Capitolo 1

Basi del calcolo delle probabilità

1.1 Eventi

Per avviare la nostra discussione dobbiamo preliminarmente delimitare il campo dei fatti (o delle situazioni o delle asserzioni) nel quale si possa o si debba parlare di *incertezza* e quindi di probabilità. Nell'ambito del linguaggio probabilistico, tali fatti vengono chiamati *eventi*.

Per (o *evento* si intende un'*affermazione verificabile,* cioè qualcosa che può assumere solo due valori: *vero* o *falso*; non è detto però che si sappia con certezza la risposta, ed allora i casi sono (almeno provvisoriamente) tre: *certo*, *impossibile*, *incerto*, a seconda che sappiamo la risposta (SI o NO) o si rimanga in uno stato di incertezza (NON SO). La probabilità è una valutazione numerica dell'incertezza mediante una quantità compresa fra zero e uno, in particolare tale quantità assume valore "zero" nel caso di eventi impossibili ed assume valore "uno" nel caso di eventi certi. Se indichiamo con X il "risultato del lancio di un dado", allora X è una quantità aleatoria che sappiamo poter assumere un valore intero fra 1 e 6. L'incertezza circa gli eventi "esce il numero 4", cioè $(X = 4)$, oppure "esce un numero pari", cioè $(X \in \{2, 4, 6\})$, viene valutata calcolando la probabilità di tali eventi.

Definizione 1.1 Un *evento* è un fatto fisico o concettuale descritto da un enunciato che ammette due soli valori logici: VERO (V) o FALSO (F).

Esempio 1.2 Se dalla produzione di una fabbrica di lampadine si prelevano tre pezzi, la proposizione:

$$A \quad = \quad \text{"almeno una delle tre lampadine dura più di 250 ore"}$$

rappresenta un evento. Fino a quando non venga effettuata una verifica, si è in condizioni di incertezza.

Esempio 1.3 La proposizione

$$A \quad = \quad \text{"Il valore di chiusura al 28 marzo 2003 dell'indice MIB30 è di 22.650 punti"}$$

rappresenta un evento la cui verità o falsità può essere valutata andando a controllare i valori delle quote del fondo per il giorno in oggetto". ♣

Esempio 1.4 La proposizione

A = "Il valore di chiusura al 25 maggio 2003 dell'indice MIB30 è superiore a 22.650 punti"

rappresenta un evento. La verità di tale evento può essere verificata solo al momento opportuno. ♣

Due eventi di particolare importanza sono *evento certo* Ω e l'*evento impossibile* $\emptyset$. Nel caso più frequente, si è in uno stato di incertezza e si parla di *evento possibile*.

Definizione 1.5 Si definisce *indicatore* di un evento A – e si indica con $|A|$ o con I_A – la quantità:

$$|A| := \begin{cases} 1 & \text{se } A \text{ è vero} \\ 0 & \text{se } A \text{ è falso} \end{cases}$$

In particolare risulta $|\Omega| = 1$ e $|\emptyset| = 0$.

Introduciamo le seguenti operazioni logiche fra due eventi A e B.

Definizione 1.6 Siano A e B due eventi. Si definisce

1. evento *somma logica* $A \vee B$ l'evento vero se e solo se almeno uno dei due eventi è vero. Le tabelle della verità (di cui la seconda è data in termini degli indicatori) sono:

A	B	$A \vee B$
V	V	V
V	F	V
F	V	V
F	F	F

$\|A\|$	$\|B\|$	$\|A \vee B\|$
1	1	1
1	0	1
0	1	1
0	0	0

   In particolare, due eventi A e B si dicono *necessari* se $A \vee B = \Omega$.

2. evento *prodotto logico* $A \wedge B$ (indicato anche con AB) è l'evento vero se e solo se entrambi gli eventi sono veri. Le tabelle della verità sono:

A	B	$A \wedge B$
V	V	V
V	F	F
F	V	F
F	F	F

$\|A\|$	$\|B\|$	$\|A \wedge B\|$
1	1	1
1	0	0
0	1	0
0	0	0

   In particolare due eventi A e B vengono detti *incompatibili* se $A \wedge B = \emptyset$.

3. evento *contrario* o *complementare* A^c o $\overline{A}$ è l'evento vero se e solo se l'evento A è falso. Le tabella della verità sono (la seconda in termini degli indicatori):

A	A^c
V	F
F	V

$\|A\|$	$\|A^c\|$
1	0
0	1

4. evento *differenza* A meno B, e si scrive $A \setminus B$ definito da $A \setminus B = A \wedge B^c$. Si noti che in generale gli eventi $A \setminus B$ e $B \setminus A$ sono diversi.

Ovviamente, per qualsiasi evento A risulta: $A \vee A^c = \Omega$ e $A \wedge A^c = \emptyset$.

Definizione 1.7 Siano A e B due eventi. Diremo che A *implica* B, e scriveremo $A \Rightarrow B$ se B risulta vero ogni volta che A è vero.

Definizione 1.8 Siano A e B due eventi. Diremo che A e B sono *logicamente indipendenti* se nessuno dei quattro eventi: $A \wedge B$, $A \wedge B^c$, $A^c \wedge B$, $A^c \wedge B^c$ è impossibile.

Esempio 1.9 Dalla produzione giornaliera di un dato macchinario si prelevano 4 pezzi e denotiamo con E_i l'evento "l'i-esimo pezzo funziona correttamente", con $i = 1, 2, 3, 4$. Sia:

$$\begin{aligned} A &= \text{“almeno uno dei 4 articoli è difettoso”}, \\ B &= \text{“almeno due dei 4 articoli sono difettosi”}. \end{aligned}$$

Ovviamente si ha:

$$\begin{aligned} A = \{&E_1^c E_2 E_3 E_4 \vee E_1 E_2^c E_3 E_4 \vee E_1 E_2 E_3^c E_4 \vee E_1 E_2 E_3 E_4^c \vee E_1^c E_2^c E_3 E_4^c \vee E_1^c E_2 E_3^c E_4 \vee \\ &E_1^c E_2 E_3 E_4^c \vee E_1 E_2^c E_3^c E_4 \vee E_1 E_2^c E_3 E_4^c \vee E_1 E_2 E_3^c E_4^c \vee E_1^c E_2^c E_3^c E_4 \vee E_1^c E_2^c E_3 E_4^c \vee \\ &E_1^c E_2 E_3^c E_4^c \vee E_1 E_2^c E_3^c E_4^c\} \end{aligned}$$

Le proposizioni che individuano gli eventi A^c e B^c si enunciano:

$$\begin{aligned} A^c &= \text{“nessuno dei 4 articoli è difettoso”}, \\ B^c &= \text{“al più uno dei 4 articoli è difettoso”}. \end{aligned}$$

e risulta, ad esempio, $A = E_1 E_2 E_3 E_4$. Calcolare gli eventi B e B^c in termini degli eventi E_i. ♣

Diagrammi di Venn. Gli eventi possono essere descritti mediante insiemi. Considerato un insieme Ω, che costituisce l'evento certo, gli eventi sono allora sottoinsiemi di Ω. Gli eventi possono quindi essere rappresentati graficamente mediante i cosidetti *diagrammi di Venn*. I concetti della logica degli eventi si possono allora far corrispondere a quelli della *teoria degli insiemi* secondo il seguente schema:

Eventi	*Simbologia*	**Insiemi**	*Simbologia*
certo	Ω	ambiente	Ω
impossibile	$\emptyset$	vuoto	$\emptyset$
implicazione	$A \Rightarrow B$	inclusione	$A \subseteq B$
contrario	$\overline{A}$	complementare	A^c
somma logica	$A \vee B$	unione	$A \cup B$
prodotto logico	$A \wedge B$	intersezione	$A \cap B$
incompatibili	$A \wedge B = \emptyset$	disgiunti	$A \cap B = \emptyset$
classe completa	$\bigvee_i A_i = \Omega$	partizione	$\bigcup_i A_i = \Omega$

Nel seguito indicheremo con lo stesso simbolo A sia l'evento che l'insieme che lo rappresenta; inoltre per uniformarci alla notazione utilizzata usualmente in letteratura, utilizzeremo la notazione insiemistica.

Definizione 1.10 Assegnato l' evento certo Ω, diremo che A è un evento di Ω se A implica Ω ($A \Rightarrow \Omega$) o, equivalentemente, se A è contenuto in Ω ($A \subseteq \Omega$).

Sia Ω l'evento certo. Per semplicità con il termine *insieme* o *famiglia di eventi di* Ω intenderemo una qualunque famiglia $\mathcal{A} = \{A_1, \ldots, A_n\}$ di eventi di Ω.

Definizione 1.11 Sia $\mathcal{A} = \{A_1, A_2, \ldots\}$ un insieme di eventi. Un evento E si dice *elementare* rispetto ad $\mathcal{A}$ – oppure *atomo* di $\mathcal{A}$ – se per ogni $A \in \mathcal{A}$ risulta

$$A \cap E \;=\; E \qquad \text{oppure} \qquad A \cap E \;=\; \emptyset \;.$$

Tale definizione stabilisce che se E è un evento elementare, allora per ogni evento A di $\mathcal{A}$ si possono avere solo due casi: E è incluso in A, oppure E è incompatibile con A. Usualmente gli eventi elementari vengono denotati col simbolo ω. L'insieme $\{\omega\}$ viene chiamato *singoletto* di Ω.

Esercizio 1.12 Una compagnia di assicurazioni è interessata alla distribuzione per età dei coniugi: se x è l'età del marito, y quella della moglie, ad ogni osservazione si può far corrispondere un punto del primo quadrante ($x > 0, y > 0$) nel piano (x, y). Determinare gli insiemi del piano che rappresentano gli eventi:

$$\begin{aligned} A &= \text{“il marito ha più di 40 anni”} \,, \\ B &= \text{“il marito è più anziano della moglie”} \,, \\ C &= \text{“la moglie ha meno di 42 anni”} \,. \end{aligned}$$

Esercizio 1.13 Uno studente sostiene tre esami in una sessione di esami. Consideriamo i seguenti eventi:

$$\begin{aligned} E_1 &= \text{“lo studente supera il primo esame”} \,, \\ E_2 &= \text{“lo studente supera il secondo esame”} \,, \\ E_3 &= \text{“lo studente supera il terzo esame”} \,. \end{aligned}$$

Calcolare i seguenti eventi: i) Lo studente supera tutti e tre gli esami; ii) lo studente supera due esami; iii) lo studente supera almeno due esami; iv)lo studente supera un esame; v) lo studente supera il primo esame; vi) lo studente supera al più due esami; vii) lo studente non supera alcun esame.

Esercizio 1.14 Si consideri il lancio di due dadi. Calcolare gli eventi tali che la somma dei due dadi sia uguale a 3, a 5, ed a 14. Qual è quel numero che corrisponde al maggior numero di coppie di risultati?

Esercizio 1.15 Un concessionario di automobili offre autoveicoli con le seguenti opzioni: i) con e senza cambio automatico; ii) con e senza climatizzatore; iii) con una o due scelte di impianto stereo; iv) tre diversi colori esterni.
Qual è l'insieme Ω? Quali sono gli eventi elementari di Ω? Quanti sono?

1.1.1 Proprietà delle operazioni fra eventi

Si può dimostrare che le operazioni fra eventi soddisfano le seguenti proprietà:

Proprietà	*Unione*	*Intersezione*
Commutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associativa	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
Idempotenza	$A \cup A = A$	$A \cap A = A$
Distributiva	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Formule di De Morgan. Un'importante relazione fra gli eventi è descritta dalle *Formule di De Morgan*. Nel caso di due eventi A e B si ha:

$$(A \cup B)^c = A^c \cap B^c \tag{1.1}$$
$$(A \cap B)^c = A^c \cup B^c \, . \tag{1.2}$$

E' possibile fare una verifica di tali relazioni in termini di tabelle di verità, rispettivamente per la (1.1) e la (1.2):

A	B	$A \cup B$	$(A \cup B)^c$	A^c	B^c	$A^c \cap B^c$
V	V	V	F	F	F	F
V	F	V	F	V	F	F
F	V	V	F	V	F	F
F	F	F	V	V	V	V

A	B	$A \cap B$	$(A \cap B)^c$	A^c	B^c	$A^c \cup B^c$
V	V	V	F	F	F	F
V	F	F	V	V	F	V
F	V	F	V	V	F	V
F	F	F	V	V	V	V

Provare per esercizio a fare la verifica utilizzando i diagrammi di Venn.

Più in generale per un insieme al più numerabile di eventi $A_1, A_2, \ldots, A_n, \ldots$ si ha:

$$\bigcup_{n=1}^{\infty} A_n = \left(\bigcap_{n=1}^{\infty} A_n \right)^c \tag{1.3}$$

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n \right)^c \tag{1.4}$$

1.1.2 Famiglie di eventi di particolare importanza.

Introduciamo ora alcune famiglie di insiemi che verranno considerate frequentemente in seguito.

Definizione 1.16 Sia Ω un insieme. Una famiglia $\mathcal{A} = \{A_1, A_2, \ldots\}$ al più numerabile di sottoinsiemi di Ω si chiama *partizione* di Ω se possiede le seguenti proprietà:

1. Se $A_i \cap A_j = \emptyset$ per ogni $A_i, A_j \in \mathcal{A}$ con $i \neq j$
2. $\bigcup_{i=1}^{\infty} A_i = \Omega$.

In particolare si parla di *partizione finita* nel caso in cui $\mathcal{A}$ contiene un numero finito di eventi e di *partizione numerabile* nel caso in cui $\mathcal{A}$ è una famiglia numerabile di eventi, cioè una famiglia costituita da un numero infinito numerabile di eventi. Quindi una partizione è una classe completa di eventi a due a due incompatibili in cui uno ed uno solo è vero. Semplici esempi di partizione possono essere costruiti considerando i risultati ottenuti lanciando un dado $\Omega = \{1, 2, 3, 4, 5, 6\}$. Possiamo avere $\mathcal{A} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$(risultati elementari) oppure $\mathcal{A} = \{\{1, 3, 5\}, \{2, 4, 6\}\}$ (risultati dispari e risultati pari), $\mathcal{A} = \{\{1, 2, 3\}, \{4, 5, 6\}\}$ (risultati inferiori a 4 e risultati maggiori o uguali a 4), e così via.

In termini di indicatori, se una famiglia $\mathcal{A}$ costituisce una partizione, allora risulta:

$$\sum_{i=1}^{n} |A_i| = |A_1| + |A_2| + \cdots + |A_n| = 1 ,$$

$$\sum_{i=1}^{\infty} |A_i| = |A_1| + |A_2| + \cdots + |A_n| + \cdots = 1 ,$$

a seconda che si tratti, rispettivamente, di una partizione finita o infinita.

Definizione 1.17 Sia Ω un insieme. Una famiglia finita $\mathcal{A} = \{A_1, \ldots, A_n\}$ di sottoinsiemi di Ω si chiama *algebra* se gode delle seguenti due proprietà:

1. Se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$;
2. Se $A, B \in \mathcal{A}$ allora $A \cup B \in \mathcal{A}$.

Si dimostra che, se $\mathcal{A}$ è un'algebra di parti di Ω, allora $\Omega, \emptyset \in \mathcal{A}$ e per ogni $A, B \in \mathcal{A}$ si ha $A \cap B \in \mathcal{A}$. Infatti:

- $\underline{\Omega \in \mathcal{A}}$: Dalla 1., se $A \in \mathcal{A}$ allora anche $A^c \in \mathcal{A}$; dalla 2. segue quindi che $A \cup A^c = \Omega \in \mathcal{A}$.
- $\underline{\emptyset \in \mathcal{A}}$: Se $\Omega \in \mathcal{A}$, per la 1. si ha che $\Omega^c = \emptyset \in \mathcal{A}$.
- $\underline{A \cap B \in \mathcal{A}}$: siano $A, B \in A$, allora per la 1. si ha che $A^c, B^c \in \mathcal{A}$ e per la 2. $A^c \cup B^c \in \mathcal{A}$. Per la formula di De Morgan si ha che $A^c \cup B^c = (A \cap B)^c \in \mathcal{A}$ e quindi ancora per la 1. segue che $A \cap B \in A$. □

Definizione 1.18 Sia Ω un insieme. Una famiglia $\mathcal{A} = \{A_1, A_2, \ldots, A_n, \ldots\}$ di parti di Ω si chiama *σ-algebra* se:

1. $\emptyset, \Omega \in \mathcal{A}$
2. Se $A \in \mathcal{A}$ allora $A^c \in \mathcal{A}$;
3. Se $A_1, A_2, \ldots \in \mathcal{A}$ allora $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ e $\bigcap_{n=1}^{\infty} A_N \in \mathcal{A}$.

Si noti in effetti che nella definizione precedente la proprietà 1. è ridondante e, per le formule di De Morgan, anche la seconda condizione della proprietà 3.

L'individuazione di Ω può essere a volte problematica in quanto, come in ogni astrazione matematica di situazioni reali, vi è un certo grado di arbitrarietà insito nella scelta dell'insieme Ω di tutti i possibili risultati di un esperimento. Tale scelta di solito implica opportune semplificazioni.

1.2 Le diverse concezioni della probabilità

La probabilità sorge nel XVII secolo con la pubblicazione dell'*Ars Conjectandi* di James Bernoulli dal tentativo di valutare e gestire quantitativamente l'incertezza a partire da studio di problemi sui giochi d'azzardo. Per quanto riguarda l'aspetto più propriamente matematico, le idee sulla probabilità non sono fortemente discordanti. Il nucleo della discordia concerne invece il significato della nozione di probabilità: vi sono concezioni secondo cui, per elevare a rango di scientifico la probabilità si deve depurarla di tutti gli elementi soggettivi che la riguardano e la caratterizzano; dall'altro vi sono quelle concezioni che non solo non ritengono tali elementi come causa di disturbo, ma ne fanno il punto di partenza per la definizione e la conseguente

teoria matematica. Le concezioni del primo tipo (*oggettive*) ritengono la probabilità come elemento del mondo fisico, esistente fuori di ciascuno di noi; le concezioni del secondo tipo (*soggettive*) ritengono che la probabilità, esprimendo in ogni caso un'opinione personale, venga ad esistenza solo nel momento in cui il giudizio del soggetto viene a contatto di date circostanze concrete, incerte.

1.2.1 La concezione classica

Il primo tentativo di una formalizzazione matematicamente rigorosa della teoria della probabilità è dovuto a P.-S. de Laplace. Assumiamo che un evento A sia costituito dall'unione di un certo numero di eventi elementari. Nella sua opera monumentale[1], viene data la definizione *classica* della probabilità di un evento A come proporzione del numero di risultati favorevoli (eventi elementari) all'evento A ed il numero (finito) di tutti i possibili risultati (eventi elementari) connessi all'esperimento in questione, dove tali risultati vengono supposti *egualmente possibili*.

Definizione 1.19 (classica) Sia $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ l'insieme degli eventi elementari (possibili risultati) concernenti un certo esperimento $\mathcal{E}$, supposti ugualmente possibili; siano $\omega_{i_1}, \omega_{i_2}, \omega_{i_p}$ i p risultati favorevoli ad un certo evento A, cioè $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega i_p\}$. Si definisce *probabilità dell'evento A* la quantità:

$$P(A) \;:=\; \frac{\#A}{\#\Omega}\,,$$

dove il simbolo $\#$ denota "numero di elementi contenuti in".

Ovviamente, se l'evento A è impossibile, allora $\#A = 0$ e quindi $P(A) = 0$.

Si noti che, in accordo a tale definizione, il calcolo delle probabilità si riduce a problemi di conteggio di eventi, sulla base di regole del calcolo combinatorio. Tale definizione è inadeguata in quanto circolare e restrittiva; essa tuttavia coglie un aspetto importante legato al fatto che esistono situazioni in cui, per esempio in base a considerazioni di tipo geometrico, si può ammettere che gli eventi elementari siano equiprobabili.

1.2.2 La concezione frequentista

La concezione frequentista proposta da von Mises[2] è quella cui aderiscono la maggior parte dei testi scientifici, specie delle scienze sperimentali. Sia $A_1, A_2, \ldots, A_n$ una successione di n prove ripetute ed indipendenti (ad esempio, lanci di una moneta) in cui un certo risultato A (ad esempio, "Testa") può o meno verificarsi. Pertanto A_1 risulta vero se l'evento A si è verificato nella prima prova (cioè se esce "testa" nel primo lancio), A_2 risulta vero se l'evento A si è verificato nella seconda prova (cioè se esce "testa" nel secondo lancio), e così via. Indichiamo

[1] P.-S. de Laplace, *Théorie analytique des probabilités*, Courcier, Paris, 1814.

[2] R. von Mises (1930), Veroyatnost i statistika (Probability and statistics), Moscow-Leningrad, (trad. inglese: Academic Press, New York, 1964).

con f_n la frazione di successi sulle n prove, cioè la frazione del numero di volte in cui si è verificato l'evento A nelle n prove:

$$f_n \;=\; \frac{|A_1| + |A_2| + \cdots + |A_n|}{n} \,.$$

In altre parole, f_n è la frequenza relativa del numero di successi nelle prime n prove.

Definizione 1.20 (frequentista) Si definisce probabilità p dell'evento A il limite della frequenza relativa f_n "a lungo andare", cioè al crescere indefinito del numero n delle prove:

$$P(A) \;:=\; \lim_{n\to\infty} f_n \,. \tag{1.5}$$

Secondo la teoria frequentista, infatti, non ha senso parlare di probabilità a proposito di un singolo evento, ma solo a proposito di una collezione di eventi dello stesso tipo.

Anche questa definizione è criticabile in quanto attiene solo ad (un numero infinito di) eventi ripetibili nelle stesse condizioni; inoltre essa non risulta adeguata dal punto di vista matematico. In base alla definizione di limite di successione, si dice che $\{a_n\}_n$ converge a l se:

$$\forall \varepsilon > 0 \quad \exists\, \nu \in \mathbb{N} \mid \text{per ogni } n > \nu \text{ si ha } \; |a_n - l| < \varepsilon \,. \tag{1.6}$$

In tal caso, notiamo che la definizione (1.5) non rispetta la (1.6) in quanto, trattandosi di eventi aleatori, non è possibile trovare l'indice ν tale che risulti $|f_n - p| < \varepsilon$ per ogni $n > \nu$.

La definizione di von Mises coglie tuttavia un aspetto interessante rispetto a quella di Laplace in quanto qui la valutazione dell'incertezza deriva dall'osservazione ripetuta di un certo fenomeno; tralascia però il fatto che la nostra conoscenza spesso non è l'esito finale di una serie di prove ripetute (per cui non si può dire nulla prima di aver fatto un numero grande di esperimenti), ma al contrario il risultato dei vari esperimenti aggiornano e modificano via via una conoscenza precedente (o una ignoranza iniziale). Al crescere del numero di esperimenti, la nostra conoscenza dipenderà sempre più dai risultati ottenuti piuttosto che dall'ipotesi iniziale.

1.2.3 La concezione soggettiva

Secondo l'approccio soggettivo, dovuto principalmente a de Finetti[3], la probabilità è una misura del *grado di fiducia* che un soggetto coerente ripone nel verificarsi di un certo evento A. In tale contesto, il termine "soggettivo" non va inteso come "mancante di oggettività" ma, al contrario, come risultante dal soppesare – con il massimo sforzo di oggettività – ogni elemento di giudizio di cui il soggetto dispone. Pertanto in questo constesto, si parla usualmente di *valutazione di probabilità* anzichè di *stima* come nel caso delle concezioni oggettive.

Un modo di valutare la probabilità di un evento è quello del cosidetto *meccanismo della scommessa*. Dato un numero reale p si dice che si effettua una *scommessa* di quota p su un evento A se, versando una somma pS (con $S > 0$ arbitrario), si riceve un *importo* S se A si verifica (scommessa *vinta*) e niente in caso contrario (perdendo la puntata pS).

[3] B. de Finetti (1937). La prévision: ses lois logiques, ses sources subjectives, *Annales de l'Institut H. Poincaré*, **7**, 1-68.

Una scommessa di quota p (su un evento A *possibile*) da luogo ad un guadagno:

$$G_1 \;=\; S - pS \;=\; (1-p)S$$

nel caso in cui l'evento A si verifica e ad un guadagno:

$$G_2 \;=\; -pS$$

Nel caso contrario. Il guadagno globale può essere scritto sinteticamente

$$G \;=\; (|A| - p)S$$

dove $G = G_1$ se l'evento A è vero (e quindi $|A| = 1$) e $G = G_2$ se l'evento A è falso (e quindi risulta $|A| = 0$).

Definizione 1.21 Una scommessa si dice *coerente* se *non* esiste alcuna somma $S \neq 0$ tale che le quantità G_1 e G_2 corrispondenti ai due possibili esiti della scommessa (cioè il verificarsi di E o del suo contrario) siano dello stesso segno.

E' questa una condizione del tutto naturale che ha lo scopo di evitare che sia garantito *a priori* – cioè indipendentemente dall'esito della scommessa – un guadagno (o una perdita). In termini analitici, una scommessa di quota p su un evento A possibile si dice *coerente* se non esiste alcuna somma S diversa da zero ($S \neq 0$) tale che la quantità $G = (|A| - p)S$ assuma lo stesso segno in corrispondenza dei due valori logici A e A^c, cioè se risulta:

$$G_1 G_2 \;=\; -pS \cdot (1-p)S \;=\; -p(1-p)S^2 \;\leq\; 0 \qquad \text{per ogni } S \neq 0\ . \tag{1.7}$$

In altre parole una scommessa è coerente se non da luogo a guadagni o perdite certe.

Proposizione 1.22 Comunque si consideri un evento A, la scommessa è coerente se e solo se la quantità p è compresa fra 0 e 1.

Dimostrazione. La tesi segue immediatamente dalla (1.7). Infatti $G_1 G_2 = -p(1-p)S^2$ risulta non negativa se e solo se si ha $0 \leq p \leq 1$. □

Definizione 1.23 (Definizione soggettiva) Si chiama *probabilità* di un evento A la misura del "grado di fiducia" in A espressa da un numero reale $p = P(A)$ tale che una scommessa di quota p su A sia coerente.

Si noti che questa definizione fornisce un *significato* del numero p, indipendentemente dal modo in cui viene valutato.

1.2.4 La teoria assiomatica

L'impostazione assiomatica è dovuta a Kolmogorov[4]. La probabilità è un particolare tipo di misura definita su un'algebra o su una σ-algebra.

Definizione 1.24 (assiomatica) Sia Ω un insieme, $\mathcal{A}$ una σ-algebra di parti di Ω. Si definisce *probabilità* una funzione $P : \mathcal{A} \to \mathbb{R}^+$ tale che:

1. $P(A) \geq 0$ per ogni $A \in \mathcal{A}$;
2. $P(\Omega) = 1$;
3. Se $\{A_n\}$ è una famiglia di elementi di $\mathcal{A}$ a due a due disgiunti, allora risulta:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \;=\; \sum_{n=1}^{\infty} P(A_n)\,. \tag{1.8}$$

La (1.8) è nota come *proprietà di σ-additività*. Si noti che nel caso di famiglie finite di eventi, la proprietà di σ-additività si riduce alla semplice proprietà di additività:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \;=\; \sum_{i=1}^{n} P(A_i)\,.$$

Definizione 1.25 Si definisce *spazio di probabilità* una terna $(\Omega, \mathcal{A}, P)$ dove:
Ω è un insieme;
$\mathcal{A}$ è una σ-algebra di parti di Ω;
P è una probabilità su $\mathcal{A}$.

Gli spazi di probabilità sono modelli matematici di fenomeni reali non deterministici. Questa modellizzazione viene effettuata basandosi su considerazioni empiriche e soggettive, tenendo conto della natura del problema. Ciò vuol dire che, in generale, dato un fenomeno aleatorio, non esiste un unico spazio di probabilità che lo descriva; è anzi possibile che persone diverse scelgano modelli basati su spazi di probabilità differenti.

Nota 1.26 Nel seguito assumeremo sempre che $\mathcal{A}$ sia un'algebra o una σ-algebra di parti di Ω e che sia sempre soddisfatta l'ipotesi di σ-additività (1.8).

Definizione 1.27 Sia $\mathcal{A}$ una famiglia di eventi di Ω. Diremo che sulla famiglia $\mathcal{A}$ è assegnata una *distribuzione di probabilità*, se ad ogni $A \in \mathcal{A}$ viene associato un numero reale $p = P(A)$ ad ogni $A \in \mathcal{A}$, tale che:

1. per ogni $A \in \mathcal{A}$ risulta $0 \leq P(A) \leq 1$;
2. $P(\emptyset) = 0$, $P(\Omega) = 1$;
3. se per ogni partizione $A_1, A_2, \ldots$ di Ω contenuta in $\mathcal{A}$ risulta:

$$\sum_{i=1}^{\infty} P(A_i) \;=\; P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots \;=\; 1\,. \tag{1.9}$$

[4] A. J. Kolmogorov (1933), *Grundbegriffe der Wahrsheinlichkeitsrechung*, Springer Verlag, Berlin (trad. inglese: Chlesea, New York, 1950)

1.3 Primi risultati del calcolo delle probabilità

Dalle tre proprietà della probabilità viste sopra discendono alcune regole per il calcolo della probabilità di vari eventi.

Teorema 1.28 (probabilità dell'evento contrario) Sia A un evento e sia A^c il suo contrario. Allora risulta:

$$P(A^c) \;=\; 1 - P(A)\,.$$

Dimostrazione. Poichè A^c è l'evento contrario di A, allora risulta ovviamente $A \cup A^c = \Omega$. Inoltre, per lo stesso motivo, A e A^c sono incompatibili. Pertanto segue:

$$P(\Omega) \;=\; P(A \cup A^c) \;=\; P(A) + P(A^c)\,.$$

Poichè $P(\Omega) = 1$, allora si ha $P(A) + P(A^c) = 1$, da cui $P(A^c) = 1 - P(A)$. □

In particolare, da questo risultato segue che $P(\emptyset) = 1 - P(\Omega) = 0$.

Teorema 1.29 Siano A, B due eventi e sia $A \setminus B$ l'evento differenza "A meno B". Allora risulta:

$$P(A \setminus B) \;=\; P(A) - P(A \cap B)\,.$$

Dimostrazione. Preliminarmente ricordiamo che si ha $A \setminus B \equiv (A \cap B^c)$. Per la proprietà distributiva dell'intersezione risulta:

$$A \;=\; A \cap \Omega \;=\; A \cap (B \cup B^c) \;=\; (A \cap B) \cup (A \cap B^c)\,. \tag{1.10}$$

I due eventi $(A \cap B)$ e $(A \cap B^c)$ sono inoltre disgiunti, pertanto dalla (1.10) segue:

$$P(A) \;=\; P(A \cap B) + P(A \cap B^c)\,. \tag{1.11}$$

da cui la tesi: $P(A \cap B^c) = P(A) - P(A \cap B)$. □
Si noti in particolare che, se $B \subseteq A$, allora $P(A \cap B) = P(B)$ e quindi $P(A \setminus B) = P(A) - P(B)$.

Teorema 1.30 (delle probabilità totali) Siano A, B due eventi. Allora risulta:

$$P(A \cup B) \;=\; P(A) + P(B) - P(A \cap B)\,. \tag{1.12}$$

Dimostrazione. Poichè risulta:

$$A \cup B \;=\; \Omega \cap (A \cup B) \;=\; (A \cup A^c) \cap (A \cup B) \;=\; A \cup (A^c \cap B)$$

e quindi, essendo gli eventi a secondo membro a due a due incompatibili, segue:

$$P(A \cup B) \;=\; P[A \cup (A^c \cap B)] \;=\; P(A) + P(A^c \cap B)\,. \tag{1.13}$$

Dalla (1.11), scambiando gli eventi A e B, segue $P(A^c \cap B) = P(B) - P(A \cap B)$ e quindi sostituendo nella (1.13), si ha la tesi.

$$P(A \cup B) \;=\; P(A) + P(A^c \cap B) \;=\; P(A) + P(B) - P(A \cap B)\ .$$

□

Si noti che, dal teorema delle probabilità totali, si ottiene immediatamente la relazione:

$$P(A \cap B) \;=\; P(A) + P(B) - P(A \cup B)\ .$$

Dal teorema delle probabilità totali segue anche $P(A \cup B) \leq P(A) + P(B)$; più in generale si ha:

$$P\left(\bigcup_n A_n\right) \;\leq\; \sum_n P(A_n)\ .$$

In particolare, se A e B sono eventi incompatibili, dalla (1.12), segue:

$$P(A \cup B) \;=\; P(A) + P(B)\ ; \tag{1.14}$$

più in generale, se $A_1, \ldots, A_n$ è un insieme di eventi a due a due incompatibili, allora dalla (1.14) segue immediatamente:

$$P\left(\bigcup_{i=1}^{n} A_i\right) \;=\; \sum_{i=1}^{n} P(A_i)\ . \tag{1.15}$$

Esercizio 1.31 La percentuale degli studenti che hanno superato gli esami A, B e C sono stati le seguenti: A: 50%, B:40%, C:30%; A e B: 35%, A e C: 25%, B e C: 20%; infine il 15% degli studenti ha superato tutti e tre gli esami.

Soluzione. La probabilità che uno studente abbia superato almeno un esame è data da:

$$\begin{aligned} P(A \cup B \cup C) \;&=\; P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad -P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &=\; 0,5 + 0,4 + 0,3 - 0,35 - 0,25 - 0,2 + 0,15 \;=\; 0,55\ . \end{aligned}$$

Calcolare la probabilità che uno studente abbia superato almeno due esami. ♣

Teorema 1.32 (proprietà di monotonia) Siano due eventi A, B con $B \subseteq A$. Allora risulta $P(B) \leq P(A)$.

Dimostrazione. In base al Teorema 1.29, si ha $P(A \backslash B) = P(A \cap B^c) = P(A) - P(A \cap B) \geq 0$. Poichè $B \subseteq A$, allora $P(A \cap B) = P(B)$ e quindi $P(A \setminus B) = P(A) - P(B) \geq 0$ da cui $P(B) \leq P(A)$. □

Esercizio 1.33 Dati n eventi arbitrari $A_1, A_2, \ldots, A_n$, dimostrare per induzione la seguente disuguaglianza (di Bonferroni):

$$P(A_1, A_2, \ldots, A_n) \;\geq\; \sum_{i=1}^{n} P(A_i) - (n-1) \;=\; 1 - \sum_{i=1}^{n} P(A_i^c)\ .$$

Soluzione. Dimostriamo la disuguaglianza dapprima nel caso $n = 2$; dal teorema delle probabilità totali, si ha:

$$P(A_1, A_2) \;=\; P(A_1) + P(A_2) - P(A_1 \cup A_2) \;\geq\; P(A_1) + P(A_2) - 1 \; .$$

Inoltre possiamo scrivere (aggiungendo e togliendo 1):

$$P(A_1) + P(A_2) - 1 - 1 + 1 \;=\; 1 - [1 - P(A_1) + 1 - P(A_2)] \;=\; 1 - P(A_1^c) - P(A_2^c) \; .$$

Ora assumiamo valida la disuguaglianza per $n = k - 1$ e dimostriamola per $n = k$. Essendo:

$$\begin{aligned} P(A_1, \ldots, A_{k-1}, A_k) &= P(A_1, \ldots, A_{k-1}) + P(A_k) - P[(A_1, \ldots, A_{k-1}) \cup A_k] \\ &\geq P(A_1, \ldots, A_{k-1}) + P(A_k) - 1 \end{aligned} \tag{1.16}$$

e, poichè la disuguaglianza si assume vera per $n = k - 1$, risulta:

$$P(A_1, \ldots, A_{k-1}) \;\geq\; \sum_{i=1}^{k-1} P(A_i) - (k - 1 - 1)$$

e quindi sostituendo nella (1.16), segue:

$$P(A_1, \ldots, A_{k-1}, A_k) \;\geq\; \sum_{i=1}^{k-1} P(A_i) - (k - 2) + P(A_k) - 1 \;=\; \sum_{i=1}^{k} P(A_i) - (k - 1) \; .$$

La seconda relazione segue come nel caso precedente:

$$\begin{aligned} \sum_{i=1}^{k} P(A_i) - (k - 1) &= \sum_{i=1}^{k} P(A_i) - (k - 1) - 1 + 1 \;=\; 1 - \sum_{i=1}^{k} [1 - P(A_i)] \\ &= 1 - \sum_{i=1}^{k} P(A_i^c) \; . \end{aligned}$$

□

Esercizio 1.34 Dati tre eventi incompatibili A, B e C determinare l'insieme delle coppie (α, β) tali che le valutazioni di probabilità $P(A) = \alpha$, $P(B) = \beta$ e $P(C) = 0.2$ siano coerenti.

Soluzione. Essendo gli eventi incompatibili, in base alle ipotesi, risulta:

$$0 \;\leq\; P\{A \cup B \cup C\} \;=\; P(A) + P(B) + P(C) \;=\; \alpha + \beta + 0,2 \;\leq\; 1 \; .$$

Pertanto l'insieme I delle valutazioni ammissibili (α, β) è dato da:

$$I \;=\; \{(\alpha, \beta) \;:\; 0 \leq \alpha + \beta \leq 0,8\} \; .$$

♣

Esercizio 1.35 Si supponga che un dado sia costruito in modo tale che la probabilità di avere un numero sia proporzionale al numero stesso. Calcolare:

1. La probabilità di ogni singolo numero;
2. la probabilità di avere un numero pari e quella di avere un numero superiore a 4.

Soluzione. Indichiamo con A_i l'evento "esce il numero i", per $i = 1, \ldots, 6$. Denotato con p la probabilità che esca la faccia A_1, cioè $P(A_1) = p$, segue:

$$P(A_2) \;=\; 2p, \qquad P(A_3) \;=\; 3p, \qquad P(A_4) \;=\; 4p, \qquad P(A_5) \;=\; 5p, \qquad P(A_6) \;=\; 6p.$$

Essendo gli eventi $A_1, \ldots, A_6$ disgiunti, per la terza condizione della definizione di distribuzione di probabilità, deve risultare:

$$P(A_1 \cup \cdots \cup A_6) \;\;=\;\; P(A_1) + \cdots + P(A_6) \;=\; 1$$

cioè

$$p + 2p + 3p + 4p + 5p + 6p \;\;=\;\; 21p \;=\; 1$$

da cui $p = 1/21$. Segue pertanto:

$$P(A_2) \;=\; 2/21, \quad P(A_3) \;=\; 3/21, \quad P(A_4) \;=\; 4/21, \quad P(A_5) \;=\; 5/21, \quad P(A_6) \;=\; 6/21$$

e quindi $P(A_2 \cup A_4 \cup A_6) = 12/21$ e $P(A_5 \cup A_6) = 11/21$. ♣

Esercizio 1.36 Dato un esperimento tale che: $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ e $P(A \cap B) = \frac{1}{4}$, calcolare:

$$\begin{array}{ccc} P(A \cup B) & P(A^c \cap B) & P(A^c \cup B) \\ P(A \cap B^c) & P(A^c \cap B^c) & P(A^c \cup B^c) \end{array}$$

Soluzione. Da applicazioni delle relazioni trovate in precedenza, segue:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{4} = \frac{7}{12} \\ P(A^c \cap B) &= P(B \setminus A) = P(B) - P(A \cap B) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \\ P(A^c \cup B) &= P(A^c) + P(B) - P(A^c \cap B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{12} = \frac{3}{4} \\ P(A \cap B^c) &= P(A \setminus B) = P(A) - P(A \cap B) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4} \\ P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) = \frac{5}{12} \\ P(A^c \cup B^c) &= P((A \cap B)^c) = 1 - P(A \cap B) = \frac{3}{4} \end{aligned}$$

♣

1.4 Valutazioni combinatorie della probabilità

Sia Ω un insieme finito di n eventi elementari, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. Essendo gli eventi ω_i a due a due disgiunti, risulta ovviamente:

$$\begin{aligned} P(\Omega) &= P(\{\omega_1\} \cup \{\omega_2\} \cup \cdots \cup \{\omega_n\}) \\ &= P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots + P(\{\omega_n\}) \;=\; 1\,. \end{aligned} \tag{1.17}$$

Se esistono ragioni di "simmetria" che inducono a ritenere tali eventi *equiprobabili*, cioè:

$$P(\{\omega_i\}) \;=\; p \qquad\qquad i = 1, \ldots, n$$

allora, dalla (1.17), si ottiene:

$$P(\{\omega_1\}) + P(\{\omega_2\}) + \cdots + P(\{\omega_n\}) \;=\; \underbrace{p + p + \cdots + p}_{n} \;=\; np \;=\; 1$$

e quindi

$$p \;=\; \frac{1}{n}\,.$$

Pertano alla probabilità p di ciascuno di essi, deve necessariamente attribuirsi, in base alla (1.9), il valore $p = 1/n$. Ne segue che se un evento A è l'unione di h degli eventi ω_i:

$$A \;=\; \{\omega_{i_1}\} \cup \{\omega_{i_2}\} \cup \cdots \cup \{\omega_{i_h}\}$$

allora la sua probabilità è:

$$\begin{aligned} P(A) &= P(\{\omega_{i_1}\} \cup \{\omega_{i_2}\} \cup \cdots \cup \{\omega_{i_h}\}) \;=\; P(\{\omega_{i_1}\}) + P(\{\omega_{i_2}\}) + \cdots + P(\{\omega_{i_h}\}) \\ &= \underbrace{\frac{1}{n} + \cdots + \frac{1}{n}}_{h} \;=\; \frac{h}{n}\,, \end{aligned}$$

cioè è uguale al rapporto fra il "numero di casi favorevoli" ed il "numero di casi possibili".

In questo modo, una valutazione *quantitativa* della probabilità si ottiene in base ad una semplice considerazione *qualitativa*: il giudizio di equiprobabilità degli eventi $A_1, A_2, \ldots, A_n$.

Il calcolo del numero di casi favorevoli e di quello dei casi possibili in certi casi è quasi immediato, attraverso ragionamenti di carattere elementare. In molte altre circostanze, invece, il calcolo di tali quantità richiede nozioni di calcolo combinatorio. Nel seguito forniamo pertanto alcune elementari nozioni del calcolo combinatorio.

Definizione 1.37 Sia $S = \{a_1, \ldots, a_n\}$ un insieme di n elementi distinti. Si dice *disposizione di classe k* estratta da S *con ripetizione* una qualunque k-upla ordinata:

$$(a_{i_1}, a_{i_2}, \ldots, a_{i_k})$$

di elementi non necessariamente distinti di S, dove $a_{i_1}, \ldots, a_{i_k} \in S$.

Pertanto due disposizioni di classe k estratte da uno stesso insieme S sono uguali quando in essa figurano gli *stessi elementi* di S nello *stesso ordine*.

Esempio 1.38 Una colonna della schedina del Totocalcio è una disposizione di classe k estratta da $S = \{1, X, 2\}$ (e quindi $n = 3$) con $k = 13$. ♣

Teorema 1.39 Il numero delle disposizioni di classe k da un insieme di n elementi è n^k.

Dimostrazione. Per formare una di queste disposizioni, ciascuno dei k elementi che la compongono può essere scelto in n modi diversi. Perciò si hanno complessivamente $n \cdot n \cdots n = n^k$ modi diversi di costruire tali disposizioni. □

Consideriamo ora una particolare disposizione di classe k da S con la condizione che i suoi elementi (estratti da S) siano tutti *distinti*. In questo caso si parla di *disposizioni semplici* (mentre nel caso precedente, come già evidenziato, si parla di disposizioni con ripetizione).

Teorema 1.40 Il numero delle disposizioni semplici di classe k da un insieme di n elementi (con $k \leq n$) è uguale a:

$$(n)_k \;=\; n(n-1)(n-2)\cdots(n-k+1) \;=\; \frac{n!}{(n-k)!}\,. \tag{1.18}$$

Dimostrazione. A differenza del caso precedente, nel formare una di queste disposizioni, si può scegliere in n modi diversi soltanto il primo dei k elementi che la compongono. Per la scelta del secondo elemento si hanno solamente $n-1$ possibilità; e così via. Per il k-esimo si hanno a disposizione $n-(k-1) = n-k+1$ possibilità. □

Nota 1.41 La scrittura $n!$ si legge "n fattoriale" e denota il prodotto dei primi n numeri naturali:

$$n! \;=\; n(n-1)(n-2)\cdots 2\cdot 1\,.$$

In particolare si pone $0! = 1$.

Esercizio 1.42 Un ufficio dispone di 10 sportelli che forniscono servizi diversi, e arrivano contemporaneamente 7 clienti che si distribuiscono fra i vari sportelli. Definita "coda" un insieme di almeno due clienti allo stesso sportello, quante sono le possibili distribuzioni dei clienti che non danno luogo al formarsi di code? Considerando poi equiprobabili tutte le possibili distribuzioni dei clienti fra i vari sportelli, calcolare la probabilità p che non si formino code.

Soluzione. Il problema viene affrontato utilizzando gli strumenti del calcolo combinatorio. In base alla definizione data, affinchè non vi siano code, è necessario che i 7 clienti si distribuiscano fra 7 dei 10 sportelli presenti nell'ufficio. Pertanto il numero di distribuzioni dei clienti

fra i 10 sportelli in maniera tale che non vi siano code è dato dal numero di disposizioni semplici di classe 7 da un insieme di 10 elementi, cioè:

$$(10)_7 = \frac{10!}{(10-7)!} = 604800\,.$$

Assumendo successivamente l'equidistribuzione fra tutte queste distribuzioni, la probabilità che non vi siano code viene valutata in base al rapporto fra in numero di casi favorevoli (cioè il numero di distribuzioni che non comportano code) ed il numero di distribuzioni possibili, che è dato dal numero di tutte le disposizioni (con possibili ripetizione) di classe 7 da un insieme di 10 elementi, che sono 10^7. Pertanto la probabilità richiesta è data da:

$$\frac{604800}{10^7} = 0.0605\,. \tag{1.19}$$

♣

Definizione 1.43 Dato un insieme S di n elementi, si chiama *permutazione* una disposizione semplice di classe n.

Corollario 1.44 Il numero delle permutazioni di un insieme di n elementi è:

$$(n)_n \;=\; \frac{n!}{0!} \;=\; n!\,. \tag{1.20}$$

Esempio 1.45 Una parola di lunghezza k è una qualsiasi successione ordinata di lettere dell'alfabeto. Se l'alfabeto ha n lettere, il numero di parole di k lettere è n^k. In particolare, quelle senza lettere ripetute sono $(n)_k$. Gli anagrammi di parole con n lettere distinte (per esempio, con $n = 4$, "rima", "mari", "rami", "armi") sono particolari permutazioni. ♣

Dato un insieme S di n elementi, finora abbiamo considerato gruppi ordinati di elementi estratti (una o più volte) da S. Consideriamo ora gruppi non ordinati di elementi di S, non necessariamente distinti.

Definizione 1.46 Sia S un insieme di n elementi distinti. Si chiama *combinazione di classe k* estratta da S un qualunque gruppo

$$\{a_1, a_2, \ldots, a_k\}$$

non ordinato di k elementi estratti da S. Diremo che un dato elemento di S ha *molteplicità* m se esso figura m volte nella combinazione.

Pertanto due combinazioni di classe k estratte da uno stesso insieme S sono uguali quando in essa figurano gli stessi elementi di S con la stessa molteplicità.

Teorema 1.47 Il numero delle combinazioni semplici di classe k estratte da un insieme di n elementi (con $k \leq n$) è

$$\binom{n}{k} \;=\; \frac{(n)_k}{(k)_k} \;=\; \frac{n!}{k!(n-k)!} \;=\; \frac{n(n-1)\cdots(n-k+1)}{k!} \tag{1.21}$$

Dimostrazione. Il numero delle disposizioni semplici di classe k di un insieme di n elementi è $(n)_k$: ciascuna di esse costituisce una k-upla ordinata, e considerando tutte le sue permutazioni si ottengono ancora delle disposizioni semplici di classe k, in numero di $k!$ e formate con gli stessi elementi. Esse individuano quindi lo stesso insieme, cioè la stessa combinazione semplice di classe k. Cioè, se consideriamo 2 disposizioni costituite dagli stessi elementi, allora esse differiranno solo per l'ordinamento. In altre parole, ad ogni combinazione corrispondono $k!$ disposizioni semplici, e quindi il numero complessivo $(n)_k$ di queste ultime si ottiene moltiplicando $k!$ per il numero delle combinazioni, e da ciò segue la (1.21). □

Teorema 1.48 Sia assegnata un'urna contenente N palline, di cui r bianche e $N - r$ nere. Se ne estraggano n (simultaneamente o successivamente, ma senza restituzione), supponendo equiprobabili tutte le combinazioni semplici di classe n delle N palline. Allora la probabilità che nella n-upla estratta vi siano k palline bianche è:

$$\frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} . \tag{1.22}$$

Dimostrazione. Una n-upla "favorevole" all'evento "k palline bianche fra le n estratte" si ottiene mettendo k palline, scelte fra le r bianche, insieme ad $n - k$ palline scelte fra le $N - r$ nere. Ogni possibile scelta del primo tipo può essere accoppiata con una qualunque scelta del secondo tipo, e quindi la probabilità richiesta è la (1.22). □

Considerata una popolazione di n elementi, il numero in cui la popolazione può essere suddivisa in k sottoinsiemi contenenti rispettivamente $r_1, r_2, \ldots, r_k$ elementi, con $r_1 + r_2 + \cdots + r_k = n$ e $0 \leq r_i \leq n$ è dato da

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \cdots r_k!} .$$

Tale quantità viene chiamate *coefficiente multinomiale*.

Esercizio 1.49 Calcoliamo in quanti modi in cui quattro giocatori di bridge possono ricevere ciascuno 13 carte delle 52 del mazzo. In termini di palline, il gioco del bridge consiste di 52 palline che devono essere poste in 4 scatole tali che in ogni scatola vi siano esattamente 13 palline. Si ha quindi:

$$\binom{52}{13, 13, 13, 13} = \frac{52!}{13!13!13!13!} \approx 2.08 \cdot 10^{48}.$$

Un altro esempio è dato dal numero di anagrammi della parola "matematica" che risulta:

$$\binom{10}{3, 1, 1, 2, 1, 2} = \frac{10!}{3!1!1!2!1!} = 151200 .$$

♣

Esercizio 1.50 Gli archivi di un ufficio sono stati memorizzati in 30 dischi da 100 Mbyte, contenenti ognuno 100 file. Un programma dovrà accedere a 28 di questi file (tutti diversi). Qual è la probabilità che esso non debba usare il primo disco?

Soluzione. La situazione in esame si modellizza con uno schema successo-insuccesso senza rimpiazzo: sono 20 prove ripetute in ognuna delle quali viene scelto un file fra i 3000 possibili. Consiedereremo "successo" la scelta di un file appartenente al primo disco (sono 100) e insuccesso la scelta di uno degli altri dischi (2900 files). Non c'è restituzione perchè un file, una volta letto, non verrà più richiamato in seguito. La probabilità richiesta è dunque

$$p \;=\; \frac{\binom{100}{0}\binom{2900}{28}}{\binom{3000}{28}} \;=\; 0,385\;.$$

♣

Esercizio 1.51 Una scatola contiene m palline numerate progressivamente, con $m \geq 5$.

1. Vengono estratte in blocco 5 palline. Qual è la probabilità α che venga estratta la pallina numero 1?

2. Vengono estratte 5 palline ogni volta con reimmissione. Qual è la probabilità β che venga estratta la pallina numero 1?

Soluzione. Indicato con A l'evento di interesse, calcoliamo la probabilità dell'evento complementare, che è più semplice. Infatti l'evento "non esce la pallina n. 1" è composto dalle cinquine che non contengono 1 e sono formate senza reimmissione, essendo l'estrazione fatta in blocco. In quanti modi posso formare una tale cinquina cioè un evento favorevole a A^c? La prima pallina è una qualsiasi delle $m-1$ (escluso 1); la seconda è scelta tra $m-2$ (cioè $m-1$ meno la pallina estratta); la terza fra $m-3$, la quarta fra $m-4$ e l'ultima fra $m-5$. In altre parole,il numero di esiti favorevoli all'evento A^c è uguale al numero di disposizioni semplici di $m-1$ oggetti presi a 5 a 5:

$$(m-1)_5 \;=\; (m-1)(m-2)(m-3)(m-4)(m-5)$$

Il numero di esiti possibili è dato dal numero di cinquine senza reimmissione, cioè al numero di disposizioni di m oggetti presi a 5 a 5:

$$(m)_5 \;=\; m(m-1)(m-2)(m-3)(m-4)\;.$$

Pertanto:

$$P(A) \;=\; 1-P(A^c) \;=\; 1-\frac{(m-1)_5}{(m)_5} \;=\; 1-\frac{m-5}{m} \;=\; \frac{5}{m}\;.$$

Per quanto riguarda il secondo punto, è anche in questo caso più semplice calcolare la probabilità dell'evento complementare. Il numero di eventi favorevoli è – nel caso di estrazioni con reimmissione – uguale al numero di disposizioni con ripetizione di $m-1$ elementi presi a 5 a 5, che è uguale a $(m-1)^5$, mentre il numero di esiti possibili è uguale al numero di disposizioni con ripetizione di m elementi presi a 5 a 5, che è uguale a m^5. Si ha pertanto:

$$P(B) \;=\; 1 - P(B^c) \;=\; 1 - \frac{(m-1)^5}{m^5} \;=\; 1 - \left(1 - \frac{1}{m}\right)^5 .$$

♣

Esercizio 1.52 Un giocatore di poker riceve all'inizio del gioco cinque carte da un normale mazzo di 52.

1. Qual è la probabilità di ricevere almeno 2 assi?

2. Qual è la probabilità di ricevere cinque carte dello stesso seme?

Per quanto riguarda il primo punto, possiamo considerare le 52 carte del mazzo sudivise in due gruppi, uno composto dai 4 assi e l'altro dalle restanti 48 carte. Indichiamo con A_k l'evento "il giocatore riceve k assi fra le sue 5 carte", con $k = 1, 2, 3, 4$. Il problema può essere considerato come un problema di estrazioni di 5 elementi da un'urna contenente 4 elementi di un tipo e 48 dell'altro tipo. In base a quanto visto in precedenza si ha:

$$P(A_k) \;=\; \frac{\binom{4}{k}\binom{48}{5-k}}{\binom{52}{5}} .$$

da cui: $P(A_2) = 0.04$, $P(A_3) = 0.0017$ e $P(A_4) = 1.847 \cdot 10^{-5}$.

L'evento "il giocatore riceve almeno due assi" è dato da $A_2 \cup A_3 \cup A_4$ e poichè gli eventi A_2, A_3 e A_4 sono a due a due disgiunti, ne segue $P(A_2 \cup A_3 \cup A_4) = P(A_2) + P(A_3) + P(A_4) = 0,042$.

Per quanto riguarda il secondo punto, indichiamo con $A_\diamondsuit$ l'evento "il giocatore riceve 5 carte di quadri", con $A_\heartsuit$ l'evento "il giocatore riceve 5 carte di cuori", con $A_\clubsuit$ l'evento "il giocatore riceve 5 carte di picche" e con $A_\spadesuit$ l'evento "il giocatore riceve 5 carte di picche". L'evento "il giocatore riceve 5 carte dello stesso seme" è pertanto $A_\diamondsuit \cup A_\heartsuit \cup A_\clubsuit \cup A_\spadesuit$.

Assumendo che le carte siano equiprobabili, ne segue che i 4 eventi $A_\diamondsuit, A_\heartsuit, A_\clubsuit$ e $A_\spadesuit$ sono equiprobabili, cioè $P(A_\diamondsuit) = P(A_\heartsuit) = P(A_\clubsuit) = P(A_\spadesuit)$. Segue pertanto $P(A_\diamondsuit \cup A_\heartsuit \cup A_\clubsuit \cup A_\spadesuit) = P(A_\diamondsuit + A_\heartsuit + A_\clubsuit + A_\spadesuit)$.

Considerato ad esempio l'evento $A_\diamondsuit$ come riferimento, si ha: $P(A_\diamondsuit \cup A_\heartsuit \cup A_\clubsuit \cup A_\spadesuit) = 4P(A_\diamondsuit)$. Calcoliamo quindi $P(A_\diamondsuit)$. Lo schema è analogo al precedente, in questo caso l'urna

di 52 elementi si considera suddivisa in un insieme di 13 elementi di un tipo (successi) e 39 elementi dell'altro (insuccessi):

$$P(A_\diamondsuit) \;=\; \frac{\binom{13}{5}\binom{39}{0}}{\binom{52}{5}} \;=\; 4.95\cdot 10^{-4}\;,$$

e quindi $P(A_\diamondsuit \cup A_\heartsuit \cup A_\clubsuit \cup A_\spadesuit) = P(A_\diamondsuit + A_\heartsuit + A_\clubsuit + A_\spadesuit) = 4P(A_\diamondsuit) = 4 \cdot 4,95 \cdot 10^{-4} = 0,198\%$. ♣

1.5 Probabilità su eventi condizionati

Dato un evento E, ad esso corrisponde, come è noto, l'alternativa "VERO o FALSO"; spesso è necessario o interessa considerare certe situazioni solo subordinatamente all'eventuale verificarsi di una data circostanza, la quale può essere espressa a sua volta mediante un evento H. Anzi, in realtà, è questo il punto di vista che andrebbe *sempre* adottato: infatti, anche quando si lancia una moneta e si considera l'evento T(= "testa") o C (= "croce"), si dà tacitamente per scontato che sia sempre vero un altro evento:

$$H \;=\; \text{"la moneta mostra una delle due facce"}\;,$$

cioè si identifica H con l'evento Ω (la moneta potrebbe infatti rotolare in un tombino, oppure restare "in piedi" appoggiata su una parete: anche se è difficile, non è però impossibile – né logicamente né praticamente – che ciò si verifichi).

Definizione 1.53 Siano E e H due eventi di Ω. Si definisce *evento condizionato* $E|H$ (da leggersi "E dato H", oppure "E condizionato a H") l'evento tale che

- $E|H$ risulta VERO se $E \cap H$ è vero
  (cioè quando essendo vero H, risulta vero anche E),
- $E|H$ risulta FALSO se $E^c \cap H$ è vero
  (cioè quando, essendo vero H, risulta falso E),
- $E|H$ risulta INDETERMINATO se H è falso
  (cioè quando H risulta falso e quindi non interessa, una volta acquisita tale informazione, sapere se E sia vero o falso).

Si noti che, a differenza dell'evento semplice (vedi Definizione 1.1), l'evento condizionato è un ente logico a 3 valori: vero, falso, indeterminato.

Per chiarire il significato della condizione $H \neq \emptyset$, si noti che essa equivale (in base alla definizione di evento impossibile) che "non è noto che H sia falso": in base ad una nuova informazione, H può successivamente risultare vero o falso. Si noti inoltre che per $H = \Omega$ si ritrova

il concetto di *evento* non condizionato, cioè $E|\Omega = E$. Infatti l'alternativa INDETERMINATO non può presentarsi (poichè Ω è sempre vero), e negli altri due casi la condizione "essendo vero H" è automaticamente soddisfatta. Si noti che l'evento $E|H$ si può scrivere come $(E \cap H)|H$; tale forma si chiama *forma ridotta*.

Nota 1.54 Si noti che, in base alla definizione di evento impossibile, la condizione $H \neq \emptyset$ equivale ad affermare che "non è noto che l'evento H sia impossibile"; ovviamente, in seguito ad ulteriori informazioni, l'evento H può risultare vero o falso. Inoltre, per $H = \Omega$, si ritrova il concetto di evento (non condizionato), cioè $E|\Omega = E$

Come ad un evento E si associa in modo naturale una scommessa (vinta quando E è vero e persa quando E è falso), ad un evento condizionato $E|H$ si può associare un tipo diverso di scommessa, da considerarsi *annullata* se H non si verifica (e vinta se si verificano H e E, persa se si verifica H ma non E). Precisamente il concetto di *scommessa condizionata* su $E|H$, di quota p, generalizza il concetto di scommessa introdotto precedentemente: versando una somma pS (con $S \neq 0$ arbitrario), si riceve un importo S se si verificano H e E, niente se si verifica H ma non E (perdendo così la puntata pS), oppure viene restituita la somma pS se H non si verifica (scommessa annullata).

Anche per gli eventi condizionati è possibile introdurre un concetto al quale attribuire il significato di "grado di fiducia" nel verificarsi di E, *supposto vero* H, per ogni evento E della famiglia $\mathcal{A}$ e quindi un opportuno spazio di probabilità su Ω. Sia $\mathcal{A}$ una famiglia di eventi e sia $H \neq \emptyset$ un evento possibile. Per ogni $H \neq \emptyset$, affinchè una scommessa condizionata su ogni $E|H$ sia *coerente*, richiediamo che $P(\cdot|H)$ sia una probabilità su $\mathcal{A}$, cioè che $P(\cdot|H)$ verifichi le tre condizioni della Definizione 1.27 di distribuzione di probabilità.

Nota 1.55 Si noti che l'evento $E|H$ è un oggetto unico, pertanto $P(E|H)$ è la probabilità dell'evento "E condizionato al verificarsi dell'evento H" (e non, come a volte si dice, la probabilità dell'evento E dopo aver appreso che l'evento H è vero).

Assegnati gli eventi E, H e $E|H$, le probabilità di $E \cap H$, H e $E|H$ sono legate dal seguente risultato (di cui si omette la dimostrazione).

Teorema 1.56 (delle probabilità composte) Siano E, H due eventi e si consideri l'evento condizionato $E|H$. Allora risulta:

$$P(E \cap H) \;=\; P(H)\, P(E|H) \;=\; P(E)P(H|E)\;. \tag{1.23}$$

Nell'impostazione assiomatica, assegnato uno spazio di probabilità $(\Omega, \mathcal{A}, P)$, e considerato $H \in \mathcal{A}$ tale che $P(H) > 0$, si definisce probabilità condizionata la quantità:

$$P(E|H) \;=\; \frac{P(E \cap H)}{P(H)}\;. \tag{1.24}$$

Nota 1.57 Si noti che la (1.28) viene spesso assunta quale definizione di probabilità condizionata; in altre parole si definisce la probabilità condizionata a partire dalla (1.28), nell'ipotesi $P(H) > 0$, e si ottiene (banalmente!) il teorema delle probabilità composte come conseguenza. Nell'impostazione soggettiva, l'evento $E|H$ viene assegnato alla sola condizione che $H \neq \emptyset$ e si ricava la (1.27) come conseguenza, che resta valida anche nel caso in cui risulta $P(H) = 0$ (o $P(E) = 0$).

Nel caso generale, la *probabilità condizionale* di H dati $A_1, \ldots, A_n$ si scrive $P(H|A_1, \ldots, A_n)$ ed è data da:

$$P(H|A_1, \ldots, A_n) \equiv P(H|A_1 \cap \ldots \cap A_n) = \frac{P(H \cap A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_n)} . \quad (1.25)$$

Se $P(A_1, \cdots, A_n) > 0$, in base a semplici operazioni algebriche si ottiene:

$$\begin{aligned} P(A_1, \cdots, A_n) &= P(A_1)\frac{P(A_1, A_2)}{P(A_1)}\frac{P(A_1, A_2, A_3)}{P(A_1, A_2)} \cdots \frac{P(A_1, A_2, \ldots, A_n)}{P(A_1, A_2, \ldots, A_{n-1})} \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1}) . \end{aligned} \quad (1.26)$$

Sia assegnato uno spazio di probabilità $(\Omega, \mathcal{A}, P)$ e sia H un evento di $\mathcal{A}$, assumendo per semplicità $P(H) > 0$. Le tre condizioni della Definizione 1.27 possono essere dimostrate immediatamente:

1. $0 \leq P(A|H) \leq 1$. Sia $A \in \mathcal{A}$, poichè $A \cap H \subseteq H$, allora $0 \leq P(A \cap H) \leq P(H)$ e quindi

$$P(A|H) = \frac{P(A \cap H)}{P(H)} \leq \frac{P(H)}{P(H)} \leq 1 .$$

2. $P(\emptyset|H) = 0$ e $P(\Omega|H) = 1$. Si ha infatti:

$$\begin{aligned} P(\emptyset|H) &= \frac{P(\emptyset \cap H)}{P(H)} = \frac{P(\emptyset)}{P(H)} = 0 , \\ P(\Omega|H) &= \frac{P(\Omega \cap H)}{P(H)} = \frac{P(\Omega)}{P(H)} = 1 . \end{aligned}$$

3. Per ogni partizione $\{A_1, \ldots, A_n, \ldots, \}$ di eventi inclusi in $\mathcal{A}$ deve risultare:

$$P(A_1|H) + P(A_2|H) + \cdots + P(A_n|H) + \cdots = 1 .$$

Poichè $A_1, A_2, \ldots, A_n \ldots$ costituisce una partizione (cioè risulta $A_i \cap A_j = \emptyset$ per $i \neq j$ e $\bigcup_{n=1}^{\infty} A_n = \Omega$) segue:

$$\begin{aligned} P(A_1|H) + \cdots + P(A_n|H) + \cdots &= \frac{P(A_1 \cap H)}{P(H)} + \cdots + \frac{P(A_n \cap H)}{P(H)} + \cdots \\ &= \frac{P[(A_1 \cap H) \cup \cdots \cup (A_n \cap H) \cup \cdots]}{P(H)} \\ &= \frac{P[(A_1 \cup \cdots \cup A_n \cup \cdots) \cap H]}{P(H)} \\ &= \frac{P(H)}{P(H)} = 1 . \end{aligned}$$

La quantità $P(E|H)$ viene chiamata *probabilità condizionale* di E dato H.

Nota 1.58 Si noti che l'evento $E|H$ è un oggetto unico, pertanto $P(E|H)$ è la probabilità dell'evento "E condizionato al verificarsi dell'evento H" (e non, come a volte si dice, la probabilità dell'evento E dopo aver appreso che l'evento H è vero).

Assegnati gli eventi E, H e $E|H$, le probabilità di $E \cap H$, H e $E|H$ sono legate dal seguente risultato (di cui si omette la dimostrazione).

Teorema 1.59 (delle probabilità composte) Siano E, H due eventi e si consideri l'evento condizionato $E|H$. Allora risulta:

$$P(E \cap H) \;=\; P(H)\, P(E|H) \;=\; P(E)P(H|E)\,. \tag{1.27}$$

Nell'impostazione assiomatica, assegnato uno spazio di probabilità $(\Omega, \mathcal{A}, P)$, e considerato $H \in \mathcal{A}$ tale che $P(H) > 0$, si definisce probabilità condizionata la quantità:

$$P(E|H) \;=\; \frac{P(E \cap H)}{P(H)}\,. \tag{1.28}$$

Nota 1.60 Si noti che la (1.28) viene spesso assunta quale definizione di probabilità condizionata; in altre parole si definisce la probabilità condizionata a partire dalla (1.28), nell'ipotesi $P(H) > 0$, e si ottiene (banalmente!) il teorema delle probabilità composte come conseguenza. Nell'impostazione soggettiva, l'evento $E|H$ viene assegnato alla sola condizione che $H \neq \emptyset$ e si ricava la (1.27) come conseguenza, che resta valida anche nel caso in cui risulta $P(H) = 0$ (o $P(E) = 0$).

Nel caso generale, la *probabilità condizionale* di H dati $A_1, \ldots, A_n$ si scrive $P(H|A_1, \ldots, A_n)$ ed è data da:

$$P(H|A_1, \ldots, A_n) \;\equiv\; P(H|A_1 \cap \ldots \cap A_n) \;=\; \frac{P(H \cap A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_n)}\,. \tag{1.29}$$

Se $P(A_1, \cdots, A_n) > 0$, in base a semplici operazioni algebriche si ottiene:

$$\begin{aligned} P(A_1, \cdots, A_n) &= P(A_1)\frac{P(A_1, A_2)}{P(A_1)}\frac{P(A_1, A_2, A_3)}{P(A_1, A_2)} \cdots \frac{P(A_1, A_2, \ldots, A_n)}{P(A_1, A_2, \ldots, A_{n-1})} \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1})\,. \end{aligned} \tag{1.30}$$

1.5.1 Indipendenza stocastica

Dati due eventi A, B, in generale, si ha $P(B) \neq P(B|A)$. Tuttavia sono molte le situazioni in cui le valutazioni di probabilità sono tali che l'apprendere il verificarsi o meno dell'evento A non porta alcuna ulteriore informazione circa il verificarsi di B, cioè la probabilità dell'evento B resta inalterata indipendentemente dal fatto che l'evento A sia vero o falso:

$$P(B|A) \;=\; P(B)\,.$$

Da qui nasce il concetto di indipendenza stocastica.

Definizione 1.61 Siano A, B due eventi di Ω. Diremo che l'evento B è *stocasticamente indipendente* dall'evento A se risulta:

$$P(A|B) \;=\; P(A)\;,$$

e, similmente, l'evento A è indipendente dall'evento B se risulta:

$$P(B|A) \;=\; P(B)\;.$$

In particolare, dalla (1.27), segue che due eventi A e B sono stocasticamente indipendenti se e solo se:

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

Nota 1.62 A differenza dell'indipendenza logica (vedi Definizione 1.8), l'indipendenza stocastica dipende non solo dagli eventi, ma anche dalla loro probabilità.

Più in generale, n eventi $A_1, \ldots, A_n$ si dicono stocasticamente indipendenti se

$$P(A_{i_1}, \cdots, A_{i_k}) \;=\; \prod_{j=1}^{k} P(A_{i_j}) \tag{1.31}$$

per ogni sottoinsieme $\{i_1, \ldots, i_k\}$ di $\{1, \ldots, n\}$. Si noti che la (1.31) è equivalente alla relazione

$$P(A_j|A_{i_1}, \cdots, A_{i_k}) \;=\; P(A_j) \tag{1.32}$$

per ogni scelta di indici j e $\{i_1, \ldots, i_k\}$ tale che $j \notin \{i_1, \ldots, i_k\}$.

Esempio 1.63 (Paradosso di De Méré) E' più probabile ottenere almeno una volta "la faccia 6" lanciando quattro volte un dado, oppure ottenere almeno una volta "un doppio 6" lanciando ventiquattro volte due dadi?

Questo problema viene chiamato "paradosso di De Méré" (in realtà si era già occupato di questo problema, un secolo prima, G. Cardano). De Méré fu un cultore di probabilità del XVII secolo, il quale scrisse al matematico B. Pascal per chiedergli come mai la frequenza relativa di "successo" risultasse sistematicamente" inferiore, nel secondo caso, a quelle del primo. Egli infatti riteneva che i due risultati avessero la stessa probabilità, probabilmente faceva il seguente ragionamento (sbagliato!):

$$\begin{aligned} \frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6} \;&=\; \frac{4}{6} \;=\; \frac{2}{3}\,, \\ \frac{1}{36}+\frac{1}{36}+\cdots+\frac{1}{36} \;&=\; \frac{24}{36} \;=\; \frac{2}{3}\,. \end{aligned}$$

Il ragionamento corretto conduce alla spiegazione dei risultati osservati. Nel primo caso consideriamo l'evento A_i ="esce la faccia 6 nell'i-esimo lancio" per $i = 1, 2, 3, 4$. La probabilità richiesta è:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4)\;.$$

Per la formula di De Morgan si ha $(A_1 \cup A_2 \cup A_3 \cup A_4) = (A_1^c \cap A_2^c \cap A_3^c \cap A_4^c)^c$ e quindi:

$$P(A_1 \cup A_2 \cup A_3 \cup A_4) \;=\; P[(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c)^c] \;=\; 1 - P(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c)\,.$$

Essendo inoltre gli eventi indipendenti, segue:

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4) &= 1 - P(A_1^c)P(A_2^c)P(A_3^c)P(A_4^c) \\ &= 1 - \left(1 - \frac{1}{6}\right)^4 \;=\; 1 - \left(\frac{5}{6}\right)^4 \;\approx\; 0.5177\,. \end{aligned}$$

Nel secondo caso, considerato l'evento B_i ="esce un doppio 6 nell'i-esimo lancio" per $i = 1, \ldots, 24$, con ragionamenti analoghi, si ha

$$\begin{aligned} P(B_1 \cup B_2 \cup \cdots \cup B_{24}) &= P[(B_1^c \cap B_2^c \cap \cdots \cap B_{24}^c)^c] \;=\; 1 - P(B_1^c \cap B_2^c \cap \cdots \cap B_{24}^c) \\ &= 1 - P(B_1^c)P(B_2^c)\cdots P(B_{24}^c) \;=\; 1 - \left(1 - \frac{1}{36}\right)^{24} \\ &= 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914\,. \end{aligned}$$

♣

Esercizio 1.64 Un'urna contiene 15 palline di cui 3 bianche, 5 nere e 7 rosse.

1. Si estrae una sola pallina: qual è la probabilità che sia rossa? oppure non nera?
2. Si estraggono due palline *con ripetizione*: qual è la probabilità di estrarre due palline bianche? una bianca e una rossa? due non nere? almeno una bianca?
3. Si estraggono due palline *senza ripetizione*: quali sono le probabilità degli eventi descritti al punto 2.?

In base ai dati del problema si ha $\#\Omega = 15$, $\#B = 3$, $\#N = 5$ e $\#R = 7$. Per quanto riguarda il primo punto si ha pertanto:

$$\begin{aligned} P(R) &= \frac{\#R}{\#\Omega} \;=\; \frac{7}{15} \;=\; 0.4667 \\ P(N^c) &= \frac{\#\Omega - \#N}{\#\Omega} \;=\; \frac{15-5}{15} \;=\; \frac{10}{15} \;=\; 0.6667\,. \end{aligned}$$

Per quanto riguarda il secondo punto indichiamo con B_1 l'evento "estrazione di una pallina bianca alla prima estrazione", con B_2 l'evento "estrazione di una pallina bianca alla seconda estrazione" e così via. L'estrazione *con ripetizione* dà luogo ad eventi indipendenti in quanto la composizione dell'urna rimane la stessa dopo ogni estrazione. Si ha pertanto:

$$\begin{aligned} P(B_1) &= P(B_2) \;=\; P(B) \;=\; \frac{\#B}{\#\Omega} \;=\; \frac{3}{15} \;=\; 0.2\,, \\ P(N_1) &= P(N_2) \;=\; P(N) \;=\; \frac{\#N}{\#\Omega} \;=\; \frac{5}{15} \;=\; 0.33\,, \\ P(R_1) &= P(R_2) \;=\; P(R) \;=\; \frac{\#B}{\#\Omega} \;=\; \frac{7}{15} \;=\; 0.4667\,. \end{aligned}$$

Segue che:

$$\begin{aligned}
P(B_1 \cap B_2) &= P(B_1)P(B_2) \,=\, [P(B)]^2 \,=\, \left(\frac{3}{15}\right)^2 \,=\, 0.04\,, \\
P[(B_1 \cap R_2) \cup (R_1 \cap B_2)] &= P(B_1 \cap R_2) + P(R_1 \cap B_2) \\
&= P(B_1)P(R_2) + P(R_1)P(B_2) \,=\, 2 \cdot P(B)P(R) \\
&= 2 \cdot \frac{3}{15} \cdot \frac{7}{15} \,=\, 0.1867\,, \\
P(N_1^c \cap N_2^c) &= P(N_1^c)P(N_2^c) \,=\, [P(N^c)]^2 \,=\, \left(1 - \frac{5}{15}\right)^2 \\
&= 0.4444\,,
\end{aligned}$$

e infine

$$\begin{aligned}
P[(B_1 \cap B_2^c) \cup (B_1^c \cap B_2) \cup (B_1 \cap B_2)] &= P(B_1 \cap B_2^c) + P(B_1^c \cap B_2) + P(B_1 \cap B_2) \\
&= P(B_1)P(B_2^c) + P(B_1^c)P(B_2) + P(B_1)P(B_2) \\
&= 2 \cdot P(B) \cdot P(B^c) + [P(B)]^2 \\
&= 2 \cdot \frac{3}{15}\left(1 - \frac{3}{15}\right) + \left(\frac{3}{15}\right)^2 \,=\, \frac{19}{75} \,=\, 0.2533\,.
\end{aligned}$$

Per quanto riguarda il terzo punto, l'estrazione *senza ripetizione* dà luogo ad eventi non indipendenti in quanto, dopo ogni estrazione, l'urna cambia la sua composizione. In questo caso si ha:

$$\begin{aligned}
P(B_1 \cap B_2) &= P(B_1)P(B_2|B_1) \,=\, \frac{3}{15} \cdot \frac{2}{14} \,=\, \frac{1}{5} \cdot \frac{1}{7} \,=\, \frac{1}{35} \,=\, 0.02857\,, \\
P[(B_1 \cap R_2) \cup (R_1 \cap B_2)] &= P(B_1)P(R_2|B_1) + P(R_1)P(B_2|R_1) \\
&= \frac{3}{15} \cdot \frac{7}{14} + \frac{7}{15} \cdot \frac{3}{14} \,=\, 2 \cdot \frac{21}{210} \,=\, 0.20\,, \\
P(N_1^c \cap N_2^c) &= P(N_1^c)P(N_2^c|N_1^c) \,=\, \left(1 - \frac{5}{15}\right)\left(1 - \frac{5}{14}\right) \,=\, \frac{10}{15} \cdot \frac{9}{14} \\
&= 0.4286
\end{aligned}$$

e infine

$$\begin{aligned}
P[(B_1 \cap B_2^c) \cup (B_1^c \cap B_2) \cup (B_1 \cap B_2)] &= P(B_1)P(B_2^c|B_1) + P(B_1^c)P(B_2|B_1^c) \\
&\quad + P(B_1)P(B_2|B_1) \\
&= \frac{3}{15}\left(1 - \frac{2}{14}\right) + \left(1 - \frac{3}{15}\right)\frac{3}{14} + \frac{3}{15} \cdot \frac{2}{14} \\
&= \frac{3}{15} \cdot \frac{12}{14} + \frac{12}{15} \cdot \frac{3}{14} + \frac{3}{15} \cdot \frac{2}{14} \,=\, 0.37143\,.
\end{aligned}$$

♣

Esercizio 1.65 Dato $\Omega = \{1,2,3\}$, $A = \{1\}$, $B = \{3\}$ e $C = \{2\}$, $P(A) = P(B) = \frac{1}{3}$, calcolare

$$\begin{array}{ccc} P(C) & P(A \cup B) & P(A^c) \\ P(B \cup C) & P(A^c \cap B^c) & P(A^c \cup B^c) \end{array}$$

Esercizio 1.66 I componenti prodotti da una certa ditta possono presentare due tipi di difetti, con percentuali rispettivamente del 3% e del 7%. I due tipi di difettosità possono presentarsi in momenti diversi della produzione per cui si può assumere che le presenze dell'uno o dell'altro siano indipendenti fra loro.

1. Qual è la probabilità che un componente presenti entrambi i difetti?
2. Qual è la probabilità che un componente sia difettoso (cioè presenti almeno uno dei due difetti)?
3. Qual è la probabilità che il componente presenti il primo difetto, sapendo che è difettoso?
4. Qualè la probabilità che esso presenti uno solo dei due difetti sapendo che esso è difettoso?

Indichiamo con A e B gli eventi corrispondenti rispettivamente alla presenza del primo e del secondo difetto, cioè $P(A) = 0.03$ e $P(B) = 0,07$; inoltre gli eventi A e B sono indipendenti. La probabilità che entrambi i difetti siano presenti è:

$$P(A \cap B) \;=\; P(A)P(B) \;=\; 0.03 \cdot 0.07 \;=\; 0.0021\ .$$

L'evento "almeno uno dei due difetti è presente" è $A \cup B$, per cui

$$P(A \cup B) \;=\; P(A) + P(B) + P(A \cap B) \;=\; 0.03 + 0.07 - 0.0021 \;=\; 0.0979\ .$$

L'evento "un pezzo presenta il primo difetto, sapendo che è difettoso" è $A|(A \cup B)$, e pertanto la sua probabilità è data da:

$$P(A|A \cup B) \;=\; \frac{P[A \cap (A \cup B)]}{P(A \cup B)} \;=\; \frac{P(A)}{P(A \cup B)} \;=\; \frac{0.03}{0.0979} \;=\; 0.0306 \;=\; 30.6\%\ .$$

Infine l'evento "è presente solo un difetto , essendo noto che il pezzo è difettoso" è dato dall'evento complementare a $A \cap B|A \cup B$ e quindi:

$$1 - P(A \cap B|A \cup B) \;=\; 1 - \frac{P(A \cap B)}{P(A \cup B)} \;=\; 1 - \frac{0.0021}{0.0979} \;=\; 0.978 \;=\; 97.8\%\ .$$

♣

Esercizio 1.67 (Il paradosso del concorrente) Questo giochino costituisce una versione aggiornata del ben più noto paradosso del carceriere. In un quiz a premi il concorrente viene posto di fronte a tre porte, indicate con A, B e C: dietro due di esse vi è una capra dietro l'altra vi è una Ferrari. Il concorrente effetta la scelta A e successivamente il conduttore apre una porta B dietro cui c'è la capra. Quest'informazione modifica la probabilità che dietro la porta A vi sia la Ferrari?

Soluzione. Indichiamo con A l'evento "dietro la porta A c'è la Ferrari" e con A^c l'evento "dietro la porta A c'è la capra"; analogamente per gli eventi B e C. Consideriamo l'evento:

$$B^*\text{: il conduttore apre la porta } B \text{ dietro cui c'è la capra.}$$

Il problema chiede di confrontare $P(A)$ e $P(A|B^*)$: conviene cambiare scelta se e solo se $P(A|B^*) > P(A)$.

Ovviamente, trattandosi di eventi equiprobabili si ha:

$$P(A) \;=\; P(B) \;=\; P(C) \;\;=\;\; \frac{1}{3}\,.$$

Passiamo al calcolo di $P(A|B^*)$. In base alla relazione della probabilità condizionata possiamo scrivere:

$$P(A|B^*) \;\;=\;\; \frac{P(B^*|A)P(A)}{P(B^*)}\,. \tag{1.33}$$

Essendo $P(A) = 1/3$, dobbiamo calcolare $P(B^*)$ e $P(B^*|A)$. Osserviamo che, in base alle regole del gioco, l'evento B^* si verifica in uno dei due seguenti casi: *i*) se dietro la porta A c'è la capra (e quindi dietro C c'è la Ferrari): in questo caso il conduttore – dovendo indicare una porta con la capra diversa da A – non può che segnalare la stanza B; *ii*) il concorrente ha indovinato, cioè dietro la porta A c'è la Ferrari: in tal caso il presentatore, potendo scegliere se aprire la porta B o la C, sceglie la C. In simboli:

$$B^* \;\;=\;\; C \cup (A \cap B^*)$$

da cui, trattandosi di eventi disgiunti, segue:

$$P(B^*) \;\;=\;\; P(A \cap B^*) + P(C) \;=\; P(B^*|A)P(A) + P(C)\,. \tag{1.34}$$

A questo punto la questione è la valutazione di $P(B^*|A)$, cioè la probabilità che il presentatore – potendo scegliere indifferentemente fra due porte dietro le quali ci sono le capre – preferisca la porta B. Nel caso in cui egli sceglie con la stessa probabilità una delle due porte si ha:

$$P(B^*|A) \;\;=\;\; P(C^*|A) \;=\; 1/2\,, \tag{1.35}$$

e quindi segue:

$$P(B^*) \;\;=\;\; P(B^*|A)P(A) + P(C) \;=\; \frac{1}{2}\cdot\frac{1}{3} + \frac{1}{3} \;=\; \frac{1}{2}\,.$$

Da cui sostituendo nella (1.33), si ottiene

$$P(A|B^*) \;\;=\;\; \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}} \;=\; \frac{1}{3}\,.$$

In altre parole, sotto l'ipotesi (1.35), la probabilità che dietro la porta A si nasconda la Ferrari non cambia e quindi è indifferente cambiare la scelta iniziale

Nota bene. Si noti che la questione sta nel fatto che NON si verifica l'evento "il conduttore apre a caso una delle altre due porte e vi trova una capra" ma l'evento "il conduttore apre una porta dietro cui c'è una capra".

In generale, non si può dire che l'informazione fornita dal presentatore aumenti la probabilità che dietro la porta A vi sia la Ferrari se non si conosce il criterio con il quale il presentatore sceglie la porta B nell'ipotesi (attenzione!) in cui il concorrente abbia scelto la porta con la Ferrari. Infatti, tenendo conto della (1.34), si ha:

$$P(A|B^*) \;=\; \frac{P(B^*|A)}{1+P(B^*|A)}\,, \tag{1.36}$$

per cui $P(A|B^*)$ dipende dal criterio di scelta del conduttore:

1. Se $P(B^*|A) < 1/2$ allora $P(A|B^*) < P(A)$ e quindi conviene cambiare;
2. Se $P(B^*|A) = 1/2$ allora $P(A|B^*) = P(A)$ e quindi è indifferente cambiare o meno;
3. Se $P(B^*|A) > 1/2$ allora $P(A|B^*) > P(A)$ e quindi conviene mantenere la scelta iniziale.

Si noti che, dalla (1.36), si vede che $P(A|B^*)$ varia fra 0 e 1/2 in dipendenza del criterio di scelta del presentatore, in particolare risulta $P(A|B^*) = 1/2$ per $P(B^*|A) = 1$ cioè nel caso in cui il presentatore ha un comportamento deterministico per cui, qualora il concorrente avesse scelto la porta con la Ferrari, egli apre sempre la porta B. ♣

Esercizio 1.68 Un'urna contiene N palline di cui pN bianche e $qN = (1-p)N$ nere, con $0 < p < 1$. Si effettua una prima estrazione senza restituzione E_1 di cui si ignora il risultato. Qual è la probabilità che la seconda estrazione E_2 fornisca una pallina bianca?

Soluzione. Il risultato della seconda estrazione dipende dal risultato della prima estrazione E_1 che è incognito. Le probabilità che la prima estrazione E_1 fornisca rispettivamente una pallina bianca $(E_1 = b)$ o una pallina nera $(E_1 = n)$, possono essere valutate in base al rapporto fra casi favorevoli e casi possibili, ottenendo i valori

$$\begin{aligned} P(E_1 = b) &= \frac{pN}{N} = p \\ P(E_1 = n) &= \frac{(1-p)N}{N} = 1-p\,. \end{aligned}$$

La probabilità che la seconda estrazione fornisca una pallina bianca è allora data da:

$$\begin{aligned} P(E_2 = b) &= P(\{E_1 = b, E_2 = b\} \cup \{E_1 = n, E_2 = b\}) \\ &= P(E_1 = b, E_2 = b) + P(E_1 = n, E_2 = b) \\ &= P(E_1 = b)P(E_2 = b|E_1 = b) + P(E_1 = n)P(E_2 = b|E_1 = n)\,. \end{aligned} \tag{1.37}$$

Se la prima estrazione ha fornito una pallina bianca allora, prima della seconda estrazione E_2, l'urna contiene $N-1$ palline di cui $pN-1$ bianche; in questo caso la probabilità che la seconda estrazione fornisca una pallina bianca è valutata mediante il rapporto fra il variabile di palline bianche ed il variabile totale di palline presenti nell'urna, che fornisce il valore:

$$P(E_2 = b|E_1 = b) \;=\; \frac{pN-1}{N-1}\,. \tag{1.38}$$

Analogamente, se la prima estrazione ha fornito una pallina nera allora, prima della seconda estrazione E_2, l'urna contiene $N-1$ palline di cui ancora pN bianche; in questo caso la probabilità che la seconda estrazione fornisca una pallina bianca è data da:

$$P(E_2 = b|E_1 = n) \;=\; \frac{pN}{N-1}\,. \tag{1.39}$$

Dalla (1.37), mediante le (1.38) e (1.39), segue la probabilità che la seconda estrazione fornisca una pallina bianca:

$$\begin{aligned} P(E_2) &= P(E_1 = b)P(E_2 = b|E_1 = b) + P(E_1 = n)P(E_2 = b|E_1 = n) \\ &= p\,\frac{pN-1}{N-1} + (1-p)\frac{pN}{N-1} \\ &= \frac{p^2N - p + pN - p^2N}{N-1} \\ &= p\,. \end{aligned}$$

Nota. In generale, iterando tale procedimento, si può dimostrare per induzione che se il risultato delle prime k estrazioni senza restituzioni $E_1, \ldots E_k$ non è noto, la probabilità che la $k+1$-esima estrazione E_{k+1} fornisca una pallina bianca è ancora p. Ciò vale in particolare per la N-esima estrazione E_N. ♣

Esercizio 1.69 Una persona deve colpire con una freccia un bersaglio quadrato di un metro di lato. L'evento A consiste nel colpire la metà destra del bersaglio, e l'evento B consiste nel colpire il triangolo inferiore sinistro. L'area di A, come quella di B, è uguale a 0.5. Calcolare

1. $P(A)$,
2. $P(A \cap B)$,
3. $P(A \cap A^c)$,
4. $P(A^c \cap B)$,
5. $P(A^c \cap B^c)$.

1.6 Il Teorema di Bayes

Il teorema di Bayes costituisce, in un certo senso, una diretta applicazione della probabilità condizionata. Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità e sia $\{H_1, H_2, \ldots, H_n\}$ una partizione dell'evento certo Ω, cioè che gli eventi H_i siano a due a due incompatibili ($H_i \cap H_j = \emptyset$ per $i \neq j$) e tali che:

$$\Omega \;=\; H_1 \cup H_2 \cup \cdots \cup H_n \;;$$

Assumiamo inoltre $P(H_i) > 0$ per ogni $i = 1, \ldots, n$. Sia E un qualunque altro evento di probabilità positiva. In base al teorema delle probabilità composte si ha:

$$P(H_i|E) \;=\; \frac{P(H_i \cap E)}{P(E)} \;=\; \frac{P(H_i)P(E|H_i)}{P(E)} \tag{1.40}$$

Poichè gli eventi $H_1, H_2, \ldots, H_n$ costituiscono una partizione dell'evento certo Ω, risulta:

$$\begin{aligned} E \;&=\; E \cap \Omega \;=\; E \cap (H_1 \cup H_2 \cup \cdots \cup H_n) \\ &=\; (E \cap H_1) \cup (E \cap H_2) \cup \cdots \cup (E \cap H_n) \;=\; \bigcup_{i=1}^{n} (E \cap H_i) \;. \end{aligned} \tag{1.41}$$

Dal principio delle probabilità totali e dalla (1.27) – o direttamente dalla (1.41) – segue:

$$P(E) \;=\; P\left(E \cap \left(\bigcup_{j=1}^{n} H_j\right)\right) \;=\; \sum_{j=1}^{n} P(E \cap H_j) \;=\; \sum_{j=1}^{n} P(H_j)P(E|H_j) \;. \tag{1.42}$$

La formula (1.42) viene chiamata *formula di disintegrazione* di $P(E)$ in base alla partizione $\{H_1, \ldots, H_n\}$ e costituisce un risultato spesso utile ai fini del calcolo poichè spesso la probabilità di un evento E è difficile a calcolarsi direttamente, ma risulta agevole determinare $P(E|H_i)$, cioè la probabilità dell'evento E condizionatamente a H_i.

Sostituendo la (1.42) nella (1.40), si ottiene la relazione:

$$P(H_i \mid E) \;=\; \frac{P(E|H_i)\, P(H_i)}{\sum_{j=1}^{n} P(E|H_j)\, P(H_j)} \;. \tag{1.43}$$

Tale relazione costituisce un importante risultato legato alle probabilità condizionali noto come *teorema di Bayes*.

Teorema 1.70 (Teorema di Bayes) Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità e sia $H_1, H_2, \ldots, H_n$ una partizione di Ω tale che $P(H_i) > 0$ per $i = 1, \ldots, n$ e E un qualunque altro evento di probabilità positiva. Allora si ha, per ogni $i = 1, 2, \ldots, n$:

$$P(H_i \mid E) \;=\; \frac{P(E|H_i)\, P(H_i)}{\sum_{j=1}^{n} P(E|H_j)\, P(H_j)} \;.$$

Tale formula è una della più importanti ed interessanti del calcolo delle probabilità. Nella terminologia statistica $H_1, H_2, \ldots, H_n$ sono dette *cause* o *ipotesi*, le probabilità $P(H_i)$, $i = 1, 2, \ldots, n$ sono dette *probabilità iniziali* (o a priori) di H_i, $P(E|H_i)$ è detta *probabilità finale* (o a posteriori) di H_i.

Il teorema di Bayes può essere visto come uno strumento che permette di correggere le informazioni a priori $P(H_i)$ sulla base delle osservazioni sperimentali $P(E|H_i)$, fornendo la probabilità a posteriori $P(H_i|E)$. In questa formula, infatti, si combinano informazioni a priori e verosimiglianze, e quanto più la probabilità a posteriori $P(H_i|E)$ è diversa dalla probabilità a priori $P(H_i)$, tanto più la verosimiglianza ha modificato le informazioni a priori sulle cause H_i.

Nel caso particolare in cui $\Omega = H \cup H^c$ e quindi $H_1 = H$ e $H_2 = H^c$, con $0 < P(H) < 1$, dalla (1.42) segue immediatamente:

$$P(E) \;=\; P(H)P(E|H) + P(H^c)P(E|H^c).$$

Pertanto la (1.43) si scrive

$$P(H_i \mid E) \;=\; \frac{P(E|H_i)\,P(H_i)}{P(H)P(E|H) + P(H^c)P(E|H^c)} \qquad i = 1, 2\,. \tag{1.44}$$

Esempio 1.71 In una fabbrica, un certo manufatto viene prodotto per il 30% da macchine di tipo H_1, per il 25% da macchine di tipo H_2 e per il rimanente 45% da macchine di tipo H_3. Le macchine H_1 producono l'1% di scarto, le H_2 l'1,2% e le H_3 il 2% di scarto. In una giornata di lavoro le tre macchine producono complessivamente 10.000 pezzi. Un pezzo scelto a caso è difettoso. Calcoliamo la probabilità che esso provenga dalle macchine H_1, H_2 o H_3. Indicato con E l'evento "il pezzo in esame è difettoso", in base ai dati risulta:

$$\begin{aligned} P(E|H_1) &= 0.01 & \qquad P(H_1) &= 0.30 \\ P(E|H_2) &= 0.012 & \qquad P(H_2) &= 0.25 \\ P(E|H_3) &= 0.02 & \qquad P(H_3) &= 0.45. \end{aligned}$$

Il problema richiede il calcolo delle quantità:

$$P(H_1|E) \qquad\qquad P(H_3|E) \qquad\qquad P(H_3|E)\,.$$

In base al teorema di Bayes, dalla (1.43), risulta: Il problema richiede il calcolo delle quantità:

$$\begin{aligned} P(H_1|E) &= \frac{P(H_1)P(E|H_1)}{P(H_1)P(E|H_1) + P(H_2)P(E|H_2) + P(H_3)P(E|H_3)}\,, \\ P(H_2|E) &= \frac{P(H_2)P(E|H_2)}{P(H_1)P(E|H_1) + P(H_2)P(E|H_2) + P(H_3)P(E|H_3)}\,, \\ P(H_3|E) &= \frac{P(H_3)P(E|H_3)}{P(H_1)P(E|H_1) + P(H_2)P(E|H_2) + P(H_3)P(E|H_3)}\,. \end{aligned}$$

In base ai dati, il denominatore di tali espressioni è dato da:

$$\begin{aligned} P(E) &= P(H_1)P(E|H_1) + P(H_2)P(E|H_2) + P(H_3)P(E|H_3) \\ &= 0.01 \cdot 0.30 + 0.012 \cdot 0.25 + 0.02 \cdot 0.45 \;=\; 0.015\,. \end{aligned}$$

E quindi, sostituendo i dati in tali espressioni si ottiene:

$$\begin{aligned} P(H_1|E) &= \frac{0.3 \cdot 0.01}{0,015} = 0.20\ , \\ P(H_2|E) &= \frac{0.25 \cdot 0.012}{0,015} = 0.20\ , \\ P(H_3|E) &= \frac{0.45 \cdot 0.02}{0,015} = 0.60\ . \end{aligned}$$

Nella seguente tabella confrontiamo le probabilità a priori e quelle a posteriori, possiamo quindi vedere come si è modificata la probabilita su ciascuna delle cause in seguito all'osservazione:

Macchine	H_1	H_2	H_3	Somma
Prob. a priori	0.30	0.25	0.45	1
Prob. a posteriori	0.20	0.20	0.60	1

Tale risultato è motivato dal fatto che la macchina H_3 è quella che ha una maggiore probabilità a priori di fornire un pezzo difettoso ed è la macchina che fornisce il maggior numero di pezzi. ♣

Esercizio 1.72 Da un lotto di N componenti, contenente al più un pezzo difettoso, si effettuano senza restituzione $n \leq N$ estrazioni. Considerati gli eventi:

H : "il lotto contiene un componente difettoso

E_i : "l'i-esimo pezzo estratto è non difettoso" con $i = 1, \ldots, N$

e posto $P(H) = p$,

1. calcolare la funzione $f(p) = P(H|E_1, \ldots, E_n)$

2. determinare, in funzione di N, i valori di n tali che $f(0.7) < 0.4$.

Soluzione. In base al teorema di Bayes si ha:

$$\begin{aligned} P(H|E_1,\ldots,E_n) &= \frac{P(E_1,\ldots,E_n|H)P(H)}{P(E_1,\ldots,E_n|H)P(H)+P(E_1,\ldots,E_n|H^c)P(H^c)} \\ &= \frac{\dfrac{\binom{N-1}{n}\binom{1}{0}}{\binom{N}{n}}\cdot p}{\dfrac{\binom{N-1}{n}\binom{1}{0}}{\binom{N}{n}}\cdot p + 1\cdot(1-p)} \end{aligned}$$

Essendo:

$$\frac{\binom{N-1}{n}\binom{1}{0}}{\binom{N}{n}} = \frac{\dfrac{(N-1)!}{n!(N-1-n)!}}{\dfrac{N!}{n!(N-n)!}} = \frac{N-n}{N}$$

segue:

$$P(E|H_1,\ldots,H_n) = \frac{\dfrac{N-n}{N}\cdot p}{\dfrac{N-n}{N}\cdot p+(1-p)} = \frac{(N-n)p}{(N-n)p+N(1-p)} = \frac{(N-n)p}{N-np}$$

e quindi

$$f(p) = \frac{(N-n)p}{N-np}\,.$$

Per quanto riguarda il secondo punto, si cercano i valori di n tali che:

$$\frac{(N-n)\cdot 0.7}{N-0.7n} < 0.4\,.$$

Svolgendo tale disuguaglianza si trova che i valori di n che soddisfano alla relazione richiesta sono:

$$n > 0.7143N\,.$$

♣

Nota 1.73 Nelle applicazioni pratiche del teorema di Bayes, la forma più semplice (1.44) è spesso quella che risulta più utile:

$$P(E|H) = \frac{P(H|E)P(E)}{P(H|E)P(E) + P(H|E^c)P(E^c)} .$$

In questo contesto, è forse più espressivo mettere in evidenza il rapporto fra le probabilità condizionali all'evento E ed alla sua negazione:

$$P(E|H) = \frac{P(H|E)P(E)}{P(H|E)P(E) + P(H|E^c)P(E^c)} = \frac{P(E)}{P(E) + (1 - P(E))\dfrac{P(H|E^c)}{P(E|H)}}$$

$$= \frac{P(E)}{P(E) + \dfrac{1 - P(E)}{R}}$$

dove

$$R = \frac{P(H|E)}{P(H|E^c)} .$$

In particolare notiamo che

$$\begin{array}{lllll} \text{se } R = 1 & \text{allora} & P(E|H) = P(E) \\ \text{se } R > 1 & \text{allora} & P(E|H) > P(E) \\ \text{se } R < 1 & \text{allora} & P(E|H) < P(E) \end{array}$$

cioè il fatto che H si verifichi fa aumentare la fiducia nell'evento E se $R > 1$; al contrario, se $R < 1$, allora il fatto che H si verifichi fa diminuire la fiducia nel verificarsi di E.

Esempio 1.74 Assumiamo che l'1% degli abitanti di una città sia affetto da una certa malattia. Un nuovo test diagnostico fornisce un risultato positivo nel 97% dei casi in cui l'individuo è affetto da tale malattia, e fornisce un risultato negativo nel 95% dei casi in cui l'individuo è sano. Un individuo scelto a caso risponde positivamente al test: calcolare la probabilità che questi sia malato. Consideriamo i seguenti eventi:

M = l'individuo scelto è malato

$T|M$ = l'individuo scelto, nel caso in cui sia malato, risponde positivamente al test

$T^c|M^c$ = l'individuo scelto, nel caso in cui sia sano, risponde negativamente al test.

Il problema fornisce le seguenti valutazioni di probabilità:

$$\begin{aligned} P(M) &= 0.01 \\ P(T|M) &= 0.97 \\ P(T^c|M^c) &= 0.95 , \end{aligned}$$

e richiede di calcolare la probabilità che un individuo sia malato, nell'ipotesi in cui egli risponda positivamente al test, cioè $P(M|T)$. In base al teorema di Bayes, dalla (1.43), si ha:

$$P(M|T) = \frac{P(M)P(T|M)}{P(M)P(T|M) + P(M^c)P(T|M^c)} \tag{1.45}$$

poichè in base ai dati del problema risulta:

$$\begin{aligned} P(M^c) &= 1 - P(M) = 0.99 \\ P(T|M^c) &= 1 - P(T^c|M^c) = 0.05 , \end{aligned}$$

la (1.45) porge:

$$P(M|T) = \frac{0.01 \cdot 0.97}{0.01 \cdot 0.97 + 0.99 \cdot 0.05} = 0.1639 .$$

Supponiamo ora che il test venge ripetuto, fornendo ancora esito positivo: calcolare la probabilità che l'individuo sia malato. Indicati ora con T_1 e T_2 rispettivamente gli eventi "l'individuo risulta positivo al test nella prima analisi" e "l'individuo risulta positivo al test nella seconda analisi", si richiede la probabilità che l'individuo sia malato, nell'ipotesi in cui due analisi forniscano esito positivo, cioè $P(M|T_1 \cap T_2)$. Ancora per il teorema di Bayes si ha:

$$P(M|T_1 \cap T_2) = \frac{P(M)P(T_1 \cap T_2|M)}{P(M)P(T_1 \cap T_2|M) + P(M^c)P(T_1 \cap T_2|M^c)} . \tag{1.46}$$

E' ragionevole assumere che, nell'ipotesi in cui l'individuo si malato, gli esiti delle due analisi siano eventi indipendenti, cioè $P(T_1 \cap T_2|M) = P(T_1|M)P(T_2|M)$, e che forniscano lo stesso risultato, cioè $P(T_1|M) = P(T_2|M) = P(T|M)$, da cui segue $P(T_1 \cap T_2|M) = [P(T|M)]^2$; analogamente nel caso in cui l'individuo sia sano si ha $P(T_1 \cap T_2|M^c) = P(T_1|M^c)P(T_2|M^c) = [P(T|M^c)]^2$. Ne segue che la (1.46) si scrive

$$\begin{aligned} P(M|T_1 \cap T_2) &= \frac{P(M)P(T_1|M)P(T_2|M)}{P(M)P(T_1|M)P(T_2|M) + P(M^c)P(T_1|M^c)P(T_2|M^c)} \\ &= \frac{0.01 \cdot 0.97^2}{0.01 \cdot 0.97^2 + 0.99 \cdot 0.95^2} \\ &= 0.7917 . \end{aligned}$$

Si vede pertanto come, dopo l'esito della prima analisi, non si può ancora ragionevolmente supporre che l'individuo sia malato; al contrario dopo l'esito della seconda analisi, anch'esso positivo, si può ragionevolmente supporre che l'individuo sia malato. ♣

1.7 Altri esercizi

Esercizio 1.75 Un'impresa produce strumenti per il controllo del corretto funzionamento di dispositivi elettronici. Considerati gli eventi:

H = il dispositivo in esame funziona correttamente,
E = lo strumento rileva un funzionamento corretto del dispositivo in esame,

l'azienda pubblicizza l'affidabilità dei propri prodotti con le seguenti specifiche: $P(E|H) = P(E^c|H^c) = 0,95$. Essendo noto che la percentuale dei dispositivi in esame che risultano difettosi è del 5%:

1. Calcolare $P(H|E)$;
2. Calcolare quale deve essere il valore di $P(E|H) = P(E^c|H^c) = p$ affinchè risulti $P(H|E) = 0,9$.

Soluzione. In base alle ipotesi si ha

$$P(E|H) = \frac{P(E \cap H)}{P(H)} = 0.95 \qquad P(E^c|H^c) = \frac{P(E^c \cap H^c)}{P(H^c)} = 0.95$$

da cui, poichè $P(H) = 0,05$, segue:

$$\begin{aligned} P(E \cap H) &= P(E|H)\,P(H) = 0.95 \cdot 0.05 = 0.0475 \\ P(E^c \cap H^c) &= P(E^c|H^c)\,P(H^c) = 0.95 \cdot 0.95 = 0.9075\,. \end{aligned}$$

Il testo richiede il calcolo di

$$P(H|E) = \frac{P(E \cap H)}{P(E)}\,, \tag{1.47}$$

in cui la quantità $P(E)$ è incognita. Per la formula di De Morgan si ha:

$$P(E^c \cap H^c) = P[(E \cup H)^c] = 1 - P(E \cup H) = 0,95\,,$$

da cui $P(E \cup H) = 0,05$. Da un'applicazione inversa del teorema delle probabilità totali, si ha:

$$P(E) = P(E \cup H) - P(H) + P(E \cap H) = 0,0975 - 0,05 + 0,0475 = 0,095$$

e sostituendo nella (1.47), si ha

$$P(H|E) = \frac{P(E \cap H)}{P(E)} = \frac{0,0475}{0,095} = 0,5\,.$$

Per quanto riguarda il secondo punto, rifacendo i calcoli come in precedenza ponendo $P(E|H) = P(E^c|H^c) = p$, si ottiene:

$$\begin{aligned} P(E \cap H) &= 0.05p \\ P(E^c \cap H^c) &= 0.95p \\ P(E \cup H) &= 1 - P(E^c \cap H^c) = 1 - 0.95p \\ P(E) &= 1 - 0.95p - 0.05 + 0.05p = 0.95 - 0.90p \end{aligned}$$

e sostituendo nell'equazione $P(H|E) = 0.9$, segue:

$$P(H|E) = \frac{P(E \cap H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)P(H^c)} = \frac{0.05p}{0.05p + 0.95(1-p)} = 0.9$$

si ottiene

$$\frac{0.05p}{0,95-0,9p} = 0.9$$

da cui $p = 0.99$. ♣

Esercizio 1.76 In una mostra di di dipinti, ci sono 12 opere di cui 10 sono originali. Un visitatore sceglie un quadro a caso e, prima di acquistarlo, chiede l'opinione di un esperto circa l'autenticità del dipinto. L'esperto fornisce mediamente un parere corretto 9 volte su 10.

1. Assumendo che l'esperto attesta l'autenticità dell'opera, qual è la probabilità che il dipinto sia effettivamente autentico?

2. Se l'esperto valuta che il dipinto non sia autentico, allora il visitatore decide di non acquistare quel dipinto ma un altro scelto a caso. Qual è la probabilità che quest'ultimo sia autentico?

Soluzione. Indichiamo con O l'evento "il dipinto scelto è originale" e con E l'evento "l'esperto attesta che il dipinto è originale". In base ai dati del problema si ha $P(O) = \frac{5}{6}$ e $P(E|O) = \frac{9}{10}$. Il problema richiede il calcolo di $P(O|E)$. In base al teorema di Bayes si ha:

$$P(O|E) = \frac{P(E|O)P(O)}{P(E|O)P(O) + P(E|O^c)P(O^c)} = \frac{\frac{9}{10}\cdot\frac{5}{6}}{\frac{9}{10}\cdot\frac{5}{6} + \frac{1}{10}\cdot\frac{1}{6}} = \frac{45}{46}.$$

Indicato con O_2 l'evento "il secondo dipinto è originale", si ha:

$$\begin{aligned} P(O_2) &= P(O_2 \cap E) + P(O_2 \cap E^c) = P(O_2|E)P(E) + P(O_2|E^c)P(E^c) \\ &= \frac{10}{11}\cdot\frac{9}{10} + \frac{9}{11}\cdot\frac{1}{10} = \frac{99}{110}. \end{aligned}$$

♣

Esercizio 1.77 Si consideri una moneta regolare e si indichino con C e T rispettivamente gli eventi "croce" e "testa" nel generico lancio.

a) Si considerino tre lanci della moneta: si calcoli la probabilità che non si presenti mai la successione CT.

b) Si considerino quattro lanci della moneta: si calcoli la probabilità che non si verifichino le successioni CT o TCC.

Soluzione. E' sufficiente considerare lo spazio dei possibili risultati:

$$\Omega = \{(CCC), (CCT), (CTC), (CTT), (TCC), (TCT), (TTC), (TTT)\}$$

e quindi l'evento richiesto A in funzione dei singoli eventi elementari:

$$A = \{(CCC), (TCC), (TTC), (TTT)\} .$$

Da cui segue:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{4}{8} = \frac{1}{2} .$$

In maniera analoga, nel secondo punto, si considera l'evento certo:

$$\begin{aligned}\Omega \;=\; \{&(CCCC), (CCCT), (CCTC), (CCTT), (CTCC), (CTCT), (CTTC), (CTTT),\\ &(TCCC), (TCCT), (TCTC), (TCTT), (TTCC), (TTCT), (TTTC), (TTTT)\} .\end{aligned}$$

e quindi l'evento A richiesto:

$$A = \{(CCCC), (TTTC), (TTTT)\}$$

Da cui segue:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{3}{16} = \frac{3}{16} .$$

♣

Esercizio 1.78 Con riferimento ad una data estrazione del lotto, si supponga che Tizia abbia giocato i numeri $\{1, 2, 3, 4, 5\}$. Considerati gli eventi: A = "Tizio fa esattamente un terno oppure una quaterna" e B = "'i numeri 1 e 2 vengono estratti", calcolare: $P(A)$, $P(B)$ e $P(A|B)$.

Soluzione.

$$P(A) \;=\; \frac{\binom{5}{3}\binom{85}{2} + \binom{5}{4}\binom{85}{1}}{\binom{90}{5}} \;=\; \frac{5 \cdot 85^2}{89 \cdot 87 \cdot 6 \cdot 66} \;\approx\; 0.0008 \;;$$

$$P(B) \;=\; \frac{\binom{2}{2}\binom{88}{3}}{\binom{90}{5}} \;=\; \frac{2}{90 \cdot 89} \;\approx\; 0.0025 \;;$$

$$P(A|B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; \frac{\binom{3}{1}\binom{85}{2} + \binom{3}{2}\binom{85}{1}}{\binom{88}{3}} \;=\; \frac{3 \cdot 85}{29 \cdot 88} \;\approx\; 0.0999 \;;$$

♣

Capitolo 2

Variabili Aleatorie Discrete

2.1 Variabili aleatorie

Uno dei concetti fondamentali della teoria della probabilità è quello di *variabile aleatoria*. Prima di darne una definizione formale, illustriamo questo concetto mediante alcuni esempi:

- il numero di chiamate effettuate ad un centralino telefonico in un certo intervallo di tempo è una quantità aleatoria ed assume valori diversi in intervalli di tempo diversi (della stessa ampiezza) in dipendenza di circostanze accidentali;

- il tempo percorso da un furgone che, da un magazzino, rifornisce una sede periferica è una quantità aleatoria ed assume valori diversi (pur effettuando lo stesso percorso) in dipendenza delle condizioni di traffico e/o di altri fattori accidentali;

- il tempo che intercorre fra due successivi guasti di una stessa apparecchiatura elettronica è una quantità aleatoria che dipende da vari fattori accidentali.

Si pone il problema di costruire una teoria matematica che consenta lo studio di fenomeni di questo tipo. Si noti che, nonostante la diversità del contenuto reale degli esempi precedenti, essi presentano nella sostanza lo stesso aspetto dal punto di vista matematico. Infatti, in ciascun caso interviene una quantità che, in un modo o nell'altro, descrive i fenomeni che si stanno studiando; sotto l'effetto di circostanze casuali, ciascuna di queste quantità può prendere una varietà di valori diversi, e non si può stabilire in anticipo quale valore la quantità in esame assumerà, perchè essa varia in modo casuale da prova a prova.

Sinonimi di variabile aleatoria sono *numero aleatorio* e *variabile casuale*. Da un punto di vista formale, diamo la seguente definizione di variabile casuale.

Definizione 2.1 Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità. Si dice *variabile aleatoria* un'applicazione $X : \Omega \to \mathbb{R}$ tale che, per ogni $t \in \mathbb{R}$, l'insieme $\{\omega \in \Omega \, : \, X(\omega) \leq t\}$ sia in $\mathcal{A}$

Una variabile aleatoria (v.a.) è dunque una funzione di ω tale che si possa calcolare $P\{\omega \, : \, X(\omega) \leq t\}$, cioè tale che abbia senso calcolare la probabilità che X assuma valori al più uguali a t, per ogni $t \in \mathbb{R}$. Più in generale è fondamentale per le v.a. il calcolo di probabilità del tipo

$P\{\omega : X(\omega) \in A\}$, dove A è un sottoinsieme di $\mathbb{R}$. Il problema è pertanto quello di studiare l'applicazione:

$$A \rightarrow P\{\omega : X(\omega) \in A\} \tag{2.1}$$

che ad ogni sottoinsieme $A \subseteq \mathbb{R}$ associa la probabilità che X assuma valori appartenenti ad A. Questa applicazione viene chiamata *legge* o *distribuzione* di X.

Nota 2.2 In generale l'applicazione (2.1) potrebbe non essere definita per ogni sottoinsieme $A \subset \mathbb{R}$, cioè può accadere che $\{\omega : X(\omega) \in A\}$ non sia un evento (in altre parole non sia un elemento della σ-algebra $\mathcal{A}$). Per gli scopi del nostro corso, supporremo nel seguito che condizione richiesta sia sempre verificata.

Se $\{\omega : X(\omega) \leq a\}$ è un evento, cioè appartiene ad $\mathcal{A}$, per ogni $a \in R$, allora segue:

1. $\{\omega : X(\omega) > a\}$ è un evento : infatti, in base alla seconda proprietà della definizione di σ-algebra, si ha $\{\omega : X(\omega) > a\} = \{\omega : X(\omega) \leq a\}^c$.

2. $\{\omega : a < X(\omega) \leq b\}$ è un evento per ogni $a, b \in R$, con $a < b$, infatti:

$$\{\omega : a < X(\omega) \leq b\} = \{\omega : X(\omega) \leq b\} \cap \{\omega : X(\omega) > a\}$$

   appartiene ad $\mathcal{A}$ in quanto intersezione di due eventi appartenenti ad $\mathcal{A}$.

3. $\{\omega : X(\omega) = x\}$ è un evento in quanto possiamo scrivere tale insieme come l'intersezione di un numero infinito di eventi

$$\{\omega : X(\omega) = x\} := \bigcap_n \left\{\omega : x - \frac{1}{n} < X(\omega) \leq x\right\} \tag{2.2}$$

   e quindi appartiene ad $\mathcal{A}$ per la terza proprietà della definizione di σ-algebra.

Nel seguito, per semplicità, scriveremo $\{X \leq t\}$, $\{X \in A\}$ e $\{X \leq x\}$ anzichè rispettivamente $\{\omega : X(\omega) \leq t\}$, $\{\omega : X(\omega) \in A\}$ e $\{\omega : X(\omega) = x\}$. Inoltre indicheremo con Ω_X il codominio di X.

2.2 Variabili aleatorie discrete

Sia $X : \Omega \rightarrow \mathbb{R}$ una v.a. ed assumiamo che il codominio Ω_X di X contenga al più un'infinità numerabile di valori, cioè $\Omega_X = \{x_1, \ldots, x_k, \ldots\}$; supponiamo inoltre che l'insieme $\{x_1, \ldots, x_k, \ldots\}$ non abbia punti di accumulazione.

Sia $t \in \mathbb{R}$ e consideriamo l'insieme dei valori di Ω_X al più uguali a t, siano essi $x_1 < \cdots < x_h \leq t$. Poichè gli insiemi $\{X = x_i\}$, $(i = 1, \ldots, h)$, sono eventi in base alla (2.2), essendo il codominio un insieme discreto, possiamo scrivere:

$$\begin{aligned} \{\omega \in \Omega : X \leq t\} &= \{\omega \in \Omega : X = x_1\} \cup \cdots \cup \{\omega \in \Omega : X = x_h\} \\ &= \bigcup_{x_i \leq t} \{\omega \in \Omega : X = x_i\} . \end{aligned} \tag{2.3}$$

Assegnata un v.a. discreta X, consideriamo la funzione $p : \mathbb{R} \to \mathbb{R}^+$ definita da

$$p(x) \;=\; P\{X = x\}\ .$$

Ovviamente tale funzione gode delle seguenti proprietà:

$$1.\quad p(x) \;\geq\; 0 \qquad \text{per ogni } x \in \mathbb{R} \tag{2.4}$$

$$2.\quad p(x) \;>\; 0 \qquad \text{per } x \in \Omega_X \tag{2.5}$$

$$3.\quad \sum_{x \in \Omega_X} p(x) \;=\; 1\ . \tag{2.6}$$

La (2.5) è ovvia; per quanto riguarda la (2.6), osserviamo che gli eventi $\{X = x_i\}$ sono a due a due disgiunti in quanto X è una funzione che associa ad ogni $\omega \in \Omega$ uno ed un solo valore $X(\omega) = x$; inoltre poichè X assume valori solo in Ω_X, ne segue che

$$\bigcup_{i=1}^{\infty} \{X = x_i\} \;=\; \{X = x_1\} \cup \cdots \cup \{X = x_i\} \cup \cdots \;=\; \{X \in \Omega_X\} \;=\; \Omega$$

e pertanto:

$$\sum_{i=1}^{\infty} p(x_i) \;=\; \sum_{i=1}^{\infty} P\{X = x_i\} \;=\; P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) \;=\; P(\Omega) \;=\; 1\ .$$

La funzione p sopra definita viene chiamata *funzione di probabilità* o *densità discreta*.

La conoscenza della densità consente di determinare la legge di X; infatti assegnato $A \subset \mathbb{R}$, la probabilità di tale evento è data da:

$$P\{X \in A\} \;=\; \sum_{x_i \in A} P\{X = x_i\} \;=\; \sum_{x_i \in A} p(x_i)\ . \tag{2.7}$$

Un'importante variabile aleatoria è la funzione indicatrice. Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità e sia $A \in \mathcal{A}$, con $P(A) = p$. Allora

$$\begin{aligned} P(|A| = 1) \;&=\; P(A) \;=\; p \\ P(|A| = 0) \;&=\; P(A^c) \;=\; 1 - p\ . \end{aligned}$$

In questo caso, pertanto, la funzione di densità assume valori positivi su 0 e 1, e zero altrove.

Esempio 2.3 Consideriamo una sequenza di tre lanci di una moneta regolare. Sia X la quantità:

X = "numero di teste in una successione di tre lanci di una moneta"

Poichè $\Omega = \{$(TTT), (TTC), (TCT), (TCC), (CTT), (CTC), (CCT), (CCC)$\}$, allora X può assumere solo i valori 0,1,2,3, cioè $\Omega_X = \{0, 1, 2, 3\}$. Ciascuna delle sequenze di risultati ha probabilità uguale a 1/8. Se indichiamo con ω la generica sequenza di risultati, allora si ha:

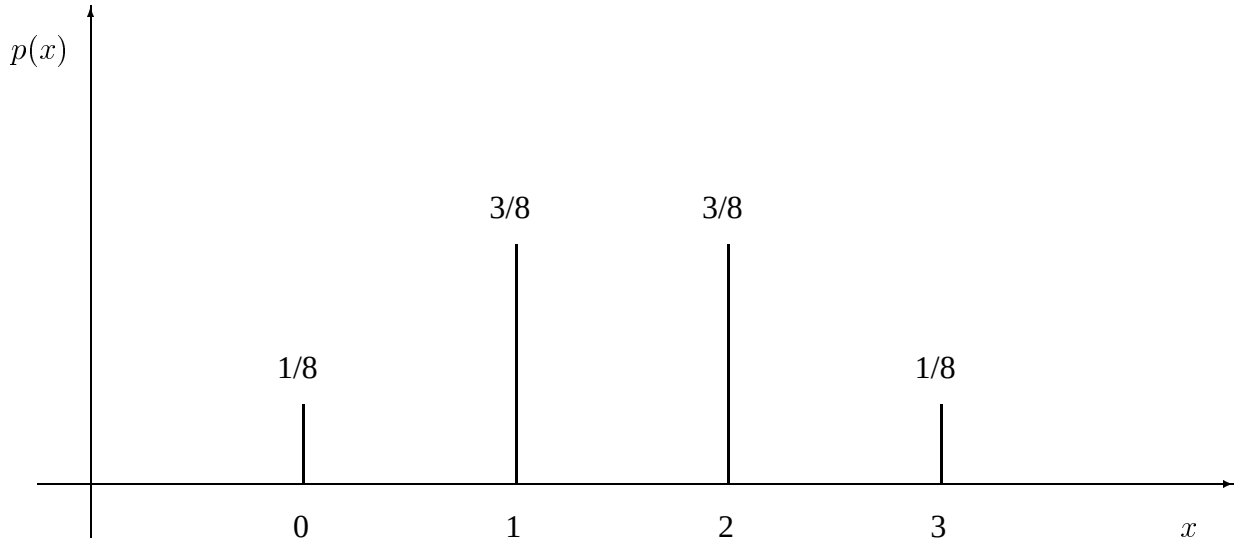


Figura 2.1: Rappresentazione grafica della distribuzione di probabilità del numero di teste uscite nel lancio di tre monete.

ω	$X(\omega) = x$	$P(\omega)$
T T T	3	1/8
T T C	2	1/8
T C T	2	1/8
T C C	1	1/8
C T T	2	1/8
C T C	1	1/8
C C T	1	1/8
C C C	0	1/8

La funzione di probabilità è data da:

$$
\begin{aligned}
p(0) &= P(X=0) = P(\text{CCC}) = 1/8 \\
p(1) &= P(X=1) = P(\text{CCT} \cup \text{CTC} \cup \text{TCC}) = P(\text{CCT}) + P(\text{CTC}) + P(\text{TCC}) = 3/8 \\
p(2) &= P(X=2) = P(\text{CTT} \cup \text{TTC} \cup \text{TCT}) = P(\text{CTT}) + P(\text{TTC}) + P(\text{TCT}) = 3/8 \\
p(3) &= P(X=3) = P(\text{TTT}) = 1/8 \, .
\end{aligned}
$$

In Figura 2.1 viene data la rappresentazione grafica della funzione di probabilità della v.a. $X = |T_1| + |T_2| + |T_3|$. E' immediato verificare la proprietà (2.6):

$$
p(0) + p(1) + p(2) + p(3) = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1 \, .
$$

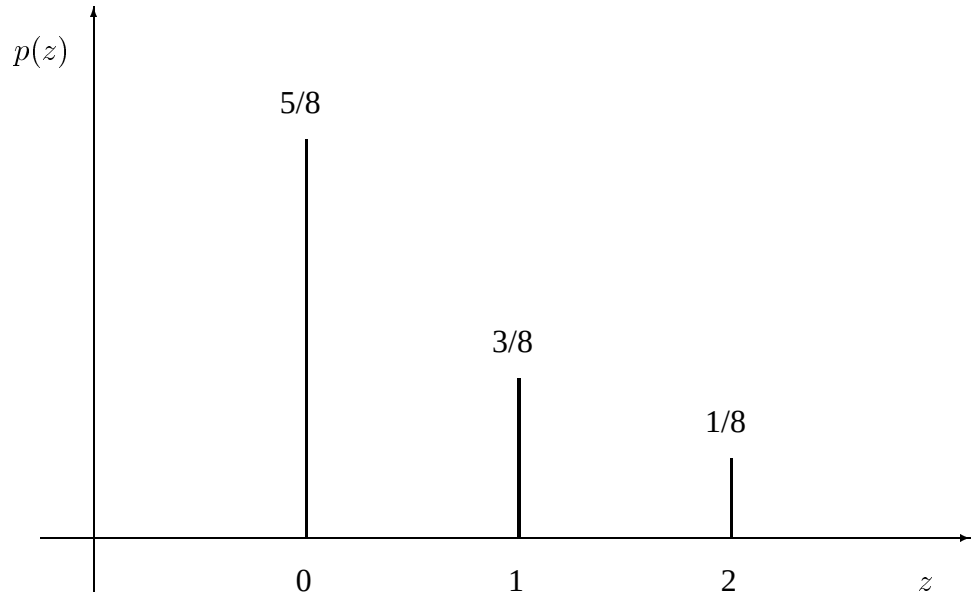


Figura 2.2: Rappresentazione grafica della distribuzione di probabilità della v.a. Z.

Sullo stesso spazio campionario Ω si può considerare la variabile aleatoria Y = "numero di coppie consecutive di teste":

$$Z \;=\; |\mathrm{T_1T_2}| + |\mathrm{T_2T_3}| \; .$$

In questo caso, come si vede dalla tabella precedente, la v.a. Z può assumere i valori $0, 1, 2$, cioè $\Omega_Z = \{0, 1, 2\}$. La funzione di probabilità è data da:

$$\begin{aligned}
p(0) \;=\; P(Z=0) \;&=\; P(\mathrm{CCC} \cup \mathrm{CCT} \cup \mathrm{CTC} \cup \mathrm{TCC} \cup \mathrm{TCT}) \\
&=\; P(\mathrm{CCC}) + P(\mathrm{CCT}) + P(\mathrm{CTC}) + P(\mathrm{TCC}) + P(\mathrm{TCT}) \;=\; 5/8 \\
p(1) \;=\; P(Z=1) \;&=\; P(\mathrm{TTC} \cup \mathrm{CTT}) \;=\; P(\mathrm{TTC}) + P(\mathrm{CTT}) \;=\; 2/8 \\
p(2) \;=\; P(Z=2) \;&=\; P(\mathrm{TTT}) \;=\; 1/8 \; .
\end{aligned}$$

Verifichiamo anche in questo caso la proprietà (2.6):

$$p(0) + p(1) + p(2) \;=\; \frac{5}{8} + \frac{2}{8} + \frac{1}{8} \;=\; 1 \; .$$

In Figura 2.2 viene data la rappresentazione grafica della funzione di probabilità della v.a. Z. ♣

Insieme alla funzione di probabilità di una variabile aleatoria si definisce un'altra funzione importante ai fini della descrizione delle principali proprietà di una variabile aleatoria: la *funzione di ripartizione* o *funzione di distribuzione*

Definizione 2.4 Sia X una variabile aleatoria discreta definita su $(\Omega, \mathcal{A}, P)$ ed a valori in $\Omega_X = \{x_1, x_2, \ldots\}$. Si definisce *funzione di ripartizione* di X la funzione $F : \mathbb{R} \to [0, 1]$ definita da:

$$F(x) := P(X \leq x) \; . \tag{2.8}$$

Sia $x_h \in \Omega_X$ allora per la (2.3) si ha:

$$F(x_h) \quad = \quad p(x_1) + p(x_2) + \cdots + p(x_h) \tag{2.9}$$

Più in generale, per $t \in \mathbb{R}$, si ha:

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t) \ . \tag{2.10}$$

L'ordinata della funzione di ripartizione calcolata nel punto $x \in \mathbb{R}$ indica pertanto la probabilità con cui la variabile aleatoria X assume valori al più uguali a x. Dalla definizione di funzione di ripartizione seguono le seguenti proprietà:

1. $F(x)$ è non decrescente: infatti per $x_1 \leq x_2$ si ha $\{X \leq x_1\} \subseteq \{X \leq x_2\}$ e quindi per la proprietà di monotonia si ha:

   $$P\{X \leq x_1\} \quad \leq \quad P\{X \leq x_2\} \ ,$$

   cioè $F(x_1) \leq F(x_2)$.

2. $F(x)$ è continua da destra, cioè $\lim_{\varepsilon \to 0^+} F(x + \varepsilon) = F(x)$: infatti considerati gli eventi $\{X \leq x\}$ e $\{X \leq x + \varepsilon\}$, risulta:

   $$\lim_{\varepsilon \to 0^+} \{X \leq x + \varepsilon\} \quad = \quad \{X \leq x\}$$

   e quindi:

   $$\lim_{\varepsilon \to 0^+} P\{X \leq x + \varepsilon\} = P\{X \leq x\} \ .$$

3. $F(x_i) - F(x_{i-1}) = p(x_i)$ per $x_i \in \Omega_X$, infatti:

   $$\begin{aligned} F(x_i) - F(x_{i-1}) &= P\{X \leq x_i\} - P\{X \leq x_{i-1}\} = P\{x_{i-1} < X \leq x_i\} \\ &= P\{X = x_i\} = p(x_i) \ . \end{aligned}$$

4. $F(x) = 0$ per $x < x_1$ e $F(x) = 1$ per $x \geq x_k$, infatti per $x < x_1$ si ha $\{X \leq x\} = \emptyset$ e per $x \geq x_k$ si ha $\{X \leq x\} = \Omega$.

In Figura 2.3 diamo la funzione di ripartizione della v.a. $X = |T_1| + |T_2| + |T_3|$.

Esercizio 2.5 Sia X la v.a. "numero di teste ottenute in due lanci consecutivi di una moneta equilibrata".

1. Descrivere lo spazio degli eventi.

2. Calcolare la funzione di densità di X e rappresentarla graficamente;

3. Calcolare la funzione di ripartizione di X e rappresentarla graficamente;

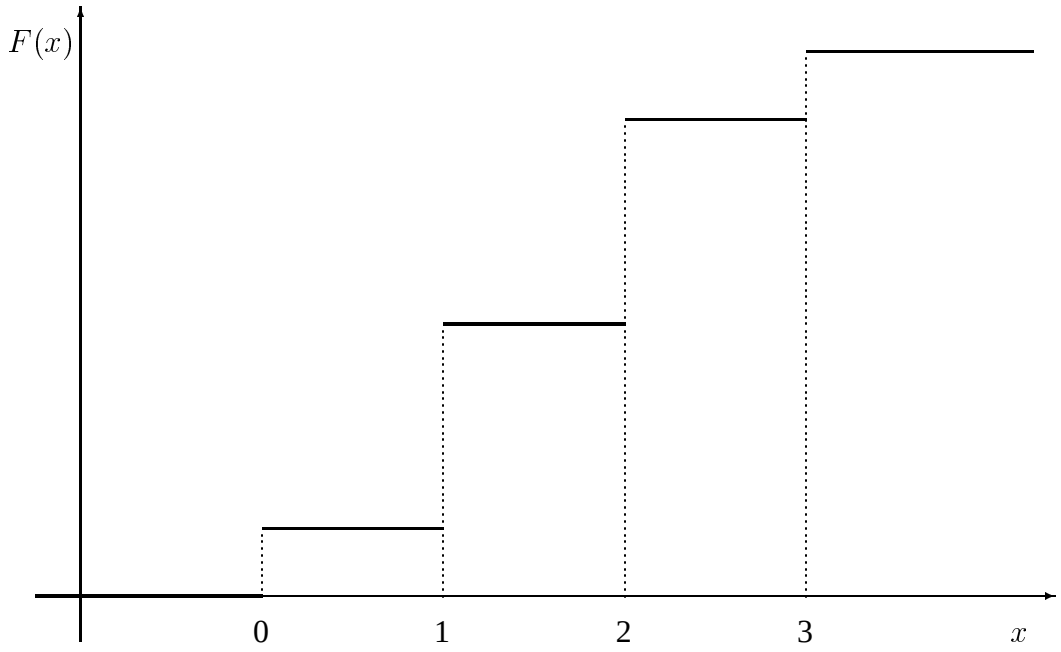


Figura 2.3: Rappresentazione grafica della funzione di ripartizione della v.a. $X = |T_1| + |T_2| + |T_3|$ concernente il numero di teste uscite nel lancio di tre monete.

Soluzione. Lo spazio degli eventi Ω è il seguente:

$$\Omega = \{TT, CT, TC, CC\} .$$

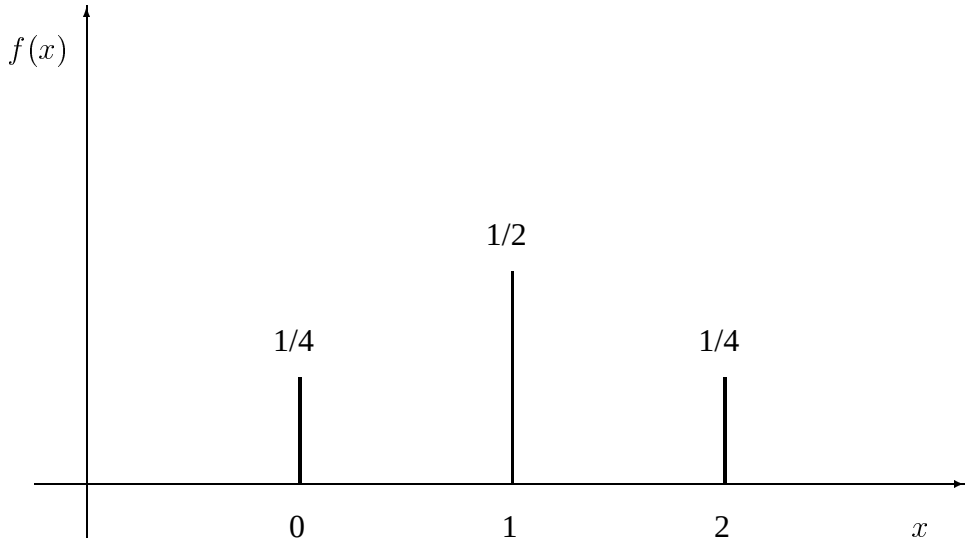
ciascuno con probabilità $\frac{1}{4}$ in quanto assumiamo l'indipendenza degli eventi elementari $\{C, T\}$ associati a ciascun lancio, cioè ad esempio $P(TT) = P(T) \cdot P(T) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$.

Considerata la v.a. X="numero di teste ottenute in due lanci consecutivi di una moneta equilibrata", ne segue che la funzione di densità assume valori strettamente positivi nell'insieme $\mathcal{X} = \{0, 1, 2\}$:

$$\begin{aligned} p(0) &= P(X = 0) = P(CC) = \frac{1}{4} = 0.25 \\ p(1) &= P(X = 1) = P(CT) + P(TC) = \frac{1}{2} = 0.5 \\ p(2) &= P(X = 2) = P(TT) = \frac{1}{4} = 0.25 . \end{aligned}$$

Si ha pertanto:

$$p(x) = \begin{cases} 0.25 & \text{per } x = 0 \\ 0,50 & \text{per } x = 1 \\ 0,25 & \text{per } x = 2 \\ 0 & \text{altrove.} \end{cases}$$



La funzione di ripartizione $F(x)$ viene costruita a partire dalla funzione di densità in base alla sua definizione:

$$F(t) = P(X \leq t).$$

Si ha pertanto:

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 0,25 & \text{per } 0 \leq x < 1 \\ 0,75 & \text{per } 1 \leq x < 2 \\ 1,00 & \text{per } 2 \leq x \end{cases}$$

e tale funzione ha la seguente rappresentazione grafica: ♣

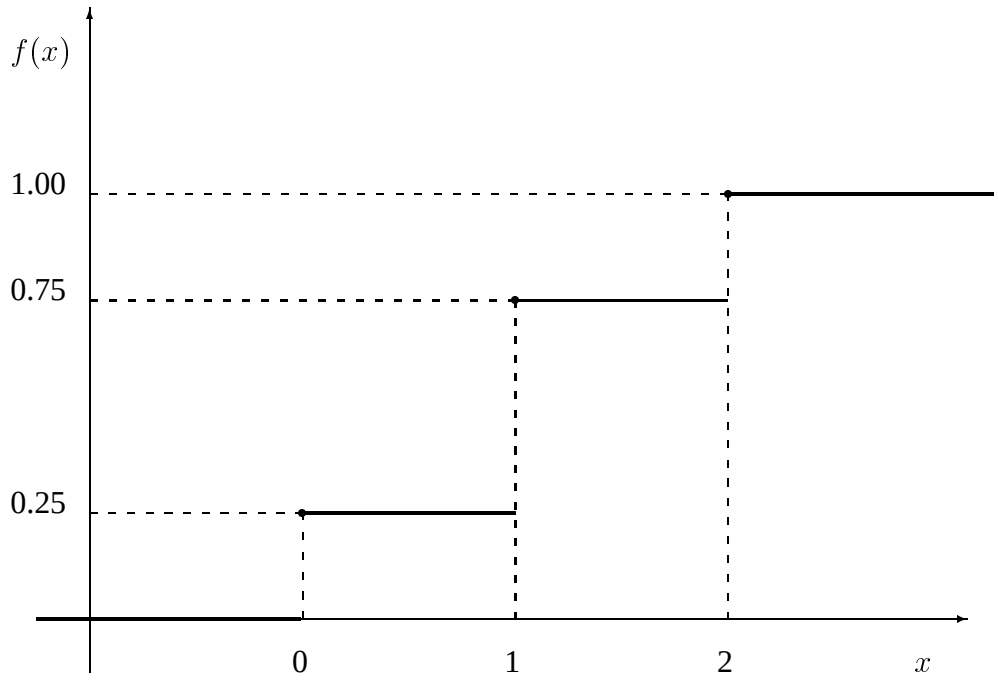
2.3 Distribuzioni congiunte

Definizione 2.6 Una v.a. m-dimensionale discreta (o *vettore aleatorio* discreto) è un'applicazione $\mathbf{X} = (X_1, \ldots, X_m) : \Omega \to \mathbb{R}$ tale che le applicazioni $X_1, \ldots, X_m$ siano delle v.a. discrete.

Ovviamente, se $\mathbf{X}$ è una v.a. discreta, allora essa può assumere al più un'infinità numerabile di valori. Infatti se $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}$, allora

$$\{\omega \in \Omega : \mathbf{X} = \mathbf{x}\} \;=\; \{\omega \in \Omega : X_1 = x_1\} \cap \cdots \cap \{\omega \in \Omega : X_m = x_m\} \qquad (2.11)$$

Tale relazione mostra che il v.a. $\mathbf{X}$ risulta uguale a $\mathbf{x}$ se e solo se X_1 assume valore x_1, X_2 assume valore x_2, ..., X_m assume valore x_m; ne segue che $\{\mathbf{X} = \mathbf{x}\}$ è un evento in quanto



intersezione di eventi. Nel seguito, anzichè la (2.11), utilizzeremo sempre la seguente scrittura sintetica:

$$\{\mathbf{X} = \mathbf{x}\} \;=\; \{X_1 = x_1\} \cap \cdots \cap \{X_m = x_m\} \; .$$

Al solito, denoteremo con Ω_X il codominio di $\mathbf{X}$.

Analogamente al caso unidimensionale, anche per le v.a. m-dimensionali si può definire una *densità discreta*, ponendo:

$$p(\mathbf{x}) := P\{\mathbf{X} = \mathbf{x}\} \; . \tag{2.12}$$

dove, come in precedenza, la funzione $p(\mathbf{x})$ verifica le tre proprietà:

$$1. \quad p(\mathbf{x}) \;\geq\; 0 \qquad \text{per ogni } \mathbf{x} \in \mathbb{R}^m \tag{2.13}$$

$$2. \quad p(\mathbf{x}) \;>\; 0 \qquad \text{per } \mathbf{x} \in \Omega_X \tag{2.14}$$

$$3. \quad \sum_{\mathbf{x} \in \Omega_X} p(\mathbf{x}) \;=\; 1 \; . \tag{2.15}$$

Analogamente al caso di variabili aleatorie, chiameremo *funzione di densità* (discreta) una funzione p che soddisfi le tre condizioni viste sopra e, se p è una funzione di densità, allora esistono certamente uno spazio di probabilità $(\Omega, \mathcal{A}, P)$ ed una v.a. m-dimensionale $\mathbf{X}$ di cui p sia la densità.

I ragionamenti che hanno portato alla (2.7) si possono ripetere esattamente per v.a. discrete m-dimensionali $\mathbf{X}$ e $A \subset \mathbb{R}^m$, e si ottiene:

$$P(\mathbf{x} \in A) \;=\; \sum_{\mathbf{x} \in A} p(\mathbf{x}) \; .$$

Se $X_1, \ldots, X_m$ sono v.a. reali, la densità della v.a. m-dimensionale $\mathbf{X} = (X_1, \ldots, X_m)$ si chiama *densità congiunta* delle v.a. $X_1, \ldots, X_m$; al contrario, se $\mathbf{X} = (X_1, \ldots, X_m)$ è una

v.a. m-dimensionale, le densità delle v.a. reali $X_1, \ldots, X_m$ – denotate rispettivamente con $p_1, \ldots, p_m$ – vengono chiamate *densità marginali*.

Un importante caso è quello $m = 2$, cioè di v.a. bidimensionali, in cui le densità congiunte e marginali possono essere assegnate mediante una *tabella a doppia entrata*. Per semplicità di notazione, in questo caso la v.a. bidimensionale viene denotata con (X, Y) anzichè con (X_1, X_2); inoltre assumiamo che il codominio delle v.a. X e Y sia costituito rispettivamente da k e h valori distinti, cioè $\Omega_X = \{x_1, \ldots, x_k\}$ e $\Omega_Y = \{y_1, \ldots, y_h\}$ (nel caso generale, dovremmo considerare tabelle con un numero infinito di righe e/o colonne). Posto $p_{ij} = p(x_i, y_j)$, $p_{i0} = p_X(x_i)$ e $p_{0j} = p_Y(y_j)$, si ottiene la seguente tabella a doppia entrata:

	y_1	$\ldots$	y_j	$\ldots$	y_h	
x_1	p_{11}	$\ldots$	p_{1j}	$\ldots$	p_{1h}	p_{10}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	p_{i1}	$\ldots$	p_{ij}	$\ldots$	p_{ih}	p_{i0}
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	p_{k1}	$\ldots$	p_{kj}	$\ldots$	p_{kh}	p_{k0}
	p_{01}	$\ldots$	p_{0j}	$\ldots$	p_{0h}	1

dove $p_{10}, \ldots, p_{k0}$ costituisce la distribuzione marginale di X e $p_{01}, \ldots, p_{0h}$ costituisce la densità marginale di Y, cioè

$$p_{i0} = p_X(x_i) := P\{X = x_i\} \qquad i = 1, \ldots, k\,,$$
$$p_{0j} = p_Y(y_j) := P\{Y = y_j\} \qquad j = 1, \ldots, h\,.$$

Ovviamente risulta:

$$p_{i0} := p_X(x_i) = P\{X = x_i\} = P\left(\bigcup_{j=1}^{h}\{X = x_i, Y = y_j\}\right) = \sum_{j=1}^{h} P\{X = x_i, Y = y_j\}$$
$$= \sum_{j=1}^{h} p(x_i, y_j) = \sum_{j=1}^{h} p_{ij}$$
$$p_{0j} := p_Y(y_j) = P\{Y = y_j\} = P\left(\bigcup_{i=1}^{k}\{X = x_i, Y = y_j\}\right) = \sum_{i=1}^{k} P\{X = x_i, Y = y_j\}$$
$$= \sum_{i=1}^{k} p(x_i, y_j) = \sum_{i=1}^{k} p_{ij}\ .$$

Si noti infine che non è possibile, in generale, conoscendo le sole densità marginali, ricostruire la densità congiunta in quanto – come si può anche vedere dalla tabella precedente – densità congiunte diverse possono avere le stesse densità marginali.

Definizione 2.7 Le m v.a. discrete $X_1, \ldots, X_m$, assegnate su uno stesso spazio di probabilità $(\Omega, \mathcal{A}.P)$, si dicono *indipendenti* se e solo se per ogni scelta di $A_1, \ldots, A_m \subset \mathbb{R}$ si ha:

$$P\{X_1 \in A_1, \ldots, X_m \in A_m\} \;=\; P\{X_1 \in A_1\} \cdots P\{X_m \in A_m\}\ . \tag{2.16}$$

Diremo che le v.a. $X_1, \ldots, X_n, \ldots$ (in numero infinito) sono indipendenti se e solo se per ogni scelta di m indici $i_1, \ldots, i_m$, le variabili $X_{i_1}, \ldots, X_{i_m}$ risultano indipendenti.

Il significato intuitivo della nozione di indipendenza di v.a. è analogo a quello di indipendenza di eventi: m v.a. sono indipendenti se la conoscenza dei valori assunti da alcune di esse non dà informazioni che modifichino la previsione di quelli assunti dalle altre. Del resto, la (2.16) scritta nel caso di due v.a., per ogni $A_1 \subseteq \Omega_{X_1}, A_2 \subseteq \Omega_{X_2}$ diviene:

$$P\{X_1 \in A_1, X_2 \in A_2\} \;=\; P\{X_1 \in A_1\}P\{X_2 \in A_2\} \tag{2.17}$$

che non significa altro che il fatto che i due eventi $\{X_1 \in A_1\}$ e $\{X_2 \in A_2\}$ sono fra loro indipendenti *per ogni scelta* degli insiemi A_1, A_2.

Nella (2.16), in particolare scelti $A_1 = \{x_1\}, \ldots, A_m = \{x_m\}$, si ottiene:

$$P\{X_1 = x_1, \ldots, X_m = x_m\} \;=\; P\{X_1 = x_1\} \cdots P\{X_m = x_m\}\ . \tag{2.18}$$

Indicato con p la densità congiunta di $X_1, \ldots, X_m$ e con $p_1, \ldots, p_m$ le rispettive densità di $X_1, \ldots, X_m$, posto $\mathbf{x} = (x_1, \ldots, x_m)$, la (2.18) si può scrivere:

$$p(\mathbf{x}) \;=\; p_1(x_1) \cdots p_m(x_m)\ . \tag{2.19}$$

Al contrario, supponiamo che valga la (2.18); se $A_1, \ldots, A_m \subset \mathbb{R}$, posto $A = A_1 \times A_2 \times \cdots \times A_m$, si ha:

$$\begin{aligned} P\{X_1 \in A_1, \ldots, X_m \in A_m\} &= \sum_{x_1 \in A_1, \ldots, x_m \in A_m} p(x_1, \ldots, x_m) \\ &= \sum_{x_1 \in A_1, \ldots, x_m \in A_m} p_1(x_1) \cdots p_m(x_m) \\ &= \sum_{x_1 \in A_1} p_1(x_1) \cdots \sum_{x_m \in A_m} p_m(x_m) \\ &= P\{X_1 \in A_1\} \cdots P\{X_m \in A_m\} \end{aligned}$$

Ciò dimostra che la (2.19) è una condizione equivalente all'indipendenza di $X_1, \ldots, X_m$. In particolare, nel caso di v.a. indipendenti, tramite la (2.19) è possibile calcolare la densità congiunta a partire dalle marginali, cosa che non è vera in generale.

2.3.1 Densità condizionali

Siano assegnato due v.a. discrete (X, Y) avente densità congiunta $p(x, y)$. Assumendo $p_X(x) > 0$, in accordo alla definizione di probabilità condizionale, si ha:

$$P(Y = y | X = x) \;=\; \frac{P(Y = y, X = x)}{P(X = x)} \;=\; \frac{p(x, y)}{p_X(x)} \qquad\qquad p_X(x) > 0\ .$$

Analogamente per $P(X = x | Y = y)$ assumendo stavolta $p_Y(y) > 0$.

Definizione 2.8 Sia (X, Y) un vettore aleatorio discreto con funzione di probabilità $p(x, y)$ e densità di probabilità marginali $p_X(x)$ e $p_Y(y)$. Si dice *densità di probabilità condizionale* di Y dato $X = x$ la funzione

$$p_{Y|X}(y|x) \;=\; P(Y = y|X = x) \;=\; \frac{p(x,y)}{p_X(x)} \qquad\qquad p_X(x) > 0\ ;$$

analogamente si dice *densità di probabilità condizionale* di X dato $Y = y$ la funzione

$$p_{X|Y}(x|y) \;=\; P(X = x|Y = y) \;=\; \frac{p(x,y)}{p_Y(y)} \qquad\qquad p_Y(y) > 0\ .$$

Con particolare riferimento ad una distribuzione doppia di probabilità, è possibile individuare $2 + h + k$ distribuzioni semplici:

- 2 distribuzioni marginali

$$X \;=\; \begin{cases} x_1 & x_2 & \cdots & x_k \\ p_{10} & p_{20} & \cdots & p_{k0} \end{cases} \qquad\qquad Y \;=\; \begin{cases} y_1 & y_2 & \cdots & y_h \\ p_{01} & p_{02} & \cdots & p_{0h} \end{cases}$$

- h distribuzioni condizionate di X dato $Y = y_j$, la cui distribuzione generica è:

$$\begin{cases} x_1 & x_2 & \cdots & x_k \\ p_{X|Y}(x_1|y_j) & p_{X|Y}(x_2|y_j) & \cdots & p_{X|Y}(x_k|y_j) \end{cases}$$

- k distribuzioni condizionate di Y dato $X = x_i$ la cui distribuzione generica è:

$$\begin{cases} y_1 & y_2 & \cdots & y_h \\ p_{Y|X}(y_1|x_i) & p_{Y|X}(y_2|x_i) & \cdots & p_{Y|X}(y_h|x_i) \end{cases}$$

Esercizio 2.9 Il responsabile di un giornale finanziario desidera conoscere quali sono le aree di investimento di maggiore interesse fra gli abbonati al giornale. Vengono scelti a caso 330 (investitori) professionisti abbonati a cui viene inviato un questionario. I risultati sono sintetizzati nella seguente tabella:

	Area di investimento				
professione	Fondi azionari	Fondi Obbligazionari	Azioni	Titoli di stato	Totale
Medico	30	25	15	0	70
Avvocato	29	34	12	6	81
Commercialista	50	35	29	15	129
Altro	21	14	10	5	50
Totale	130	108	66	26	330

1. Valutare la probabilità che un investitore non sia né un avvocato né un commercialista;
2. Valutare la probabilità che l'investitore sia un commercialista avente maggiore interesse nei fondi obbligazionari;
3. Valutare la probabilità che il professionista sia un avvocato, nel caso in cui si abbia maggiore interesse nell'acquisto di azioni;
4. Valutare la probabilità che l'investimento principale del professionista non sia nei titoli di stato;
5. Sia A l'evento che l'investitore sia un avvocato e B l'evento che l'area di maggiore interesse siano i titoli di stato. Verificare se gli venti A e B sono mutuamente esclusivi.

Esercizio 2.10 Siano X e Y due variabili aleatorie discrete, a valori rispettivamente in $\Omega_X = \{0, 2\}$ e $\Omega_Y = \{1, 2\}$. Considerate le seguenti probabilità congiunte in funzione del parametro α:

$$\begin{aligned} P(X = 0, Y = 1) &= \alpha \\ P(X = 0, Y = 2) &= 1 - 4\alpha \\ P(X = 1, Y = 1) &= \alpha\,. \end{aligned}$$

Calcolare l'insieme dei valori ammissibili per α. Verificare se esistono valori di α tali che X e Y risultano stocasticamente indipendenti ed in tal caso calcolarli.

Soluzione. Trattandosi di una distribuzione di probabilità deve ovviamente risultare:

$$\begin{aligned} P(X = 0, Y = 1) + P(X = 0, Y = 2) + P(X = 1, Y = 1) + P(X = 1, Y = 2) = \\ = \alpha + 1 - 4\alpha + \alpha + P(X = 1, Y = 2) = 1 \end{aligned}$$

da cui si ottiene $P(X = 1, Y = 2) = 2\alpha$ e quindi la seguente tabella:

		X		
		0	1	
Y	1	α	α	2α
	2	$1 - 4\alpha$	2α	$1 - 2\alpha$
		$1 - 3\alpha$	3α	1

Affinchè i valori in tabella siano delle probabilità devono essere verificate preliminarmente le condizioni:

$$0 \le \alpha \le 1/2\,; \qquad 0 \le 1 - 2\alpha \le 1\,; \qquad 0 \le 1 - 3\alpha \le 1\,; \qquad 0 \le 1 - 4\alpha \le 1$$

da cui segue $0 \le \alpha \le 1/4$.

Le v.a. X e Y risultano stocasticamente indipendenti se e solo se è soddisfatto il seguente sistema ($p_{ij} = p_{i0}p_{0j}$ per ogni i, j):

$$\begin{aligned} \alpha &= 2\alpha(1-3\alpha) \\ \alpha &= 6\alpha^2 \\ 1-4\alpha &= (1-2\alpha)(1-3\alpha) \\ 2\alpha &= 3\alpha(1-2\alpha) \end{aligned}$$

Le quattro equazioni conducono alla equazione seguente:

$$6\alpha^2 - \alpha = 0$$

soddisfatto per $\alpha_1 = 0$ e $\alpha_2 = 1/6$, valori entrambi compatibili con l'insieme dei valori ammissibili. In particolare, per $\alpha = 0$, la distribuzione è concentrata sulla coppia di valori $(X = 0, Y = 2)$. ♣

2.4 Speranza matematica

Sia X una v.a. definita su uno spazio di probabilità $(\Omega, \mathcal{A}, P)$.

Definizione 2.11 Sia X una v.a. (discreta) che assuma i valori $x_1, x_2, \ldots$ ed indichiamo con $p(x)$ la sua densità. Assumiamo che:

$$\sum_i |x_i| p(x_i) \quad < \quad +\infty \, .$$

In questo caso si definisce *speranza matematica* di X, la quantità

$$\mathbb{E}[X] := \sum_i x_i \, p(x_i) = \sum_i x_i P\{X = x_i\} \tag{2.20}$$

In altre parole la speranza matematica di X è data dalla (2.20) a condizione che la serie dei valori assoluti converga. Sinonimi di speranza matematica sono i termini *media, valore medio, attesa, valore atteso, previsione*. In alcuni testi, più raramente, la speranza matematica viene denotata col simbolo $\mathbb{P}[X]$

In effetti, osservando che i termini che intervengono nella (2.20) non sono altro che i possibili valori di X moltiplicati per la probabilità con cui essi vengono assunti, il significato intuitivo della speranza matematica è quello della media dei possibili valori di X. Nel caso in cui $\mathbb{E}[X] = 0$, si dice che X è *centrata*.

Si noti che la speranza matematica di una v.a. X dipende unicamente dalla sua funzione di densità: se due v.a. X e Y hanno la stessa densità allora avranno anche la stessa speranza matematica, cioè $\mathbb{E}[X] = \mathbb{E}[Y]$.

In generale, se $\mathbf{X} = (X_1, \ldots, X_m)$ è una v.a. m-dimensionale, chiameremo *speranza matematica* di $\mathbf{X}$ il vettore avente per componenti le speranze matematiche delle componenti di $\mathbf{X}$, cioè:

$$\mathbb{E}[\mathbf{X}] := (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_m])^t \, .$$

Posto $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ e $\mu_i = \mathbb{E}[X_i]$, per $i = 1, \ldots, m$, la relazione precedente possiamo scriverla:

$$\boldsymbol{\mu} \;=\; (\mu_1, \ldots, \mu_m)^t \; .$$

dove ^{t} indica vettore trasposto.

Esempio 2.12 Una *slot machine* dispone di due quadranti: in ogni quadrante possono comparire 3 diversi tipi di figure: mele, fragole, ciliege. La macchina è strutturata in modo che i due quadranti girino in modo indipendente l'uno dall'altro, e quando si arrestano compaia una delle tre figure in ognuno dei due quadranti. Indichiamo con M, F e C rispettivamente gli eventi "mela", "fragola" e "ciliegia"

Dopo aver osservato attentamente il gioco per un lungo periodo, si stabilisce che le probabilità di uscita di ogni figura sono diverse, e supponiamo che la distribuzione di probabilità sui tre eventi $\Omega = \{M, F, C\}$ sia la seguente:

$$P(M) \;=\; 0,1 \qquad\qquad P(F) \;=\; 0,4 \qquad\qquad P(C) \;=\; 0,5 \; .$$

Il gioco consiste nell'abbassare la leva che permette di far girare le figure; il risultato che comparirà sarà una delle 9 possibili coppie di figure: $\{(M,M), (M,F), (M,C), (F,M), (F,F), (F,C), (C,M), (C,F), (C,C)\}$. Il giocatore inserisce 10 centesimi di euro nella macchina e, in corrispondenza di uno dei possibili eventi, riceve le seguenti somme:

Risultato	Somma corrisposta
(M, M)	10 euro
(F, F)	2 euro
(C, C)	1 euro
figure diverse	0 euro

In base a quanto detto sopra, i risultati non sono equiprobabili; inoltre poichè gli eventi sono indipendenti si ha:

Risultato	Probabilità
(M, M)	$P(M,M) = P(M) \cdot P(M) = 0,1 \cdot 0,1 = 0,01$
(F, F)	$P(F,F) = P(F) \cdot P(F) = 0,4 \cdot 0,4 = 0,16$
(C, C)	$P(C,C) = P(C) \cdot P(C) = 0,5 \cdot 0,5 = 0,25$
figure diverse	$1 - \{P(M,M) + P(F,F) + P(C,C\} = 1 - (0,01 + 0,16 + 0,25) = 0,58$

Considerata la v.a. $X =$"guadagno in una giocata", essa è data da:

$$X \;=\; 1 \cdot |(M,M)| + 0.20 \cdot |(F,F)| + 0.10 \cdot |(C,C)| - 0.10 \; ,$$

da cui si ottiene la seguente distribuzione di probabilità di X:

$$\begin{array}{lcllcl} x_1 & = & 9 & p_1 & = & P(M,M) = 0,01 \\ x_2 & = & 1 & p_2 & = & P(F,F) = 0,16 \\ x_3 & = & 0 & p_3 & = & P(C,C) = 0,25 \\ x_4 & = & -1 & p_4 & = & 1 - \{P(M,M) + P(F,F) + P(C,C)\} = 0,58 \; . \end{array}$$

Pertanto il valore atteso di X è dato da:

$$\begin{aligned}\mathbb{E}[X] &= p_1x_1 + p_2x_2 + p_3x_3 + p_4x_4 \\ &= 0.01 \cdot 9 + 0.16 \cdot 1 + 0.25 \cdot 0 - 0.58 \cdot 1 \;=\; -0.33\ .\end{aligned}$$

Il valore ottenuto può essere interpretato dicendo che, disponendo di un capitale iniziale di 10 euro ed effettuando 100 partite, il guadagno finale sarà di -3.3 euro, cioè il capitale finale sarà all'incirca di 6.7 euro, cioè la nostra somma si sarà ridotta di 1/3. ♣

2.4.1 Proprietà della speranza matematica.

La speranza matematica possiede le seguenti proprietà:

1. Per ogni $\beta \in \mathbb{R}$, la speranza della variabile aleatoria costante $X = \beta$ coincide con la costante stessa:
$$\mathbb{E}(\beta) = \beta\ .$$

2. Se X è una v.a. avente speranza matematica finita e $\alpha, \beta \in \mathbb{R}$, allora anche la v.a. $\alpha X + \beta$ ha speranza matematica finita e risulta:
$$\mathbb{E}[\alpha X + \beta] = \alpha\,\mathbb{E}[X] + \beta\ . \tag{2.21}$$
Infatti, dalla definizione di valore atteso segue immediatamente:
$$\mathbb{E}[\alpha X + \beta] = \sum_i (\alpha x_i + \beta)p_i \;=\; \alpha \sum_i x_i p_i + \beta \sum_i p_i \;=\; \alpha\mathbb{E}[X] + \beta$$

3. Se X e Y sono due v.a. aventi speranza matematica finita, allora anche la v.a. $X \pm Y$ ha speranza matematica finita e si ha:
$$\mathbb{E}[X \pm Y] = \mathbb{E}[X] \pm \mathbb{E}[Y]\ . \tag{2.22}$$
Infatti, in base alla definizione di distribuzione marginali, nel caso discreto si ha:
$$\begin{aligned}\mathbb{E}[X \pm Y] &= \sum_i \sum_j (x_i + y_j)p_{ij} = \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij} \\ &= \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i x_i p_i + \sum_j y_j p_j \\ &= \mathbb{E}[X] + \mathbb{E}[Y]\ .\end{aligned}$$

4. Le due precedenti proprietà possono essere sintetizzate affermando che se due v.a. X e Y hanno speranza matematica finita, allora anche la v.a. $\alpha X + \beta Y$ ha speranza matematica finita e si ha:
$$\mathbb{E}[\alpha X + \beta Y] = \alpha\,\mathbb{E}[X] + \beta\,\mathbb{E}[Y]\ . \tag{2.23}$$
dove α, β sono numeri reali arbitrari.

5. Se X e Y sono v.a. indipendenti aventi speranza matematica finita, allora si ha:

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \tag{2.24}$$

Infatti, nel caso di v.a. indipendenti si ha $p_{ij} = p_i p_j$ e quindi

$$\begin{aligned}\mathbb{E}[XY] &= \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_i p_j \\ &= \sum_i x_i p_i \sum_j y_j p_j = \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

La media non è l'unico parametro che caratterizza una distribuzione; nella maggior parte dei casi, infatti, essa è insufficiente a descrivere la distribuzione ed a rappresentare le caratteristiche peculiari.

2.5 Momenti e varianza

In modo generale si può affermare che tutte le caratteristiche di una distribuzione di probabilità sono rappresentate da una famiglia di parametri chiamati *momenti*.

Definizione 2.13 Sia k un intero positivo, sia c una costante ed assumiamo che v.a. $(X-c)^k$ abbia speranza matematica finita. Si definisce *momento di ordine k* intorno al punto c la quantità:

$$\mathbb{E}[(X-c)^k] := \sum_i (x_i - c)^k p(x_i) \tag{2.25}$$

Due casi notevoli. Il primo si ottiene per $c = 0$, in tal caso si ha il *momento di ordine k*, indicato con $\mathbb{E}[X^k]$; il secondo si ottiene per $c = \mathbb{E}[X] = \mu$, in tal caso si ottiene il *momento centrale di ordine k* intorno alla media, denotato con $\mathbb{E}[(X-\mu)^k]$, sempre nell'ipotesi in cui esso esista finito.

Il momento centrale di ordine 2 costituisce un caso di particolare importanza.

Definizione 2.14 Sia X una variabile aleatoria definita su un dato spazio di probabilità $(\Omega, \mathcal{A}, P)$ avente momento secondo $\mathbb{E}[X^2]$ finito. Si definisce *varianza* di X la quantità $\text{Var}(X) := \sigma^2 = \mathbb{E}[(X-\mu)^2]$. La quantità $\sigma := \sqrt{\text{Var}(X)}$ viene chiamata *scarto quadratico medio* o *deviazione standard* della v.a. X.

Si noti che in particolare risulta:

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2 - 2\mu X + \mu^2] = \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2\ ,\end{aligned} \tag{2.26}$$

essendo $\mathbb{E}[X] = \mu$.
Sia $\mathbb{E}[X^2] < \infty$. Allora si definisce la *v.a. standardizzata* Z

$$Z := \frac{X - \mathbb{E}[X]}{\sqrt{\text{Var}(X)}} = \frac{X-\mu}{\sigma}\ . \tag{2.27}$$

Si noti che risulta $\mathbb{E}[Z] = 0$ e $\text{Var}(Z) = 1$. Inoltre la quantità $\sigma = \sqrt{\text{Var}(X)}$ viene chiamata *deviazione standard*.

2.5.1 Proprietà della varianza.

1. Se α è una costante, allora:
$$\text{Var}(\alpha) = 0 \ .$$

2. Se X è una v.a. e α, β sono delle costanti, allora la v.a. $\alpha X + \beta$ ha momento secondo finito e si ha:
$$\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X) \ . \tag{2.28}$$
Infatti, posto $\mathbb{E}[X] = \mu$, risulta:
$$\begin{aligned}\text{Var}(\alpha X + \beta) &= \mathbb{E}\{[(\alpha X + \beta) - \mathbb{E}(\alpha X + \beta)]^2\} = \mathbb{E}\{[(\alpha X + \beta) - \alpha\mu - \beta]^2\} \\ &= \mathbb{E}\{[\alpha(X - \mu)]^2\} = \mathbb{E}\{\alpha^2 (X - \mu)^2\} = \alpha^2 \text{Var}(X) \ .\end{aligned}$$

3. Se X è una v.a. avente momento secondo finito e valore atteso μ e c è una costante qualunque, allora si dimostra la seguente relazione:
$$\mathbb{E}[(X - \mu)^2] \leq \mathbb{E}[(X - c)^2] \qquad \text{per ogni } c \in \mathbb{R} \ .$$

4. Se X e Y sono due v.a. indipendenti aventi momento secondo finito, allora la v.a. $X + Y$ ha momento secondo finito e risulta:
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \tag{2.29}$$

2.6 Disuguaglianze di Markov e Cebicev

La varianza, oltre a dare un'idea della dispersione intorno al valor medio, consente di valutare quantitativamente, mediante una maggiorazione certe probabilità, anche in mancanza di ulteriori informazioni sulla distribuzione della v.a. X.

Teorema 2.15 (Disuguaglianza di Markov) Sia X una v.a. a valori non negativi. Posto $\mu = \mathbb{E}[X]$ si ha per ogni $\alpha > \mu$:
$$P(X \geq \alpha) \leq \frac{\mu}{\alpha} \ . \tag{2.30}$$

Dimostrazione. Per semplicità diamo la dimostrazione nel caso in cui X sia una v.a. discreta che assume valori $\{x_1, x_2, \ldots, x_n\}$. Se $\alpha \leq \mu$ l'asserto è banale, per cui possiamo considerare il caso $\alpha > \mu$. La tesi segue dalla definizione di valore atteso tenendo conto che
$$\mu = \sum_{i=1}^{n} p_i x_i \geq \sum_{i \,:\, x_i > \alpha} p_i x_i \geq \alpha \sum_{i \,:\, x_i > \alpha} p_i = \alpha P(X \geq \alpha) \ .$$

□

Corollario 2.16 (Disuguaglianza di Cebicev) Sia X una v.a. avente valore atteso $\mu = \mathbb{E}[X]$ e varianza $\sigma^2 = \text{Var}(X)$. Allora si ha per ogni $k \geq 1$:
$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \ .$$

Dimostrazione. La dimostrazione segue applicando la disuguaglianza di Markov (2.30) alla v.a. $Y = (X - \mu)^2$ con $\alpha = (k\sigma)^2$. Tenendo conto che i due eventi $\{|X - \mu| \geq k\sigma\}$ e $\{(X - \mu)^2 \geq (k\sigma)^2\}$ coincidono e che $\mathbb{E}[Y] = \mathbb{E}[(X - \mu)^2] = \sigma^2$, segue:

$$P(|X - \mu| \geq k\sigma) \equiv P((X - \mu)^2 \geq (k\sigma)^2) \leq \frac{\mathbb{E}[Y]}{k^2\sigma^2} = \frac{1}{k^2}\,.$$

□

Esempio 2.17 In un grande supermercato vi sono 30 casse. In un qualunque giorno di sabato, l'incasso globale di una di tali casse può essere assimilato ad una variabile aleatoria avente distribuzione non nota con valor medio $\mu = 2000$ euro e scarto quadratico medio $\sigma = 250$ euro (stimati in base a dati ricavati in precedenza). Valutare la probabilità che l'incasso globale della cassa n. 6 sia compreso fra 1500 e 2500 euro.

Sia X_i la quantità aleatoria "incasso di una cassa durante la giornata della i-esima cassa", con $i = 1, \ldots, 30$. In mancanza di ulteriori informazioni, si può ritenere che le variabili X_i siano indipendenti ed identicamente distribuite, anche se la distribuzione non è nota. Il quesito si risolve mediante un'applicazione della disuguaglianza di Cebycev:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}\,.$$

Infatti si richiede la probabilità che l'incasso sia compreso fra 1500 e 2500 euro, cioè $P(1500 \leq X \leq 2500)$. Tale quantità può essere scritta come

$$P(1500 \leq X \leq 2500) = P(|X - 2000| \leq 500) \tag{2.31}$$

e poichè $\sigma = 250$ euro, allora dalla (2.31), segue $500 = 2 \cdot \sigma$. Pertanto risulta

$$P(|X - 2000| \leq 500) = P(|X - 2000| \leq 2 \cdot \sigma) \geq 1 - \frac{1}{2^2} = 0.75\,.$$

♣

2.7 Covarianza e Correlazione.

Definizione 2.18 Siano X, Y due v.a. aventi varianza finita, ed indichiamo con μ_X e μ_Y le rispettive speranze matematiche. Si definisce *covarianza* di X e Y, e si indica col simbolo $\mathrm{Cov}(X, Y)$, la quantità

$$\mathrm{Cov}(X, Y) := \mathbb{E}\{[X - \mu_X][Y - \mu_Y]\}\,. \tag{2.32}$$

Si nota che la covarianza è uguale alla speranza del prodotto degli scarti delle variabili aleatorie X e Y dalle rispettive speranze, analogamente al caso statistico. In particolare si ha:

$$\mathrm{Cov}(X, X) = \mathbb{E}\{[X - \mu_X][X - \mu_X]\} = \mathbb{E}\{[X - \mu_X]^2\} = \mathrm{Var}(X)\,.$$

Dalla definizione seguono subito due proprietà:

$$\begin{aligned} \text{Cov}(X, \alpha) &= \mathbb{E}\{[X - \mu_X][\alpha - \alpha]\} = 0 \qquad (2.33) \\ \text{Cov}(\alpha X, \beta Y) &= \mathbb{E}\{[\alpha X - \alpha\mu_X][\beta Y - \beta\mu_Y]\} \\ &= \alpha\beta\mathbb{E}\{[X - \mu_X][Y - \mu_Y]\} = \alpha\beta\text{Cov}(X, Y)\ . \qquad (2.34) \end{aligned}$$

Inoltre, svolgendo la (2.32), si ottiene:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}\{[X - \mu_X][Y - \mu_Y]\} = \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= \mathbb{E}[XY] - \mu_X\mathbb{E}[Y] - \mu_Y\mathbb{E}[X] + \mu_X\mu_Y \\ &= \mathbb{E}[XY] - \mu_X\mu_Y\ . \qquad (2.35) \end{aligned}$$

Dalla (2.24), segue che se X e Y sono stocasticamente indipendenti, allora $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ e quindi si ha $\text{Cov}(X, Y) = 0$.

La covarianza di X e Y è definita se le due v.a. hanno varianza finita, ed in tal caso si dimostra che sussiste la seguente relazione:

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\,\text{Var}(Y)} \qquad (2.36)$$

dove l'uguaglianza vale solo in uno dei due seguenti casi:

1. se almeno una delle due v.a., X o Y, è costante (ed in questo caso si ha $0 = 0$);
2. se X è funzione lineare di Y, cioè sono legati da una relazione del tipo $Y = \alpha X + \beta$.

Infatti, se $Y = \alpha$ =costante, allora la prima affermazione segue dalla (2.33). Nell'altro caso, se consideriamo la v.a. $\alpha X + Y$, si ha:

$$\text{Var}(\alpha X + Y) = \alpha^2\text{Var}(X) + \text{Var}(Y) + 2\alpha\text{Cov}(X, Y)$$

e dovendo risultare $\text{Var}(\alpha X + Y) \geq 0$ per ogni α, deve risultare:

$$\text{Cov}^2(X, Y) - \text{Var}(X)\text{Var}(Y) \leq 0$$

da cui segue la (2.36).

2.7.1 Varianza della somma di variabili aleatorie

Siano X, Y due v.a. aventi varianza finita e speranze matematiche rispettivamente uguali a μ_X e μ_Y. Si ha allora:

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[\{(X + Y) - (\mu_X + \mu_Y)\}^2] = \mathbb{E}[\{(X - \mu_X) + (Y - \mu_Y)\}^2] \\ &= \mathbb{E}[(X - \mu_Y X)^2] + \mathbb{E}[Y - \mu_Y)^2] + \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)\ . \qquad (2.37) \end{aligned}$$

Più in generale, si può dimostrare la relazione:

$$\text{Var}(\alpha X + \beta Y) = \alpha^2\text{Var}(X) + \beta^2\text{Var}(Y) + 2\alpha\beta\text{Cov}(X, Y)\ . \qquad (2.38)$$

In particolare, nel caso in cui X e Y sono v.a. indipendenti risulta $\text{Cov}(X,Y) = 0$ e quindi si ritrova la (2.29).

La (2.37) si estende al caso n dimensionale come segue. Se $X, \ldots, X_m$ sono v.a. aventi varianza finita, allora si ha

$$\text{Var}(X_1 + \cdots + X_m) = \sum_{i=1}^{m} \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j) \ . \tag{2.39}$$

In particolare, se le m v.a. sono indipendenti, allora risulta $\text{Cov}(X_i, X_j) = 0$ per $i \neq j$ e quindi dalla (2.39) segue:

$$\text{Var}(X_1 + \cdots + X_m) = \sum_{i=1}^{m} \text{Var}(X_i) \ . \tag{2.40}$$

Tale relazione generalizza la (2.29).

In generale, posto $\sigma_{ij} = \text{Cov}(X_i, X_j)$ e $\sigma_{ii} = \text{Var}(X_i)$, per $i, j = 1, \ldots, m$, si definisce *matrice di varianze e covarianze* la matrice:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{pmatrix} \ .$$

Definizione 2.19 Date due v.a. X e Y, aventi varianza non nulla, si definisce *coefficiente di correlazione lineare* di X e Y, e si indica col simbolo ρ, la quantità

$$\rho \ = \ \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \ . \tag{2.41}$$

dove $\sigma_X = \sqrt{\text{Var}(X)}$ e $\sigma_Y = \sqrt{\text{Var}(Y)}$.

In altre parole, in base alla (2.36), il coefficiente di correlazione è uguale alla covarianza normalizzata, cioè la covarianza diviso il valore massimo teoricamente possibile. Ne segue che si ha $-1 \leq \rho \leq 1$.

Due proprietà del coefficiente di correlazione:

1. $\rho(X,Y)$ è invariante per scala, cioè risulta $\rho(\alpha X, \beta Y) = \rho(X,Y)$ per ogni $\alpha, \beta \in \mathbb{R}$. Infatti, in base alle (2.28) e (2.34), segue:

$$\begin{aligned} \rho(\alpha X, \beta Y) \ &= \ \frac{\text{Cov}(\alpha X, \beta Y)}{\sqrt{\text{Var}(\alpha X)\text{Var}(\beta Y)}} \ = \ \frac{\alpha\beta \cdot \text{Cov}(X,Y)}{\sqrt{\alpha^2 \text{Var}(X) \cdot \beta^2 \text{Var}(Y)}} \\ &= \ \frac{\alpha\beta \cdot \text{Cov}(X,Y)}{\alpha\beta\sqrt{\text{Var}(X)\text{Var}(Y)}} \ = \ \rho(X,Y) \ . \end{aligned}$$

2. Risulta $|\rho(X,Y)| = 1$ se e solo se esistono $\alpha, \beta \in \mathbb{R}$ tali che $Y = \alpha X + \beta$.
   Infatti, sia $\rho(X,Y) = 1$, allora in base alla definizione vuol dire che $\text{Cov}^2(X,Y) = \text{Var}(X)\text{Var}(Y)$ e quindi in base a quanto detto prima, segue l'asserto.
   Al contrario, se $Y = \alpha X + \beta$, allora $\text{Cov}^2(X,Y) = \alpha^2 \text{Var}^2(X)$ e $\text{Var}(Y) = \alpha^2 \text{Var}(X)$.

Ovviamente si può considerare la *matrice di correlazione*. Posto

$$r_{ij} := \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

possiamo scrivere:

$$\mathbf{R} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{pmatrix} ,$$

dove ovviamente risulta:

$$r_{11} = r_{22} = \cdots = r_{mm} = 1 .$$

Esempio 2.20 Si prendano in considerazione due diverse possibilità di giocare 100 euro al lotto, supponendo che per ogni euro giocato si abbia una speranza pari a 80 centesimi con una deviazione standard di 50 centesimi. La prima possibilità, indicata con A, consiste nel giocare rispettivamente 20, 30 e 50 euro su tre ruote s, t, u scelte a caso. La seconda possibilità, indicata con B, consiste nel giocare tutta la somma di 100 euro su una sola ruota v estratta a caso. Qual è, fra le due possibilità A e B, quella meno rischiosa?

Innanzitutto è bene osservare che le due strategie di gioco A e B hanno la stessa speranza matematica. Infatti, denotato con X la v.a. "guadagno in una giocata per una puntata pari a un euro", in base ai dati del problema risulta $\mathbb{E}[X] = 0.8$ e $\text{Var}(X) = 0.5$. Denotate poi con A e B i guadagni ottenuti rispettivamente in base alle due strategie A e B, si ha:

$$\begin{aligned} A &:= 20X_s + 30X_t + 50X_u \\ B &:= 100X_v \end{aligned}$$

dove X_s, X_t, X_u e X_v sono le quantità aleatorie "guadagno ricavato da una giocata sulla ruota considerata (s, t, u, v)" aventi tutte la stessa speranza matematica e la stessa varianza. da cui

$$\begin{aligned} \mathbb{E}[A] &= 20\mathbb{E}[X_s] + 30\mathbb{E}[X_t] + 50\mathbb{E}[X_u] = 20 \cdot 0.8 + 30 \cdot 0.8 + 50 \cdot 0.8 = 80 \\ \mathbb{E}[B] &= 100 \cdot \mathbb{E}[X_v] = 100 \cdot 0.8 = 80 . \end{aligned}$$

Sono invece diverse invece le varianze. Dalla (2.40) si ha infatti, tenento conto che $\text{Var}(X) = 0.5^2 = 0.25$:

$$\begin{aligned} \text{Var}(A) &= \text{Var}(20X_s + 30X_t + 50X_u) = 20^2\text{Var}(X_s) + 30^2\text{Var}(X_t) + 50^2\text{Var}(X_u) \\ &= 400 \cdot 0.25 + 900 \cdot 0.25 + 2500 \cdot 0,25 = 950 \\ \text{Var}(B) &= \text{Var}(100X_v) = 10000\text{Var}(X_v) = 10000 \cdot 0.25 = 2500 . \end{aligned}$$

Quindi, sebbene i due tipi di gioco tendano, per un elevato numero di giocate, a far perdere 20 euro per ogni 100 euro giocati, questa perdita si realizza in modo diverso nella strategia A da quella che si realizzarebbe nella strategia B. Il metodo consiste nel dividere la somma che si vuole giocare e giocarla in più volte riduce considerevolmente la varianza del gioco,

ed è inferiore per puntate multiple di somme piccole piuttosto che una sola puntata elevata, riducendo quindi il rischio del gioco.

Come caso limite, se si dividono i 100 euro in 100 parti uguali, si avrà una varianza di vincita 100 volte più piccola, cioè una deviazione standard di 10 volte più piccola di quella che si avrebbe giocando tutta la somma in un sol colpo. ♣

Esercizio 2.21 Sia $\Omega = \{a, b, c\}$ e P la distribuzione di probabilità su Ω definita da:

$$P(a) \,=\, \frac{1}{2}\,, \qquad\qquad P(b) \,=\, P(c) \,=\, \frac{1}{4}\,.$$

Siano X e Y la variabili aleatorie definite da:

$$X \,=\, \mathbb{1}_{\{a\}} - \mathbb{1}_{\{b,c\}}\,, \qquad\qquad Y \,=\, \mathbb{1}_{\{b\}} - \mathbb{1}_{\{c\}}\,,$$

dove $\mathbb{1}_A(\omega)$ è la funzione indicatrice, cioè $\mathbb{1}_A(\omega) = 1$ se $\omega \in A$ e $\mathbb{1}_A(\omega) = 0$ se $\omega \notin A$. Dimostrare che X e Y non sono stocasticamente indipendenti e che si ha $\mathrm{Cov}(X, Y) = 0$.

Soluzione. In base alle ipotesi del problema, si ha $X \in \{-1, 1\}$ e $Y \in \{-1, 0, 1\}$, in particolare:

$$\begin{aligned}
\text{evento } a &\;\Rightarrow\; X = 1\,, Y = 0 \\
\text{evento } b &\;\Rightarrow\; X = -1\,, Y = 1 \\
\text{evento } c &\;\Rightarrow\; X = -1\,, Y = -1\,.
\end{aligned}$$

Si ha pertanto la seguente distribuzione di probabilità della variabile congiunta (X, Y):

		Y		
X	-1	0	$+1$	
-1	1/4	0	1/4	1/2
$+1$	0	1/2	0	1/2
	1/4	1/2	1/4	1

Ovviamente X e Y non sono stocasticamente indipendenti in quanto

$$P(X = -1, Y = 0) \,=\, 0 \;\neq\; P(X = -1) \cdot P(Y = 0) \,=\, \frac{1}{2} \cdot \frac{1}{2} \,=\, \frac{1}{4}\,.$$

Invece X e Y sono non correlate in quanto risulta $\mathrm{Cov}(X, Y) = 0$. Infatti:

$$\begin{aligned}
\mathbb{E}[XY] \;&=\; \frac{1}{4} \cdot (-1) \cdot (-1) + \frac{1}{2} \cdot 1 \cdot 0 + \frac{1}{4} \cdot (-1) \cdot 1 \,=\, 0 \\
\mathbb{E}[X] \;&=\; -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} \,=\, 0 \\
\mathbb{E}[Y] \;&=\; -1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} \,=\, 0
\end{aligned}$$

da cui

$$\mathrm{Cov}(X, Y) \;=\; \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \,=\, 0\,.$$

♣

2.8 Altri esercizi

Esercizio 2.22 n commensali si ripartiscono a caso intorno ad una tavola rotonda. Sia X il più piccolo numero (aleatorio) di commensali che separa Pietro e Paolo. Calcolare la legge di X e la speranza matematica di X.

Soluzione. Sia n il numero totale di commensali, distinguiamo i due casi: n dispari e n pari.

Se n è dispari, allora possiamo scrivere $n = 2p + 1$ dove p è un numero dispari, con $p \geq 1$. In questo caso X è una v.a. con legge uniforme su $\{0, 1, \ldots, p-1\}$ e pertanto:

$$\mathbb{E}[X] \;=\; \frac{p}{2} \qquad\qquad \text{Var}(X) \;=\; \frac{p(p-2)}{12}\,.$$

Se n è pari, allora possiamo scrivere $n = 2p$, dove p è un numero dispari; in questo caso si ha:

$$\begin{aligned} P(X = p-1) &= \frac{1}{n-1} \;=\; \frac{1}{2p-1} \\ P(X = k) &= \frac{2}{2p-1} \;=\; \frac{2}{2p-1} \qquad \text{per } 1 \leq k \leq p-2 \end{aligned}$$

da cui, tenendo conto che

$$1 + 2 + \cdots + (p-2) \;=\; \frac{(p-1)(p-2)}{2}$$

$$\mathbb{E}[X] \;=\; \frac{p-1}{2p-1} + \frac{(p-1)(p-2)}{2}\frac{2}{2p-1} \;=\; \frac{(p-1)^2}{2p-1}\,.$$

♣

Esercizio 2.23 Completare la seguente distribuzione di probabilità congiunta nell'ipotesi di indipendenza stocastica delle v.a. X e Y:

	x_1	x_2	x_3	
y_1	*	*	*	*
y_2	0.06	*	*	0.3
y_3	*	0.18	*	*
	*	0.3	*	1

Soluzione.

	x_1	x_2	x_3	
y_1	0.02	0.03	0.05	0.1
y_2	0.06	0.09	0.15	0.3
y_3	0.12	0.18	0.3	0.6
	0.2	0.3	0.5	1

♣

Esercizio 2.24 Completare la seguente distribuzione di probabilità congiunta nell'ipotesi di indipendenza stocastica delle v.a. X e Y:

	x_1	x_2	x_3	
y_1	*	0.2	*	0.4
y_2	*	*	0.105	*
y_3	*	*	*	*
	*	*	0.3	1