$ $(0 \notin \mathrm{rg}(X))$ 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 $\{X \leq x\} = \varnothing$, se $x < c$. In ogni caso quindi $\{X \leq x\} \in \mathcal{A}$.

Dato $E \in \mathcal{A}$, sia $X = |E|$. Allora, $\{X \leq x\} = \varnothing$, 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, denotiamo, adottando le stesse notazioni della Sezione 1.12.1, con $f_1 = f^{(\mathcal{P}_1)}$ una versione di X, con $g_1 = g^{(\widetilde{\mathcal{P}}_1)}$ una versione di Y e con $\mathcal{P}^{(1)}$ la partizione costituita dai casi elementari $\omega^{(1)} = \omega_1 \wedge \widetilde{\omega}_1 \in \mathcal{P}_1 \wedge \widetilde{\mathcal{P}}_1$. Allora, le funzioni $f^{(1)}$, $g^{(1)}$ su $\mathcal{P}^{(1)}$ tali che $f^{(1)}(\omega^{(1)}) = f_1(\omega_1)$ e $g^{(1)}(\omega^{(1)}) = g_1(\widetilde{\omega}_1)$ sono versioni, rispettivamente, di f_1 e g_1.

Verifichiamo intanto che la somma $X + Y$ è una v.a. Poiché la funzione che associa ad ogni $\omega^{(1)}$ il valore $f_1(\omega_1) + g_1(\widetilde{\omega}_1)$ è la versione di $X + Y$ relativa a $\mathcal{P}^{(1)}$, risulta

$$\{X + Y < t\} = \bigvee_{\omega^{(1)}: f_1(\omega_1) + g_1(\widetilde{\omega}_1) < t} \omega^{(1)} = \bigvee_{\omega^{(1)}: f^{(1)}(\omega^{(1)}) + g^{(1)}(\omega^{(1)}) < t} \omega^{(1)}.$$

è relativa alla σ-algebra presa in considerazione; infatti, nell'esempio, X sarebbe diventato una v.a. se avessimo posto $\mathcal{A} = \mathcal{E}_{\mathrm{L}}(\mathcal{P}_{sec})$.

[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 $|X| = X^+ + X^-$.

Osservato che, per ogni $\omega^{(1)}$ e per ogni t reale, si ha $f^{(1)}(\omega^{(1)}) + g^{(1)}(\omega^{(1)}) < t \Leftrightarrow f^{(1)}(\omega^{(1)}) < r < t - g^{(1)}(\omega^{(1)}) \Leftrightarrow f^{(1)}(\omega^{(1)}) < r$ e $g^{(1)}(\omega^{(1)}) < t - r$ per qualche razionale r, tramite il Teorema 1.8.5(i),(iii),(iv), otteniamo

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

e quindi, per il Lemma 3.1.1, $\{X+Y<t\} = \bigvee_{r\in\mathbb{Q}}[\{X<r\}\wedge\{Y<t-r\}] \in \mathcal{A}$. Ne segue, per il lemma appena citato, che $X+Y$ 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,

$$\mathrm{set}(\{\tau(X)<t\}) = \{\omega_1 : \tau(f_1(\omega_1)) \in]-\infty, t[\,\} = \{\omega_1 : f_1(\omega_1) \in \tau^{-1}(]-\infty, t[)\}.$$

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 $r_x, r'_x \in \mathbb{Q}$ tali che $x \in [r_x, r'_x] \subset W$. Conseguentemente, $W = \bigcup_{x\in W}[r_x, r'_x]$ da cui otteniamo

$$\begin{aligned}
\mathrm{set}(\{\tau(X)<t\}) &= \{\omega_1 : f_1(\omega_1) \in \bigcup_{x\in W}[r_x, r'_x]\} = \bigcup_{x\in W}\{\omega_1 : f_1(\omega_1) \in [r_x, r'_x]\} \\
&= \bigcup_{x\in W}\{\omega_1 : r_x \le f_1(\omega_1) \le r'_x\} = \bigcup_{x\in W} \mathrm{set}(\{r_x \le X \le r'_x\}) \\
&= \mathrm{set}(\bigvee_{x\in W}\{r_x \le X \le r'_x\})
\end{aligned}$$

e quindi, per il Teorema 3.1.2, $\{\tau(X)<t\}) = \bigvee_{x\in W}\{r_x \le X \le r'_x\} \in \mathcal{A}$, 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 il quadrato di ogni v.a. è una v.a.(in quanto ottenuto tramite la funzione continua $\tau(x) = x^2$) e che $XY = \frac{1}{4}[(X+Y)^2 - (X-Y)^2]$. □

Osservazione 3.1.4. Date le v.a. X e Y, il teorema appena provato assicura che il numero aleatorio $Z = X - Y$ è una v.a.. Conseguentemente, per il Teorema 3.1.2, sono elementi di $\mathcal{A}$ gli eventi $\{X = Y\} = \{Z = 0\}$, $\{X < Y\} = \{Z < 0\}$ e $\{X \le Y\} = \{Z \le 0\}$. △

3.2 Funzione 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.

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, rileva 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:*

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

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

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

[3] Come nella nota 25 a piè di pagina 66, usiamo le notazioni più snelle $\Pr(X \leq x)$, $\Pr(a < X \leq b), \ldots$ al posto delle $\Pr(\{X \leq x\})$, $\Pr(\{a < X \leq b\}), \ldots$. Inoltre, adottiamo le notazioni più usuali $\Pr(X \geq x)$, $\Pr(X > x)$ al posto delle $\Pr(x \leq X)$ e $\Pr(x < X)$.

[4] Ove, $\inf X$ e $\sup X$ sono, nell'ordine, l'estremo inferiore e superiore del rango di 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) Discende dalla monotonia, osservato che, per (1.3), $\{X \leq x\} \to \{X \leq x\} \vee \{x < X \leq x'\} = \{X \leq x'\}$ per ogni $x < x'$.

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

Le proposizioni (iv) $\div$ (vii) seguono da (ii), dal Teorema di rappresentazione 1.8.5 e dalla continuità della probabilità[5]. Precisiamo che denotiamo con f la versione di X relativa alla partizione $\mathcal{P}$.

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

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

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

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

(vi) Data la successione non crescente $(\{X \leq x + \frac{1}{n}\})_{n\geq 1}$, da (1.4) si ha

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

(vii) Data la successione non decrescente $(\{X \leq x - \frac{1}{n}\})_{n\geq 1}$, da (1.4) si ha

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

[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 $(x_n)_{n\geq 1}$ convergente a t.

(viii) Conseguenza di (vi), (vii) e del Teorema 3.1.2, osservato che, per (1.3), $\Pr(X \leq x) = \Pr(X < x) + \Pr(X = x)$.

(ix) Conseguenza immediata del Teorema 3.1.2, di (1.3) e (vi), (vii).
L'ultima parte della tesi, segue banalmente da (viii) e dal Teorema 2.1.3. $\square$

Nell'osservazione seguente forniamo l'espressione della f.r. di v.a. con rango discreto; inoltre, formalizziamo l'idea intuitiva di *scelta a caso di un numero reale in un intervallo limitato* tramite la nozione di distribuzione uniforme in un intervallo.

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. Posto $p_i = \Pr(X = x_i)$ per ogni $i \leq n$, si ha

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

In particolare, se X è un numero certo, risulta $n = 1$ e quindi $p_1 = \Pr(X = x_1) = 1$; riesce allora

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

Se invece $X = |E|$ $(E \in \mathcal{A})$, si ha $x_1 = 0$, $x_2 = 1$ e quindi $p_1 = \Pr(\overline{E})$ e $p_2 = \Pr(E)$; risulta allora

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

(ii) Data un v.a. X con rango numerabile, siano $x_1 < x_2 < \dots$ i suoi valori possibili. Posto $p_i = \Pr(X = x_i)$ $(i \geq 1)$, si ha

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

(iii) Distribuzione uniforme in un intervallo Una v.a. X con rango incluso nell'intervallo $I = [a, b]$ $(a < b)$ è **distribuita uniformemente in** I se:

$$F_X(x) = \begin{cases} 0, & \text{se } x < a \\ \dfrac{x-a}{b-a}, & \text{se } a \le x \le b \\ 1, & \text{se } x > b \end{cases}.^{6}$$

Che questa particolare f.r. continua sia una possibile formalizzazione dell'idea intuitiva di "scelta a caso nell'intervallo di estremi $a < b$", 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:*

(i) $\Pr(X = x) = 0$, *per ogni* $0 \le x \le 1$;

(ii) Per ogni $n \ge 1$, *sia* $0 = x_0 < x_1 < \cdots < x_{n-1} < x_n = 1$ *un'arbitraria suddivisione di* $[0, 1]$ *in intervalli di ampiezza* $\frac{1}{n}$. *Posto* $E_1 = \{x_0 \le X \le x_1\}, E_2 = \{x_1 < X \le x_2\}, \ldots, E_n = \{x_{n-1} < X \le x_n\}$, *risulta* $\Pr(E_k) = \frac{1}{n}$ *per ogni* $k \le n$.[7]

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

Assumiamo ora (i) + (ii). Sia $x \le 0$; allora, per (1.3), $\{X \le 0\} = \{X < 0\} \vee \{X = 0\}$ e quindi, per (i), $F(0) = \Pr(X < 0) + \Pr(X = 0) = 0$. Sia $x \ge 1$; allora $\{0 \le X \le 1\} \to \{X \le x\}$ e quindi, per la monotonia, $F(x) = 1$. Supposto infine $0 < x < 1$, sia intanto x razionale. Allora $x = \frac{m}{n}$ con $1 \le m < n$. Ne segue (procedendo per induzione su n a partire dall'ultima formula di (1.3))

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

e quindi, considerata la suddivisione $x_i = \frac{i}{n}$ $(i = 0, \ldots, n)$, da (ii) otteniamo

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

---

[6] L'espressione della f.r. assicura che $\inf X = a$ e $\sup X = b$. Infatti, se fosse $\inf X > a$, si perverrebbe, tramite il Teorema 3.2.1(ix), alla contraddizione $0 = \Pr(a \le X < \inf X)) = F_X(\inf X) - F_X(a) = \frac{\inf X - a}{b-a} > 0$; se fosse $\sup X < b$, si avrebbe la contraddizione $0 = \Pr(\sup X < X \le b) = F_X(b) - F_X(\sup X) = 1 - \frac{\sup X - a}{b-a} = \frac{b - \sup X}{b-a} > 0$.

[7] Cioè, per ogni ripartizione dell'intervallo unitario in intervalli di uguale ampiezza, sono uguali le probabilità che la v.a. assuma un valore appartenente ad uno qualsiasi di essi.

Sia ora x 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 continue 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 continua strettamente monotona in un intervallo includente il rango di X. Allora, la trasformata $Y = \tau(X)$ è una v.a. tale che:*

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

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

Dimostrazione. Poniamo $F = F_X$, $G = F_Y$, $\alpha = \inf Y$, $\beta = \sup Y$ e $\iota = \tau^{-1}$.

Per il Teorema 3.2.1(iii), $G(y) = 0$, se $y < \alpha$, e $G(y) = 1$, se $y \geq \beta$.

Sia ora $\alpha < y < \beta$. Esistono allora $x', x'' \in \mathrm{rg}(X)$ tali che $\tau(x') < y < \tau(x'')$. Ne segue, per il teorema di connessione delle funzioni continue[8], $y = \tau(x)$ per qualche x nel dominio di τ. Dunque, $\{\tau(X) \leq y\} = \{\tau(X) \leq \tau(x)\} = \{\tau(X) \leq \tau(\iota(y))\}$ e quindi $\{\tau(X) \leq y\} = \{X \leq \iota(y)\}$, se τ è crescente, e $\{\tau(X) \leq y\} = \{X \geq \iota(y)\}$, se τ è decrescente. Dunque, $G(y) = F(\iota(y))$, se τ è crescente, e $G(y) = 1 - F(\iota(y)^-)$, se τ è decrescente (Teorema 3.2.1(ix)).

Sia infine $y = \alpha$. Esiste allora una successione decrescente $(y_n = \tau(x_n))_{n\geq 1}$ convergente ad α con $x_n \in \mathrm{rg}(X)$ per ogni $n \geq 1$. Supponiamo intanto che τ sia crescente. Ne segue che la successione $(x_n)_{n\geq 1}$ 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 $x, x' \in \mathrm{rg}(X)$ tali che $x_n \geq x_0 > x > x' > \inf X$ per

[8]Una funzione continua in un intervallo ha come insieme-imagine un intervallo.

ogni $n \geq 1$. Risulta pertanto $y_n > \tau(x) > \tau(x')$ per ogni $n \geq 1$ e quindi, passando al limite per $n \to +\infty$, si ha la contraddizione $\alpha \geq \tau(x) > \tau(x') \geq \alpha$. Dunque $x_0 = \inf X$. Allora, se $\alpha \in \mathbb{R}$, tramite (1.4), si ha

$$\begin{aligned} \{Y \leq y\} &= \{Y \leq \alpha\} = \bigwedge_{n\geq 1} \{Y \leq y_n\} = \bigwedge_{n\geq 1} \{Y \leq \tau(x_n)\} \\ &= \bigwedge_{n\geq 1} \{X \leq x_n\} = \begin{cases} \{X \leq \inf X\} & \text{se } \inf X \in \mathbb{R} \\ \varnothing & \text{se } \inf X = -\infty \end{cases} \end{aligned}$$

e quindi $G(y) = F(\inf X)$, tenuto conto del Teorema 3.2.1(iv), Se invece $\alpha = -\infty$, risulta

$$G(y) = \lim_{t\to\alpha} G(t) = \lim_{n\to+\infty} G(\tau(x_n)) = \lim_{n\to+\infty} F(x_n) = F((\inf X)^+) = F(\inf X),$$

tenuto conto del Teorema 3.2.1(iv),(vi). Nel caso τ decrescente, si procede in modo analogo, tenendo presente anche il Teorema 3.2.1(ix). □

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 $\tau(x) = x^2$, risulta, per 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 $y \geq 0$. △

Esempio 3.2.5. (i) Sia X una v.a. distribuita uniformemente nell'intervallo I di estremi $a < b$ (Osservazione 3.2.2(iii)) e $\tau(x) = cx + d$ ($c \neq 0$) per ogni $x \in [a, b]$. Allora $\iota(y) = \tau^{-1}(y) = \frac{y-d}{c}$ per ogni $y \in \tau([a, b])$. Considerata la v.a. trasformata $Y = cX + d$, sia intanto $c > 0$. Allora, $\tau([a, b]) = [\tau(a), \tau(b)]$ da cui otteniamo $\inf Y = \tau(a)$, $\sup Y = \tau(b)$ e quindi $F_Y(y) = 0$, se $y < \tau(a)$, e $F_Y(y) = 1$, se $y \geq \tau(b)$. Poiché $\inf X = a$ (nota 6 a piè di pagina 83), $F_Y(\tau(a)) = F_X(a) = 0$. Sia infine $\tau(b) < y < \tau(a)$. Allora $a < \iota(Y) < b = \sup X$ e quindi

$$F_Y(y) = F_X(\iota(y)) = \frac{\frac{y-d}{c} - a}{b-a} = \frac{y - \tau(a)}{c(b-a)} = \frac{y - \tau(a)}{\tau(b) - \tau(a)}.$$

Conseguentemente, Y è distribuito uniformemente nell'intervallo $[\tau(a), \tau(b)]$.

Supponiamo ora $c < 0$. Allora, $\tau([a, b]) = [\tau(b), \tau(a)]$ da cui otteniamo $\inf Y = \tau(b)$, $\sup Y = \tau(a)$ e quindi $F_Y(y) = 0$, se $y < \tau(b)$, e $F_Y(y) = 1$, se $y \geq \tau(a)$. Risulta inoltre $F_Y(\tau(b)) = 1 - F_X(b) = 0$. Sia infine $\tau(b) < y < \tau(a)$. Allora $a < \iota(Y) < b$ e quindi

$$F_Y(y) = 1 - F_X(\iota(y)) = 1 - \frac{\frac{y-d}{c} - a}{b-a} = \frac{\tau(b) - y}{c(b-a)} = \frac{y - \tau(b)}{\tau(a) - \tau(b)}.$$

Dunque, Y è distribuito uniformemente in $[\tau(b), \tau(a)]$.

(ii) Sia X una v.a. distribuita uniformemente nell'intervallo $]-\frac{\pi}{2}, \frac{\pi}{2}[$ e $\tau(x) = \operatorname{tg} x$ per ogni $x \in]-\frac{\pi}{2}, \frac{\pi}{2}[$. Allora, $\iota(y) = \operatorname{tg}^{-1}(y) = \operatorname{arctg} y$ per ogni y reale. Si ha quindi la f.r.:

$$F_Y(y) = F_X(\iota(y)) = \frac{\operatorname{arctg} y + \frac{\pi}{2}}{\frac{\pi}{2} + \frac{\pi}{2}} = \frac{\operatorname{arctg} y}{\pi} + \frac{1}{2}$$

detta **distribuzione di Cauchy**.

(iii) Considerati un evento E di probabilità $\frac{1}{2}$ e una v.a. Z distribuita uniformemente in $I = [0,1]$, sia $X = -|E| + Z|\overline{E}| = Z - (Z+1)|E|$[9]. Poiché $\operatorname{rg}(X) = \{-1\} \cup I$, risulta $F_X(x) = 0$, se $x < -1$, e $F_X(x) = 1$, se $x > 1$. Inoltre, se $-1 \leq x < 0$, dalla formula di disintegrazione si ha

$$\begin{aligned} F_X(x) &= \Pr(X \leq x) = \Pr(\{X \leq x\}|E)\Pr(E) + \Pr(\{X \leq x\}|\overline{E})\Pr(\overline{E}) \\ &= \frac{1}{2}\left[\Pr(\{-|E| \leq x\}|E) + \Pr(\{Z \leq x\}|\overline{E}) = \frac{1}{2}\right. \end{aligned}$$

Infine, se $0 \leq x \leq 1$, risulta

$$F_X(x) = \frac{1}{2}\left[\Pr(\{-|E| \leq x\}|E) + \Pr(\{Z \leq x\}|\overline{E})\right] = \frac{1}{2}(1+x).$$

Dunque

$$F_X(x) = \begin{cases} 0, & \text{se } x < -1 \\ \frac{1}{2}, & \text{se } -1 \leq x < 0 \\ \frac{1}{2}(x+1), & \text{se } 0 \leq x < 1 \\ 1, & \text{se } x \geq 1 \end{cases}.$$

Ne segue, considerata la v.a. $Y = 1 - X$,

$$F_Y(y) = \begin{cases} 0, & \text{se } y < 0 \\ \frac{y}{2}, & \text{se } 0 \leq y < 1 \\ \frac{1}{2}, & \text{se } 1 \leq y < 2 \\ 1, & \text{se } y \geq 2 \end{cases},$$

osservato che $\inf Y = 0$ e $\sup Y = 2$.

Considerata invece la v.a. $U = X^2$, otteniamo

$$F_U(t) = \begin{cases} \frac{\sqrt{t}}{2}, & \text{se } 0 \leq t < 1 \\ 1, & \text{se } t \geq 1 \end{cases},$$

ricordato che, per l'Osservazione 3.2.4, $F_U(t) = F_X(\sqrt{t}) - F_X((-\sqrt{t})^-)$. ◊

[9] La v.a. X può interpretarsi come il risultato della scommessa relativa al lancio di una moneta equilibrata: si perde 1, se esce testa, e si riceve un importo pari al numero estratto a caso in $[0,1]$, altrimenti.

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 è una funzione integrale, nel senso che esiste una funzione (detta (**funzione di**) **densità** di X) $f \geq 0$ (Riemann) integrabile su ogni intervallo limitato[10] 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[11]. Per il teorema di additività dell'integrale[12] si ha

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

e quindi $F_X(x) = F_X(x^-)$. Ne segue, per 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(iv),(v),(ix), si ha

$$\Pr(\alpha \lhd X \prec \beta) = F_X(\beta) - F_X(\alpha) = \int_\alpha^\beta f(t)dt \tag{3.1}$$

per ogni α, β tali che $-\infty \leq \alpha < \beta \leq +\infty$, avendo denotato, come al solito, con $\lhd, \prec$ uno qualunque dei simboli $<, \leq$[13]. Dunque $\int_{-\infty}^{+\infty} f(t)dt = 1$.

Sia ora x_0 un punto di continuità di f. Dati $\delta, \delta' \geq 0$ tali che $\Delta = \delta + \delta' >$

---

[10]E quindi limitata su ogni intervallo limitato. Ricordiamo che esempi di funzioni integrabili su ogni intervallo limitato sono le continue, le generalmente continue (cioè con un numero finito di punti di discontinuità) e, più in generale, le monotone limitate.

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

[12]Se $-\infty \leq \alpha < a < b$, allora $\int_\alpha^b f(t)dt = \int_\alpha^a f(t)dt + \int_a^b f(t)dt$.

[13]Ricordiamo che, data una funzione g integrabile su ogni intervallo limitato dell'asse reale, si ha (per definizione) $\int_{-\infty}^{+\infty} g(x)dx = \lim_{(a,b) \to (-\infty, +\infty)} \int_a^b g(x)dx$. Per quanto riguarda il limite, ricordiamo che $\lim_{(a,b) \to (-\infty, +\infty)} h(x, y) = \lambda \in \mathbb{R}$ significa che per ogni $\epsilon > 0$ esiste $k > 0$ tale che per ogni $x < -k$ e $y > k$ risulta $|h(x, y) - \lambda| < \epsilon$.

0, otteniamo, tenuto conto di (3.1) e del teorema della media[14]

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

con λ compreso tra gli estremi inferiore e superiore di f su $[x_0 - \delta', x_0 + \delta]$. Conseguentemente, per la continuità di f in x_0, si ha $\lambda = f(x_0) + o(\Delta)$[15]. Allora

$$\Pr(x_0 - \delta' < X < x_0 + \delta) = f(x_0)\Delta + o(\Delta)\Delta$$

e quindi

$$\lim_{\Delta\to 0^+} \frac{\Pr(x_0 - \delta' < X < x_0 + \delta)}{\Delta} = f(x_0). \tag{3.2}$$

Dunque, 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.

Ora, tramite (3.1), ponendo in (3.2) $\delta' = 0$, risulta

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

e ponendo $\delta = 0$, otteniamo

$$\begin{aligned} f(x_0) &= \lim_{\Delta\to 0^+} \frac{\Pr(x_0 - \delta' < X < x_0)}{\Delta} = \lim_{\delta'\to 0^+} \frac{F_X(x_0) - F_X(x_0 - \delta')}{\delta'} \\ &= \lim_{t\to 0^-} \frac{F_X(x_0 + t) - F_X(x_0)}{t}. \end{aligned}$$

Dunque, le derivate destra e sinistra della funzione di ripartizione coincidono con il valore della densità nel punto x_0. Ne segue, tenuto conto dell'arbitrarietà dl x_0, che la funzione di ripartizione è derivabile in ogni punto di contiuità x_0 della funzione di densità e riesce

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

[14] Sia f una funzione integrabile su $[a, b]$. Esiste allora un numero reale λ compreso tra gli estremi inferiore e superiore di f su $[a, b]$ tale che $\int_a^b f(x)dx = \lambda(b - a)$.

[15] Ove, usando la *notazione di Landau*, $o(\Delta)$ denota un infinitesimo d'ordine superiore al primo rispetto a Δ.

Osservazione 3.3.1. La funzione di densità non è unica (l'integrale di Riemann non cambia se si modifica 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(iii)), è una densità di X la funzione:

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

qualunque sia $\alpha \geq 0$. △

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

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

Dimostrazione. Poichè τ è continua (essendo derivabile), per il Teorema 3.1.3, Y è una v.a.. Posto $F = F_X$, $G = F_Y$, $a = \inf X$, $b = \sup X$, $\alpha = \inf Y$, $\beta = \sup Y$ e $\iota = \tau^{-1}$, dal Teorema 3.2.1(iii) otteniamo $F(a) = 0$ e $F(b) = 1$, tenuto conto della continuità di F.

Assumiamo ora τ crescente. Ne segue $a = \iota(\alpha)$ e $b = \iota(\beta)$ e quindi, per il Teorema 3.2.3[16], $G(y) = 0 = \int_{-\infty}^{y} f_Y(y)dy$ per ogni $y \leq \alpha$. Sia pertanto $y \in]\alpha, \beta]$. Allora, $G(y) = F(\iota(y))$. Notato che $a < \iota(y)$, dal Teorema 3.2.1(ix) e dalla continuità di F, si ha $F(\iota(y)) = F(a) + \Pr(a \leq X < \iota(y)) = \Pr(a \leq X < \iota(y))$ 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[17], ι è derivabile e riesce $\iota'(z) = \tau'(i(z))^{-1}$ per ogni $z \in I$. Notato che, per il teorema di continuità della funzione inversa[18],

[16] Che non verrà più citato esplicitamente nel corso della dimostrazione.

[17] Sia g una funzione strettamente monotona derivabile in x_0 con derivata diversa da zero. Allora, la funzione inversa è derivabile nel punto $z_0 = g(t_0)$ e risulta $(g^{-1})'(z_0) = g'(t_0)^{-1}$.

[18] È continua la funzione inversa di una funzione strettamente monotona su un intervallo chiuso e limitato.

ι è continua otteniamo, per la continuità della composta, che anche la derivata ι' è continua. Effettuando nell'ultimo integrale il cambio di variabile $t \rightsquigarrow \iota(z)$ risulta, tramite il teorema d'integrazione per sostituzione[19],

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

tenuto conto che la derivata ι' è positiva, essendo ι crescente. Ne segue, in particolare, $\int_{-\infty}^{\beta} f_Y(z)dz = 1$ da cui, se $y > \beta$, si ha

$$\int_{-\infty}^{y} f_Y(z)dz = \int_{-\infty}^{\beta} f_Y(z)dz + \int_{\beta}^{y} f_Y(z)dz = \int_{-\infty}^{\beta} f_Y(z)dz = 1 = G(y).$$

Assumiamo infine τ decrescente. Ne segue $a = \iota(\beta)$ e $b = \iota(\alpha)$. Poiché nel caso $y < \alpha$ l'uguaglianza $G(y) = \int_{-\infty}^{y} f_Y(y)dy$ è banale, assumiamo $y = \alpha$. Allora, per la continuità di F, si ha $G(y) = 1 - F(b) = 0 = \int_{-\infty}^{y} f_Y(y)dy$. Sia ora $y \in]\alpha, \beta]$. Se $y = \beta$, allora, per (3.1), risulta

$$G(\beta) = 1 = F(b) - F(a) = \int_{a}^{b} f(t)dt = \int_{\iota(\beta)}^{\iota(\tau(b))} f(t)dt.$$

Se $y < \beta$, si ha $a < \iota(y)$ da cui, per il Teorema 3.2.1(ix), otteniamo $F(\iota(y)) = \Pr(a \le X < \iota(y))$ e quindi, per (3.1),

$$G(y) = 1 - \int_{a}^{\iota(y)} f(t)dt = \int_{a}^{b} f(t)dt - \int_{a}^{\iota(y)} f(t)dt = \int_{\iota(y)}^{b} f(t)dt = \int_{\iota(y)}^{\iota(\tau(b))} f(t)dt.$$

Risulta dunque $G(y) = \int_{\iota(y)}^{\iota(\tau(b))} f(t)dt$ per ogni $y \in]\alpha, \beta]$. Poiché ι è derivabile con derivata prima continua, tramite il cambio di variabile $t \rightsquigarrow \iota(z)$ nell'integrale, otteniamo

$$\begin{aligned} G(y) &= \int_{y}^{\tau(b)} f(\iota(z))\iota'(z)dz = \int_{y}^{\alpha} f(\iota(z))\iota'(z)dz = \int_{\alpha}^{y} f(\iota(z))(-\iota'(z))dz \\ &= \int_{\alpha}^{y} f(\iota(z))|\iota'(z)|dz = \int_{-\infty}^{y} f_Y(z)dz, \end{aligned}$$

tenuto conto che la derivata ι' è negativa, essendo ι decrescente. Ne segue, in particolare, $\int_{-\infty}^{\beta} f_Y(z)dz = 1$ da cui, se $y > \beta$, si ha $\int_{-\infty}^{y} f_Y(z)dz = \int_{-\infty}^{\beta} f_Y(z)dz = 1 = G(y)$. La dimostrazione è coì conclusa. $\square$

Esempio 3.3.3. Siano X una v.a. assolutamente continua con densità f continua in un intervallo chiuso e limitato I includente il rango di X e $\tau(x) = cx + d$ ($c \neq 0$)

---

[19] Siano g una funzione continua nell'intervallo $I = [a, b]$ e h una funzione derivabile con derivata continua nell'intervallo $J = [c, d]$ tale che $h(J) \subseteq I$. Allora, $\int_{h(u)}^{h(v)} g(t)dt = \int_{u}^{v} g(h(z))h'(z)dz$ per ogni $u, v \in J$ ($u < v$).

per ogni $x \in I$. Allora $\iota(y) = \tau^{-1}(y) = \frac{y-d}{c}$ e quindi $\iota'(y) = c^{-1}$. Risulta pertanto

$$f_Y(y) = \begin{cases} \dfrac{1}{|c|} f\left(\dfrac{y-d}{c}\right), & \text{se } \inf Y \leq y \leq \sup Y \\ 0, & \text{altrimenti} \end{cases}.$$

Conseguentemente, se X è una v.a. **simmetrica** (cioè ha una funzione di densità pari[20]), possiamo assumere come densità di $-X$ quella di X. $\Diamond$

3.4 Speranza matematica

Considerata la famiglia $\mathcal{V}_b$ delle variabili aleatorie **limitate** (cioè con rango limitato)[21], si chiama **speranza matematica** (o **valor medio**) **su** $\mathcal{V}_b$ (relativa alla probabilità Pr prefissata) ogni applicazione $\mathrm{E} : \mathcal{V}_b \to \mathbb{R}$ che verifica gli assiomi:

E1 $\quad \mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ (additività).

E2 $\quad \mathrm{E}(X) \leq \mathrm{E}(Y)$, se $X \leq Y$ (monotonia).

E3 $\quad \mathrm{E}(|E|) = \Pr(E)$ (compatibilità).[22]

Conveniamo che E denoti una speranza matematica su $\mathcal{V}_b$ e che X, Y, Z (dotati o no di apici o pedici) rappresentino, salvo avviso contrario, variabili aleatorie limitate.

Nell'osservazione seguente forniamo una giustificazione degli assiomi proposti mediante lo schema della scommessa.

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 pari a $S(v|E| - p)$, ove S vale 1, se si scommette su E, e -1, se si scommette contro E. Ora, poiché nell'espressione del guadagno compare l'importo aleatorio $Y = v|E|$, possiamo rivedere la scommessa come un'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.

[20]Ricordiamo che una funzione g è *pari* se $g(-x) = g(x)$ per ogni x del suo dominio.

[21]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).

[22]La prova dell'esistenza di questo tipo di applicazioni esula dai limiti imposti al corso.

Alla luce di questa osservazione chiamiamo, dato un importo aleatorio X arbitrario, *scommessa relativa a* X un'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 i valori di 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 degli 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 $\mathbf{S} = (S_1, \dots, S_n) \in \{-1, 1\}^n$.

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

$$\begin{aligned} G^{(X,Y,X+Y)}(\mathbf{p};\mathbf{S}) &= S'(X-p') + S''(Y-p'') + S(X+Y-p) \\ &= [(S'+S)X + (S''+S)Y\,] - (S'p' + S''p'' + Sp). \end{aligned}$$

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 gli eventi sono incompatibili (Teorema 1.12.1(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)[23]; 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 $\mathcal{A}$. $\triangle$

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, (x) assicura che una v.a. non negativa con valor medio nullo assume quasi certamente il valore zero.

Teorema 3.4.2. *Sussistono le proposizioni:*

(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* $X \geq 0$*;*

[23] 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").

(iv) $|\mathrm{E}(X)| \leq \mathrm{E}(|X|)$;

(v) *Se* X *è il numero certo di valore* α*, allora* $\mathrm{E}(X) = \alpha$;

(vi) *Siano* X *una v.a. con* $\mathrm{rg}(X) = \{x_1, \ldots, x_n\}$ *e* τ *una funzione definita su* $\mathrm{rg}(X)$*. Allora,* $\tau(X)$ *è una v.a. limitata e risulta:*

$$\mathrm{E}(\tau(X)) = \sum_{i=1}^{n} \tau(x_i) \Pr(X = x_i).$$

In particolare, se τ *è l'identità,* $\mathrm{E}(X) = \sum_{i=1}^{n} x_i \Pr(X = x_i)$;

(vii) *Se* $X \geq 0$ *e* $a > 0$*, allora*

$$\Pr(X \geq a) \leq \frac{\mathrm{E}(X)}{a} \qquad \text{(disuguaglianza di Markov)};$$

(viii) *Se* $X \neq 0$ *e* $a > 0$*, allora*

$$\Pr(|X| > a) \geq \frac{\mathrm{E}(X^2) - a^2}{\sup(X^2)} \qquad \text{(disuguaglianza di Kolmogorov)};$$

(ix) *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) dalla monotonia E2, (i) (osservato che $-|X| \leq X \leq |X|$).

(i) Dall'additività E1 risulta, tramite induzione su n,

$$\mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{E}(X_i) \tag{3.4}$$

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 $\mathrm{E}(0 \cdot X) = \mathrm{E}(|\varnothing|) = \Pr(\varnothing) = 0$. Se $\alpha = -1$, da E1 otteniamo $\mathrm{E}(X) + \mathrm{E}(-X) = \mathrm{E}(X + (-X)) = \mathrm{E}(|\emptyset|) = 0$ e quindi

$$\mathrm{E}(-X) = -\mathrm{E}(X). \tag{3.5}$$

Supponiamo ora $\alpha > 0$ e procediamo per casi. Sia $\alpha = n$. Allora, tramite (3.4) risulta

$$\mathrm{E}(nX) = \mathrm{E}(\overbrace{X + \cdots + X}^{n \text{volte}}) = \overbrace{\mathrm{E}(X) + \cdots + \mathrm{E}(X)}^{n \text{volte}} = n\mathrm{E}(X).$$

Sia $\alpha = \frac{1}{n}$. 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 $\alpha = \frac{m}{n}$. 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 α irrazionale. Esistono 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 $r_nX \leq \alpha X \leq q_nX$ da cui, per la monotonia E2, $r_n\mathrm{E}(X) = \mathrm{E}(r_nX) \leq \mathrm{E}(\alpha X) \leq \mathrm{E}(q_nX) = q_n\mathrm{E}(X)$ e quindi, passando al limite per $n \to +\infty$, si ha $\mathrm{E}(\alpha X) = \alpha\mathrm{E}(X)$. Sia ora X arbitrario. Considerate le parti positiva e negativa di X, otteniamo, tramite (3.4), (3.5) e quanto appena provato, $\mathrm{E}(\alpha X) = \mathrm{E}(\alpha(X^+ - X^-)) = \mathrm{E}(\alpha X^+ + (-\alpha X^-)) = \mathrm{E}(\alpha X^+) + \mathrm{E}(-(\alpha X^-)) = \mathrm{E}(\alpha X^+) - \mathrm{E}(\alpha X^-) = \alpha\mathrm{E}(X^+) - \alpha\mathrm{E}(X^-) = \alpha[\mathrm{E}(X^+) - \mathrm{E}(X^-)] = \alpha\mathrm{E}(X^+ - X^-) = \alpha\mathrm{E}(X)$.

Supponiamo infine $\alpha < 0$. Allora, per (3.5), $\mathrm{E}(\alpha\, X) = \mathrm{E}(-|\alpha|\, X) = -\mathrm{E}(|\alpha|\, X) = -|\alpha|\,\mathrm{E}(X) = \alpha\mathrm{E}(X)$.

(ii) Posto $\alpha = \inf X$, $\beta = \sup X$, dalla compatibilità E3, (i) e la monotonia 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, per il Teorema 3.1.3, $\tau(X) = \sum_{i=1}^n \tau(x_i)\,|\{X = x_i\}| \in \mathcal{V}_b$, da (i) e la compatibilità E3 otteniamo $\mathrm{E}(\tau(X)) = \sum_{i=1}^n \tau(x_i)\,\mathrm{E}(|\{X = x_i\}|) = \sum_{i=1}^n \tau(x_i)\,\mathrm{Pr}(X = x_i)$.

(vii) Poiché $a\,|\{X \geq a\}| \leq X$, dalla compatibilità, (i) e la monotonia si ha $a\,\mathrm{Pr}(X \geq a) = a\,\mathrm{E}(|\{X \geq a\}|) = \mathrm{E}(a|\{X \geq a\}|) \leq \mathrm{E}(X)$ e quindi la tesi.

(viii) Osservato che, per il Teorema 1.12.1(iii), risulta

$$\begin{aligned} X^2 &= X^2(|\{|X| \leq a\}| + |\{|X| > a\}|) = X^2|\{|X| \leq a\}| + X^2|\{|X| > a\}| \\ &\leq a^2 + \sup(X^2)|\{|X| > a\}| \end{aligned}$$

dalla monotonia, (i) e la compatibilità otteniamo

$$\mathrm{E}(X^2) \leq a^2 + \sup(X^2)\mathrm{E}(|\{|X| > a\}|) = a^2 + \sup(X^2)\,\mathrm{Pr}(|X| > a)$$

e quindi la tesi ($\sup(X^2) > 0$!).

(ix) Data la successione non crescente di eventi $(\{X < \frac{1}{n}\})_{n\geq 1}$, tramite (1.4) e la continuità dall'alto della probabilità, risulta

$$\begin{aligned} \mathrm{Pr}(X = 0) &= \mathrm{Pr}(X \leq 0) = \mathrm{Pr}(\bigwedge_{n\geq 1}\{X < \frac{1}{n}\}) \\ &= \lim_{n\to+\infty} \mathrm{Pr}(X < \frac{1}{n}) = 1 - \lim_{n\to+\infty} \mathrm{Pr}(X \geq \frac{1}{n}) = 1, \end{aligned}$$

osservato che, per (vii), $\mathrm{Pr}(X \geq \frac{1}{n}) \leq \mathrm{E}(X)\,(\frac{1}{n})^{-1} = 0$. □

Il prossimo risultato fornisce l'espressione del valor medio di trasformate certe di v.a. con rango numerabile.

Teorema 3.4.3. *Sia X una v.a. con $\mathrm{rg}(X) = \{x_1, x_2, \dots\}$ numerabile e τ una funzione limitata su $\mathrm{rg}(X)$ tale che $\tau(X) \in \mathcal{V}_b$. Allora, la serie $\sum_{n\geq 1} \tau(x_n) \Pr(X = x_n)$ è assolutamente convergente e riesce:*

$$\mathrm{E}(\tau(X)) = \sum_{n\geq 1} \tau(x_n) \Pr(X = x_n).$$

In particolare, se τ è l'identità, $\mathrm{E}(X) = \sum_{n\geq 1} x_n \Pr(X = x_n)$.

Dimostrazione. Posto $Y = \tau(X)$, $y_n = \tau(x_n)$ e $E_n = \{X = x_n\}$ $(n \geq 1)$, risulta, tramite il Teorema 1.12.1(v), $\sum_{n\geq 1} |E_n| = |\bigvee_{n\geq 1} E_n| = |\Omega|$, cioè:

$$\sum_{n\geq 1} |E_n| = 1. \tag{3.6}$$

Supposto $y_n \in [a, b]$ per ogni $n \geq 1$ (τ è limitata su $\mathrm{rg}(X)$!), la convergenza assoluta della serie discende immediatamente dalla

$$\sum_{n\geq 1} |y_n| \Pr(E_n) \leq \sum_{n\geq 1} \max(|a|, |b|) \Pr(E_n) = \max(|a|, |b|) \sum_{n\geq 1} \Pr(E_n) = \max(|a|, |b|).$$

Per provare la seconda parte della tesi, poniamo $F_n = \bigvee_{i=1}^n E_i$ e consideriamo le v.a. semplici (teoremi 3.1.2 e 3.1.3):

$$\underline{Y}_n = \sum_{i=1}^n y_i\, |E_i| + a\, |\overline{F}_n|, \qquad \overline{Y}_n = \sum_{i=1}^n y_i\, |E_i| + b\, |\overline{F}_n|$$

per ogni $n \geq 1$. Risulta allora

$$\underline{Y}_n \leq Y \leq \overline{Y}_n. \tag{3.7}$$

Notato che $Y = \sum_{n\geq 1} y_n\, |E_n|$ e $\sum_{i=1}^n |E_i| = |\bigvee_{i=1}^n E_i| = |F_n| = 1 - |\overline{F}_n|$, procediamo per casi. Sia intanto $|\overline{F}_n| = 0$. Allora, per (3.6), $\sum_{i\geq n+1} |E_i| = 0$, cioè $|E_i| = 0$ per ogni $i \geq n$; ne segue, $\sum_{i\geq n+1} y_i\, |E_i| = 0$ da cui $Y = \sum_{i=1}^n y_i |E_i|$ e quindi (3.7). Sia ora $|\overline{F}_n| = 1$. Allora, per (3.6), $\sum_{i=1}^n |E_i| = 0$ e quindi $\sum_{i\geq n+1} |E_i| = 1$; ne segue $Y = \sum_{i\geq n+1}^n y_i\, |E_i|$ da cui otteniamo $b \sum_{i\geq n+1} |E_i| = \sum_{i\geq n+1} b\, |E_i| \geq Y \geq \sum_{i\geq n+1} a\, |E_i| = a \sum_{i\geq n+1} |E_i| = a$ e quindi (3.7).

Tramite (3.7) e la monotonia E2, risulta $\mathrm{E}(\underline{Y}_n) \leq \mathrm{E}(Y) \leq \mathrm{E}(\overline{Y}_n)$ e quindi, per il Teorema 3.4.2(vi) e la compatibilità E3,

$$\sum_{i=1}^n y_i \Pr(E_i) + a \Pr(\overline{F}_n) \leq \mathrm{E}(Y) \leq \sum_{i=1}^n y_i \Pr(E_i) + b \Pr(\overline{F}_n). \tag{3.8}$$

Ora, poiché $\bigvee_{n\geq 1} F_n = \bigvee_{n\geq 1} E_n = \Omega$ e la successione $(F_n)_{n\geq 1}$ è non decrescente, si ha, per la continuità dal basso della probabilità, $\Pr(F_n) \uparrow \Pr(\Omega) = 1$ e quindi $\Pr(\overline{F}_n) = 1 - \Pr(F_n) \downarrow 0$. Conseguentemente, passando al limite per $n \to +\infty$ (la serie è convergente in quanto assolutamente convergente!), da (3.8) otteniamo $\sum_{n\geq 1} y_n \Pr(E_n) \leq \mathrm{E}(Y) \leq \sum_{n\geq 1} y_n \Pr(E_n)$, cioè la tesi. $\square$

Poiché la serie considerata nel teorema precedente è permutabile (essendo assolutamente convergente), la sua somma non dipende dal modo in cui vengono numerati gli elementi del rango 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. Precisiamo che tale notazione verrà adottata anche nel caso che il rango sia finito.

Il teorema seguente fornisce l'espressione del valor medio di trasformate certe continue di v.a. assolutamente continue.

Teorema 3.4.4. *Sia X una v.a. assolutamente continua con rango incluso in $[a, b]$ e densità f continua in $[a, b]$. Sia inoltre τ una funzione continua in $[a, b]$. Risulta allora:*

$$\mathrm{E}(\tau(X)) = \int_a^b \tau(x) f(x) dx.$$

In particolare, se τ è la funzione identica, $\mathrm{E}(X) = \int_a^b x f(x) dx$.

Dimostrazione. Osservato che ha senso considerare il valor medio di $\tau(X)$, in quanto, per il Teorema 3.1.3 e il teorema di Weierstrass[24], $\tau(X) \in \mathcal{V}_b$, proviamo l'uguaglianza.

Per il teorema da Heine[25], 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 $|f(x') - f(x'')| < \epsilon$ e $|\tau(x')f(x') - \tau(x'')f(x'')| < \epsilon$ per ogni $x', x'' \in J_i = [x_{i-1}, x_i]$ $(i = 1, \ldots, n)$. Ricorrendo al teorema di Weierstrass, poniamo $\alpha_i = \min(\tau f)(J_i)$, $\beta_i = \max(\tau f)(J_i)$ e $\Delta_i = x_i - x_{i-1}$ $(i = 1, \ldots, n)$. Allora, per la definizione dell'integrale di Riemann, si ha

$$\sum_{i=1}^n \alpha_i \Delta_i \leq \int_a^b \tau(x) f(x) dx \leq \sum_{i=1}^n \beta_i \Delta_i = \sum_{i=1}^n \alpha_i \Delta_i + \sum_{i=1}^n (\beta_i - \alpha_i)\Delta_i,$$

osservato che le sommatorie sono, rispettivamente, una somma inferiore e una somma superiore relative alla suddivisione considerata. Poiché

$$\sum_{i=1}^n (\beta_i - \alpha_i)\Delta_i < \epsilon \sum_{i=1}^n \Delta_i = (b - a)\,\epsilon \tag{3.9}$$

[24] Ogni funzione continua definita su un insieme compatto ammette minimo e massimo.

[25] Ogni funzione continua definita su un insieme compatto K è uniformemente continua, cioè, per ogni $\epsilon > 0$ esiste $\delta > 0$ tale che: $x_1, x_2 \in K$ e $|x_1 - x_2| < \delta \Rightarrow |f(x_1) - f(x_2)| < \epsilon$.

risulta

$$\sum_{i=1}^{n} \alpha_i \Delta_i \leq \int_a^b \tau(x) f(x) dx < \sum_{i=1}^{n} \alpha_i \Delta_i + (b-a)\,\epsilon. \tag{3.10}$$

Considerata infine la partizione dell'evento certo:

$$E_1 = \{x_0 \leq X \leq x_1\}, \;\; E_i = \{x_{i-1} < X \leq x_i\}\,(i = 2, \ldots, n),$$

ricorriamo ancora una volta al teorema di Weierstrass, indicando con $\underline{x}_i, \overline{x}_i$, rispettivamente, un punto di minimo e uno di massimo di τ nell'intervallo J_i $(i = 1, \ldots, n)$. Risulta allora $\sum_{i=1}^{n} \tau(\underline{x}_i)|E_i| \leq \tau(X) \leq \sum_{i=1}^{n} \tau(\overline{x}_i)|E_i|$ e quindi, tramite il Teorema 3.4.2(i), la compatibilità E3 e la monotonia E2, si ha

$$\sum_{i=1}^{n} \tau(\underline{x}_i) \Pr(E_i) = \mathrm{E}(\sum_{i=1}^{n} \tau(\underline{x}_i)|E_i|) \leq \mathrm{E}(\tau(X)) \leq \mathrm{E}(\sum_{i=1}^{n} \tau(\overline{x}_i)|E_i|) = \sum_{i=1}^{n} \tau(\overline{x}_i) \Pr(E_i).$$

Ricordato che la f.r. F_X è continua, dal Teorema 3.2.1(ix) otteniamo $\Pr(E_i) = F_X(x_i) - F_X(x_{i-1})$ $(i = 1, \ldots, n)$ e quindi

$$\sum_{i=1}^{n} \tau(\underline{x}_i)[F_X(x_i) - F_X(x_{i-1})] \leq \mathrm{E}(\tau(X)) \leq \sum_{i=1}^{n} \tau(\overline{x}_i)[F_X(x_i) - F_X(x_{i-1})].$$

Essendo la densità continua in $[a, b]$, la f.r. è, per (3.3), derivabile in ogni punto dell'intervallo e risulta $F'_X = f$. Ne segue, per il teorema di Lagrange[26], $F_X(x_i) - F_X(x_{i-1}) = f(\xi_i)\Delta_i$ con $\xi_i \in]x_{i-1}, x_i[\subset J_i$ $(i = 1, \ldots, n)$. Si ha allora

$$\sum_{i=1}^{n} \tau(\underline{x}_i)[f(\underline{x}_i) + (f(\xi_i) - f(\underline{x}_i))]\Delta_i \leq \mathrm{E}(\tau(X)) \leq \sum_{i=1}^{n} \tau(\overline{x}_i)[f(\overline{x}_i) + (f(\xi_i) - f(\overline{x}_i))]\Delta_i$$

da cui otteniamo

$$\begin{aligned}
\sum_{i=1}^{n} \alpha_i \Delta_i \quad &+ \quad \sum_{i=1}^{n} \tau(\underline{x}_i)(f(\xi_i) - f(\underline{x}_i))\Delta_i \\
&\leq \quad \sum_{i=1}^{n} \tau(\underline{x}_i) f(\underline{x}_i)\Delta_i + \sum_{i=1}^{n} \tau(\underline{x}_i)(f(\xi_i) - f(\underline{x}_i))\Delta_i \\
&= \quad \sum_{i=1}^{n} \tau(\underline{x}_i)[f(\underline{x}_i) + (f(\xi_i) - f(\underline{x}_i))]\Delta_i \leq \mathrm{E}(\tau(X)) \\
\sum_{i=1}^{n} \beta_i \Delta_i \quad &+ \quad \sum_{i=1}^{n} \tau(\overline{x}_i)(f(\xi_i) - f(\overline{x}_i))\Delta_i \\
&\geq \quad \sum_{i=1}^{n} \tau(\overline{x}_i) f(\overline{x}_i)\Delta_i + \sum_{i=1}^{n} \tau(\overline{x}_i)(f(\xi_i) - f(\overline{x}_i))\Delta_i \\
&= \quad \sum_{i=1}^{n} \tau(\overline{x}_i)[f(\overline{x}_i) + (f(\xi_i) - f(\overline{x}_i))]\Delta_i \geq \mathrm{E}(\tau(X)).
\end{aligned}$$

[26] Sia f continua in $[a, b]$ e derivabile in tutti i punti interni. Allora, esiste un punto ξ interno all'intervallo tale che $f(b) - f(a) = f'(\xi)(b - a)$.

Dunque

$$\begin{aligned}\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i &\le \mathrm{E}(\tau(X))\\ &\le \sum_{i=1}^{n}\beta_i\Delta_i+\sum_{i=1}^{n}\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))\Delta_i. \quad (3.11)\end{aligned}$$

Ora, tramite (3.9), si ha

$$\begin{aligned}&\sum_{i=1}^{n}\beta_i\Delta_i+\sum_{i=1}^{n}\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))\Delta_i-\left[\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i\right]\\ &=\sum_{i=1}^{n}(\beta_i-\alpha_i)\Delta_i+\sum_{i=1}^{n}[\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))-\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))]\Delta_i\\ &<(b-a)\,\epsilon+\sum_{i=1}^{n}[\,|\tau(\overline{x}_i)|\,|f(\xi_i)-f(\overline{x}_i)|+|\tau(\underline{x}_i)|\,|f(\xi_i)-f(\underline{x}_i)|\,]\,\Delta_i\\ &<(b-a)\,\epsilon+2(\max|\tau|)\,\epsilon\sum_{i=1}^{n}\Delta_i=(1+2\max|\tau|)(b-a)\,\epsilon\end{aligned}$$

e quindi

$$\begin{aligned}&\sum_{i=1}^{n}\beta_i\Delta_i+\sum_{i=1}^{n}\tau(\overline{x}_i)(f(\xi_i)-f(\overline{x}_i))\Delta_i\\ &<\left[\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i\right]+(1+2\max|\tau|)(b-a)\,\epsilon.\end{aligned}$$

Allora, per (3.11),

$$\begin{aligned}&\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i\le\mathrm{E}(\tau(X))\\ &<\sum_{i=1}^{n}\alpha_i\Delta_i+\sum_{i=1}^{n}\tau(\underline{x}_i)(f(\xi_i)-f(\underline{x}_i))\Delta_i+(1+2\max|\tau|)(b-a)\,\epsilon\end{aligned}$$

da cui, tramite (3.10), otteniamo

$$\begin{aligned}\left|\mathrm{E}(\tau(X))-\int_a^b\tau(x)f(x)dx\right| &< \sum_{i=1}^{n}|\tau(\underline{x}_i)||f(\xi_i)-f(\underline{x}_i)|\Delta_i+(1+2\max|\tau|)(b-a)\,\epsilon\\ &\le \max|\tau|\,\epsilon\,(b-a)+(1+2\max|\tau|)(b-a)\\ &= (1+3\max|\tau|)(b-a).\end{aligned}$$

Ne segue, per l'arbitrarietà di ϵ, la tesi. $\square$

Osservazione 3.4.5. (i) Il Teorema 3.4.3 fornisce un'espressione del valor medio 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* (anche estese) 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 (Sezione 2.6.2), considerare la speranza matematica della v.a. estesa T tempo di attesa del primo successo ottenendo l'uguaglianza $\mathrm{E}(T) = p^{-1}$.

Infatti, 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[27] di due serie geometriche di ragione $q > 0$ e quindi, per il teorema di Mertens[28],

$$\sum_{n\geq 1} nq^{n-1} = [\sum_{m\geq 0} q^m]\,[\sum_{m'\geq 1} q^{m'}] = \left(\frac{1}{p}\right)^2.$$

Conseguentemente, per (2.14) e la regola aritmetica $\pm\infty \cdot 0 = 0$,

$$\mathrm{E}(T) = \sum_{n\geq 1} n \Pr(T = n) = \sum_{n\geq 1} npq^{n-1} = p \sum_{n\geq 1} nq^{n-1} = \frac{1}{p}.$$

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 in quanto assicura che si devono effettuare *in media* m estrazioni per vedere estratta per la prima volta una pallina bianca; se poi si continuano le estrazioni, ci si attenderà, essendo il processo privo di memoria, di estrarre *in media* una pallina bianca ogni qualvolta si fanno m estrazioni.

[27]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 $c_m = a_0 b_m + a_1 b_{m-1} + \cdots + a_{m-1} b_1 + a_m b_0$.

[28]Se la serie $\sum_{m\geq 0} a_m$ converge assolutamente ad a e 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.

Facciamo ora vedere che, nel caso delle estrazioni con contagio simmetrico (Sezione 2.6.3), non sempre è possibile considerare il tempo medio di attesa del primo successo. Infatti, con riferimento alle estrazioni con contagio unitario ($a = 1$) da un'urna di due palline, tramite (2.17), otteniamo

$$p_n = \Pr(T = n) = \frac{\binom{-1}{1}\binom{-1}{n-1}}{\binom{-2}{n}\binom{n}{1}} = \frac{1}{n(n+1)}$$

per ogni $n \geq 1$. Osservato che

$$\sum_{n\geq 1} p_n = \sum_{n\geq 1}(\frac{1}{n} - \frac{1}{n+1}) = \lim_{n\to+\infty}\sum_{i=1}^{n}(\frac{1}{n} - \frac{1}{n+1}) = \lim_{n\to+\infty}(1 - \frac{1}{n+1}) = 1,$$

possiamo considerare, tramite il Teorema 2.4.1, una probabilità numerabilmente additiva sulla σ-algebra $\mathcal{E}_{\mathrm{L}}(\mathcal{T})$ tramite la distribuzione:

$$\Pr(T = t) = \begin{cases} \frac{1}{t(t+1)}, & \text{se } t = 1, 2, \dots \\ 0, & \text{se } t = +\infty \end{cases}.$$

Ne segue che, in questo contesto, non si può considerare il tempo medio di attesa del primo successo in quanto

$$\sum_{n\geq 1} n \Pr(T = n) = \sum_{n\geq 1} \frac{1}{n+1} = +\infty,$$

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 continua tale che l'integrale improprio $\int_{-\infty}^{+\infty} |x|f(x)dx$ sia finito, ponendo:

$$\mathrm{E}(U) = \int_{-\infty}^{+\infty} xf(x)dx.^{29}$$

In questo modo è possibile considerare, ad esempio, la speranza matematica di una v.a. U distribuita secondo la normale di densità $f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$.

[29] Osserviamo che la condizione di assoluta integrabilità $\int_{-\infty}^{+\infty} |g(x)|dx < +\infty$ assicura l'esistenza finita dell'integrale $\int_{-\infty}^{+\infty} g(x)dx$.

Infatti, dati a, b $(a < b)$, 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

$$\int_{-\infty}^{+\infty} x\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx = c\,\mu\int_{-\infty}^{+\infty} e^{-t^2}dt = c\,\mu\sqrt{\pi} = \sqrt{2\pi\sigma^2}\mu,$$

ricordato che $\int_{-\infty}^{+\infty} e^{-t^2}dt = \sqrt{\pi}$. Conseguentemente,

$$\mathrm{E}(U) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{+\infty} x\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx^{30}.$$

Concludiamo provando che, come nel caso numerabile, anche in questo contesto ci sono v.a. prive di speranza matematica. Infatti, sia X una v.a. distribuita secondo la distribuzione di Cauchy (Esempio 3.2.5(ii)). Allora, X è assolutamente continua con densità $f(x) = \frac{1}{\pi\,(1+x^2)}$. Conseguentemente,

$$\int_a^b xf(x)dx = \int_a^b \frac{x}{\pi(1+x^2)}\,dx = \left.\frac{\ln(1+x^2)}{2\pi}\right|_a^b = \frac{1}{2\pi}\,\ln\left(\frac{1+b^2}{1+a^2}\right)$$

e quindi non è possibile considerare la speranza matematica di X, in quanto la funzione $\frac{1+b^2}{1+a^2}$ non ammette limite per $(a, b) \to (-\infty, +\infty)$[31]. $\triangle$

3.5 Varianza e covarianza

Data una v.a. limitata X, per il Teorema 3.1.3, $(X - \mathrm{E}(X))^2 \in \mathcal{V}_b$ e quindi possiamo considerare la speranza matematica:

$$\mathrm{Var}(X) = \mathrm{E}\big((X - \mathrm{E}(X))^2\big),$$

[30] Abbiamo tralasciato, per non appesantire l'esposizione, la verifica della finitezza dell'integrale $\int_a^b |x|\exp(-\frac{(x-\mu)^2}{2\sigma^2})dx$.

[31] A tal fine, basta osservare che il limite vale k^2, se $b = -ka$ $(k > 0)$.

chiamata **varianza** (o **momento centrale secondo**) di X[32].

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:*

(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* $\mathrm{Pr}(X = \mathrm{E}(X)) = 1$*;*

(iv) Se $\epsilon > 0$*, allora*

$$\mathrm{Pr}(|X - \mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2} \text{ (disuguaglianza di Bienaymé-Čebičev);}$$

(v) $\mathrm{Var}(|E|) = \mathrm{Pr}(E)\,\mathrm{Pr}(\overline{E})$*;*

(vi) Se X *è una v.a. con rango discreto limitato, allora:*

$$\mathrm{Var}(X) = \sum_{x \in \mathrm{rg}(X)} (x - \mathrm{E}(X))^2 \,\mathrm{Pr}(X = x);$$

(vii) Se X *è una v.a. assolutamente continua con rango incluso in* $[a, b]$ *e densità* f *continua in* $[a, b]$*, allora:*

$$\mathrm{Var}(X) = \int_a^b (x - \mathrm{E}(X))^2 f(x)dx.$$

Dimostrazione. La proposizioni (iii) è conseguenza del Teorema 3.4.2(ix); le proposizioni (vi), (vii) seguono dai teoremi 3.4.2(vi), 3.4.3 e 3.4.4 (ponendo $\tau(x) = (x - \mathrm{E}(X))^2$).

(i) Per il Teorema 3.4.2(i),(v) si ha $\mathrm{Var}(X) = \mathrm{E}\big(X^2 - 2\mathrm{E}(X)X + \mathrm{E}(X)^2\big) = \mathrm{E}(X^2) - 2\mathrm{E}(X)\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

[32]Per indicarla si usa anche la notazione σ^2_X.

(ii) Risulta $\mathrm{Var}(\alpha X+\beta)=\mathrm{E}\big([(\alpha X+\beta)-\mathrm{E}(\alpha X+\beta)]^2\big)=\mathrm{E}\big([\alpha X+\beta-(\alpha\,\mathrm{E}(X)+\beta)]^2\big)=\mathrm{E}\big(\alpha^2\,(X-\mathrm{E}(X))^2\big)=\alpha^2\,\mathrm{Var}(X)$.

(iv) Dalla disuguaglianza di Markov (Teorema 3.4.2(vii)) 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) e la compatibilità 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) Considerata una v.a. X non coincidente con $\mathrm{E}(X)$ e fissato $\epsilon>0$, dalle disuguaglianze di Bienaymé-Čebičev e di Kolmogorov (Teorema 3.4.2(viii)) otteniamo

$$\frac{\mathrm{Var}(X)-\epsilon^2}{\sup[(X-\mathrm{E}(X))^2]}\leq\Pr(|X-\mathrm{E}(X)|>\epsilon)\leq\frac{\mathrm{Var}(X)}{\epsilon^2}$$

e quindi la varianza può essere intesa come un *indice di dispersione*, poiché fornisce una limitazione inferiore (significativa se $\mathrm{Var}(X)>\epsilon^2$) e una limitazione superiore (significativa se $\mathrm{Var}(X)<\epsilon^2$) della probabilità che la v.a. assuma un valore che disti per più di ϵ dalla speranza matematica (che come sappiamo è invece un indice di posizione).

(ii) Detta **deviazione standard** di X la quantità $\mathrm{Ds}(X)=\sqrt{\mathrm{Var}(X)}$[33], poniamo $n\,\mathrm{Ds}(X)$ 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. △

Date infine le v.a. limitate X e Y, risulta $(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\in\mathcal{V}_b$ (Teorema 3.1.3) e quindi possiamo considerare la speranza matematica:

$$\mathrm{Cov}(X,Y)=\mathrm{E}\big((X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\big)$$

[33]Denotata, in accordo con la nota 32 a piè di pagina 103, anche con σ_X.

chiamata **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))[34]; infine, la disuguaglianza di Cauchy-Schwarz fornisce un importante collegamento tra la covarianza di due v.a. e le rispettive varianze.

Teorema 3.5.3. *Sussistono le proposizioni:*

(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) $$\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} \mathrm{Cov}(X_i,X_j);$$

(v) $\mathrm{Cov}(|E|,|F|) = \Pr(E\wedge F) - \Pr(E)\Pr(F)$*;*

(vi) Risulta:

$$\mathrm{Cov}(X,Y)^2 \le \mathrm{Var}(X)\,\mathrm{Var}(Y) \quad \text{(Disuguaglianza di Cauchy-Schwarz).}$$

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. Sia $\mu = \mathrm{E}(X)$, $\nu = \mathrm{E}(Y)$ e $\mu_i = \mathrm{E}(X_i)$ $(i = 1,\dots,n+1)$.

(i) Per il Teorema 3.4.2(i),(v), $\mathrm{E}\big((X-\mu)(Y-\nu)\big) = \mathrm{E}\big(XY - \nu X - \mu Y + \mu\nu\big) = \mathrm{E}(XY) - \nu\mu - \mu\nu + \mu\nu$ e quindi la tesi.

(ii) Da (i) e dal Teorema 3.4.2(i) otteniamo $\mathrm{Cov}(\alpha X,Y) = \mathrm{E}(\alpha XY) - \mathrm{E}(\alpha X)\nu = \alpha[\mathrm{E}(XY) - \mu\nu = \alpha\,\mathrm{Cov}(X,Y)$.

[34]Evidentemente non vale il viceversa, in quanto gli indicatori sono non correlati anche quando gli eventi sono entrambi trascurabili.

(iii) Poiché per $n = 1$ l'uguaglianza è banale, assumiamo che sussista per n e proviamola per $n + 1$. Posto $Z = \sum_{i=1}^{n} X_i$, da (i) risulta

$$\begin{aligned}
\mathrm{Cov}(\sum_{i=1}^{n+1} X_i, Y) &= \mathrm{E}(Y \sum_{i=1}^{n+1} X_i) - \mathrm{E}(\sum_{i=1}^{n+1} X_i) = \mathrm{E}(YZ + YX_{i+1}) - \mathrm{E}(Z + X_{n+1})\nu \\
&= \mathrm{E}(YZ) + \mathrm{E}(YX_{n+1}) - (\mathrm{E}(Z) + \mu_{n+1})\nu \\
&= [\mathrm{E}(YZ) - \mathrm{E}(Z)\nu] + [\mathrm{E}(YX_{n+1}) - \mu_{n+1}\nu\,] \\
&= \mathrm{Cov}(Z, Y) + \mathrm{Cov}(X_{n+1}, Y) \\
&= \sum_{i=1}^{n} \mathrm{Cov}(X_i, Y) + \mathrm{Cov}(X_{n+1}, Y) = \sum_{i=1}^{n+1} \mathrm{Cov}(X_i, Y).
\end{aligned}$$

(iv) Sia intanto $n = 2$. Risulta

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= \mathrm{E}\big([(X_1 + X_2) - \mathrm{E}(X_1 + X_2)]^2\big) = \mathrm{E}\big(\big[\sum_{i=1}^{2}(X_i - \mu_i)\big]^2\big) \\
&= \mathrm{E}\big(\sum_{i=1}^{2}(X_i - \mu_i)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)\big) \\
&= \sum_{i=1}^{2} \mathrm{E}\big((X_i - \mu_i)^2\big) + 2\,\mathrm{E}\big((X_1 - \mu_1)(X_2 - \mu_2)\big) \\
&= \sum_{i=1}^{2} \mathrm{Var}(X_i) + 2\,\mathrm{Cov}(X_1, X_2).
\end{aligned}$$

Assumiamo quindi che l'uguaglianza sussista per $n \geq 2$ e proviamola per $n + 1$. Posto $Z = \sum_{i=1}^{n} X_i$, da (iii) si ha

$$\begin{aligned}
\mathrm{Var}\big(\sum_{i=1}^{n+1} X_i\big) &= \mathrm{Var}\big(Z + X_{n+1}\big) = \mathrm{Var}\big(Z\big) + \mathrm{Var}(X_{n+1}) + 2\,\mathrm{Cov}\big(Z, X_{n+1}\big) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n} \mathrm{Cov}(X_i, X_j) \\
&\qquad + \mathrm{Var}(X_{n+1}) + 2 \sum_{i=1}^{n} \mathrm{Cov}(X_i, X_{n+1})
\end{aligned}$$

e quindi la tesi.

(v) Per (i), il Teorema 1.12.1(i) e la compatibilità 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

$$\mathrm{E}(UV)^2 \leq \mathrm{E}(U^2)\mathrm{E}(V^2). \tag{3.12}$$

A tal fine, dato un numero reale t qualsiasi, consideriamo la v.a. $tU + V$. Dalla linearità (Teorema 3.4.2(i)) otteniamo

$$\mathrm{E}\big((tU+V)^2\big) = \mathrm{E}(t^2U^2+2t\,UV+V^2) = \mathrm{E}(U^2)\,t^2+2\,\mathrm{E}(UV)\,t+\mathrm{E}(V^2). \tag{3.13}$$

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 $\mathrm{E}(UV) = 0$ [35] e quindi (3.12) è verificata. Se invece $\mathrm{E}(U^2) \neq 0$, l'equazione di secondo grado

$$\mathrm{E}(U^2)\,t^2 + 2\,\mathrm{E}(UV)\,t + \mathrm{E}(V^2) = 0 \tag{3.14}$$

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 (3.12).

Assumiamo infine $\mathrm{E}(UV) \neq 0$ e che (3.12) sia un'uguaglianza. Allora, $\mathrm{E}(U^2) \neq 0$ e quindi l'equazione (3.14) ammette l'unica soluzione:

$$t^* = -\frac{\mathrm{E}(UV)}{\mathrm{E}(U^2)}.$$

Ne segue, per (3.13), $\mathrm{E}\big((t^*U + V)^2\big) = 0$ e quindi, per il Teorema 3.4.2(ix), $1 = \Pr(V = -t^*U) = \Pr(Y - \nu = -t^*(X - \mu)) = \Pr(Y = -t^*X + (t^*\mu + \nu))$, cioè Y è quasi certamente 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 quasi certamente una trasformata affine di Y crescente, se $\mathrm{E}(UV) > 0$, e decrescente, se $\mathrm{E}(UV) < 0$. □

Osservazione 3.5.4. Nel caso particolare che X, Y siano v.a. con varianza positiva, la disuguaglianza di Cauchy-Schwarz può essere riformulata dichiarando che l'**indice di correlazione** (di Bravais):

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Ds}(X)\,\mathrm{Ds}(Y)}$$

[35] 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 $\mathrm{E}(UV) < 0$.

è compreso tra -1 e 1 e che sussiste quasi certamente una relazione affine tra X e Y decrescente, se $\rho_{X,Y} = -1$, e crescente, se $\rho_{X,Y} = 1$.
Il vantaggio dell'indice di correlazione rispetto alla covarianza è che risulta, per i teoremi 3.5.1(ii) e 3.5.3(ii), invariante rispetto ai cambiamenti di scala, nel senso che $\rho_{aX,bY} = \rho_{X,Y}$ per ogni $a, b > 0$. $\triangle$

Osservazione 3.5.5. (i) Con riferimento ai modelli di estrazione considerati nella Sezione 2.6, determiniamo l'espressione della speranza matematica e della varianza di S_n. Per quanto riguarda il valor medio, dalla linearità (Teorema 3.4.2(i)) e compatibilità E3, si ha

$$\mathrm{E}(S_n) = \sum_{i=1}^{n} \mathrm{E}(|E_i|) = \sum_{i=1}^{n} \Pr(E_i) = n\Pr(E_1) = np.$$

Passando alla varianza, per i teoremi 3.5.3(iv),(v) e 3.5.1(v), risulta

$$\begin{aligned}
\mathrm{Var}(S_n) &= \sum_{i=1}^{n} \mathrm{Var}(|E_i|) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} \mathrm{Cov}(|E_i|, |E_j|) \\
&= \sum_{i=1}^{n} \Pr(E_i)\Pr(\overline{E}_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} [\Pr(E_i \wedge E_j) - \Pr(E_i)\Pr(E_j)] \\
&= npq + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n} [\Pr(E_i \wedge E_j) - p^2] \\
&= npq + 2\binom{n}{2}[\Pr(E_1 \wedge E_2) - p^2]
\end{aligned}$$

e quindi

$$\mathrm{Var}(S_n) = npq + n(n-1)[\Pr(E_1 \wedge E_2) - p^2].$$

- Estrazioni senza rimessa Da (2.11) (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 (2.13) risulta $\Pr(E_1 \wedge E_2) - p^2 = 0$ e quindi

$$\mathrm{Var}(S_n) = npq.$$

• Estrazioni con contagio simmetrico Da (2.17) 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

$$\mathrm{Var}(S_n) = npq + n(n-1)\left[a\,\frac{pq}{s+a}\right] = npq\left[1 + \frac{(n-1)a}{s+a}\right] = npq\,\frac{s+an}{s+a}.$$

(ii) Sia X distribuita uniformemente nell'intervallo $[a, b]$ (Osservazione 3.2.2(iii)). Allora, X è assolutamente continua con densità $f(x) = (b-a)^{-1}$, se $a \le x \le b$, 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. 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 $\mathrm{Var}(X) = \frac{35}{12} = 29.1\overline{6}$. △

Osservazione 3.5.6. La disuguaglianza di Markov (Teorema 3.4.2(vii)) fornisce, supposto $X \geq 0$ e $a > 0$, una limitazione superiore della probabilità dell'evento $\{X \geq a\}$ tramite la speranza matematica. Un'ulteriore limitazione superiore si può ottenere ricorrendo alla varianza, se a è diverso dalla speranza matematica di X. Infatti, posto $\alpha = \mathrm{E}(X) \neq a$, risulta $\{X \geq a\} = \{X - \alpha \geq a - \alpha\} \to \{|X - \alpha| \geq a - \alpha\}$ e quindi, per la monotonia, $\Pr(X \geq a) \leq \Pr(|X - \alpha| \geq a - \alpha)$. Ne segue, per la disuguaglianza di Bienaymé-Čebičev(Teorema 3.5.1(iv)),

$$\Pr(X \geq a) \leq \frac{\mathrm{Var}(X)}{(a - \alpha)^2}. \tag{3.15}$$

Per illustrare la portata di questa disuguaglianza, consideriamo il problema seguente. Con riferimento ai 600 lanci di un dado equilibrato, fornire una limitazione significativa della probabilità che almeno per 200 volte appaia il numero cinque. A tal fine, introduciamo gli eventi E_i : "esce 5 al lancio i-simo" ($1 \leq i \leq 600$) e la frequenza di successo $S_{600} = \sum_{i=1}^{600} |E_i|$. Risulta allora $\Pr(E_i) = \frac{1}{6}$ per ogni i e, per i teoremi 3.5.3(iv),(v) e 3.5.1(v),

$$\begin{aligned} \mathrm{Var}(S_{600}) &= \sum_{i=1}^{600} \mathrm{Var}(|E_i|) + 2 \sum_{i=1}^{599} \sum_{j>i}^{600} \mathrm{Cov}(|E_i|, |E_j|) \\ &= \frac{250}{3} + 2 \sum_{i=1}^{599} \sum_{j>i}^{600} [\Pr(E_i \wedge E_j) - \Pr(E_i) \Pr(E_j)] = \frac{250}{3} . \end{aligned}$$

Notato che, per la linearità (Teorema 3.4.2(i)), $\mathrm{E}(S_{600}) = 100$, da (3.15) otteniamo la maggiorazione $\Pr(S_{600} \geq 200) \leq \frac{1}{120}$, che è molto più accurata della $\Pr(S_{600} \geq 200) \leq \frac{1}{2}$ che si ricava mediante la disuguaglianza di Markov. △

Esempio 3.5.7. Con riferimento ai lanci successivi di un dado equilibrato, consideriamo il gioco che consiste nel ricevere, in ogni lancio, un importo uguale al numero uscito, se è pari, e di perderlo, se è dispari. È allora interessante individuare un numero di lanci che assicuri una probabilità di almeno 0.98 che la vincita cumulata sia positiva.

A tal fine, introduciamo gli eventi $E_h^{(n)}$: "esce il numero h nel lancio n-simo" ($i = 1, \ldots, 6$) e la v:a.:

$$X_n = \sum_{h=1}^{6} (-1)^h \, h \, |E_h^{(n)}|$$

che esprime il risultato del gioco al lancio n-simo. Dalla linearità (Teorema 3.4.2(i)) e compatibilità E3 si ha $\mathrm{E}(X_n) = \sum_{h=1}^{6}(-1)^h\, h\, \Pr(E_h^{(n)}) = \frac{1}{6}\sum_{h=1}^{6}(-1)^h = \frac{1}{2}$; inoltre, per i teoremi 3.5.3(iv),(ii),(v) e 3.5.1(v), riesce

$$\begin{aligned}
\mathrm{Var}(X_n) &= \sum_{h=1}^{6}\mathrm{Var}(|E_h^{(n)}|) + 2\sum_{h=1}^{5}\sum_{k>h}^{6}\mathrm{Cov}((-1)^h\,|E_h^{(n)}|, (-1)^k\,|E_k^{(n)}|) \\
&= \frac{5}{6} + 2\sum_{h=1}^{5}\sum_{k>h}^{6}(-1)^{h+k}\,\mathrm{Cov}(|E_h^{(n)}|, |E_k^{(n)}|) \\
&= \frac{5}{6} + \sum_{h=1}^{5}\sum_{k>h}^{6}(-1)^{h+k}\,[\Pr(E_h^{(n)} \wedge E_k^{(n)}) - \Pr(E_h^{(n)})\Pr(E_k^{(n)})] = \frac{5}{6}.
\end{aligned}$$

Considerate le v.a. X_i, X_j $(i \neq j)$, dal Teorema 3.5.3(iii), otteniamo

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}\left(\sum_{h=1}^{6}(-1)^h\, h\,|E_h^{(i)}|, \sum_{k=1}^{6}(-1)^k\, k\,|E_k^{(j)}|\right) \\
&= \sum_{h,k\in\{1,\dots,6\}} (-1)^{h+k}\, hk\,\mathrm{Cov}(|E_h^{(i)}|, |E_k^{(j)}|) \\
&= \sum_{h,k\in\{1,\dots,6\}} (-1)^{h+k}\, hk\,[\Pr(E_h^{(i)} \wedge E_k^{(j)}) - \Pr(E_h^{(i)})\Pr(E_k^{(j)})] = 0
\end{aligned}$$

e quindi le due v.a. sono non correlate. Ne segue, per il Teorema 3.5.3(iv), $\mathrm{Var}(S_n) = \sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{5}{6}n$.

Siamo ora in grado di risolvere il problema in esame, cioè di individuare un n tale che $\Pr(S_n > 0) \geq 0.98$. Notato che $\left\{\left|\frac{S_n}{n} - \frac{1}{2}\right| < \frac{1}{2}\right\} = \left\{0 < \frac{S_n}{n} < 1\right\} \to \left\{\frac{S_n}{n} > 0\right\} = \{S_n > 0\}$, dalla disuguaglianza di Bienaymé-Čebičev (Teorema 3.5.1 (iv)), otteniamo

$$\begin{aligned}
\Pr(S_n > 0) &\geq \Pr\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| < \frac{1}{2}\right) = 1 - \Pr\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \frac{1}{2}\right) \\
&\geq 1 - \frac{4}{n^2}\mathrm{Var}(S_n) = 1 - \frac{10}{3n}.
\end{aligned}$$

A questo punto, per individuare n, basta risolvere la disuguaglianza $1 - \frac{10}{3n} \geq 0.98$ ottenendo $n \geq \frac{500}{3}$. È sufficiente quindi fare 17 lanci. $\diamondsuit$

3.6 Il Teorema di Bernoulli

A partire dall'opera di Gerolamo Cardano *De ludo aleæ* (1526), si riteneva che ci fosse uno stretto legame tra la probabilità e la 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 un gran numero di volte nelle stesse condizioni, ciascuno degli eventi possibili si manifesta con una frequenza (relativa) che è presso a poco uguale alla sua probabilità. L'approssimazione cresce ordinariamente col crescere del numero delle prove [36]

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[37].

Teorema 3.6.1 (di Bernoulli)**.** *La successione $(E_n)_{n\geq 1}$ sia costituita da eventi di medesima probabilità $p \in]0,1[$. Allora, posto:*

$$S_n = \sum_{i=1}^{n} |E_i| \quad (n \geq 1)^{38}$$

e dato un numero reale $\epsilon > 0$ arbitrario, sussistono le proposizioni:

(i) Se gli eventi della successione sono a due a due non correlati, allora

$$\lim_{n\to+\infty} \Pr\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) = 1;$$

[36] Guido Castelnuovo, *Calcolo delle Probabilità* (1919).

[37] 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."

[38] Numero aleatorio che rappresenta, analogamente al caso delle estrazioni da un'urna, la **frequenza** (**di successo**) (numero di eventi che si verificano tra gli n considerati). In questo ordine di idee, $\frac{S_n}{n}$ rappresenta la **frequenza relativa** (**di successo**) (proporzione di successi sul totale n di eventi considerati).

(ii) Se la probabilità si fattorizza, per ogni $n \geq 2$, sulla partizione generata $\mathcal{P}_{\mathrm{G}}(E_1, \dots .E_n)$, allora

$$\lim_{n \to +\infty} \Pr\left(|\, S_n - np\,| \leq \epsilon\right) = 0.$$[39]

Dimostrazione. (i) Per l'ipotesi di non correlazione degli eventi E_i, E_j $(i \neq j)$, dal Teorema 2.5.1 otteniamo $\Pr(E_i \wedge E_j) = p^2$ e quindi, per i teoremi 3.5.3(iv),(v) e 3.5.1(v), $\mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{Var}(|E_i|) = npq$; inoltre, per la linearità (Teorema 3.4.2(i)), si ha $\mathrm{E}(\frac{S_n}{n}) = p$. Ne segue, tramite la disuguaglianza di Bienaymé-Čebičev (Teorema 3.5.1(iv)) e il Teorema 3.5.1(ii),

$$\Pr\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}(\frac{S_n}{n})}{\epsilon^2} = \frac{\frac{1}{n^2}\, npq}{\epsilon^2} = \frac{pq}{n\epsilon^2}.$$

e quindi

$$\Pr\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) \geq 1 - \frac{pq}{n\epsilon^2}. \tag{3.16}$$

Passando infine al limite per $n \to +\infty$, otteniamo la tesi.

(ii) Dato $n \geq 2$, per l'ipotesi di fattorizzabilità, risulta

$$\Pr(S_n = h) \geq Pr(S_n = h-1) \Leftrightarrow \binom{n}{h} p^h q^{n-h} \geq \binom{n}{h-1} p^{h-1} q^{n-h+1} \Leftrightarrow h \leq (n+1)p.$$

Poiché il valore massimo della probabilità $p_n(\cdot) = \Pr(S_n = \cdot)$ si ottiene in corrispondenza ad un valore h_n^* tale che $p_n(h_n^* - 1) \leq p_n(h_n^*)$ e $p_n(h_n^* + 1) \leq p_n(h_n^*) = p_n((h_n^* + 1) - 1)$, ne segue, per quanto provato, $np - q \leq h_n^* \leq (n+1)p$. Posto allora $d_n = h_n^* - np$ si ha $-q \leq d_n \leq p$ e quindi $|d_n| \leq 1$; riesce inoltre

$$\Pr(S_n = h_n^*) = \binom{n}{np + d_n} p^{np+d_n} q^{nq-d_n}. \tag{3.17}$$

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 l'identità $\binom{n}{h} = \frac{n!}{h!\,(n-h)!}$, otteniamo, posto

$$a_n = \sqrt{\frac{n}{2\pi(np + d_n)(nq - d_n)}}, \qquad b_n = \frac{1 + \frac{\theta_n}{12n}}{(1 + \frac{\theta_{np+d_n}}{12(np+d_n)})(1 + \frac{\theta_{nq-d_n}}{12(nq-d_n)})},$$

---

[39] Quindi, da un punto di vista interpretativo, con alta probabilità la frequenza relativa sarà asintoticamente molto vicina alla probabilità comune p, mentre la frequenza sarà molto lontana dalla sua media np.

l'uguaglianza

$$\binom{n}{np+d_n} = a_n \left(\frac{n}{np+d_n}\right)^{np+d_n} \left(\frac{n}{nq-d_n}\right)^{nq-d_n} b_n.$$

Riesce dunque, tramite (3.17),

$$\Pr(S_n = h_n^*) = a_n \left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} b_n. \tag{3.18}$$

Notato che $e^x \geq 1 + x$ per ogni x reale, risulta

$$\begin{aligned}\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} &= \left(1 - \frac{d_n}{np+d_n}\right)^{np+d_n} \left(1 + \frac{d_n}{nq-d_n}\right)^{nq-d_n} \\ &\leq e^{-d_n} e^{d_n} = 1\end{aligned}$$

$$\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n} = \left(1 + \frac{d_n}{np}\right)^{-(np+d_n)} \left(1 - \frac{d_n}{nq}\right)^{-(nq-d_n)}$$

$$\geq \exp\left(-\frac{d_n}{np}(np+d_n)\right) \exp\left(\frac{d_n}{nq}(nq-d_n)\right) = \exp\left(-\frac{d_n^2}{npq}\right)$$

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 presente che $|d_n| \leq 1$,

$$\lim_{n\to+\infty} \left[\left(\frac{np}{np+d_n}\right)^{np+d_n} \left(\frac{nq}{nq-d_n}\right)^{nq-d_n}\right] = 1.$$

Ricorrendo ancora a $|d_n| \leq 1$, otteniamo che $b_n \to 1$, per $n \to +\infty$, e

$$\lim_{n\to+\infty} \sqrt{2\pi npq}\; a_n = \lim_{n\to+\infty} \sqrt{\frac{pq}{(p + \frac{d_n}{n})(q - \frac{d_n}{n})}} = 1.$$

Ne segue, tramite (3.18),

$$\lim_{n\to+\infty} \sqrt{2\pi npq}\,\Pr(S_n = h_n^*) = 1. \tag{3.19}$$

Siamo ora finalmente in grado di provare la proposizione (ii). 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, si ha

$$\Pr(|S_n - np| \leq \epsilon) = \sum_{h\in[np-\epsilon,np+\epsilon]} \Pr(S_n = h) \leq (2\epsilon + 1)\Pr(S_n = h_n^*)$$

e quindi, per (3.19),

$$\Pr(|S_n - np| \leq \epsilon) \leq \frac{(2\epsilon + 1)}{\sqrt{2\pi npq}} \left(\sqrt{2\pi npq}\ \Pr(S_n = h_n^*) \right) \to 0, \text{ per } n \to +\infty.$$

La dimostrazione è così conclusa. □

Osservazione 3.6.2. (i) È di fondamentale importanza tenere presente che la prima parte del 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 non correlazione siano ragionevoli.[40]

(ii) La disuguaglianza (3.16) assicura che la prima parte del teorema di Bernoulli sussiste anche nel caso di probabilità p estrema. △

La prima parte del Teorema di Bernoulli ha dato inizio a tutta una serie di indagini volte a trovare delle condizioni sufficienti per la validità della cosiddetta **legge debole dei grandi numeri** riguardante successioni $(X_n)_{n\geq 1}$ di variabili aleatorie che, nel caso di v.a. limitate[41], afferma:

$$\lim_{n\to+\infty} \Pr\left(\left| \frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n \mathrm{E}(X_i)}{n} \right| < \epsilon \right) = 1.$$

qualunque sia $\epsilon > 0$[42].

In questo ordine di idee, di particolare interesse per le applicazioni, è il risultato seguente che fornisce una condizione sufficiente per la validità della legge debole dei grandi numeri nel caso di successioni di v.a. limitate non correlate positivamente.

[40] A tale proposito, Corrado Gini in *Rileggendo Bernoulli* (1946) scrive: "Quell'indebito passaggio dalla probabilità diretta alla probabilità inversa*[omissis]* costituisce il peccato originale del calcolo delle probabilità".

[41] Come supposto nelle sezioni 3.4 e 3.5.

[42] Cioè, da un punto di vista interpretativo, che con alta probabilità la media aritmetica delle prime v.a. è asintoticamente molto vicina alla media aritmetica delle relative speranze matematiche.

Teorema 3.6.3. *Sia* $(X_n)_{n\geq 1}$ *una successione di v.a. limitate tale che*

$$\lim_{n\to+\infty} \frac{\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)}{n^2} = 0^{43}$$

e $\mathrm{Cov}(X_i,X_j)\leq 0$ $(i\neq j)$. *Allora, la successione verifica la legge debole dei grandi numeri.*

Dimostrazione. Posto $S_n = \sum_{i=1}^n X_i$, dalla disuguaglianza di Bienaymé-Čebičev (Teorema 3.5.1(iv)), dalla linearità (Teorema 3.4.2(i)) e dai teoremi 3.5.1(ii), 3.5.3(iv), otteniamo

$$\begin{aligned}
\Pr\left(\left|\frac{S_n}{n} - \frac{\sum_{i=1}^n \mathrm{E}(X_i)}{n}\right| \geq \epsilon\right) &= \Pr\left(\left|\frac{S_n}{n} - \mathrm{E}\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right) \leq \frac{\mathrm{Var}(\frac{S_n}{n})}{\epsilon^2} \\
&= \frac{\displaystyle\sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j>i}^{n}\mathrm{Cov}(X_i,X_j)}{\epsilon^2 n^2} \\
&\leq \frac{\sum_{i=1}^n \mathrm{Var}(X_i)}{\epsilon^2 n^2},
\end{aligned}$$

osservato che gli eventi non sono correlati positivamente. Ne segue la tesi. □

Esempio 3.6.4. Con riferimento all'Esempio 3.5.7, risulta

$$\lim_{n\to+\infty} \frac{\mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)}{n^2} = \lim_{n\to+\infty}\frac{5}{6n} = 0.$$

Siamo dunque nelle ipotesi del teorema precedente e quindi

$$\lim_{n\to+\infty} \Pr\left(\left|\frac{\sum_{i=1}^n X_i}{n} - \frac{1}{2}\right| < \epsilon\right) = 1,$$

qualunque sia $\epsilon > 0$. Pertanto, con alta probabilità la media aritmetica delle vincite (positive o negative) è asintoticamente molto vicina a $\frac{1}{2}$. ♢

3.7 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

[43]Condizione che è verificata se $\frac{\mathrm{Var}(X_n)}{n} \to 0$ al divergere di n.

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:

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

denotiamo, come al solito, con $E'_1 \wedge \cdots \wedge E'_n$ il generico costituente della partizione generata $\mathcal{P}_2 = \mathcal{P}_{\mathrm{G}}(E_1, \ldots, E_n)$.

Al fine d'introdurre sulla famiglia degli eventi logicamente dipendenti dalla partizione:

$$\mathcal{P} = \{H_j \wedge (E'_1 \wedge \cdots \wedge E'_n) \neq \varnothing : H_j \in \mathcal{P}_1 \text{ e } E'_1 \wedge \cdots \wedge E'_n \in \mathcal{P}_2\}$$

una probabilità "naturale", consideriamo su $\mathcal{P}_1$ una distribuzione (di probabilità) $p_0, p_1, \ldots, p_s$ e, tenuto conto di (2.13), su $\mathcal{P}_2$ la distribuzione:

$$p^{(j)}_{\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}}} = \begin{cases} \theta_j^k (1-\theta_j)^{n-k}, & \text{se } H_j \wedge (E'_1 \wedge \cdots \wedge E'_n) \neq \varnothing \\ 0, & \text{altrimenti} \end{cases},$$

per ogni $j = 0, \ldots, s$. 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 $\mathcal{E}_{\mathrm{L}}(\mathcal{P})$ tale che $p_j = \Pr(H_j)$ $(j = 0. \ldots, s)$ e

$$\Pr(\underbrace{E'_1 \wedge \cdots \wedge E'_n}_{k \text{ affermati}} \mid H_j) = \theta_j^k (1-\theta_j)^{n-k} \qquad (\Pr(H_j) > 0).$$

Conseguentemente, gli eventi $E_1, \ldots, E_n$ sono scambiabili con riferimento alla probabilità condizionata $\mathrm{P}(\cdot \mid H_j)$; pertanto, dato un costituente $E'_{i_1} \wedge \cdots \wedge E'_{i_m}$ con k affermazioni e $i_h \neq i_l$ $(h \neq l)$, $m \leq n$, dal Teorema 2.6.1 otteniamo

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

e quindi

$$\Pr(\underbrace{E'_{i_1} \wedge \cdots \wedge E'_{i_m}}_{k \text{ affermati}} \mid H_j) = \theta_j^k (1-\theta_j)^{m-k} \qquad (\Pr(H_j) > 0) \qquad (3.20)$$

qualunque sia $j = 1, \dots, s$[44].

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 $(j = 0, \dots, s)$ valori possibili, che rappresenta, da un punto di vista interpretativo, la sconosciuta "composizione dell'urna". Osservato che $\{Z = \theta_j\} = H_j \in \mathcal{E}_{\mathrm{L}}(\mathcal{P})$, risulta $\Pr(Z = \theta_j) = p_j$ $(j = 0, \dots, s)$. Dalla formula di disintegrazione e (3.20) si ha allora

$$\Pr(E_i) = \sum_{j:p_j>0} \Pr(E_i \mid H_j)\, p_j = \sum_{j:p_j>0} \theta_j\, p_j = \sum_{j=0}^{s} \theta_j \Pr(Z = \theta_j)$$

e quindi, per il Teorema 3.4.2(vi),

$$\Pr(E_i) = \mathrm{E}(Z) \qquad (i = 1, \dots, n). \tag{3.21}$$

Dati ora gli eventi E_h, E_l $(h \neq l)$, da (3.20) e dal Teorema 3.4.2(vi) (con $\tau(x) = x^2$), otteniamo

$$\begin{aligned}\Pr(E_h \wedge E_l) &= \sum_{j:p_j>0} \Pr(E_h \wedge E_l \mid H_j) p_j = \sum_{j:p_j>0} \theta_j^2\, p_j = \sum_{j=0}^{s} \theta_j^2 \Pr(Z = \theta_j) \\ &= \mathrm{E}(Z^2)\end{aligned}$$

e quindi, per (3.21) e il Teorema 3.5.1(i),

$$\Pr(E_h \wedge E_l) - \Pr(E_h)\Pr(E_l) = \mathrm{E}(Z^2) - \mathrm{E}(Z)^2 = \mathrm{Var}(Z).$$

Conseguentemente:
− se $\mathrm{Var}(Z) > 0$, si ha $\Pr(E_h \wedge E_l) > 0$ da cui otteniamo $\Pr(E_h) > 0$, $\Pr(E_l) > 0$, $\Pr(E_h \mid E_l) = \frac{\Pr(E_h \wedge E_l)}{\Pr(E_l)} > \Pr(E_h)$ e quindi gli eventi sono *correlati positivamente*;

[44] Per dimostrarla, procediamo per induzione a ritroso su m, osservato che, posto $\mathrm{P}(\cdot) = \Pr(\cdot \mid H_j)$, $\theta = \theta_j$ e considerato un costituente $E_1' \wedge \cdots \wedge E_m'$ con k affermazioni, l'uguaglianza

$$\mathrm{P}(E_1' \wedge \cdots \wedge E_m') = \theta^k (1-\theta)^{m-k}$$

sussiste per $m = n$. Assumiamo quindi che sussista per $m \leq n$ e proviamola per $m-1$. Sia dunque $E = E_1' \wedge \cdots \wedge E_{m-1}'$ con k affermati. Allora, $E = (E \wedge E_m) \vee (E \wedge \overline{E}_m)$, con $E \wedge E_m$ e $E \wedge \overline{E}_m$ aventi, rispettivamente, $k+1$ e k affermazioni. Ne segue $\mathrm{P}(E) = \mathrm{P}(E \wedge E_m) + \mathrm{P}(E \wedge \overline{E}_m) = \theta^{k+1}(1-\theta)^{m-k-1} + \theta^k (1-\theta)^{m-k} = \theta^k (1-\theta)^{m-k-1}[\theta + (1-\theta)] = \theta^k (1-\theta)^{(m-1)-k}$.

– se $\mathrm{Var}(Z) = 0$ e se $\mathrm{E}(Z) > 0$, allora, per (3.21), $\Pr(E_h) > 0$, $\Pr(E_l) > 0$, $\Pr(E_h \mid E_l) = \Pr(E_h)$ e quindi gli eventi sono *non correlati*.[45]

Proviamo infine che, con alta probabilità, la frequenza relativa di successo tende, al procedere delle estrazioni, a coincidere con la composizione incognita dell'urna, cioè:

$$\lim_{n \to +\infty} \Pr\left(\left| \frac{S_n}{n} - Z \right| < \epsilon \right) = 1. \tag{3.22}$$

Dalla formula di disintegrazione e (1.2) otteniamo

$$\begin{aligned} \Pr\left(|\frac{S_n}{n} - Z \,| < \epsilon \right) &= \sum_{j:p_j>0} \Pr\left(\{|\frac{S_n}{n} - Z \,| < \epsilon\} |\{Z = \theta_j\} \right) p_j \\ &= \sum_{j:p_j>0} \Pr\left(\{|\frac{S_n}{n} - \theta_j \,| < \epsilon\} |\{Z = \theta_j\} \right) p_j. \end{aligned} \tag{3.23}$$

Ora, dato $p_j > 0$, si ha, per (3.20), $\Pr(E_i \mid \{Z = \theta_j\}) = \theta_j > 0$ $(j = 1, \ldots, n)$ e quindi, ancora per (3.20), $E_1, \ldots, E_n$ sono, con riferimento alla probabilità condizionata $\Pr(\cdot \mid \{Z = \theta_j\})$, a due e due non correlati. Allora, per la prima parte del Teorema di Bernoulli (Teorema 3.6.1(i)) e l'Osservazione 3.6.2(ii), risulta

$$\lim_{n \to +\infty} \Pr\left(\{|\frac{S_n}{n} - \theta_j \,| < \epsilon\} |\{Z = \theta_j\} \right) = 1.$$

Ne segue (3.22), passando al limite per $n \to +\infty$ nella (3.23).

Osservazione 3.7.1. Supponiamo di effettuare le n estrazioni realizzando la storia non trascurabile $E'_1 \wedge \cdots \wedge E'_n$ avente k affermazioni. Assumiamo inoltre $\Pr(H_j) > 0$ $(j = 0, \ldots, s)$. Viene allora naturale chiedersi quale sia la composizione dell'urna più probabile vista tale evidenza, cioè per quale evento H_j è massima la probabilità condizionata

$$\begin{aligned} \Pr(H_j \mid E'_1 \wedge \cdots \wedge E'_n) &= \frac{\Pr(E'_1 \wedge \cdots \wedge E'_n \mid H_j)}{\Pr(E'_1 \wedge \cdots \wedge E'_n)} \Pr(H_j) \\ &= \frac{\theta_j^k (1-\theta_j)^{n-k}}{\Pr(E'_1 \wedge \cdots \wedge E'_n)} \, p_j, \end{aligned}$$

[45]Si ottiene così un'ulteriore conferma che la probabilità dipende *essenzialmente* dallo stato di conoscenza; infatti, a fronte della medesima situazione "fisica" (estrazioni con rimessa da un'urna, e quindi composizione che rimane invariata), la non correlazione sussiste, nel caso della conoscenza quasi certa della composizione ($\mathrm{Var}(Z) = 0$; Teorema 3.5.1(iii)), mentre diviene positiva in caso contrario (l'uscita di bianca (rossa) fa aumentare la probabilità di vedere ancora bianca (rossa)).

ove le due uguaglianze sono ottenute, nell'ordine, ricorrendo alle formule di Bayes e (3.20). 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 problema di massimo:

$$\max_{j=0,\dots,s} \theta_j^k (1-\theta_j)^{n-k}.$$

Ora, la funzione $g(x) = x^k(1-x)^{n-k}$ assume valore massimo su $[0,1]$ 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 crescente prima di x_0 e 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. △

3.8 Esercizi

1. Sia X distribuito in $[0,\pi]$ con densità proporzionale alla funzione $g(x) = a \sin x$. Determinare: $*$ la f.r.; $*$ la f.r. di X^3+1.

2. Dato X distribuito uniformemente in $[-2,1]$, determinare: $*$ la f.r. di $|X|$; $*$ la f.r. di $(X-1)^2$; $*$ una limitazione (inferiore o superiore) significativa di $\Pr(|X-\frac{3}{2}| \geq \frac{1}{5})$.

3. Dato X distribuito in $[0,3]$ con densità proporzionale alla funzione $g(x) = x^2$, se $0 \leq x \leq 1$, e $g(x) = 2x-1$, se $1 \leq x \leq 3$, determinare: $*$ la f.r.; $*$ la speranza matematica; $*$ la f.r. di $e^{X+1}+1$.

4. Dato X con funzione di ripartizione F continua e crescente, determinare la f.r. di $Y = F(X)$.

5. Dato X con speranza matematica 5 e varianza 10, individuare un intorno di 5 tale che la probabilità che X appartenga a tale intorno sia almeno 0.9.

6. Dato X distribuito in $[0,1]$ con f.r. $F(x) = \frac{2x+1}{4}$ $(0 \leq x < 1)$, calcolare la probabilità dell'evento $\bigvee_{n\geq 1}\{\frac{1}{n+1} < X < \frac{1}{n}\}$.

7. Con riferimento al lancio simultaneo di due dadi equilibrati distinguibili, calcolare la covarianza delle v.a. "minimo dei due numeri che escono nel lancio" e "massimo dei due numeri che escono nel lancio".

8. Sia X distribuito in $\{-2, -1, 3\} \cup [0, 1]$. Sapendo che $\Pr(X = -2) = \frac{1}{12}$, $\Pr(X = -1) = \frac{3}{12}$, $\Pr(X = 3) = \frac{4}{12}$ e che il resto della probabilità è distribuito in modo diffuso con distribuzione uniforme nell'intervallo unitario, determinare la f.r. e disegnarne un grafico[46]. Calcolare inoltre $\Pr(-2 \leq X < -1)$, $\Pr(\frac{1}{2} \leq X < 3)$, $\Pr(X < 2)$, $\Pr(X \geq -1)$ e $\lim\limits_{n \to +\infty} \Pr(-1 - \dfrac{1}{n} \leq X \leq -1 + \dfrac{1}{n})$. Disegnare infine il grafico della funzione $h(x) = \Pr(X = x)$.

9. Siano X, Y tali che $\Pr(X \leq Y) = 1$. Provare che $F_Y \leq F_X$.

10. Individuare un numero di lanci di un dado equilibrato che assicuri una probabilità non inferiore a 0.95 che appaia almeno una volta 6.

11. Dato X con densità $f(x) = \frac{3}{8}$, se $-1 \leq x \leq 1$, e $f(x) = \frac{3}{8}x^{-4}$ altrimenti, determinare $a > 0$ tale che $\Pr(-a \leq X \leq a)$.

12. Con riferimento ai lanci di una moneta equilibrata, sia E_n : " esce testa nel lancio n-simo". Calcolare la varianza: $*$ di $X = |\overline{E}_1| - |E_2| + |E_1 \wedge E_3|$; $*$ di $X = \max(|\overline{E}_1|, |E_3| + |E_1 \wedge E_2|)$.

13. Dato X con media 15, si dica per quali valori della varianza risulta $\Pr(X \leq 12 \vee X \geq 25) \leq \frac{1}{3}$.

14. Con riferimento ai lanci di una dado equilibrato, sia X_n il numero che esce nel lancio n-simo. Determinare un numero di lanci da effettuare affinchè $\Pr(\sum_{i=1}^{n} X_i < n^2) > \frac{1}{2}$.

15. Dato X con $\mathrm{E}(X) = 8$, $\Pr(X < 6) = 0.2$ e $\Pr(X \geq 12) = 0.3$, fornire una limitazione inferiore positiva della sua varianza.

16. Sia X distribuito uniformemente nell'intervallo $[a, b]$ con speranza matematica e varianza unitarie. Determinare a, b e la funzione di ripartizione di X.

17. Data l'urna $U(6, 4)$ si lanciano due dadi equilibrati distinguibili e si procede a estrazioni senza rimessa, se la somma dei numeri usciti è maggiore di 6, e con contagio unitario altrimenti. Determinare: $*$ la f.r., la speranza

[46]Dato X distribuito in $[a, b]$ con masse concentrate nei punti x_i $(i = 1, \ldots, n)$, con la frase "il resto della probabilità è distribuito in modo diffuso con distribuzione $g \geq 0$ in $[a, b]$" intendiamo dire che la funzione g è tale che $\int_a^b g(x)dx = 1 - \sum_{i=1}^{n} \Pr(X = x_i)$ e $F_X(x) = \sum_{i: x_i \leq x} \Pr(X = x_i) + \int_a^x g(t)dt$, per ogni x.reale.

matematica e la varianza del numero di palline bianche presenti nell'urna dopo cinque estrazioni; $*$ la correlazione degli eventi E :“ la somma dei numeri usciti non supera 6” e F :“ nelle prime 3 estrazioni sono uscite 2 bianche”.

18. Sia X distribuito in $\{0, 2, 8, \dots\}$ con $\Pr(X = 0) = \frac{3}{4}$ e $\Pr(X = 2^n) = (\frac{1}{5})^n$ per ogni $n \geq 1$. Determinare il minimo n tale che $\Pr(X \leq 2^n) \geq 0.999$.

19. Supposto che la probabilità di uscita di testa nel lancio di una moneta truccata sia $\frac{1}{3}$, si consideri il gioco nel quale si riceve 2, se esce testa, e si perde 1 altrimenti. Indicato con G_n il guadagno cumulato alla fine del lancio n-simo, dire per quali n risulta $\Pr(G_n < n) \geq 0.9$.

20. Da un'urna formata da una pallina bianca, una nera e una rossa si estrae una pallina riponendo poi nell'urna la pallina estratta con due altre aventi colori differenti e diversi da quella estratta. Si procede poi a estrazioni con rimessa dall'urna così ottenuta. Posto E_n :“ esce pallina bianca o nera nell'estrazione n-sima”, si consideri il guadagno cumulato G_n nelle prime n estrazioni qualora, per ogni $m \geq 1$, si riceva 3, se E_m si verifica, e si perda 1 altrimenti. Calcolare: speranza matematica e varianza di G_n; $*$ f.r. di G_3; $*$ $\Pr(G_3 \geq 0)$, $\Pr(1 \leq G_3 \leq \pi)$ e $\Pr(G_{11} = 0)$.

21. La v.a. X è distribuita sui primi sei numeri naturali secondo la tabella:

i =	1	2	3	4	5	6
$\Pr(X = i)$	0.1	0.2	0.1	0.1	0.2	0.3

Calcolare la speranza matematica a fronte dell'informazione “X è un numero pari”.

22. Supposto $\text{rg}(X) = [-2, 1]$ e $\text{E}(X^4) = \frac{1}{2}$, fornire una limitazione significativa della probabilità $\Pr(|X| > \frac{1}{2})$.

Capitolo 4

Valutazione dell'incertezza III

> Si estendono le nozioni di funzione di ripartizione e di funzione di densità, introdotte nel capitolo precedente per le variabili aleatorie, alle coppie aleatorie.
>
> Tenendo conto della conoscenza limitata che gli studenti del primo anno di corso hanno dell'integrazione di funzioni di due o più variabili reali, la trattazione della densità congiunta è condotta entro i limiti di una esposizione atta a fornire un primo approccio.

In questo capitolo, denotiamo con $\mathcal{A}$ una σ-algebra di eventi, con Pr una probabilità numerabilmente additiva su $\mathcal{A}$ e con X_1, X_2 variabili aleatorie.

4.1 Funzione di ripartizione congiunta

Dati due numeri reali arbitrari x e y, gli eventi$\{X_1 \leq x\}$, $\{X_2 \leq y\}$ appartengono ad $\mathcal{A}$ (in quanto X_1, X_2 sono v.a.) e quindi $\{X_1 \leq x\} \wedge \{X_2 \leq y\} \in \mathcal{A}$. Possiamo dunque considerare la funzione $F_{\underline{X}}$ definita su $\mathbb{R}^2$:

$$F_{\underline{X}}(x, y) = \Pr(\{X_1 \leq x\} \wedge \{X_2 \leq y\}),$$

chiamata **funzione di ripartizione congiunta** (in breve f.r. congiunta) della coppia aleatoria $\underline{X} = (X_1, X_2)$.

Il teorema seguente, analogo al Teorema 3.2.1, assicura che la f.r. congiunta è una funzione continua a destra e non decrescente rispetto a ciascuno dei suoi argomenti; che tende a zero se un qualsiasi argomento tende a $-\infty$ e tende a uno se entrambi gli argomenti tendono a $+\infty$. Inoltre, la proposizione

(i) mostra che $F_{\underline{X}}(x,y)$ è la probabilità che la coppia aleatoria $\underline{X}$ assuma un valore appartenente al terzo quadrante del piano di vertice il punto (x,y):

$$Q_{x,y} =]-\infty, x] \times]-\infty, y] = \{(u,v) \in \mathbb{R}^2 : u \leq x \,\mathrm{e}\, v \leq y\}.$$

Teorema 4.1.1. *Sussistono le proposizioni:*

(i) $F_{\underline{X}}(x,y) = \Pr(\underline{X} \in Q_{x,y})$*;*

(ii) $0 \leq F_{\underline{X}} \leq 1$*;*

(iii) $F_{\underline{X}}(x,y) \leq F_{\underline{X}}(x',y')$*, se* $x \leq x'$ *e* $y \leq y'$*;*

(iv) $\lim_{x\to-\infty} F_{\underline{X}}(x,y) = 0 = \lim_{y\to-\infty} F_{\underline{X}}(x,y)$*;*

(v) $\lim_{(x,y)\to(+\infty,+\infty)} F_{\underline{X}}(x,y) = 1$*;*

(vi) $\lim_{u\to x^+} F_{\underline{X}}(u,y) = F_{\underline{X}}(x,y) = \lim_{v\to y^+} F_{\underline{X}}(x,v)$*;*

(vii) Per ogni a,b,c,d *tali che* $a < b$ *e* $c < d$*, risulta:*

$$\begin{aligned}\Pr(\{a < X_1 \leq b\} \wedge \{c < X_2 \leq d\}) &= F_{\underline{X}}(b,d) + F_{\underline{X}}(a,c) \\ &\quad - F_{\underline{X}}(a,d) - F_{\underline{X}}(b,c).\end{aligned}$$

Dimostrazione. (i) Tramite (1.6) otteniamo

$$F_{\underline{X}}(x,y) = \Pr(\{X_1 \in]-\infty,x]\} \wedge \{X_2 \in]-\infty,y]\}) = \Pr(\underline{X} \in Q_{x,y}).$$

(iii) Segue dalla monotonia della probabilità, notato che $\{\underline{X} \in Q_{x,y}\} \to \{\underline{X} \in Q_{x',y'}\}$ per ogni $x \leq x'$ e $y \leq y'$.

Le proposizioni (iv)÷(vi) seguono da (iii), dal Teorema di rappresentazione 1.8.5 e dalla continuità della probabilità. Con riferimento alla Sezione 1.12.3, denotiamo con f la versione di $\underline{X}$ relativa alla partizione $\mathcal{P}^{(2)}$.

(iv) Considerata la successione non crescente $(\{\underline{X} \in Q_{-n,y}\})_{n\geq1}$, risulta

$$\lim_{n\to+\infty} F_{\underline{X}}(-n,y) = \lim_{n\to+\infty} \Pr(\underline{X} \in Q_{-n,y}) = \Pr(\bigwedge_{n\geq1}\{\underline{X} \in Q_{-n,y}\}) = \Pr(\varnothing) = 0,$$

osservato che $\mathrm{set}(\bigwedge_{n\geq1}\{\underline{X} \in Q_{-n,y}\}) = \bigcap_{n\geq1}\{\omega^{(2)} : f(\omega^{(2)}) \in Q_{-n,y}\} = \emptyset$. In modo analogo si procede per l'altro limite.

(v) Considerata la successione non decrescente $(\{\underline{X} \in Q_{n,n}\})_{n\geq 1}$, risulta

$$\lim_{n\to+\infty} F_{\underline{X}}(n,n) = \lim_{n\to+\infty} \Pr(\underline{X} \in Q_{n,n}) = \Pr(\bigvee_{n\geq 1}\{\underline{X} \in Q_{n,n}\}) = \Pr(\Omega) = 1,$$

osservato che $\mathrm{set}(\bigvee_{n\geq 1}\{\underline{X} \in Q_{n,n}\}) = \bigcup_{n\geq 1}\{\omega^{(2)} : f(\omega^{(2)}) \in Q_{n,n}\} = \mathcal{P}$. Dato $\epsilon > 0$, esiste quindi m tale che $1 - F_{\underline{X}}(m,m) < \epsilon$. Allora, per ogni $x \geq m$ e $y \geq m$, da (iii) otteniamo $1 - F_{\underline{X}}(x,y) \leq 1 - F_{\underline{X}}(m,m) < \epsilon$. Ne segue la tesi.

(vi) Considerata la successione non crescente $(\{X_1 \leq x + \frac{1}{n}\})_{n\geq 1}$, da (1.4) otteniamo $\{X_1 \leq x\} \wedge \{X_2 \leq y\} = (\bigwedge_{n\geq 1}\{X_1 \leq x + \frac{1}{n}\}) \wedge \{X_2 \leq y\} = \bigwedge_{n\geq 1}(\{X_1 \leq x + \frac{1}{n}\} \wedge \{X_2 \leq y\})$ e quindi

$$\begin{aligned}
\lim_{n\to+\infty} F_{\underline{X}}(x + \frac{1}{n}, y) &= \lim_{n\to+\infty} \Pr(\{X_1 \leq x + \frac{1}{n}\} \wedge \{X_2 \leq y\}) \\
&= \Pr(\bigwedge_{n\geq 1}(\{X_1 \leq x + \frac{1}{n}\} \wedge \{X_2 \leq y\}) \\
&= \Pr(\{X_1 \leq x\} \wedge \{X_2 \leq y\} = F_{\underline{X}}(x,y).
\end{aligned}$$

In modo analogo si procede per l'altro limite.

(vii) Tramite (1.6), (1.5) risulta

$$\begin{aligned}
\{X_1 \leq b\} \wedge \{X_2 \leq d\} &= \{\underline{X} \in]-\infty, b] \times]-\infty, d]\} \\
&= \{\underline{X} \in Q_{a,d} \cup Q_{b,c} \cup]a,b] \times]c,d]\} \\
&= \{\underline{X} \in Q_{a,d}\} \vee \{\underline{X} \in Q_{b,c}\} \vee \{\underline{X} \in]a,b] \times]c,d]\} \\
&= \{\underline{X} \in Q_{a,d}\} \vee \{\underline{X} \in Q_{b,c}\} \\
&\qquad \vee (\{a < X_1 \leq b\} \wedge \{c < X_2 \leq d\})
\end{aligned}$$

da cui, per la formula d'inclusione-esclusione, si ha

$$\begin{aligned}
F_{\underline{X}}(b,d) &= \Pr(\underline{X} \in Q_{a,d}) + \Pr(\underline{X} \in Q_{b,c}) \\
&\qquad + \Pr(\{a < X_1 \leq b\} \wedge \{c < X_2 \leq d\}) - \Pr(\underline{X} \in Q_{a,c}),
\end{aligned}$$

osservato che gli eventi $\{\underline{X} \in Q_{a,d}\}$, $\{\underline{X} \in Q_{b,c}\}$ sono incompatibili con l'evento $\{a < X_1 \leq b\} \wedge \{c < X_2 \leq d\}$. Ne segue, tramite (i),

$$F_{\underline{X}}(b,d) = F_{\underline{X}}(a,d) + F_{\underline{X}}(b,c) + \Pr((a < X_1 \leq b) \wedge (c < X_2 \leq d)) - F_{\underline{X}}(a,c).$$

La dimostrazione è così conclusa. $\square$

Il prossimo teorema collega le funzioni di ripartizione delle v.a. X_1 e X_2 (dette **funzioni di ripartizione marginali**) con la funzione di ripartizione congiunta.

Teorema 4.1.2. *Sussistono le proposizioni:*

(i) $F_{X_1}(x) = \lim_{y\to+\infty} F_{\underline{X}}(x,y)$*;*

(ii) $F_{X_2}(y) = \lim_{x\to+\infty} F_{\underline{X}}(x,y)$.

Dimostrazione. Poiché le due proposizioni si provano allo stesso modo, dimostriamo la prima. Osservato che la successione di eventi $(\{\underline{X} \in Q_{x,n}\})_{n\geq 1}$ è non decrescente, dalla continuità dal basso della probabilità, da (1.6) e dal Teorema 4.1.1(i), otteniamo

$$\begin{aligned} F_{X_1}(x) = \Pr(X_1 \leq x) &= \Pr(\{X_1 \leq x\} \wedge \Omega) = \Pr(\{X_1 \leq x\} \wedge \bigvee_{n\geq 1}\{X_2 \leq n\}) \\ &= \Pr(\bigvee_{n\geq 1}(\{X_1 \leq x\} \wedge \{X_2 \leq n\})) \\ &= \Pr(\bigvee_{n\geq 1}\{\underline{X} \in Q_{x,n}\}) = \lim_{n\to+\infty}\Pr(\underline{X} \in Q_{x,n}) \\ &= \lim_{n\to+\infty} F_{\underline{X}}(x,n) \end{aligned}$$

e quindi, per il Teorema 4.1.1(iii), $F_{X_1}(x) = \lim_{y\to+\infty} F_{\underline{X}}(x,y)$. □

Osservazione 4.1.3. Il teorema appena provato assicura che la *conoscenza congiunta* implica la *conoscenza marginale*, nel senso che la f.r. congiunta permette di conoscere le f.r. marginali. Facciamo ora vedere che non vale il viceversa, cioè che le f.r. marginali non individuano la f.r. congiunta. A tal fine, consideriamo il lancio simultaneo di due dadi equilibrati distinguibili A, B e indichiamo con X_1, X_2, rispettivamente, il numero che esce con il dado A e quello con il dado B. Ovviamente, $\Pr(X_1 = i \wedge X_2 = j) = \frac{1}{36}$ $(i, j = 1, \ldots, 6)$. Indichiamo infine con $\widetilde{X}_2$ il numero che esce con il dado B "opposto" al numero X_1 che esce con il dado A[1]. Poiché i numeri impressi sui dadi sono messi in modo tale che numeri opposti hanno somma sette, risulta $\Pr(X_1 = i \wedge \widetilde{X}_2 = j) = \frac{1}{6}$, se $i + j = 7$, e $\Pr(X_1 = i \wedge \widetilde{X}_2 = j) = 0$ altrimenti. Conseguentemente, le distribuzioni congiunte $F_{\underline{X}}$, $F_{(X_1,\widetilde{X}_2)}$ sono diverse pur avendo le medesime marginali. △

[1]Quindi il numero che è inciso sulla faccia di A a contatto con il tavolo (sul quale cadono i dadi).

4.2 Variabili aleatorie indipendenti

Prendendo spunto dalla nozione d'indipendenza stocastica data per gli eventi (tralasciando la condizione d'indipendenza logica), verrebbe spontaneo ritenere indipendenti le v.a. X_1, X_2 se ogni evento riguardante determinazioni numeriche dell'una è non correlato con qualunque evento che faccia riferimento alle determinazioni numeriche dell'altra. Quindi, in particolare, dovrebbero essere non correlati gli eventi del tipo $\{X_1 = x\}$ e $\{X_2 = y\}$ qualunque siano x, y reali. Conseguentemente, non si potrebbe, per il Teorema 3.2.1(viii), parlare d'indipendenza nel caso di v.a. continue. Per ovviare a questo inconveniente, osserviamo che, data una qualsiasi v.a. X, gli eventi del tipo $\{X \leq x\}$ non sono, per il Teorema 3.2.1(v), tutti trascurabili e che forniscono i "mattoni" con i quali costruire la f.r. di X. Viene allora naturale richiedere che la non correlazione sussista solo con riferimento a questo tipo particolare di eventi.

Chiamiamo dunque X_1, X_2 **indipendenti** se $\Pr(X_i \leq u \,|\, X_j \leq v) = \Pr(X_i \leq u)$, per ogni u, v reali tali che $\Pr(X_j \leq v) > 0$ $(i \neq j; i, j = 1, 2)$[2].

Una conseguenza notevole dell'indipendenza è che (come ora proveremo) consente di determinare l'espressione della f.r. congiunta tramite le f.r marginali, identificando così i due tipi di conoscenza: congiunta e marginale.

Teorema 4.2.1. *Sono equivalenti le proposizioni:*

(i) Le v.a. X_1, X_2 sono indipendenti;

(ii) $F_{\underline{X}}(x, y) = F_{X_1}(x)F_{X_2}(y)$, per ogni x, y reali.

Dimostrazione. *(i)* $\Rightarrow$ *(ii)*. Sia intanto uno dei fattori $F_{X_1}(x), F_{X_2}(y)$ nullo. Allora, per la monotonia, $F_{\underline{X}}(x, y) = \Pr(\{X_1 \leq x\} \wedge \{X_2 \leq y\}) \leq \min(\Pr(X_1 \leq x), \Pr(X_2 \leq y)) = \min(F_{X_1}(x), F_{X_2}(y))$ e quindi $F_{\underline{X}}(x, y) = 0$. Siano infine non nulli entrambi i fattori. Allora, da (i) otteniamo $F_{\underline{X}}(x, y) = \Pr(\{X_1 \leq x\} \wedge \{X_2 \leq y\}) = \Pr(X_1 \leq x \,|\, X_2 \leq y) \Pr(X_2 \leq y) = \Pr(X_1 \leq x) \Pr(X_2 \leq y) = F_{X_1}(x)F_{X_2}(y)$.

(ii) $\Rightarrow$ *(i)*. Sia $\Pr(X_j \leq v) > 0$ e $i \neq j$. Allora, da (ii) segue

$$\Pr(X_i \leq u \,|\, X_j \leq v) = \frac{\Pr(\{X_i \leq u\} \wedge \{X_j \leq v\})}{\Pr(X_j \leq v)} = \frac{F_{X_i}(u)F_{X_j}(v)}{F_{X_j}(v)} = F_{X_i}(u),$$

qualunque sia u. Pertanto, le due v.a. sono indipendenti. $\square$

[2] Avendo indicato con $\Pr(X_i \leq u \,|\, X_j \leq v)$ la probabilità dell'evento condizionato $\{X_i \leq u\} \,|\, \{X_j \leq v\}$.

Il prossimo risultato assicura, in particolare, che la speranza matematica del prodotto di due v.a. semplici e indipendenti coincide con il prodotto delle loro speranze matematiche.

Teorema 4.2.2. *Siano X_1, X_2 v.a. indipendenti con ranghi, rispettivamente, $\{x_1, \ldots, x_i, \ldots\}$ e $\{y_1, \ldots, y_j, \ldots\}$ finiti o numerabili. Sussistono allora le proposizioni:*

(i) $\Pr(\{X_1 = x_i\} \wedge \{X_2 = y_j\}) = \Pr(X_1 = x_i)\Pr(X_2 = y_j)$ *per ogni x_i, y_j, se i ranghi sono costituiti da punti isolati;*

(ii) $\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2)$*, se le v.a. sono semplici.*

Dimostrazione. (i) Sia $x_h^{(i)} \uparrow x_i$, $y_h^{(j)} \uparrow y_j$ con $x_h^{(i)} \notin \mathrm{rg}(X_1)$ e $y_h^{(j)} \notin \mathrm{rg}(X_2)$ per ogni $h \geq 1$. Allora, per h sufficientemente grande, dai teoremi 4.1.1(vii) e 4.2.1 si ha

$$\begin{aligned}
\Pr(\{X_1 = x_i\} \wedge \{X_2 = y_j\}) &= \Pr(\{x_h^{(i)} < X_1 \leq x_i\} \wedge \{y_h^{(j)} < X_2 \leq y_j\}) \\
&= F_{\underline{X}}(x_i, y_j) + F_{\underline{X}}(x_h^{(i)}, y_h^{(j)}) - F_{\underline{X}}(x_h^{(i)}, y_j) - F_{\underline{X}}(x_i, y_h^{(j)}) \\
&= F_{X_1}(x_i)F_{X_2}(y_j) + F_{X_1}(x_h^{(i)})F_{X_2}(y_h^{(j)}) \\
&\quad - F_{X_1}(x_h^{(i)})F_{X_2}(y_j) - F_{X_1}(x_i)F_{X_2}(y_h^{(j)}) \\
&= [F_{X_1}(x_i) - F_{X_1}(x_h^{(i)})]\,[F_{X_2}(y_j) - F_{X_2}(y_h^{(j)})]
\end{aligned}$$

da cui, passando al limite per $h \to +\infty$, otteniamo

$$\Pr(\{X_1 = x_i\} \wedge \{X_2 = y_j\}) = [F_{X_1}(x_i) - F_{X_1}(x_i^-)]\,[F_{X_2}(y_j) - F_{X_2}(y_j^-)]$$

e quindi, per il Teorema 3.2.1(viii), la tesi.

(ii) Osservato che

$$X_1X_2 = \sum_{(i,j)\in \mathrm{rg}(X_1)\times \mathrm{rg}(X_2)} x_i\, y_j\, |\{X_1 = x_i\} \wedge \{X_2 = y_j\}|,$$

dal Teorema 3.4.2(vi) otteniamo

$$\mathrm{E}(X_1X_2) = \sum_{(i,j)\in \mathrm{rg}(X_1)\times \mathrm{rg}(X_2)} x_i\, y_j\, \Pr(\{X_1 = x_i\} \wedge \{X_2 = y_j\})$$

e quindi, tramite (i),

$$\begin{aligned}
\mathrm{E}(X_1X_2) &= \sum_{(i,j)\in \mathrm{rg}(X_1)\times \mathrm{rg}(X_2)} x_i\, y_j\ \Pr(X_1 = x_i)\Pr(X_2 = y_j) \\
&= \left[\sum_{i\in \mathrm{rg}(X_1)} x_i\ \Pr(X_1 = x_i)\right]\left[\sum_{j\in \mathrm{rg}(X_2)} y_j\ \Pr(X_2 = y_j)\right].
\end{aligned}$$

Ne segue, ricorrendo ancora al Teorema 3.4.2(vi), la tesi. □

Osservazione 4.2.3. La proposizione (ii) del teorema precedente assicura che, nel caso di v.a. semplici, l'indipendenza implica la non correlazione (Teorema 3.5.3(i)). Facciamo ora vedere, con un esempio, che non vale l'implicazione opposta. Considerata la v.a. X_1 distribuita uniformemente in $\{-1, 0, 1\}$, sia $X_2 = X_1^2$. Allora $\mathrm{E}(X_1) = 0 = \mathrm{E}(X_1X_2)$ da cui otteniamo $\mathrm{E}(X_1X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2)$ e quindi la non correlazione. D'altra parte, $\Pr(X_2 \leq 0 \mid X_1 \leq -1) = \Pr(X_2 = 0 \mid X_1 = -1) = 0 < \frac{1}{3} = \Pr(X_2 = 0) = \Pr(X_2 \leq 0)$ e quindi la non validità dell'indipendenza. △

Esempio 4.2.4. (i) Considerate le v.a. indipendenti X_1, X_2 si ha, per il Teorema 4.2.1, $\Pr(\{X_1 \leq t\} \wedge \{X_2 \leq t\}) = F_{\underline{X}}(t.t) = F_{X_1}(t)F_{X_2}(t)$ per ogni t reale. Ciò osservato, determiniamo la f.r. della v.a.:

• $Y = \min(X_1, X_2)$. Riesce $F_Y(t) = \Pr(\{X_1 \leq t\} \vee \{X_2 \leq t\})$ da cui otteniamo $F_Y(t) = \Pr(X_1 \leq t) + \Pr(X_2 \leq t) - \Pr(\{X_1 \leq t\} \wedge \{X_2 \leq t\}) = F_{X_1}(t) + F_{X_2}(t) - F_{X_1}(t)F_{X_2}(t) = F_{X_1}(t)(1 - F_{X_2}(t)) + 1 - (1 - F_{X_2}(t)) = 1 + (F_{X_1}(t) - 1)(1 - F_{X_2}(t))$ e quindi

$$F_Y(t) = 1 - (1 - F_{X_1}(t))(1 - F_{X_2}(t)).$$

• $Y = \max(X_1, X_2)$. Risulta $F_Y(t) = \Pr(\{X_1 \leq t\} \wedge \{X_2 \leq t\})$ e quindi

$$F_Y(t) = F_{X_1}(t)F_{X_2}(t).$$

(ii) Con riferimento ai lanci simultanei ripetuti di una moneta e di un dado equilibrati, determiniamo la probabilità che esca testa prima di un numero minore di 5. Indicati con X_1 il tempo di attesa del primo successo (esce testa) nel lancio della moneta e con X_2 quello del primo successo (esce un numero minore di 5) nel lancio del dado, risulta, tramite (2.14), $\Pr(X_1 = n) = (\frac{1}{2})^n$ e $\Pr(X_2 = m) = \frac{2}{3}\,(\frac{1}{3})^{m-1}(n, m \geq 1)$. Osservato che, per la monotonia e (2.14), $\Pr(\{X_i = +\infty\} \wedge \{X_j = t\}) = 0$ $(t = 1, 2, \ldots, +\infty)$, otteniamo

$$\Pr(X_1 < X_2) = \sum_{n\geq 1}\sum_{m\geq n+1} \Pr(\{X_1 = n\} \wedge \{X_2 = m\}).$$

Supposto, come è naturale in questo contesto, l'indipendenza dei tempi di attesa[3], risulta, per il Teorema 4.2.2(i), $\Pr(\{X_1 = n\} \wedge \{X_2 = m\}) = (\frac{1}{2})^n \frac{2}{3} (\frac{1}{3})^{m-1} (n, m \geq 1)$. Conseguentemente, la probabilità richiesta è data dalla:

$$\Pr(X_1 < X_2) = \sum_{n \geq 1} \left(\frac{1}{2}\right)^n \left[\sum_{m \geq n+1} \frac{2}{3} \left(\frac{1}{3}\right)^{m-1} \right].$$

Osservato che

$$\sum_{m \geq n+1} \frac{2}{3} \left(\frac{1}{3}\right)^{m-1} = \frac{2}{3} \left(\frac{1}{3}\right)^n \sum_{h \geq 0} \left(\frac{1}{3}\right)^h = \frac{2}{3} \left(\frac{1}{3}\right)^n \frac{1}{1 - \frac{1}{3}} = \left(\frac{1}{3}\right)^n,$$

otteniamo

$$\Pr(X_1 < X_2) = \sum_{n \geq 1} \left(\frac{1}{2}\right)^n \left(\frac{1}{3}\right)^n = \sum_{n \geq 1} \left(\frac{1}{6}\right)^n = \frac{1}{6} \sum_{h \geq 0} \left(\frac{1}{6}\right)^h = \frac{1}{6} \frac{1}{1 - \frac{1}{6}}$$

e quindi $\Pr(X_1 < X_2) = \frac{1}{5}$. $\Diamond$

4.3 Funzioni di densità congiunte

Nel caso bidimensionale (o multidimensionale) l'introduzione della funzione di densità richiede molte precisazioni di natura analitica. Per tale ragione, consideriamo questa nozione limitandoci al caso di coppie aleatorie $\underline{X}$ con ranghi inclusi in *domini normali*, cioè insiemi chiusi $D \neq \emptyset$ del tipo:

$$\begin{aligned} D &= \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ e } \varphi_1(x) \leq y \leq \varphi_2(x)\} \\ &= \{(x, y) \in \mathbb{R}^2 : c \leq y \leq d \text{ e } \psi_1(y) \leq x \leq \psi_2(y)\}, \end{aligned}$$

con φ_1, φ_2 funzioni continue in $[a, b]$ $(a < b)$ e ψ_1, ψ_2 funzioni continue in $[c, d]$ $(c < d)$ (come lo sono, ad esempio, i cerchi, i poligoni convessi e gli insiemi aventi come frontiera ellissi). Risulta quindi $D \subseteq R = [a, b] \times [c, d]$.

Nell'osservazione seguente percorriamo, per completezza d'esposizione, a grandi linee la costruzione dell'integrale doppio su domini normali.

[3]La nozione d'indipendenza per v.a. estese si ottiene da quella data per le v.a. consentendo a u, v di assumere anche i valori $\pm\infty$. Osserviamo che la proposizione (i) del Teorema 4.2.2 continua a sussistere anche in questo ambito.

Osservazione 4.3.1. RICHIAMI SUGLI INTEGRALI DOPPI. Sia f una funzione limitata su D. Sia intanto $D = R$. La definizione dell'integrale doppio ricalca, in questo caso, quella dell'integrale di Riemann di una funzione di una sola variabile. Date le suddivisioni:

$$\mathfrak{S}_1 : a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$$
$$\mathfrak{S}_2 : c = y_0 < y_1 < \cdots < y_{m-1} < y_m = d,$$

rispettivamente, di $[a, b]$ e $[c, d]$, i prodotti cartesiani $I_{hk} = [x_h, x_{h+1}] \times [y_k, y_{k+1}]$ $(h = 0, \ldots, n-1, k = 0, \ldots, m-1)$ forniscono una suddivisione $\mathfrak{S}$ di D. Possiamo allora introdurre la $\mathfrak{S}$-*somma inferiore*:

$$\underline{\mathrm{S}}(\mathfrak{S}) = \sum_{\substack{0 \le h \le n-1 \\ 0 \le k \le m-1}} \inf_{(x,y) \in I_{hk}} f(x, y)\,(x_{h+1} - x_h)(y_{k+1} - y_k)$$

e la $\mathfrak{S}$-*somma superiore*:

$$\overline{\mathrm{S}}(\mathfrak{S}) = \sum_{\substack{0 \le h \le n-1 \\ 0 \le k \le m-1}} \sup_{(x,y) \in I_{hk}} f(x, y)\,(x_{h+1} - x_h)(y_{k+1} - y_k).$$

Se ora si considerano tutte le possibili suddivisioni $\mathfrak{S}_1$ di $[a, b]$ e $\mathfrak{S}_2$ di $[c, d]$, l'insieme $\underline{\mathrm{S}}$ delle relative somme inferiori $\underline{\mathrm{S}}(\mathfrak{S})$ e quello $\overline{\mathrm{S}}$ delle relative somme superiori $\overline{\mathrm{S}}(\mathfrak{S})$ formano una coppia di classi separate. Qualora, le classi risultino contigue, cioè $\sup \underline{\mathrm{S}} = \lambda = \inf \overline{\mathrm{S}}$, la funzione f viene detta *integrabile su* D e si pone:

$$\iint_D f(x, y)\,dx\,dy = \lambda.$$

Sia infine D un dominio normale qualsiasi. Considerata la funzione g di dominio R così definita:

$$g(x, y) = \begin{cases} f(x, y), & \text{se } (x, y) \in D \\ 0, & \text{se } (x, y) \in R \setminus D \end{cases},$$

la funzione f viene detta *integrabile su* D se g è integrabile su R e si pone:

$$\iint_D f(x, y)\,dx\,dy = \iint_R g(x, y)\,dx\,dy.$$

Supponiamo ora che la funzione f sia continua in D. Allora. f risulta integrabile su D e sussistono le seguenti *formule di riduzione*:

$$\iint_D f(x, y)\,dx\,dy = \int_a^b dx \int_{\varphi_1(x)}^{\varphi_2(x)} f(x, y)\,dy = \int_c^d dy \int_{\psi_1(y)}^{\psi_2(y)} f(x, y)\,dx.^4$$

[4] Avendo posto $\int_u^{u'} dx \int_v^{v'} f(x, y)\,dy = \int_u^{u'} [\int_v^{v'} f(x, y)\,dy]\,dx$ e $\int_v^{v'} dy \int_u^{u'} f(x, y)\,dx = \int_v^{v'} [\int_u^{u'} f(x, y)\,dx]\,dy$, qualunque siano i numeri reali u, u' e v, v'.

Ricordiamo infine che si pone $\iint_\emptyset f(x,y)\,dx\,dy = 0$ e che si chiama *area di D* il valore dell'integrale $\mathrm{m}(D) = \iint_D dx\,dy$ della funzione costante di valore 1 su D; inoltre che, se D è unione di domini normali $D_1, \dots, D_n$ privi di punti interni a comune, risulta $\iint_D f(x,y)\,dx\,dy = \sum_{i=1}^n \iint_{D_i} f(x,y)\,dx\,dy$ (*additività*). $\triangle$

Prendendo spunto dal caso unidimensionale, chiamiamo **assolutamente continua** la coppia aleatoria $\underline{X}$ con rango incluso nel dominio normale D, se esiste una funzione $f \geq 0$ *continua* in D (detta (**funzione di**) **densità congiunta** di $\underline{X}$) tale che $f(x,y) = 0$, se $(x,y) \notin D$, e

$$F_{\underline{X}}(x,y) = \iint_{D\cap Q_{x,y}} f(u,v)\,du\,dv$$

per ogni x, y reali. Risulta allora (la dimostrazione è omessa):

$$\Pr(\underline{X} \in D') = \iint_{D\cap D'} f(x,y)\,dx\,dy = \iint_{R\cap D'} f(x,y)\,dx\,dy, \tag{4.1}$$

per ogni dominio normale D'.[5] Riesce quindi $\iint_D f(x,y)\,dx\,dy = 1$ e, come facilmente si verifica usando (1.6), $\Pr(X_1 \in [a,b]) = 1 = \Pr(X_2 \in [c,d\,])$.

Il teorema seguente, di notevole valenza applicativa, rileva che l'esistenza di una densità congiunta assicura l'assoluta continuità delle due v.a. X_1, X_2 e inoltre che la conoscenza delle loro densità consente di calcolare, in modo semplice via moltiplicazione, quella congiunta (identificando in questo modo la conoscenza congiunta con quella marginale anche al livello delle densità).

Conveniamo che, nel seguito, f denoti una densità congiunta della coppia aleatoria $\underline{X}$.

Teorema 4.3.2. *Sussistono le proposizioni:*

(i) Le v.a. X_1, X_2 sono assolutamente continue con densità rispettive (chiamate **densità marginali***):*

$$f_{X_1}(x) = \begin{cases} \displaystyle\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy, & \text{se } x \in [a,b] \\ 0, & \text{se } x \notin [a,b] \end{cases}$$

$$f_{X_2}(y) = \begin{cases} \displaystyle\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)\,dx, & \text{se } y \in [c,d] \\ 0, & \text{se } y \notin [c,d] \end{cases};$$

[5] Per quanto riguarda la seconda uguaglianza basta notare che, se $c' \leq \varphi_1(x) \leq \varphi_2(x) \leq d'$, allora $\int_{c'}^{d'} f(x,y)\,dy = \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy$ per ogni x di $[a,b]$.

(ii) *Le densità marginali sono funzioni uniformemente continue, rispettivamente, in $[a,b]$ e $[c,d]$;*

(iii) *Se $D = R$ e X_1, X_2 sono indipendenti, allora il prodotto $f_{X_1} f_{X_2}$ delle densità marginali è una densità congiunta di $\underline{X}$.*

Dimostrazione. (i) Sia $x \in [a,b]$. Allora, da (1.6) e (4.1) otteniamo

$$\begin{aligned} F_{X_1}(x) &= \Pr(X_1 \le x) = \Pr(\{a \le X_1 \le x\} \wedge \{c \le X_2 \le d\}) \\ &= \Pr(\underline{X} \in [a,x] \times [c,d]) = \iint_{[a,x]\times[c,d]} f(u,y)\,du\,dy \\ &= \int_a^x du \int_c^d f(u,y)\,dy = \int_a^x du \int_{\varphi_1(u)}^{\varphi_2(u)} f(u,y)\,dy. \end{aligned}$$

Osservato che $\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy \ge 0$ per ogni $x \in [a,b]$, possiamo interpretare f_{X_1} come una funzione di densità di X_1. In modo analogo si prova che f_{X_2} è una funzione di densità di X_2.

(ii) Con riferimento alla densità marginale f_{X_1}, assumiamo intanto che φ_1 sia la costante nulla. Dato $\epsilon > 0$, dalla continuità di f nell'insieme chiuso D e di φ_2 in $[a,b]$, esiste, per il teorema di Heine, $\delta > 0$ tale che $|f(x',y) - f(x'',y)| < \epsilon$ e $\varphi_2(x') - \varphi_2(x'')| < \epsilon$ per ogni x', x'' tali che $(x',y), (x'',y) \in D$ e $|x' - x''| < \delta$. Scelti $x' < x'' \in [a,b]$ tali che $|x' - x''| < \delta$, sia $0 \le y \le \min\limits_{x' \le x \le x''} \varphi_2(x)$. Allora, $(x',y), (x'',y) \in D$ e quindi, osservato che

$$\begin{aligned} f_{X_1}(x') - f_{X_1}(x'') &= \int_0^{\varphi_2(x')} f(x',y)\,dy - \int_0^{\varphi_2(x'')} f(x'',y)\,dy \\ &= \int_0^{\varphi_2(x')} f(x',y)\,dy - \int_0^{\varphi_2(x')} f(x'',y)\,dy \\ &\qquad + \int_0^{\varphi_2(x')} f(x'',y)\,dy - \int_0^{\varphi_2(x'')} f(x'',y)\,dy \\ &= \int_0^{\varphi_2(x')} (f(x',y) - f(x'',y))\,dy + \int_{\varphi_2(x'')}^{\varphi_2(x')} f(x'',y)\,dy, \end{aligned}$$

otteniamo

$$\begin{aligned} |f_{X_1}(x') - f_{X_1}(x'')| &\le \int_0^{\varphi_2(x')} |f(x',y) - f(x'',y)|\,dy + \int_{\varphi_2(x'')}^{\varphi_2(x')} |f(x'',y)|\,dy \\ &\le \varphi_2(x')\,\epsilon + \max|f|\;|\varphi_2(x') - \varphi_2(x'')| \le (\max\varphi_2 + \max|f|)\epsilon. \end{aligned}$$

Dall'arbitrarietà di $\epsilon > 0$ segue la tesi.

Sia ora φ_1 arbitrario. Considerato il cambio di variabile $t \rightsquigarrow y - \varphi_1(x)$, dal teorema d'integrazione per sostituzione otteniamo

$$f_{X_1}(x) = \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)\,dy = \int_0^{\varphi_2(x)-\varphi_1(x)} f(x, t+\varphi_1(x))\,dt,$$

qualunque sia x in $[a,b]$. Osservato che $\widetilde{\varphi}_2 = \varphi_2 - \varphi_1$ è continua in $[a,b]$ e $\widetilde{f}(x,y) = f(x, y+\varphi_1(x))$ è continua in $\widetilde{D} = \{(x,y) \in \mathbb{R}^2 : a \le x \le b \text{ e } 0 \le y \le \widetilde{\varphi}_2(x)\}$, dalla prima parte della dimostrazione otteniamo che f_{X_1} è uniformemente continua. In modo analogo si prova l'uniforme continuità dell'altra densità marginale.

(iii) Siano $D = R$ e X_1, X_2 indipendenti. Posto $\overline{f}(x,y) = f_{X_1}(x) f_{X_2}(y)$, risulta $\overline{f}(x,y) \ge 0$ e $\overline{f}(x,y) = 0$, se $(x,y) \notin D$; inoltre, per (ii), $\overline{f}$ è continua in D. Sia intanto $(x,y) \in D$. Allora, dal Teorema 4.2.1 e (i) otteniamo

$$\begin{aligned} F_{\underline{X}}(x,y) &= F_{X_1}(x) F_{X_2}(y) = \left[\int_a^x f_{X_1}(u)du\right]\left[\int_c^y f_{X_2}(v)dv\right] \\ &= \int_a^x \left[\int_c^y f_{X_1}(u) f_{X_2}(v)\,dv\right] du = \iint_{[a,x]\times[c,y]} \overline{f}(u,v)\,du\,dv \\ &= \iint_{D \cap Q_{x,y}} \overline{f}(u,v)\,du\,dv. \end{aligned}$$

Negli altri casi basta ricorrere a quanto appena provato e osservare che $D \cap Q_{x,y} = \emptyset$, se $x < a$ o $y < c$, e $D \cap Q_{x,y} = D = D \cap Q_{b,d}$, se $x > b$ e $y > d$, e $D \cap Q_{x,y} = D \cap Q_{x,d}$, se $a \le x \le b$ e $y > d$, e $D \cap Q_{x,y} = D \cap Q_{b,y}$, se $x > b$ e $c \le y \le d$. □

Esempio 4.3.3. (i) Date le v.a. X_1, X_2 indipendenti e distribuite uniformemente, rispettivamente, nell'intervallo $[a,b]$ e $[c,d]$, la f.r. congiunta è il prodotto delle f.r. marginali e la densità congiunta è il prodotto delle densità marginali (teoremi 4.2.1 e 4.3.2(iii)). Quindi, $F_{\underline{X}}(x,y) = \frac{(x-a)(y-c)}{(b-a)(d-c)}$ e $f(x,y) = \frac{1}{(b-a)(d-c)}$, per ogni $(x,y) \in D = [a,b] \times [c,d]$. Conseguentemente, se $b - a = d - c = 1$, la probabilità che $\underline{X}$ appartenga a un dominio normale D' è, per (4.1), l'area del dominio $D' \cap D$.

(ii) Date le v.a. X_1, X_2 indipendenti e distribuite uniformemente nell'intervallo $[0,1]$, consideriamo le v.a. $X = \min(X_1, X_2)$, $Y = |X_1 - X_2|$ e $Z = 1 - \max(X_1, X_2)$[6] e verifichiamo che hanno la medesima funzione di ripartizione. Chiaramente, se V denota una qualsiasi di queste v.a., risulta $\Pr(V \le t) = 0$, se $t < 0$, e $\Pr(V \le t) = 1$, se $t \ge 1$. Sia dunque $0 \le t \le 1$. Allora, dall'Esempio 4.2.4(i) otteniamo intanto $F_X(t) = 1 - (1-t)^2$.

[6]Da un punto di vista interpretativo, le v.a. introdotte rappresentano le lunghezze dei tre intervalli in cui $[0,1]$ viene suddiviso dai punti X_1, X_2 scelti a caso.

Per quanto riguarda Z, osservato che $Z = \min(1 - X_1, 1 - X_2)$, verifichiamo che le v.a. $U_i = 1 - X_i$ $(i = 1, 2)$ sono distribuite uniformemente e indipendenti. Dall'Esempio 3.2.5(i) (con $c = -1, d = 1$) risulta intanto che sono distribuite uniformemente in $[0, 1]$. Allora, $\Pr(U_i = x_i) = 0$ $(i = 1, 2)$ e quindi, per la monotonia della probabilità,

$$\begin{aligned} F_{\underline{U}}(x, y) &= \Pr(\{U_1 \le x\} \wedge \{U_2 \le y\}) \\ &= \Pr(\{U_1 = x\} \wedge \{U_2 \le y\}) + \Pr(\{U_1 < x\} \wedge \{U_2 = y\}) \\ &\qquad\qquad\qquad + \Pr(\{U_1 < x\} \wedge \{U_2 < y\}) \\ &= \Pr(\{U_1 < x\} \wedge \{U_2 < y\}). \end{aligned}$$

Ciò osservato, passiamo all'indipendenza. Posto $F_i = F_{X_i}$ $(i = 1, 2)$, sia $(x, y) \in D = [0, 1]^2$. Tramite i teoremi 4.1.1(vii) e 4.2.1 risulta

$$\begin{aligned} F_{\underline{U}}(x, y) &= \Pr(\{1 - X_1 < x\} \wedge \{1 - X_2 < y\}) \\ &= \Pr(\{1 - x < X_1 \le 1\} \wedge \{1 - y < X_2 \le 1\}) \\ &= F_{\underline{X}}(1, 1) + F_{\underline{X}}(1 - x, 1 - y) - F_{\underline{X}}(1 - x, 1) - F_{\underline{X}}(1, 1 - y) \\ &= F_1(1)F_2(1) + F_1(1 - x)F_2(1 - y) - F_1(1 - x)F_2(1) - F_1(1)F_2(1 - y) \\ &= 1 + (1 - x)(1 - y) - (1 - x) - (1 - y) = xy = F_{U_1}(x)F_{U_2}(y). \end{aligned}$$

Per gli altri casi basta ricorrere a quanto appena provato e osservare che $D \cap Q_{x,y} = \emptyset$, se $x < 0$ o $y < 0$, e $D \cap Q_{x,y} = D = D \cap Q_{1,1}$, se $x > 1$ e $y > 1$, e $D \cap Q_{x,y} = D \cap Q_{x,1}$, se $0 \le x \le 1$ e $y > 1$, e $D \cap Q_{x,y} = D \cap Q_{1,y}$, se $x > 1$ e $0 \le y \le 1$. Dunque, in ogni caso si ha $F_{\underline{U}}(x, y) = F_{U_1}(x)F_{U_2}(y)$. Conseguentemente, per il Teorema 4.2.1, U_1 e U_2 sono indipendenti. Possiamo allora ricorrere all'Esempio 4.2.4(i) ottenendo $F_Z(t) = 1 - (1 - F_{U_1}(t))(1 - F_{U_2}(t)) = 1 - (1 - t)^2 = F_X(t)$.

Passando infine alla v.a. Y otteniamo

$$\begin{aligned} \Pr(Y > t) = \Pr(|X_1 - X_2| > t) &= \Pr(\{X_1 > X_2 + t\} \vee \{X_2 > X_1 + t\}), \\ &= \Pr(X_1 > X_2 + t) + \Pr(X_2 > X_1 + t). \end{aligned}$$

Considerati il triangolo T_1 di vertici $(t, 0), (1, 0), (1, 1-t)$ e il triangolo T_2 di vertici $(0, t), (0, 1), (1 - t, 1)$, riesce $\{\underline{X} \in T_1\} = \{X_1 > X_2 + t\} \cap D$, $\{\underline{X} \in T_2\} = \{X_2 > X_1 + t\} \cap D$ da cui, tramite (i), si ha

$$\begin{aligned} \Pr(X_1 > X_2 + t) &= \Pr(\{X_1 > X_2 + t\} \cap D) = \Pr(\underline{X} \in T_1) \\ &= \mathrm{m}(T_1) = \frac{1}{2}(1 - t)^2 = \mathrm{m}(T_2) = \Pr(X_2 > X_1 + t). \end{aligned}$$

Allora, $\Pr(Y > t) = (1 - t)^2$ da cui otteniamo $F_Y(t) = 1 - (1 - t)^2 = F_X(t)$.

(iii) Suddiviso il triangolo D di vertici (-1, 0), (1, 0), (0,1) nei triangoli D_1 di vertici (0,0), (-1, 0), (0,1) e D_2 di vertici (0,0), (1, 0), (0,1), sia la coppia aleatoria $\underline{X}$ distribuita in D con densità congiunta proporzionale alla funzione $g(x,y) = y - x$, se $(x,y) \in D_1$, e $g(x,y) = y$, se $(x,y) \in D_2$. Risulta

$$\begin{aligned}
\iint_D g(x,y)\,dx\,dy &= \int_0^1 dy \int_{y-1}^{1-y} g(x,y)\,dx \\
&= \int_0^1 dy \left[\int_{y-1}^{0} (y-x)\,dx + \int_0^{1-y} y\,dx\right] \\
&= \int_9^1 dy \left[\frac{1-y^2}{2} + y(1-y)\right] = \int_9^1 \frac{1-3y^2+2y}{2}\,dy = \frac{1}{2}
\end{aligned}$$

da cui, tramite la condizione di normalizzazione $\iint_D f(x,y)\,dx\,dy = 1$, otteniamo

$$f(x,y) = 2\,g(x,y) = \begin{cases} 2(y-x), & \text{se } (x,y) \in D_1 \\ 2y, & \text{se } (x,y) \in D_2 \end{cases}.$$

Ne segue, per quanto concerne le densità marginali,

$$\begin{aligned}
f_{X_2}(y) &= 2\int_{y-1}^{1-y} g(x,y)\,dx = \begin{cases} 1-3y^2+2y, & \text{se } 0 \le y \le 1 \\ 0, & \text{altrimenti} \end{cases} \\
f_{X_1}(x) &= \begin{cases} 2\,\int_0^{1+x} g(x,y)\,dy = 2\,\int_0^{1+x} (y-x)\,dy = 1-x^2, & \text{se } -1 \le x \le 0 \\ 2\,\int_0^{1-x} g(x,y)\,dy = 2\,\int_0^{1-x} y\,dy = (1-x)^2, & \text{se } 0 \le x \le 1 \\ 0, & \text{altrimenti} \end{cases}.
\end{aligned}$$

Individuate le densità marginali, procediamo calcolando la f.r. congiunta nei punti del rettangolo $R = [-1,1] \times [0,1]$. Sia intanto $-1 \le x \le 0$ e $0 \le y \le x+1$. Considerato il trapezio T di vertici $(-1,0), (x,0), (x,y), (y-1,y)$, risulta $T = D \cap Q_{x,y}$ e quindi

$$F_{\underline{X}}(x,y) = 2\int_0^y dv \int_{v-1}^{x} (v-u)\,du = \frac{y\,[3xy + 3(1-x^2) - y^2]}{3}.$$

Sia ora $-1 \le x \le 0$ e $x+1 \le y \le 1$. Considerato il triangolo T di vertici $(-1,0)$, $(x,0)$, $(x,x+1)$, risulta $T = D \cap Q_{x,y} = D \cap Q_{x,x+1}$ e quindi dal caso precedente (con $y = x+1$) otteniamo

$$F_{\underline{X}}(x,y) = \frac{(x+1)^2(2-x)}{3}.$$

Sia ora $0 \le x \le 1$ e $0 \le y \le 1-x$. Considerati il trapezio T di vertici $(-1,0)$, $(x,0)$, $(x,y), (y-1,y)$ e il rettangolo R' di vertici $(0,0), (x,0), (x,y), (0,y)$, risulta

$T = (D \cap Q_{0,y}) \cup R'$ e quindi dal primo caso (con $x = 0$) otteniamo

$$\begin{aligned} F_{\underline{X}}(x,y) &= \iint_{D\cap Q_{0,y}} f(x,y)\,dx\,dy + \iint_{R'} f(x,y)\,dx\,dy = F_{\underline{X}}(0,y) + 2\int_0^y dv \int_0^x v\,du \\ &= \frac{y(3-y^2)}{3} + xy^2 = \frac{y\,(3-y^2+3xy)}{3}. \end{aligned}$$

Sia infine $0 \le x \le 1$ e $1 - x \le y \le 1$. Considerati i trapezi T di vertici $(-1,0)$, $(x,0)$, (x,y), $(y-1,y)$ e T' di vertici $(1-y,0)$, $(x,0)$, $(x,1-x)$, $(1-y,y)$, risulta $T = (D \cap Q_{1-y,y}) \cup T'$ e quindi dal caso precedente (con $x = 1-y$) otteniamo

$$\begin{aligned} F_{\underline{X}}(x,y) &= \iint_{D\cap Q_{1-y,y}} f(x,y)\,dx\,dy + \iint_{T'} f(x,y)\,dx\,dy \\ &= F_{\underline{X}}(1-y,y) + 2\int_{1-y}^x du \int_0^{1-u} v\,dv = \frac{y\,(3+3y-4y^2)}{3} + \frac{(x-1)^3+y^3}{3}. \end{aligned}$$

Consideriamo ora la v.a. trasformata $Z = X_2 - X_1$ e determiniamo la f.r. di F_Z. Denotato con S_z il semipiano di equazione $y - x \le z$, risulta $F_Z(z) = \Pr(Z \le z) = \Pr(X_2 - X_1 \le z) = \Pr(\underline{X} \in S_z \cap D)$ e quindi $F_Z(z) = 0$, se $z \le -1$, e $F_Z(z) = 1$, se $z \ge 1$; inoltre, se $-1 \le z \le 1$, l'intersezione $D \cap S_z$ è il triangolo di vertici $(-z,0),(1,0),(\frac{1-z}{2},\frac{1+z}{2})$. Sia intanto $-1 \le z \le 0$. Allora

$$F_Z(z) = 2\int_0^{\frac{1+z}{2}} dy \int_{y-z}^{1-y} y\,dx = 2\int_0^{\frac{1+z}{2}} y(1-2y+z)\,dy = \frac{(z+1)^3}{12}.$$

Sia infine $0 \le z \le 1$. Considerati i triangoli T di vertici $(0,0),(1,0),(\frac{1}{2},\frac{1}{2})$ e T_1 di vertici $(-z,0),(0,0),(0,z)$ e T_2 di vertici $(\frac{1-z}{2},\frac{1-z}{2}),(\frac{1}{2},\frac{1}{2}),(\frac{1-z}{2},\frac{1+z}{2})$ e il parallelogramma P di vertici $(0,0),(\frac{1-z}{2},\frac{1-z}{2}),(\frac{1-z}{2},\frac{1+z}{2}),(0,z)$, otteniamo $D \cap S_z = T \cup T_1 \cup T_2 \cup P = (D \cap S_0) \cup T_1 \cup T_2 \cup P$ e quindi

$$\begin{aligned} F_Z(z) &= \iint_{D\cap S_0} f(x,y)\,dx\,dy \\ &\quad + \iint_{T_1} f(x,y)\,dx\,dy + \iint_{T_2} f(x,y)\,dx\,dy + \iint_P f(x,y)\,dx\,dy \\ &= F_Z(0) \\ &\quad + 2\left[\int_{-z}^0 dx \int_0^{x+z} (y-x)\,dy + \int_{\frac{1-z}{2}}^{\frac{1}{2}} dx \int_x^{1-x} y\,dy + \int_0^{\frac{1-z}{2}} dx \int_x^{x+z} y\,dy\right] \\ &= \frac{1}{12} + \int_{-z}^0 (z^2-x^2)\,dx + \int_{\frac{1-z}{2}}^{\frac{1}{2}} (1-2x)\,dx + \int_0^{\frac{1-z}{2}} z(z+2x)\,dx \\ &= \frac{1}{12} + \frac{2}{3}z^3 + \frac{1}{4}z^2 + \frac{1}{4}z(1-z^2). \end{aligned}$$

Concludiamo con la probabilità condizionata $\Pr(X_2 \geq X_1 \,|\, Z \leq \frac{1}{2})$. Risulta

$$
\begin{aligned}
\Pr(X_2 \geq X_1 \,|\, Z \leq \frac{1}{2}) &= \frac{\Pr(\{X_2 \geq X_1\} \wedge \{X_2 \leq X_1 + \frac{1}{2}\})}{\Pr(Z \leq \frac{1}{2})} \\
&= \frac{\Pr(X_1 \leq X_2 \leq X_1 + \frac{1}{2})}{\Pr(Z \leq \frac{1}{2})} = \frac{\Pr(0 \leq Z \leq \frac{1}{2})}{\Pr(Z \leq \frac{1}{2})} \\
&= \frac{F_Z(\frac{1}{2}) - F_Z(0)}{F_Z(\frac{1}{2})} = \frac{23}{31},
\end{aligned}
$$

ove la penultima uguaglianza deriva dal Teorema 3.2.1(ix). $\Diamond$

Il prossimo risultato (che non dimostriamo) fornisce l'espressione della speranza matematica di trasformazioni continue di coppie aleatorie assolutamente continue.

Teorema 4.3.4. *Siano f e τ funzioni continue in R. Risulta allora:*

$$\mathrm{E}(\tau(\underline{X})) = \int_a^b dx \int_c^d \tau(x,y) f(x,y)\, dy = \int_c^d dy \int_a^b \tau(x,y) f(x,y)\, dx.$$

Il teorema seguente fornisce una condizione sufficiente affinchè la speranza matematica del prodotto di due v.a. indipendenti sia, come nel caso di v.a. semplici, il prodotto delle rispettive speranze matematiche.

Teorema 4.3.5. *Sia $D = R$ e f continua in D. Allora, se X_1, X_2 sono indipendenti, risulta* $\mathrm{E}(X_1 X_2) = \mathrm{E}(X_1)\mathrm{E}(X_2)$.

Dimostrazione. Dai teoremi 4.3.4 (con $\tau(x,y) = xy$) e 4.3.2(iii) otteniamo

$$
\begin{aligned}
\mathrm{E}(X_1 X_2) &= \int_a^b dx \int_c^d x\, y\, f_{X_1}(x) f_{X_2}(y)\, dy = \int_a^b x f_{X_1}(x) \left[\int_c^d y f_{X_2}(y)\, dy\right] dx \\
&= \left[\int_a^b x f_{X_1}(x) dx\right] \left[\int_c^d y f_{X_2}(y)\, dy\right].
\end{aligned}
$$

Ne segue, per i teoremi 4.3.2(i),(ii) e 3.4.4, la tesi. $\square$

Proviamo infine un semplice criterio che consente di decidere di primo acchito se le variabili aleatorie non sono indipendenti.

Teorema 4.3.6. *Siano $D^+ = \{(x,y) \in D : f(x,y) > 0\}$ l'insieme di positività della densità congiunta, $D_1^+ = \{x : \exists y((x,y) \in D^+)\}$ la proiezione di D^+ sull'asse delle ascisse e $D_2^+ = \{y : \exists x((x,y) \in D^+)\}$ la proiezione di D^+ sull'asse delle ordinate. Sia, inoltre, la densità congiunta continua in R. Allora, se le v.a. X_1, X_2 sono indipendenti, risulta $D^+ = D_1^+ \times D_2^+$.*

Dimostrazione. Assumiamo, per assurdo, $D^+ \neq D_1^+ \times D_2^+ = R'$. Poiché $D^+ \subset R'$, esiste $(x_0, y_0) \in R' \setminus D^+$. Allora, $x_0 \in D_1^+$ e quindi esiste y' tale che $(x_0, y') \in D^+$. Essendo D^+ un insieme aperto (f è continua in R'!), esiste un intorno rettangolare U di (x_0, y') incluso in D^+. Conseguentemente, $f(x_0, y) > 0$ per ogni $(x_0, y) \in U$ e quindi

$$f_{X_1}(x_0) = \int_{\varphi_1(x_0)}^{\varphi_2(x_0)} f(x_0, y)\, dy \geq \int_{\{y:(x_0,y)\in U\}} f(x_0, y)\, dy > 0.$$

Poiché in modo analogo si prova $f_{X_2}(y_0) > 0$, risulta

$$f_{X_1}(x_0) f_{X_2}(y_0) - f(x_0, y_0) > 0 \tag{4.2}$$

ricordato che $(x_0, y_0) \notin D^+$.

Notato che $x_0 \in [a, b]$ e $y_0 \in [c, d]$, la funzione $f_{X_1}(x) f_{X_2}(y) - f(x, y)$ è, per il Teorema 4.3.2(ii), continua in (x_0, y_0) e quindi, per (4.2), esiste un intorno rettangolare $V \subset R$ di (x_0, y_0) tale che $f_{X_1}(x) f_{X_2}(y) > f(x, y)$, per ogni $(x, y) \in V$.

Ora, l'ipotesi di continuità della densità congiunta in R e (4.1) ci consentono di assumere come dominio normale di riferimento R al posto di D. Ne segue allora, tramite il Teorema 4.3.2(iii) e (4.1),

$$\Pr(\underline{X} \in V) = \iint_V f(x, y)\, dx\, dy < \iint_V f_{X_1}(x) f_{X_2}(y)\, dx\, dy = \Pr(\underline{X} \in V),$$

ottenendo così una contraddizione. □

Esempio 4.3.7. Sia la coppia aleatoria distribuita nel cerchio D di raggio unitario centrato nell'origine con densità congiunta proporzionale alla funzione $g(x, y) = \sqrt{1 - x^2 - y^2}$. Allora, il coefficiente di proporzionalità è $\frac{3}{2\pi}$ (reciproco del volume della semisfera di raggio unitario) e quindi

$$f(x, y) = \begin{cases} \frac{3}{2\pi} \sqrt{1 - x^2 - y^2}, & \text{se } (x, y) \in D \\ 0, & (x, y) \notin D \end{cases}.$$

Poiché f è continua in $R = [-1, 1] \times [-1, 1]$ e D^+ è il cerchio aperto, dal teorema appena provato segue che le v.a. X_1, X_2 non sono indipendenti. ◊

4.4 Esercizi

1. Siano X_1, X_2 uniformemente distribuite in $[0, 1]$ e indipendenti e sia $Z = X_1 X_2$. Determinare: $*$ la probabilità che la loro distanza sia maggiore di $\frac{1}{3}$; $*$ per quali n risulta $\Pr(Z \leq \frac{1}{n} | X_1 > \frac{1}{2}) < \frac{1}{10}$); $*$ la f.r. di Z e la covarianza tra Z e $X_1 + X_2$; $*$ la speranza matematica di Z e di $X_2 - X_1^2$.

2. Dati X_1, X_2 con rango i primi 6 numeri naturali e distribuzione congiunta:

$$\Pr(\{X_1 = i\} \wedge \{X_2 = j\}) = \begin{cases} 0, & \text{se } i = j \\ \frac{1}{30}, & \text{se } i \neq j \end{cases},$$

determinare $\mathrm{E}(X_1 + X_2)$, $\mathrm{Var}(X_1 + X_2)$ e $\Pr(X_1 X_2 \text{ pari} \mid X_1 + X_2 \leq 5)$.

3. Siano X_1 il numero che esce nel lancio di un dado equilibrato e X_2 il numero di teste che escono in 6 lanci di una moneta equilibrata. Determinare: $*$ la f.r. congiunta; $*$ $\mathrm{E}(X_1 + X_2)$; $*$ $\Pr(X_1 = X_2)$; $*$ $\Pr(X_1 > X_2 \mid X_1 \neq X_2)$; $*$ la speranza matematica di X_1 sapendo che $X_1 + X_2 = 5$.

4. Dati X_1, X_2 indipendenti, sia X_1 distribuito uniformemente nell'intervallo $[0, 1]$ e X_2 distribuito nell'intervallo $[0, 2]$ secondo la *densità di Simpson*:

$$f_{X_2}(y) = \begin{cases} y, & \text{se } 0 \leq y \leq 1 \\ 2 - y, & \text{se } 1 \leq y \leq 2 \end{cases}.$$

Determinare la densità della v.a. $X_1 + X_2$.

5. Dati X_1, X_2 indipendenti, sia X_1 distribuito nell'intervallo $[1, 4]$ con densità proporzionale alla funzione $g(x) = x^2 + 1$ e X_2 distribuito uniformemente nell'intervallo $[-1, 2]$. Determinare $\mathrm{E}(X_1)$, $\mathrm{Var}(X_2)$, $\mathrm{E}(X_1^{X_2})$ e $\mathrm{E}(\frac{X_2}{X_1})$.

6. Da un'urna contenente 4 palline numerate da 1 a 4 si estraggono due palline. Indicato con X_1 il massimo numero estratto e con X_2 quello minimo, determinare: $*$ $\Pr(X_1 \leq 3 \mid |X_1 - 2X_2| = 2)$; $*$ $\Pr(X_2 \geq 2 \mid X_1 \geq 3)$; $*$ il grafico della funzione $g(t) = F_{\underline{X}}(t, t)$.

7. Data $\underline{X}$ distribuita nel triangolo D di vertici $(1, 0), (1, 1), (0, 1)$ con densità proporzionale alla funzione $g(x, y) = x^2$, determinare: $*$ le densità marginali; $*$ la speranza matematica e la varianza di X_2; $*$ la probabilità che l'equazione aleatoria di secondo grado $X_2\, t^2 + 2X_1\, t + 1 = 0$ ammetta soluzioni reali.

8. Data $\underline{X}$ distribuita nel parallelogramma D di vertici $(1, 0), (2, -1), (3, 0), (2, 1)$ con densità proporzionale alla funzione $g(x, y) = x(y+1)$, determinare: $*$ la densità marginale e la speranza matematica di X_1; $*$ la probabilità condizionata $\Pr(X_1 + 3X_2 \mid X_1 \leq 2)$.

9. Data $\underline{X}$ distribuita nel rettangolo D di vertici $(2, 0), (1, 1), (-1, -1), (0, -2)$ con densità proporzionale alla funzione $g(x, y) = |x\, y|$, determinare: $*$ le f.r. marginali; $*$ n tale che $\Pr(X_1^n - X_2 \geq 0) \geq \frac{1}{30}$.

10. Data $\underline{X}$ distribuita nel quadrato D di vertici $(-1, 0)$, $(0, -1)$, $(1, 0)$, $(0, 1)$ con densità proporzionale alla funzione $g(x, y) = x^2 + |y|$, determinare: $*$ le densità marginali; $*$ la probabilità condizionata $\Pr(X_2 - X_1 \leq 0 \mid X_1 \geq 0)$.

11. Data $\underline{X}$ distribuita uniformemente[7] nel quadrato D di vertici $(1, 1)$, $(2, 1)$, $(2, 2)$, $(1, 1)$, determinare: $*$ le f.r. e la speranza matematica delle v.a. X_1+X_2, X_1X_2 e $\frac{X_2}{X_1}$; $*$ una limitazione significativa della $\Pr(|X_1 + X_2 - \mathrm{E}(X_1) - \mathrm{E}(X_2)| > 2)$.

12. Sia la v.a. X_1 a valori nell'intervallo $[-1, 1]$ e X_2 tale che $0 \leq X_2 \leq 2 - |X_1|$. Indicato con D il rango della coppia aleatoria, sia $\underline{X}$ distribuito in D con densità congiunta proporzionale alla funzione $g(x, y) = |x|y$. Sono le v.a. indipendenti?

13. Data $\underline{X}$ distribuita nel dominio D delimitato dall'ellisse di equazione $x^2 + 4y^2 = 4$ con densità congiunta proporzionale alla funzione $g(x, y) = |x| + |y|$, determinare $\Pr(X_2 - |X_1| \geq 0)$.

14. Sia $\underline{X}$ distribuita nel quadrato D di vertici $(-1, 0)$, $(0, -1)$, $(1, 0)$, $(0, 1)$ con densità congiunta proporzionale alla funzione

$$g(x, y) = \begin{cases} 1 - |xy|, & \text{se } xy \leq 0 \\ 1, & \text{altrimenti} \end{cases}.$$

Indicata con F la f.r. della v.a. $X_1 + X_2$, calcolare $F(0)$ e risolvere le disequazioni $F(t) \leq 0$ e $F(t) \leq 1$. Sono le v.a. indipendenti?

15. Con riferimento al lancio simultaneo di due dadi distinguibili, Siano X_1 il massimo dei due numeri usciti e X_2 il valore assoluto della loro differenza. Determinare: $*$ la distribuzione di probabilità della coppia aleatoria $\underline{X}$; $*$ la correlazione esistente tra X_1 e X_2.

16. Con riferimento ai tre lanci di una moneta truccata tale che la probabilità che esca testa sia pari a $\frac{1}{3}$, siano X_1 il numero di teste che escono nei primi due lanci e X_2 il numero di croci che escono nel secondo e terzo lancio. Determinare: $*$ la distribuzione di probabilità della coppia aleatoria $\underline{X}$; $*$ la correlazione esistente tra X_1 e X_2; $*$ la speranza matematica e la varianza del loro prodotto.

---

[7] Cioè, con densità congiunta costante nel dominio D.

Bibliografia

[1] Baclawski, K., Cerasoli, M., Rota, G.C., Introduzione alla probabilità, Unione Matematica Italiana, Bologna, 1984.

[2] Baldi, P., Calcolo delle probabilità (seconda edizione), McGraw-Hill, Milano, 2011.

[3] Crisma, L., Introduzione alla teoria delle probabilità coerenti, Volume I, EUT, Trieste, 2006.

[4] Daboni, L., Calcolo delle probabilità ed elementi di statistica, UTET, Torino, 1970.

[5] Dall'Aglio, G., Calcolo delle probabilità, Zanichelli, Bologna, 1987.

[6] De Finetti, B. Calcolo delle probabilità (dattiloscritto 1937-'38), Atti dell'XI Convegno A.M.A.S.E.S., Bologna, 1987.

[7] Gnedenko, B.V., Teoria della probabilità, Editori Riuniti, Edizioni Mir, Roma, 1979.

Finito di stampare nel mese di marzo 2017
EUT Edizioni Università di Trieste