

SAPIENZA Università di Roma
Anno Accademico 2009-10

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI
CORSO DI LAUREA TRIENNALE IN MATEMATICA

**INTRODUZIONE AL
CALCOLO DELLE PROBABILITÀ**

(Versione del 31 maggio 2010, ancora incompleta e provvisoria)

Fabio Spizzichino e Giovanna Nappo

A.A. 2009/10

Indice

Introduzione	iv
1 Fenomeni aleatori; spazio dei risultati elementari di un esperimento	1
1.1 Operazioni logiche su eventi e interpretazione insiemistica	1
1.2 Esercizi di verifica	4
2 Spazi finiti di probabilità	6
2.1 Prime definizioni e proprietà	6
2.2 Esercizi di verifica	12
3 Probabilità “classiche” e calcolo combinatorio	13
3.1 Probabilità “classiche”	13
3.2 Calcolo combinatorio: primi elementi	14
3.3 Alcuni classici esempi	15
3.4 Alcune proprietà dei coefficienti binomiali.	20
3.5 Approfondimenti sul calcolo combinatorio	23
3.6 Esercizi di verifica	27
4 Probabilità condizionate	29
4.1 Definizione di probabilità condizionata	29
4.2 Conseguenze immediate della definizione di probabilità condizionata	31
4.2.1 Formula delle probabilità composte	31
4.2.2 Formula delle probabilità totali	32
4.2.3 Formula di Bayes	34
4.3 Esercizi di verifica	38
5 Correlazione e indipendenza fra eventi	39
5.1 Il caso di due eventi: correlazione positiva, negativa e indipendenza	39
5.2 Indipendenza fra partizioni e fra algebre di eventi	41
5.3 Indipendenza completa e prove bernoulliane	43
5.4 Indipendenza completa di partizioni	45
5.5 Esercizi di verifica	47
6 Probabilità binomiali e ipergeometriche; estrazioni casuali da urne	49
6.1 Probabilità binomiali	49
6.2 Estrazioni casuali da urne con reiserimento	50
6.3 Estrazioni casuali da urne senza reiserimento e Probabilità ipergeometriche	52
6.4 Esercizi di verifica	56
7 Variabili aleatorie e distribuzioni di probabilità	57
7.1 Esercizi di verifica	63
8 Distribuzioni congiunte di più variabili aleatorie	65
8.1 Indipendenza stocastica fra variabili aleatorie.	71
8.2 Esercizi di verifica	73
8.2.1 Soluzione di alcuni esercizi importanti	73

9	Valore atteso di una variabile aleatoria e relative proprietà	76
9.1	Valore atteso condizionato e valore atteso ATTENZIONE Sezione NUOVA	90
9.2	Esercizi di verifica	92
10	Varianza, Covarianza e comportamento delle medie aritmetiche di variabili aleatorie	95
10.1	Diseguaglianza di Cauchy e coefficiente di correlazione	104
10.2	Appendice: Covarianza della somma di n variabili aleatorie	106
10.3	Esercizi di verifica	107
11	Campionamento da popolazioni con composizione incognita; indipendenza condizionata	108
11.1	Caso estrazioni casuali senza reinserimento e con distribuzione binomiale per R	111
11.2	Esempi	113
11.3	Indipendenza condizionata	115
11.4	Esercizi di verifica	121
12	Modelli di occupazione e schemi di estrazioni da urne	123
12.1	Modello di Maxwell-Boltzmann	126
12.2	Modello di Bose-Einstein	127
12.3	Modello di Fermi-Dirac	127
12.4	Schemi di estrazioni da urne	127
12.5	Alcuni esempi	130
12.6	Distribuzione multinomiali	132
12.7	Distribuzioni marginali e condizionate nei modelli di occupazione	134
12.8	Distribuzioni marginali e condizionate per la distribuzione multinomiale	136
12.9	Esercizi di verifica	138
13	Spazi di probabilità e variabili aleatorie in casi più generali	140
13.0.1	Definizione generale di spazio di probabilità	140
13.1	Definizione generale di variabile aleatoria	146
13.2	Distribuzioni di probabilità, funzioni di distribuzione	148
13.3	Funzioni di distribuzione continue, funzioni di densità di probabilità	156
13.4	Valori attesi per variabili aleatorie generali	162
13.5	Esempi svolti	165
13.6	Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini	169
13.6.1	Il caso delle trasformazioni affini	170
	Grafico della funzione di distribuzione e della densità di una gaussiana standard	176
	Tavola della funzione di distribuzione gaussiana standard	177
14	Variabili aleatorie in casi più generali: indipendenza, Legge dei Grandi Numeri e Teorema Centrale del Limite.	179
14.1	Famiglie di variabili aleatorie indipendenti	179
14.2	Legge dei Grandi Numeri	181
14.2.1	Approfondimenti sull'utilizzo della disuguaglianza di Chebyshev	181
14.2.2	Formulazione della Legge dei Grandi Numeri	185
14.3	Somma di variabili aleatorie indipendenti e Teorema Centrale del Limite	186
14.3.1	Esempi di calcolo della somma di variabili aleatorie indipendenti	186
14.3.2	Approssimazione normale e Teorema Centrale del Limite	187

14.3.3 Altre conseguenze del Teorema Centrale del Limite e relazioni con la legge dei grandi numeri 190

Alfabeto greco **194**

Introduzione

L'introduzione sarà scritta in seguito.

Si tratta di un primo aggiustamento della versione di febbraio 2010. Le correzioni rispetto alla versione precedente sono evidenziate con vari colori.

Si ricorda comunque che questi appunti non sono completi, anche se abbiamo cominciato a colmare alcune lacune: ad esempio è stata scritta una parte sul valore atteso di variabili aleatorie con densità tuttavia manca ancora un'introduzione alle catene di Markov. Comunque gli argomenti mancanti e ulteriori approfondimenti si possono trovare sui testi consigliati:

Mauro Piccioni, "Probabilità di base "(Aracne 2010)

Giorgio Dall'Aglio, "Calcolo delle Probabilità"

(Zanichelli 2002)

Paolo Baldi, "Calcolo delle Probabilità e Statistica"(Mac Graw Hill 200?)

Su tali testi è possibile trovare anche altri esempi, applicazioni e dimostrazioni degli argomenti presentati.

NOTAZIONI.

Si è cercato di mettere sempre le parentesi graffe per gli eventi e per le combinazioni (per distinguerle dalle disposizioni, che invece mantengono le parentesi tonde).

Inoltre si è cercato di evitare la scrittura che usa il simbolo $|$ per significare *tale che*, del tipo

$$\{\omega \in \Omega | X(\omega) \leq x\}$$

sostituendola con la scrittura che usa i due punti, ovvero con

$$\{\omega \in \Omega : X(\omega) \leq x\},$$

per evitare possibili confusioni con il segno $|$ che invece si riferisce alle probabilità condizionate.

Il complementare di un evento E , pensato come sottoinsieme di Ω , viene indicato sempre (si spera) con \bar{E} .

Va infine detto che le correzioni ed alcune lezioni sono a cura di Giovanna Nappo, che Fabio Spizzichino non ha avuto tempo di rivedere tutte le correzioni fatte e che ovviamente le correzioni potrebbero contenere altri errori....

Le segnalazioni degli errori sono sempre **molto gradite**, grazie.

1 Fenomeni aleatori; spazio dei risultati elementari di un esperimento

1.1 Operazioni logiche su eventi e interpretazione insiemistica

Iniziamo con una discussione euristica mirante a giustificare la successiva definizione della nozione di *spazio finito di probabilità*, che verrà data nella prossima lezione in un caso particolare.

Come punto di partenza, pensiamo ad un esperimento che possa dar luogo a diversi risultati possibili. I risultati verranno chiamati “eventi”.

Possiamo vedere un evento come una proposizione relativa al modo di risultare di tale esperimento.

Esempio 1.1. Consideriamo l'esperimento consistente nell'osservazione dei punti ottenuti dal lancio di una coppia di dadi a sei facce.

Indichiamo tali punti con i simboli X_1, X_2 .

Esempi di possibili eventi sono: $A \equiv \{X_1 \leq X_2\}$, $B \equiv \{X_1 + X_2 \text{ pari}\}$, $C \equiv \{X_1 > 3\}$, ...

Indichiamo, per il momento, con il simbolo \mathcal{E} la famiglia dei possibili eventi distinti in un esperimento. Come è facile rendersi conto (e verificheremo presto), la famiglia \mathcal{E} costituita da tutti gli eventi nel precedente Esempio 1.1 è una famiglia finita. In quanto immediatamente segue, ci limiteremo ancora a considerare esperimenti per cui \mathcal{E} è una famiglia finita; successivamente, tale limitazione verrà eliminata.

È anche facile rendersi conto che, all'interno della famiglia \mathcal{E} , è naturale introdurre le operazioni di *somma logica* (oppure *or*), di *prodotto logico* (oppure *and*) e di *negazione* (oppure *not*), che verranno rispettivamente indicate (per il momento) con i simboli $\vee, \wedge, \tilde{\cdot}$; siano E_1, E_2, E eventi appartenenti ad \mathcal{E} , allora la somma logica $E_1 \vee E_2$ coincide con l'evento:

$$\{\text{si è verificato almeno uno dei due eventi } E_1 \text{ ed } E_2\}$$

il prodotto logico $E_1 \wedge E_2$ coincide con l'evento:

$$\{\text{si sono verificati entrambi i due eventi } E_1 \text{ ed } E_2\}$$

la negazione \tilde{E} coincide con l'evento:

$$\{\text{non si è verificato l'evento } E\}.$$

Definizione 1.1. Un evento $E \in \mathcal{E}$ si dice **composto** se esistono almeno due eventi $E_1, E_2 \in \mathcal{E}$, tali che

$$E = E_1 \vee E_2, E \neq E_1, E \neq E_2.$$

Un evento che non sia composto si dice **semplice** o **elementare**.

Esempio 1.2. *Nell'esperimento del lancio di un dado a sei facce, l'evento $E = \{X_1 > 3\}$ è un evento composto. In tale esperimento gli eventi semplici sono dati da*

$$\{X_1 = 1\}, \{X_1 = 2\}, \dots, \{X_1 = 6\},$$

e l'evento E si riscrive come $\{X_1 > 3\} \vee \{X_1 = 4\} \vee \{X_1 = 5\} \vee \{X_1 = 6\}$.

Nell'esperimento del lancio di una coppia di dadi gli eventi semplici sono invece quelli del tipo

$$\{X_1 = h, X_2 = k\} \quad h = 1, 2, \dots, 6; k = 1, 2, \dots, 6$$

ed un evento del tipo $\{X_1 = h\}$ risulta essere un evento composto, in quanto possiamo scrivere

$$\{X_1 = h\} = \bigvee_{k=1}^6 \{X_1 = h, X_2 = k\}.$$

Osservazione 1. È facile verificare che $E \in \mathcal{E}$ composto si decompone in uno ed in un sol modo (a meno dell'ordine) come somma logica di un numero finito di eventi elementari.

Indichiamo ora con i simboli $\omega_1, \dots, \omega_N$ gli eventi elementari in un esperimento.

Definizione 1.2. $\Omega \equiv \{\omega_1, \dots, \omega_N\}$ che ha come punti gli eventi elementari di un esperimento viene detto **spazio campione**, per quell'esperimento.

Indichiamo con il simbolo $\mathcal{P}(\Omega)$ la **famiglia delle parti** di Ω (ossia la famiglia dei sottoinsiemi di Ω) e, per $E \in \mathcal{P}(\Omega)$, indichiamo con $|E|$ la **cardinalità** di E .

Esempio 1.3. *Un'urna inizialmente contiene quattro oggetti numerati da 1 a 4. Vuotiamo l'urna facendo quattro successive estrazioni senza reinserimento, osservando di volta in volta il numero indicato sull'oggetto estratto.*

Si ha $\Omega = \{\text{permutazioni}^1 \text{ di } \{1, 2, 3, 4\}\}$, e $|\Omega| = 24$.

Noi vogliamo analizzare quei casi in cui vi sia una *situazione di incertezza* (cioè di mancanza di completa informazione) circa il modo di risultare dell'esperimento stesso. Ciò significa che non sappiamo a priori quale effettivamente si realizzerà fra i diversi risultati elementari possibili. In tali casi parleremo di fenomeni aleatori o di *esperimenti aleatori*.

Parleremo dunque di esperimento aleatorio quando non sappiamo quali eventi saranno verificati e quali risulteranno falsi.

In tale ambito, un evento composto E è verificato se e solo se si verifica un evento elementare ω_i che si presenti nella decomposizione di E .

Esempio 1.4. *Si lancia un dado a sei facce; si ha $\Omega = \{\omega_1, \dots, \omega_6\}$. Supponiamo si verifichi ω_4 , allora saranno anche verificati, ad esempio, gli eventi composti: $\{X \leq 5\}$, $\{X \text{ pari}\}$, $\{X > 2\}$ e non sono verificati, ad esempio, gli eventi $\{X > 5\}$, $\{X \text{ dispari}\}$, $\{X \leq 3\}$, $\{X \text{ numero primo}\}$, ...*

¹Per la definizione di permutazione vedere più avanti la Sezione 3.2

Dati due eventi $E_1, E_2 \in \mathcal{E}$, dunque,

(a) $E_1 \vee E_2$ si verifica se e solo se è verificato un evento elementare ω_i che si presenti nella decomposizione di E_1 e/oppure di E_2

(b) $E_1 \wedge E_2$ si verifica se e solo se è verificato un evento elementare ω_i che si presenti sia nella decomposizione di E_1 che in quella di E_2

(c) \tilde{E}_1 si verifica se e solo se è verificato un evento elementare ω_i che non sia presente nella decomposizione di E_1

Osservazione 2 (Eventi come sottoinsiemi di Ω). Per definizione, i punti dello spazio Ω sono gli “eventi semplici” o “risultati elementari” dell’esperimento considerato.

Notiamo ora che sussiste una corrispondenza biunivoca fra sottoinsiemi di Ω , costituiti da più di un elemento, e gli eventi composti: basta infatti associare, ad un evento composto, l’insieme costituito dagli eventi semplici che lo compongono; viceversa ad un sottoinsieme di Ω possiamo associare l’evento composto che si ottiene come somma logica degli elementi (eventi semplici) in esso contenuti.

Ad un evento semplice $\omega_i \in \Omega$, facciamo corrispondere il singleton $\{\omega_i\} \in \mathcal{P}(\Omega)$.

Dato un evento $E \in \mathcal{E}$, indichiamo per comodità con $\mathcal{H}(E)$ il sottoinsieme di Ω individuato secondo quanto appena detto.

Osservazione 3 (Operazioni su eventi e operazioni su sottoinsiemi). Consideriamo di nuovo la corrispondenza biunivoca \mathcal{H} fra eventi e sottoinsiemi di Ω , stabilita nella precedente **Osservazione 2**:

$$\mathcal{E} \xleftrightarrow{\mathcal{H}} \mathcal{P}(\Omega).$$

Ci si rende facilmente conto, da quanto detto sopra, che, in tale corrispondenza biunivoca fra eventi e sottoinsiemi, le operazioni \vee, \wedge, \sim (definite su \mathcal{E}) vengono rispettivamente trasformate nelle operazioni booleane di unione \cup , di intersezione \cap , e di passaggio al complementare \complement (definite su $\mathcal{P}(\Omega)$, la famiglia delle parti di Ω); infatti, traducendo “in formule” i precedenti punti (a), (b) e (c) potremo scrivere, per degli arbitrari $E_1, E_2, E \in \mathcal{E}$,

$$\mathcal{H}(E_1 \vee E_2) = \mathcal{H}(E_1) \cup \mathcal{H}(E_2),$$

$$\mathcal{H}(E_1 \wedge E_2) = \mathcal{H}(E_1) \cap \mathcal{H}(E_2),$$

$$\mathcal{H}(\tilde{E}) = \complement(\mathcal{H}(E)).$$

Da questo momento in poi, quindi, potremo identificare “eventi” e sottoinsiemi di Ω e dunque lasceremo cadere l’uso dei simboli $\mathcal{E}, \vee, \wedge, \sim, \mathcal{H}(\cdot)$; continueremo la trattazione utilizzando solo le nozioni di sottoinsieme di Ω e di operazioni booleane fra sottoinsiemi.

Dovremo però continuare ad aver presente il significato di tipo “logico” che stiamo dando a tali nozioni, nel contesto dell’analisi di fenomeni aleatori. In tale ambito, risulterà naturale attribuire un’interpretazione di tipo “logico” a varie semplici nozioni di tipo insiemistico; a tale proposito vediamo intanto lo specchio presentato qui di seguito.

Interpretazione “logica” di nozioni di tipo insiemistico:

· $A \subseteq B$ significa che ogni evento elementare che rende verificato A rende verificato anche B e dunque interpretiamo la relazione $A \subseteq B$ come “ A **implica** B ”

· Ω è un evento vero qualunque evento elementare si verifichi, in quanto esso contiene tutti gli eventi elementari e dunque interpretiamo Ω come l’**evento certo**

... \emptyset , **l'insieme vuoto**, non contenendo alcuno degli eventi elementari possibili, è un evento che non è mai verificato; dunque interpretiamo \emptyset come **l'evento impossibile**

... $A \cup B = \Omega$ significa che l'evento costituito dal verificarsi di almeno uno dei due eventi A o B coincide con l'evento certo Ω ; dunque interpretiamo tale condizione come **A e B sono esaustivi** (è certo che se ne verifichi **almeno** uno dei due)

... $A \cap B = \emptyset$ significa che l'evento costituito dal verificarsi di entrambi gli eventi A e B coincide con l'evento impossibile \emptyset ; dunque interpretiamo la condizione $A \cap B = \emptyset$ come **A e B sono incompatibili** (è certo che se ne verifichi **al più** uno dei due).

Terminiamo questa lezione analizzando il significato "logico" della nozione di **partizione dell'evento certo**.

Consideriamo una collezione di sottoinsiemi H_1, \dots, H_m dello spazio Ω ($H_l \in \mathcal{P}(\Omega), l = 1, \dots, m$). Tale collezione costituisce una **partizione di Ω** se e solo se

$$\bigcup_{l=1}^m H_l = \Omega; \quad H_{l_1} \cap H_{l_2} = \emptyset, \text{ per } l_1 \neq l_2.$$

Interpretando H_1, \dots, H_m come eventi, abbiamo che essi sono a **due a due incompatibili** (cioè è impossibile che se ne possano verificare due contemporaneamente) e, d'altra parte, essi sono **esaustivi** (è certo che se ne verifichi almeno uno); dunque: è certo che si verifichi uno ed uno soltanto degli eventi H_1, \dots, H_m .

1.2 Esercizi di verifica

Esercizio 1.1. Consideriamo l'esperimento consistente nel lancio di una pallina nel gioco della roulette. In tale esperimento è naturale porre

$$\Omega \equiv \{0, 1, \dots, 36\}$$

e vedere i risultati *manque, passe, noir, rouge, pair, unpair*, come altrettanti eventi composti. Supponiamo che nell'esperimento si verifichi l'evento elementare $\{16\}$. Quale degli eventi composti sopra elencati è verificato e quale no?

Come funziona la roulette

Si ricorda che la Roulette è una ruota con trentasette settori numerati da zero a trentasei. Una pallina viene fatta girare e alla fine si ferma su uno di questi numeri. Inoltre puntare su *manque* significa puntare su un numero tra 1 e 18, che puntare su *passe* significa puntare su un numero tra 19 e 36, puntare su *noir* significa puntare su un numero nero, puntare su *rouge* significa puntare su un numero rosso, ed analogamente per *pair*, ovvero pari e *unpair*, ovvero dispari. Ai fini della soluzione dell'esercizio, è importante sapere che il 16 è rosso.

Esercizio 1.2. Dati due eventi A e B , scrivete, in termini di operazioni booleane, l'espressione dell'evento:

{si verifica esattamente un solo evento fra A e B }.

Esercizio 1.3. Siano A , B e C eventi. Scrivete le espressioni degli eventi:

- Almeno due tra questi si verificano;
- Esattamente due tra questi si verificano;
- Al più due tra questi si verificano;
- Esattamente uno tra questi si verifica.

Esercizio 1.4. Un'urna contiene oggetti di tipo A ed oggetti di tipo B ; si eseguono due successive estrazioni dall'urna e si definiscono, per $i = 1, 2$, gli eventi:

$$E_i = \{\text{oggetto di tipo } A \text{ alla } i\text{-esima estrazione}\}.$$

In termini di operazioni booleane su E_1, E_2 , scrivete l'espressione per l'evento

{gli oggetti risultanti dalle due successive estrazioni sono dello stesso tipo}.

Esercizio 1.5. Un'urna contiene esattamente quattro elementi di tipo A e tre elementi di tipo B ; da tale urna si effettuano tre successive estrazioni senza reinserimento, registrando il tipo dell'elemento via via estratto.

- Elencate gli eventi elementari in questo esperimento e contate quanti sono.
- Quanti sono, fra tali eventi elementari, quelli che realizzano l'evento

{almeno due elementi di tipo B fra i tre elementi estratti}?

- Quali e quanti sono, fra tali eventi elementari, quelli che realizzano l'evento

{almeno due elementi di tipo B } \cup \{\text{l'elemento estratto alla seconda estrazione è di tipo } B\}?

Esercizio 1.6. Consideriamo di nuovo l'urna di cui nell'esercizio precedente. Se ne effettuano sette successive estrazioni senza reinserimento (cioè l'urna viene progressivamente svuotata), registrando anche in questo caso soltanto il tipo dell'elemento via via estratto (tutti gli elementi di tipo A sono indistinguibili fra di loro, e tutti gli elementi di tipo B sono indistinguibili fra di loro).

Elencate gli eventi elementari in questo esperimento e contate quanti sono.

Esercizio 1.7. Considerate di nuovo l'urna come nel precedente Esercizio 1.5. Questa volta però le tre estrazioni sono effettuate con reinserimento.

- Elencate anche in questo caso gli eventi elementari.
- Dove risiede la differenza fra le due situazioni di estrazioni con e senza reinserimento? (Per rispondere a tale domanda servono degli elementi non ancora studiati in questa prima lezione²).

Esercizio 1.8. Consideriamo ora il caso in cui vengono effettuate sette estrazioni con reinserimento dalla stessa urna di cui nei precedenti esercizi.

Quanti sono gli elementi elementari?

Esercizio 1.9. Una moneta viene lanciata due volte, registrando ogni volta se il risultato sia stato testa o croce.

- Elencate gli eventi elementari possibili, in questo esperimento, e contate quanto vale $|\Omega|$, la cardinalità di Ω .
- Qual è la cardinalità di $\mathcal{P}(\Omega)$, l'insieme delle parti di Ω ? Cioè, quanti sono in tutto gli eventi, contando sia quelli semplici, quelli composti e quelli "banali" \emptyset e Ω ?

²Questo tipo di problemi sarà esaminato in generale nella Sezione 6

2 Spazi finiti di probabilità

2.1 Prime definizioni e proprietà

Introduciamo ora il concetto di **probabilità**. Come vedremo, tale concetto permette di formalizzare il problema di esprimere uno stato di incertezza circa il modo di risultare di un esperimento aleatorio

Sia dato uno spazio campione Ω e sia $\mathcal{P}(\Omega)$ la famiglia delle sue parti.

Definizione 2.1. (provvisoria) Una misura di probabilità o, più semplicemente, una probabilità su $(\Omega, \mathcal{P}(\Omega))$ è una funzione che soddisfa i seguenti assiomi³

i) $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$

ii) $P(\Omega) = 1$ (**condizione di normalizzazione**)

iii) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, per $E_1 \cap E_2 = \emptyset$ (**proprietà di additività**).

Definizione 2.2. Uno spazio finito di probabilità è una terna $(\Omega, \mathcal{P}(\Omega), P)$ dove Ω è uno insieme finito, $\mathcal{P}(\Omega)$ è la famiglia delle parti di Ω e P è una misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$.

Osservazione 1. Prima di proseguire è opportuno citare il fatto che esistono diverse possibili interpretazioni del termine probabilità: ad esempio probabilità classiche, frequentistiche, soggettivistiche, etc...

Non rientra nei nostri scopi soffermarci sul significato e la portata di tali interpretazioni; per quanto ci riguarda ci basta accennare al fatto che, all'interno di ciascuna di dette interpretazioni, è giustificato imporre che la probabilità soddisfi le condizioni i), ii), iii) della Definizione 2.1.

Su tale base possiamo imporre tale condizioni come assiomi e procedere in modo appunto assiomatico; e di tali assiomi vedremo fra poco alcune conseguenze immediate.

Esercizio proposto 2.1. Pensiamo all'esperimento del lancio di un dado a sei facce con

$$\Omega = \{\omega_1, \dots, \omega_6\}.$$

e poniamo

$$P(E) = \frac{|E|}{6}.$$

Verificate che $P(\cdot)$ soddisfa gli assiomi i), ii), iii) e calcolare

$$P(\{X \leq 2\} \cup \{X \geq 5\}), \quad P(\{X \geq 3\} \cap \{X \leq 4\}).$$

In quanto segue consideriamo ancora il caso di spazio campione finito:

$$\Omega = \{\omega_1, \dots, \omega_N\}.$$

Elenchiamo ora alcune proprietà della probabilità, che risultano conseguenze immediate degli assiomi i), ii), iii) della Definizione 2.1.

³Su alcuni testi la proprietà i) è sostituita dalla **proprietà di non negatività**, ossia:

i') $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$, con la proprietà che $P(E) \geq 0$ per ogni $E \in \mathcal{P}(\Omega)$.

È facile vedere che le proprietà i'), ii) e iii) implicano che $P(E) \in [0, 1]$.

Per brevità d'ora in poi scriveremo \bar{E} invece di $\complement(E)$, per indicare il complementare (o la “negazione”) di un evento $E \in \mathcal{P}(\Omega)$.

Prima di tutto notiamo che la proprietà *iii*) di additività si generalizza al caso di n eventi disgiunti a due a due

iii) Siano $E_1, \dots, E_n \in \mathcal{P}(\Omega)$ disgiunti (o incompatibili) a due a due, ovvero tali che

$$E_i \cap E_j = \emptyset, \quad \text{per } i, j \in \{1, 2, \dots, n\}, \text{ con } i \neq j; \quad (1)$$

allora si ha la condizione

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i). \quad (2)$$

(la dimostrazione si ottiene facilmente per induzione su n)

Le relazioni elencate qui di seguito costituiscono delle immediate conseguenze degli assiomi della probabilità. Si invita il lettore a verificarle per esercizio.

(a) Per $E \in \mathcal{P}(\Omega)$, ponendo

$$p(\omega_i) = P(\{\omega_i\}), \quad i = 1, \dots, N,$$

risulta

$$P(E) = \sum_{\substack{i=1 \\ i: \omega_i \in E}}^N p(\omega_i) = \sum_{\omega \in E} p(\omega). \quad (3)$$

(b) Per ogni $E \in \mathcal{P}(\Omega)$ risulta

$$P(\bar{E}) = 1 - P(E). \quad (4)$$

(c) L'evento impossibile ha probabilità nulla, ovvero

$$P(\emptyset) = 0 \quad (5)$$

(d) (**proprietà di monotonia**) Siano $A, B \in \mathcal{P}(\Omega)$ tali che $A \subseteq B$; allora risulta

$$P(A) \leq P(B). \quad (6)$$

(e) Per arbitrari $A, B \in \mathcal{P}(\Omega)$ risulta

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (7)$$

(f) Siano $H_1, \dots, H_n \in \mathcal{P}(\Omega)$ tali che

$$\bigcup_{i=1}^n H_i = \Omega, \quad H_i \cap H_j = \emptyset \quad \text{per } i \neq j, \quad (8)$$

ossia la collezione H_1, \dots, H_n è una partizione dell'evento certo; allora si ha la condizione

$$\sum_{i=1}^n P(H_i) = 1. \quad (9)$$

In particolare, ricordando che $p(\omega_i) = P(\{\omega_i\})$, $i = 1, \dots, N$, e prendendo $H_i = \{\omega_i\}$, deve risultare

$$p(\omega_i) \geq 0 \quad e \quad \sum_{i=1}^N p(\omega_i) = 1. \quad (10)$$

Ulteriori conseguenze degli assiomi della probabilità verranno viste in seguito, **dopo aver introdotto la probabilità condizionata (si veda la Lezione 4).**

Le precedenti proprietà (a)–(f) si possono dimostrare a partire dalla definizione (3) di probabilità in uno spazio finito.

Alcune dimostrazioni con la definizione di probabilità in uno spazio finito

A titolo di esempio dimostriamo la proprietà (a):

Chiaramente $E = \cup_{\omega \in E} \{\omega\}$, ovvero, se indichiamo con $n = |E|$, e con $\{j_1, j_2, \dots, j_n\}$ il sottoinsieme degli indici $\{1, 2, \dots, N\}$, tali che

$$E = \{\omega_{j_1}, \omega_{j_2}, \dots, \omega_{j_n}\},$$

si ha

$$E = \bigcup_{i=1}^n \{\omega_{j_i}\}.$$

Gli insiemi $\{\omega_{j_i}\}$ sono chiaramente incompatibili a due a due e quindi dalla proprietà di additività *iii'*) si ottiene che

$$P(E) = \sum_{i=1}^n P(\{\omega_{j_i}\}).$$

Tuttavia molte delle altre proprietà possono essere dimostrate sia a partire dalla (3) che dalla seguente relazione: per ogni A e B si ha

$$P(A) = P(A \cap B) + P(A \cap \bar{B}). \quad (11)$$

A sua volta (11) deriva dall'osservazione che

$$A = (A \cap B) \cup (A \cap \bar{B}),$$

che gli eventi $A \cap B$ e $A \cap \bar{B}$ sono incompatibili e infine dalla proprietà *iii*).

Suggerimenti per le dimostrazioni dei punti (b) – (f)

Ad esempio la proprietà (b) si ottiene prendendo $A = \Omega$ e $B = E$, per cui $A \cap B = \Omega \cap E = E$ e $A \cap \bar{B} = \Omega \cap \bar{E} = \bar{E}$, da cui

$$P(\Omega) = P(E) + P(\bar{E}) \stackrel{ii)}{\iff} 1 = P(E) + P(\bar{E})$$

che è equivalente alla (4).

La proprietà (c) a sua volta si può derivare dalla (b) osservando che l'evento impossibile è il complementare dell'evento certo, ossia $\emptyset = \bar{\Omega}$, oppure osservando che $E = E \cup \emptyset$ e che, ovviamente, $E \cap \emptyset = \emptyset$, per cui dalla proprietà *ii)*,

$$P(E) = P(E \cup \emptyset) = P(E) + P(\emptyset).$$

La proprietà (d) (monotonia della probabilità) invece deriva dall'osservare che, se $A \subset B$, allora $A \cap B = A$, $A \cap \bar{B} = \emptyset$, la proprietà di base (11) diviene

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A \cap B),$$

e infine dall'osservare che, scambiando il ruolo di A e B nella (11), si ha

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) \geq P(A \cap B) = P(A).$$

La proprietà (e) si può dimostrare a partire sempre dalla (11) (utilizzata anche scambiando il ruolo di A e B) e osservando che

$$A \cup B = (A \cap B) \cup (A \cap \bar{B}) \cup (B \cap \bar{A}).$$

Infine la proprietà (f) è immediata conseguenza degli assiomi, e precisamente delle proprietà *i)* e *iii')*.

Osservazione 2. Nel caso in cui Ω è un insieme finito, possiamo guardare alla probabilità nei due modi, apparentemente diversi ma sostanzialmente equivalenti, che verranno illustrati qui di seguito (teniamo presente il fatto che ciascun punto di Ω può essere visto come un particolare sottoinsieme, cioè come un sottoinsieme composto da un solo elemento) :

1) Prima definiamo P come una funzione di insieme, cioè

$$P : \mathcal{P}(\Omega) \rightarrow [0, 1]; \quad E \mapsto P(E)$$

che soddisfi gli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1, e poi definiamo la funzione di punto

$$p : \Omega \rightarrow [0, 1]; \quad \omega_i \mapsto p(\omega_i) := P(\{\omega_i\}).$$

Questa funzione dovrà soddisfare le condizioni

$$p(\omega_i) \geq 0, \quad \sum_{\omega \in \Omega} p(\omega) = 1. \quad \text{ovvero} \quad \sum_{i=1}^N p(\omega_i) = 1. \quad (12)$$

2) Prima definiamo una funzione di punto

$$p : \Omega \rightarrow [0, 1]; \quad \omega_i \mapsto p(\omega_i),$$

che soddisfi le condizioni (12) e poi definiamo una funzione di insieme

$$E \mapsto P(E) := \sum_{\omega \in E} p(\omega),$$

ossia attraverso la precedente formula (3).

Proposizione 2.1. *La funzione di insieme definita come sopra, ossia dalla formula (3), è una probabilità sull'insieme finito Ω , cioè soddisfa gli assiomi i), ii), iii) della Definizione 2.1.*

Esercizio proposto 2.2. *Dimostrare la precedente Proposizione 2.1.*

Esempio 2.1 (Un esempio concreto di spazio di probabilità finito). *Sia $\Omega = \{a, b, c, d\}$ siano $p(a) = 1/8$, $p(b) = 1/4$, $p(c) = 1/2$, $p(d) = 1/8$. Chiaramente $p(a), p(b), p(c), p(d) \geq 0$ e $p(a) + p(b) + p(c) + p(d) = 1$. Allora la probabilità definita dalla precedente formula sull'insieme delle parti di $\Omega = \{a, b, c, d\}$ è data dalla funzione che è specificata nella seguente tabella.*

**Tabella della Probabilità definita nell'esempio 2.1
(un esempio concreto di spazio di probabilità finito)**

$$\begin{aligned} \emptyset &\mapsto P(\emptyset) = 0 \\ \{a\} &\mapsto P(\{a\}) = p(a) = 1/8 \\ \{b\} &\mapsto P(\{b\}) = p(b) = 1/4 \\ \{c\} &\mapsto P(\{c\}) = p(c) = 1/2 \\ \{d\} &\mapsto P(\{d\}) = p(d) = 1/8 \\ \{a, b\} &\mapsto P(\{a, b\}) = p(a) + p(b) = 1/8 + 1/4 = 3/8 \\ \{a, c\} &\mapsto P(\{a, c\}) = p(a) + p(c) = 1/8 + 1/2 = 5/8 \\ \{a, d\} &\mapsto P(\{a, d\}) = p(a) + p(d) = 1/8 + 1/8 = 1/4 \\ \{a, b, c\} &\mapsto P(\{a, b, c\}) = p(a) + p(b) + p(c) = 1/8 + 1/4 + 1/2 = 7/8 \\ \{a, b, d\} &\mapsto P(\{a, b, d\}) = p(a) + p(b) + p(d) = 1/8 + 1/4 + 1/8 = 1/2 \\ \{a, c, d\} &\mapsto P(\{a, c, d\}) = p(a) + p(c) + p(d) = 1/8 + 1/2 + 1/8 = 3/4 \\ \{b, c, d\} &\mapsto P(\{b, c, d\}) = p(b) + p(c) + p(d) = 1/4 + 1/2 + 1/8 = 7/8 \\ \{a, b, c, d\} &\mapsto P(\{a, b, c, d\}) = p(a) + p(b) + p(c) + p(d) = 1/8 + 1/4 + 1/2 + 1/8 = 1 \end{aligned}$$

Osservazione 3 (Probabilità definite a meno di un fattore di proporzionalità; normalizzazione). Una misura di probabilità sullo spazio $\Omega = \{\omega_1, \dots, \omega_N\}$ è individuata quando vengano assegnati i numeri $p(\omega_i) = P(\{\omega_i\})$, ($i = 1, 2, \dots, N$) soddisfacenti le condizioni (12). Supponiamo ora che $p(\omega_i)$, ($i = 1, 2, \dots, N$) siano assegnati a meno di una costante di proporzionalità; supponiamo cioè che siano assegnati dei numeri g_i ($i = 1, 2, \dots, N$), tali che

$$p(\omega_i) = K \cdot g_i \quad (13)$$

essendo K un'opportuna costante positiva. Dalla condizione di normalizzazione (10), si ricava

$$K = \frac{1}{\sum_{j=1}^N g_j}; \quad p(\omega_i) = \frac{g_i}{\sum_{j=1}^N g_j}.$$

Notiamo che si usa esprimere brevemente la condizione (13) usando il seguente simbolismo:

$$p(\omega_i) \propto g_i.$$

Esempio 2.2 (dado non equilibrato). *Un dado ha sei facce numerate da 1 a 6; esso è pesato in modo tale che ciascuna faccia abbia una probabilità di presentarsi (in un singolo lancio) proporzionale al suo valore. Sia*

$$A \equiv \{\text{si presenta un numero pari}\}.$$

Trovare $P(A)$.

Soluzione. Si ha $\Omega = \{\omega_1, \dots, \omega_6\}$ e vogliamo imporre

$$p(\omega_i) = K \cdot i, \quad i = 1, \dots, 6,$$

essendo K una costante positiva da determinare, imponendo la condizione di normalizzazione (10); si ottiene dunque

$$p(\omega_i) = \frac{i}{21}, \quad i = 1, \dots, 6$$

e

$$P(A) = p(\omega_2) + p(\omega_4) + p(\omega_6) = \frac{12}{21}.$$

2.2 Esercizi di verifica

Esercizio 2.1. Un dado è pesato in modo tale che la probabilità di avere un punto pari è il doppio della probabilità di avere un punto dispari. Qual è la probabilità di avere punto pari?

Esercizio 2.2. Siano A e B due eventi tali che

$$P(A \cap B) = P(A \cap \bar{B}) = P(\bar{A} \cap B) = P(\bar{A} \cap \bar{B}).$$

- Quanto vale $P(A \cap B)$?
- Qual è la probabilità che, fra A e B , se ne verifichi almeno uno?
- Qual è la probabilità che se ne verifichi esattamente uno?

Esercizio 2.3. Una moneta viene lanciata due volte e poniamo

$$E_i \equiv \{\text{testa all}'i\text{-esimo lancio}\}, \quad i = 1, 2.$$

Mostrare che la condizione $P(E_1 \cap \bar{E}_2) = P(\bar{E}_1 \cap E_2)$ implica $P(E_1) = P(E_2)$.

Esercizio 2.4. Siano A , B , e C tre eventi. Dimostrate che vale la formula

$$\begin{aligned} P(A \cup B \cup C) &= \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

(questa formula costituisce un caso particolare della formula di inclusione - esclusione, che verrà vista in una delle lezioni successive).

Esercizio 2.5. Siano A , B , e C tre eventi tali che

$$\begin{aligned} P(A \cap B \cap C) &= P(A \cap \bar{B} \cap C) = 0.1, \\ P(\bar{A} \cap B \cap C) &= P(A \cap B \cap \bar{C}) = 0.15, \\ P(A \cap \bar{B} \cap \bar{C}) &= P(\bar{A} \cap B \cap \bar{C}) = P(\bar{A} \cap \bar{B} \cap C) = 0.05. \end{aligned}$$

Calcolare

- $P(A)$, $P(B)$, $P(C)$
- $P(A \cup B)$, $P(A \cup C)$
- $P(A \cup B \cup C)$.

3 Probabilità “classiche” e calcolo combinatorio

3.1 Probabilità “classiche”

Qui ci soffermiamo a trattare alcuni casi particolari, ma molto rilevanti, di spazi di probabilità finiti. Sia dunque

$$\Omega = \{\omega_1, \dots, \omega_N\},$$

e supponiamo che si voglia porre

$$p(\omega_i) = K, \quad \forall \omega_i \in \Omega \quad (14)$$

per un’opportuna costante positiva K . Si vuole cioè imporre che tutti i risultati elementari siano, fra di loro, equiprobabili.

Ci riferiremo a tale caso dicendo che si ha una **distribuzione di probabilità uniforme sugli eventi elementari**.

Confrontando la posizione (14) con la condizione di normalizzazione (10), otteniamo immediatamente

$$p(\omega_i) = \frac{1}{N}, \quad i = 1, \dots, N$$

e da ciò segue, ricordando la formula (3) del precedente paragrafo,

$$P(E) = \frac{|E|}{N}, \quad \forall E \in \mathcal{P}(\Omega). \quad (15)$$

Esempio 3.1. *L’addetto ad un guardaroba restituisce a caso n ombrelli che gli sono stati consegnati; qual è la probabilità che il secondo cliente abbia indietro il suo proprio ombrello?*

Soluzione. Si ha che Ω è costituito dalle permutazioni⁴ di n elementi; dunque $|\Omega| = n!$. L’evento

$$E \equiv \{\text{Il secondo cliente riceve indietro il suo ombrello}\}$$

è un evento composto, costituito da tutte le permutazioni che tengono fisso il secondo elemento; tali permutazioni sono in numero di $(n-1)!$, corrispondente al numero delle possibili permutazioni dei restanti $(n-1)$ elementi. L’espressione “a caso” vuole significare che tutte le permutazioni sono da considerare equiprobabili fra di loro. Dunque

$$P(E) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Osservazione 1. La formula (15) esprime il fatto che, nel caso in cui tutti gli eventi elementari di uno spazio finito sono equiprobabili, la probabilità di un generico evento composto si calcola quale **rapporto fra casi favorevoli e casi possibili**.

Si faccia attenzione al fatto che la (15) *non costituisce una definizione del concetto di probabilità, ma soltanto una formula per il suo calcolo in un caso particolare:*

nel precedente paragrafo abbiamo già introdotto tale concetto in modo assiomatico;

la suddetta formula è stata ottenuta come immediata conseguenza degli assiomi stessi, nel caso particolare di eventi elementari equiprobabili.

Osservazione 2. Nel caso in cui si imponga la condizione (14), il calcolo della probabilità di un evento composto E si riduce al problema, **combinatorio**, di individuare $N = |\Omega|$ e $|E|$.

⁴Per la definizione di permutazione vedere più avanti la Sezione 3.2

3.2 Calcolo combinatorio: primi elementi

Facendo seguito alle precedenti *Osservazione 1* e *Osservazione 2*, ci rivolgiamo ora a richiamare succintamente alcune nozioni basilari di calcolo combinatorio, che risultano indispensabili per affrontare i primi problemi di calcolo delle probabilità.

Le formule che verranno presentate si ricavano facilmente tramite applicazione del **principio di induzione finita**.

Iniziamo innanzitutto ricordando due fatti fondamentali:

a) Due insiemi finiti hanno la stessa cardinalità se e solo se fra essi è possibile stabilire una **corrispondenza biunivoca**.

b) Dati due arbitrari insiemi A e B , si definisce **prodotto cartesiano** di A per B l'insieme costituito dalle coppie ordinate (a, b) dove $a \in A$ e $b \in B$; indichiamo tale insieme con il simbolo $A \times B$. Nel caso in cui A e B sono insiemi finiti, risulta

$$|A \times B| = |A| \cdot |B|.$$

In quanto immediatamente segue supponiamo di aver fissato un arbitrario insieme A costituito da n elementi:

$$A \equiv \{a_1, \dots, a_n\}.$$

Disposizioni con ripetizione di classe k di n elementi.

Una disposizione con ripetizione di classe k degli n elementi di A non è altro che una k -upla ordinata degli elementi stessi.

Tali disposizioni costituiscono dunque l'insieme $A^k = \overbrace{A \times A \times \dots \times A}^{k \text{ volte}}$, e si ha $|A^k| = n^k$.

Disposizioni senza ripetizione di classe k di n elementi e permutazioni di n elementi

Le disposizioni senza ripetizione di classe k degli n elementi sono le k -uple costituite da elementi di A , **tutti diversi fra loro**.

Tali disposizioni costituiscono un sottoinsieme, dell'insieme A^k , di cardinalità⁵

$$\overbrace{n(n-1)\dots(n-(k-1))}^{k \text{ fattori}} = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!},$$

dove si è usata la notazione n fattoriale, ovvero $n! = n(n-1)\dots 3 \cdot 2 \cdot 1$.

Nel caso in cui si ponga $k = n$, si ottengono le **permutazioni** degli elementi di A . Di conseguenza il numero delle permutazioni di n elementi è $n!$.

Combinazioni di classe k di n elementi

Si tratta di classi di equivalenza di disposizioni senza ripetizione di classe k di n elementi, modulo la relazione di equivalenza costituita dal considerare equivalenti due disposizioni che contengono gli stessi elementi, eventualmente in ordine diverso.

Alternativamente una combinazione di classe k di n elementi si può definire come **un sottoinsieme di cardinalità k di un insieme di cardinalità n** .

Se C_k^n indica il numero delle combinazioni di classe k di n elementi, D_k^n indica il numero delle disposizioni senza ripetizione di classe k di n elementi, e P_k indica il numero delle permutazioni di k elementi, è facile convincersi⁶ che

$$D_k^n = C_k^n \cdot P_k,$$

⁵Si veda la nota di approfondimento a pagina 26

⁶Si veda la nota di approfondimento pagina 27

da cui, tenendo conto che $P_k = k!$ e $D_k^n = \frac{n!}{(n-k)!}$, si ricava immediatamente $\frac{n!}{(n-k)!} = C_k^n \cdot k!$.
 Il numero complessivo di tali combinazioni è dunque dato da

$$C_k^n = \frac{1}{k!} \cdot \frac{n!}{(n-k)!}.$$

Si pone

$$\binom{n}{k} \equiv \frac{n!}{k!(n-k)!}.$$

Il numero $\binom{n}{k}$ prende il nome di **coefficiente binomiale n sopra k** (o anche n su k).

Esempio 3.2. Consideriamo un circolo costituito da n persone e supponiamo di dover eleggere un presidente, un segretario e un tesoriere.

Se pensiamo di scegliere tre persone diverse, ognuna con la sua specifica carica, ciascuna scelta coincide con una disposizione senza ripetizione di classe 3 degli n elementi; abbiamo $n(n-1)(n-2)$ possibili scelte.

Se pensiamo che ogni carica è assegnata con una votazione indipendente dalle altre, si possono avere anche delle ripetizioni (cioè è ammesso un cumulo delle cariche); in tal caso ciascuna possibile scelta coincide con una disposizione di classe 3, con ripetizione, degli n elementi; abbiamo n^3 possibili scelte.

Se pensiamo di eleggere complessivamente una terna di persone diverse, senza attribuire una specifica carica a ciascuna di loro, ma incaricandoli complessivamente dei compiti di presidente, di segretario e di tesoriere, ciascuna possibile scelta coincide con una combinazione di classe 3 degli n elementi; abbiamo, in tal caso $\binom{n}{3} = \frac{n(n-1)(n-2)}{6}$ possibili scelte.

A proposito di coefficienti binomiali è utile introdurre la seguente convenzione: per ogni numero naturale n , si pone

$$\binom{n}{0} = 1,$$

come è ovvio, sia tenendo presente la convenzione $0! = 1$, sia tenendo presente il fatto che l'unico sottoinsieme di cardinalità zero è l'insieme vuoto.

Esercizio proposto 3.1. Dimostrare che $\sum_{k=0}^n C_n^k = 2^n$, senza usare la formula della potenza del binomio di Newton⁷.

3.3 Alcuni classici esempi

Consideriamo ora qualche semplice e classico esempio di probabilità combinatorie.

Esempio 3.3 (Problema del compleanno). Qual è la probabilità che, fra M persone scelte a caso, ve ne siano almeno due che festeggiano il compleanno nello stesso giorno? (Si supponga l'anno costituito da 365 giorni e che vi sia una situazione di simmetria rispetto alle nascite).

Soluzione. Calcoliamo la probabilità dell'evento complementare

$$\bar{E} = \{\text{Le } M \text{ persone festeggiano il compleanno in tutti giorni diversi}\}$$

⁷Si veda anche la nota di approfondimento a pagina 23

Lo spazio Ω è costituito dalle disposizioni con ripetizione di classe M di 365 elementi (i giorni dell'anno solare). \bar{E} è un evento composto costituito da tutte le disposizioni senza ripetizione di classe M di 365 elementi. Quindi

$$P(\bar{E}) = \frac{365 \cdot 364 \cdot \dots \cdot (365 - M + 1)}{(365)^M}$$

e la probabilità cercata è fornita da $P(E) = 1 - P(\bar{E})$. Indichiamo ora tale probabilità con $P_M(E)$ per mettere in evidenza la sua dipendenza dal valore di M . Ovviamente $P_M(E)$ è una funzione crescente di M ed è interessante notare che si ha $P_M(E) > \frac{1}{2}$ per $M > 22$, in particolare si ha $P_{22}(E) \simeq 0.4756$ mentre $P_{23}(E) \simeq 0.5072$.

Esempio 3.4 (“Paradosso del Cavalier De Méré”). *È più probabile ottenere almeno un asso in 4 lanci consecutivi di un dado o un doppio asso in 24 lanci consecutivi di una coppia di dadi?*

Soluzione. Anche qui conviene calcolare le probabilità dei due eventi complementari:

$$\begin{aligned} P(\{\text{almeno un asso in 4 lanci}\}) &= 1 - P(\{\text{nessun asso in 4 lanci}\}) \\ P(\{\text{almeno un doppio asso in 24 lanci}\}) &= \\ &= 1 - P(\{\text{nessun doppio asso in 24 lanci}\}). \end{aligned}$$

I risultati possibili nei 4 lanci del dado sono rappresentati dalle disposizioni con ripetizione di classe 4 di 6 elementi; in altre parole, possiamo rappresentare Ω come lo spazio delle quaterne ordinate (x_1, x_2, x_3, x_4) con $x_i \in \{1, \dots, 6\}$. Dunque $|\Omega| = 6^4$. Gli eventi elementari che costituiscono l'evento composto $\{\text{nessun asso in 4 lanci}\}$ corrispondono, invece, alle disposizioni con ripetizione di classe 4 dei 5 elementi $\{2, \dots, 6\}$. Si ha quindi

$$P(\{\text{almeno un asso in 4 lanci}\}) = 1 - \frac{5^4}{6^4} \simeq 0.52.$$

Analogamente si ottiene

$$P(\{\text{almeno un doppio asso in 24 lanci}\}) = 1 - \frac{35^{24}}{36^{24}} \simeq 0.49.$$

Esempio 3.5. *Un gruppo di $4N$ persone comprende $2N$ ragazzi e $2N$ ragazze. Vengono formate a caso due squadre di $2N$ persone ciascuna.*

a) *Qual è la probabilità che tutte le ragazze si trovino nella stessa squadra e tutti i ragazzi nella squadra avversaria?*

b) *Qual è la probabilità che ciascuna squadra sia, all'opposto, composta esattamente da N ragazzi ed N ragazze?*

Soluzione. Qui il generico evento elementare è specificato da un modo di scegliere $2N$ oggetti (i componenti della prima squadra) da un insieme di $4N$ oggetti; dunque la cardinalità dello spazio degli eventi elementari Ω è data da $\binom{4N}{2N}$.

Suggerimento: in effetti si può pensare che le ragazze siano numerate come f_1, f_2, \dots, f_{2N} ed i ragazzi come m_1, m_2, \dots, m_{2N} . Per specificare la prima squadra basta prendere un sottoinsieme di $R = \{f_1, f_2, \dots, f_{2N}, m_1, m_2, \dots, m_{2N}\}$ di cardinalità $2N$, ovvero

$$\Omega = \{\text{combinazioni dei } 4N \text{ elementi di } R \text{ di classe } 2N\}, \quad \text{con } |\Omega| = \binom{4N}{2N}$$

Nel caso a) due soli eventi elementari sono favorevoli e dunque la probabilità cercata è $\frac{2}{\binom{4N}{2N}}$.

Suggerimento: i due casi favorevoli sono le due combinazioni $\{f_1, f_2, \dots, f_{2N}\} \in \{m_1, m_2, \dots, m_{2N}\}$.

Nel caso b) gli eventi elementari favorevoli⁸ sono in numero di $\binom{2N}{N}\binom{2N}{N}$, corrispondente al numero dei modi in cui si possono scegliere N ragazze dal gruppo di tutte le $2N$ e N ragazzi dal gruppo di tutti i $2N$. La probabilità cercata è dunque data da

$$\frac{\binom{2N}{N}\binom{2N}{N}}{\binom{4N}{2N}}.$$

Esempio 3.6. Supponiamo che una moneta perfetta venga lanciata n volte. Per $h \leq n$, qual è la probabilità di nessuna testa sui primi h lanci?

Soluzione. Possiamo schematizzare gli eventi elementari in questo esperimento come gli elementi dell'insieme $\{0, 1\}^n$ cioè come n -uple con elementi uguali a 0 (croce) o uguali a 1 (testa), (ad esempio l'evento elementare $\omega \equiv (0, 0, 1, 1, 0, \dots, 0)$ coincide con il fatto che i primi due lanci danno croce, poi si hanno consecutivamente due risultati testa, e poi in tutti i successivi lanci si ottiene ancora croce); dunque si ha $|\Omega| = 2^n$.

L'evento $\{\text{nessuna testa sui primi } h \text{ lanci}\}$ è allora l'evento composto

$$E \equiv \{\omega \in \Omega : \omega = (0, 0, \dots, 0, \omega_{h+1}, \dots, \omega_n), \text{ con } (\omega_{h+1}, \dots, \omega_n) \in \{0, 1\}^{n-h}\}.$$

Traduciamo la condizione che la moneta sia perfetta con la posizione

$$p(\omega) = \frac{1}{2^n}, \quad \forall \omega \in \Omega.$$

Si ha $|E| = 2^{n-h}$ e dunque $P(E) = \frac{2^{n-h}}{2^n} = \left(\frac{1}{2}\right)^h$.

Esempio 3.7. Qual è la probabilità di k risultati testa negli n lanci di una moneta?

Soluzione. Si ha lo stesso spazio di probabilità dell'esercizio precedente; questa volta $|E| = \binom{n}{k}$ e dunque $P(E) = \frac{\binom{n}{k}}{2^n}$; come vedremo in seguito si tratta di un caso particolare di probabilità **binomiali**.

Esempio 3.8. Trovare la probabilità di k voti per lo schieramento A in un sondaggio elettorale di ampiezza n in un gruppo di M elettori di cui è noto che m_1 votano per A e $m_2 = M - m_1$ votano per B .

Soluzione. Si è sottinteso che gli n elettori siano stati selezionati senza reinserimento. L'esperimento consiste dunque nel selezionare un sottoinsieme di cardinalità n (il campione) dall'insieme degli M elettori (la popolazione) e quindi $|\Omega| = \binom{M}{n}$. Si sottintende che il sondaggio sia condotto in modo casuale, cioè che ogni "campione" abbia uguale probabilità $\frac{1}{\binom{M}{n}}$ di essere estratto. Fra tali "campioni", ve ne sono $\binom{m_1}{k} \cdot \binom{m_2}{n-k}$ che contengono k elettori per A e $(n-k)$ per B . Infatti, ci sono $\binom{m_1}{k}$ modi di selezionare k elettori fra i votanti per A , ci sono $\binom{m_2}{n-k}$ modi di selezionare $(n-k)$ elettori fra i votanti per B e, inoltre, una qualunque scelta di k elettori fra i votanti per A e di $(n-k)$

⁸Per convincersene si consiglia il lettore di considerare il caso $N = 1$ ed $N = 2$, elencando esplicitamente sia tutti i casi possibili che tutti i casi favorevoli.

elettori fra i votanti per B dà luogo ad una n -upla di elettori (un campione) che contiene k elettori per A .

Dunque la probabilità cercata è data da

$$\frac{\binom{m_1}{k} \cdot \binom{m_2}{n-k}}{\binom{M}{n}} = \frac{\binom{m_1}{k} \cdot \binom{m_2}{n-k}}{\binom{m_1+m_2}{n}}.$$

Come vedremo in seguito, si tratta di un caso particolare di probabilità *ipergeometriche*. Osserviamo che i valori possibili per k devono rispettare la condizione

$$0 \leq k \leq m_1, \quad 0 \leq n - k \leq m_2, \quad \text{con} \quad n \leq M = m_1 + m_2,$$

che, dopo semplici passaggi, diviene

$$0 \vee (n - m_2) = \max(0, n - m_2) \leq k \leq \min(n, m_1) = n \wedge m_1.$$

Esercizio proposto 3.2. *Siano M , m_1 , n e k numeri assegnati e tali che*

$$m_1 < M, \quad n < M, \quad \max(0, m_1 + n - M) \leq k \leq \min(n, m_1).$$

Verificate l'identità

$$\frac{\binom{m_1}{k} \binom{M-m_1}{n-k}}{\binom{M}{n}} = \frac{\binom{n}{k} \binom{M-n}{m_1-k}}{\binom{M}{m_1}}.$$

Riuscite a darne un'interpretazione probabilistica?

Interpretazione dell'identità

$$\frac{\binom{m_1}{k} \binom{M-m_1}{n-k}}{\binom{M}{n}} = \frac{\binom{n}{k} \binom{M-n}{m_1-k}}{\binom{M}{m_1}},$$

del precedente esercizio-proposto 3.2

Pensate ad esempio di avere un'urna e di estrarre tutte le palline in ordine: si suppone come al solito che l'urna contenga m_1 palline di tipo A ed $m_2 = M - m_1$ di tipo B. Siamo interessati alla probabilità che tra le prime n (con $1 \leq n \leq M$) estratte ce ne siano (esattamente) k di tipo A (e quindi $n - k$ di tipo B). Un esperimento si può descrivere annotando solo l'ordine con cui vengono estratti i tipi. Ad esempio se $m_1 = 3$ ed $m_2 = 2$, i risultati possibili sono

(A, A, A, B, B)	(A, A, B, A, B)	(A, A, B, B, A)	(A, B, A, A, B)	(A, B, A, B, A)
(A, B, B, A, A)	(B, A, A, A, B)	(B, A, A, B, A)	(B, A, B, A, A)	(B, B, A, A, A)

Non è difficile convincersi che tutti i casi possibili sono equiprobabili e che sono in tutto quante le combinazioni di 5 elementi di classe 3, infatti basta specificare le posizioni in cui sono uscite le palline di tipo A. Con questa corrispondenza i casi precedenti sono in corrispondenza con

{1, 2, 3}	{1, 2, 4}	{1, 2, 5}	{1, 3, 4}	{1, 3, 5}
{1, 4, 5}	{2, 3, 4}	{2, 3, 5}	{2, 4, 5}	{3, 4, 5}

Più in generale si ottiene che i casi possibili sono appunto

$$C_{m_1}^M = \binom{M}{m_1},$$

in quanto per elencare tutti i casi possibili basta specificare solo le m_1 posizioni occupate dalle A su tutte le M posizioni.

Per quanto riguarda i casi favorevoli, iniziamo di nuovo con il caso particolare dell'urna precedente e con $n = 3$ e $k = 2$. I casi favorevoli sono

(A, A, B, A, B)	(A, A, B, B, A)
(A, B, A, A, B)	(A, B, A, B, A)
(B, A, A, A, B)	(B, A, A, B, A)

e si ottengono nel seguente modo: basta specificare quali sono le due A tra le prime 3 posizioni, e dove si trova la terza A tra le ultime $2 = 5 - 3$ posizioni. Il numero totale risulta quindi $\binom{3}{2} \cdot \binom{2}{1} = 3 \cdot 2 = 6$.

Più in generale per elencare tutti i casi favorevoli all'evento che tra le prime n estratte ce ne siano k di tipo A, basta specificare le k posizioni tra le prime n in cui si trovano le A (e questo si può fare in $C_k^n = \binom{n}{k}$ modi, e inoltre specificare le $m_1 - k$ posizioni occupate dalle A, nelle ultime $M - n$ posizioni (e questo si può fare in $C_{m_1-k}^{M-n} = \binom{M-n}{m_1-k}$ modi). In totale quindi si hanno

$$\binom{n}{k} \cdot \binom{M-n}{m_1-k}$$

casi favorevoli.

Utilizzando (16) è anche facile, per $0 \leq k \leq n$, ottenere la seguente uguaglianza

$$\sum_{r=k}^n \binom{r}{k} = \binom{n+1}{k+1}.$$

È ben noto (e comunque si verifica immediatamente **per induzione, utilizzando la (16)**) che i coefficienti binomiali intervengono come segue nello sviluppo della potenza di un binomio: siano a, b due arbitrari numeri reali non nulli e sia n un numero naturale; allora risulta

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k \cdot b^{n-k}. \quad (18)$$

Ponendo nella (18) $a = x, b = 1$, otteniamo l'identità

$$(x+1)^n = \sum_{k=0}^n \binom{n}{k} x^k. \quad (19)$$

In particolare ponendo $a = b = 1$, otteniamo

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

Tenendo presente che $\binom{n}{k}$ coincide con il numero di sottoinsiemi di cardinalità k contenuti in un insieme composto da n elementi, otteniamo che 2^n è uguale alla cardinalità della famiglia delle parti di un insieme di n elementi. Dunque se in un esperimento vi sono n eventi elementari, vi sono allora in tutto 2^n eventi fra elementari, composti e contando anche l'evento certo e quello impossibile (**si veda anche la nota di approfondimento a pagina 23**).

Un'altra utile identità si può ottenere facilmente per una qualunque terna di numeri naturali r, s, n con $n \leq r + s$:

$$\sum_{k=0}^{n \wedge r} \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}, \quad (20)$$

dove la somma è estesa a tutti gli indici k per i quali $0 \leq k \leq r$ e $0 \leq n - k \leq s$.

Quest'ultima **uguaglianza** può essere verificata ad esempio osservando intanto quanto segue: **per ogni x reale**

$$(1+x)^{r+s} = (1+x)^r \cdot (1+x)^s.$$

Da una parte, utilizzando la (19) con $n = r$, e $n = s$, e la proprietà che il prodotto di sommatorie è la sommatoria del prodotto (si veda il richiamo nel riquadro qui sotto)

Richiamo: va ricordato che, dati R ed S interi ($R, S \geq 1$), e a_1, \dots, a_R e b_1, \dots, b_S , la sommatoria di tutti i prodotti del tipo $a_r b_s$ coincide con il prodotto delle somme di a_r per le somme di b_s :

$$\sum_{r=1}^R \sum_{s=1}^S a_r b_s = \sum_{r=1}^R a_r \left(\sum_{s=1}^S b_s \right) = \left(\sum_{s=1}^S b_s \right) \left(\sum_{r=1}^R a_r \right)$$

$$\begin{aligned}
(1+x)^r \cdot (1+x)^s &= \left(\sum_{k=0}^r \binom{r}{k} x^k \right) \cdot \left(\sum_{h=0}^s \binom{s}{h} x^h \right) \\
&= \sum_{k=0}^r \sum_{h=0}^s \binom{r}{k} \binom{s}{h} x^{k+h} = \sum_{n=0}^{r+s} \sum_{\substack{0 \leq k \leq r, 0 \leq h \leq s \\ k+h=n}} \binom{r}{k} \binom{s}{h} x^{k+h} \\
&= \sum_{n=0}^{r+s} \sum_{0 \leq k \leq r, 0 \leq n-k \leq s} \binom{r}{k} \binom{s}{n-k} x^n = \sum_{n=0}^{r+s} \left(\sum_{k=0}^{r \wedge n} \binom{r}{k} \binom{s}{n-k} \right) x^n.
\end{aligned}$$

D'altra parte, per la (19) con $n = r + s$,

$$(1+x)^{r+s} = \sum_{n=0}^{r+s} \binom{r+s}{n} x^n,$$

e la (20) si ottiene confrontando termine a termine tali due sviluppi. Ponendo in particolare, ad esempio, $r \leq s$ e $n = s$, e utilizzando il fatto che $\binom{s}{s-k} = \binom{s}{k}$ si ottiene

$$\sum_{k=0}^r \binom{r}{k} \binom{s}{k} = \binom{r+s}{s}$$

e potremo dunque anche scrivere

$$\sum_{k=0}^{r \wedge s} \binom{r}{k} \binom{s}{k} = \binom{r+s}{s} = \binom{r+s}{r}, \quad (21)$$

che per $r = s = n$ diviene

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}. \quad (22)$$

Esercizio proposto 3.3. Una moneta perfetta viene lanciata r volte da Renato e s volte da Stefano ($r > 3, s > 3$). Si ponga

X = numero dei lanci in cui Renato ottiene il risultato testa

Y = numero dei lanci in cui Stefano ottiene il risultato testa.

Calcolare la probabilità dei seguenti eventi.

- (a) $\{X = 3\} \cap \{Y = 3\}$ (b) $\{X = 3\} \cup \{Y = 3\}$ (c) $\{\max(X, Y) = 3\}$ (d) $\{X = Y\}$.

Suggerimento per la soluzione dell'Esercizio proposto 3.3

Per calcolare la probabilità dell'evento $\{X = Y\}$ si consiglia di considerare il fatto che la moneta è perfetta e quindi, ai fini del calcolo delle probabilità è equivalente scambiare successo (esce testa) con insuccesso (esce croce) nel caso di uno dei due giocatori, ad esempio Stefano.

In altre parole si sta considerando che la probabilità di $\{X = Y\}$ è la stessa di $\{X = Y'\}$, dove $Y' = s - Y$ è il numero di croci ottenute da Stefano.

Ma allora $\{X = Y'\} = \{X = s - Y\} = \{X + Y = s\}$ e ciò permette di ottenere immediatamente, e con un ragionamento probabilistico, la relazione (21) (nota anche come formula di Vandermonde).

3.5 Approfondimenti sul calcolo combinatorio

Una dimostrazione elementare di

$$\sum_{k=0}^n C_k^n = \sum_{k=0}^n \binom{n}{k} = 2^n$$

La relazione

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

si può dedurre anche nel seguente modo (senza conoscere la formula del binomio di Newton): l'insieme delle parti $\mathcal{P}(\{a_1, \dots, a_n\})$ di $\{a_1, \dots, a_n\}$ è l'unione delle famiglie dei sottoinsiemi di cardinalità k al variare di k da 0 ad n , per cui la cardinalità dell'insieme delle parti, $|\mathcal{P}(\{a_1, \dots, a_n\})|$, soddisfa la relazione seguente

$$\sum_{k=0}^n C_k^n = |\mathcal{P}(\{a_1, \dots, a_n\})|.$$

Già sappiamo che $C_k^n = \binom{n}{k}$, inoltre si può ricavare facilmente che $\mathcal{P}(\{a_1, \dots, a_n\})$ ha cardinalità 2^n : infatti $\mathcal{P}(\{a_1, \dots, a_n\})$ è in corrispondenza biunivoca con l'insieme delle funzioni $g : \{a_1, \dots, a_n\} \mapsto \{0, 1\}$. La corrispondenza è data da $G \longleftrightarrow g = \mathbf{1}_G$, dove

$$\mathbf{1}_G(a_i) = 1, \quad \text{se } a_i \in G$$

$$\mathbf{1}_G(a_i) = 0, \quad \text{se } a_i \notin G.$$

Queste ultime sono tante quanti sono gli elementi del prodotto cartesiano di $\{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\}$ (n volte) e sono quindi 2^n . Per capire meglio la corrispondenza poniamo $n = 4$ e consideriamo il sottoinsieme $G = \{a_1, a_3\}$:

$$\{a_1, a_3\} \iff \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 0 & 1 & 0 \end{pmatrix} \iff (1, 0, 1, 0) \in \{0, 1\}^4.$$

Ancora, ad esempio $(0, 0, 0, 0)$ corrisponde all'insieme vuoto, e $(1, 1, 1, 1)$ a tutto $A = \{a_1, a_2, a_3, a_4\}$. Nel caso generale di $A = \{a_1, a_2, \dots, a_n\}$, la corrispondenza tra i sottoinsiemi di A e gli elementi di $\{0, 1\}^n$ dovrebbe essere quindi chiara, e di conseguenza l'uguaglianza $|\mathcal{P}(\{a_1, \dots, a_n\})| = 2^n$.

Dimostrazione della formula di Stiefel, senza l'uso dei fattoriali

Prendendo come interpretazione di $\binom{n}{k}$ il numero C_k^n dei sottoinsiemi di cardinalità k di un insieme $\{a_1, a_2, \dots, a_{n-1}, a_n\}$ di cardinalità n , la formula di Stiefel diviene

$$C_k^n = C_{k-1}^{n-1} + C_k^{n-1}$$

Per dimostrare la precedente uguaglianza si può ragionare anche nel seguente modo (attenzione, si noti che **non è necessario sapere il valore esplicito di C_k^n** , cioè nella dimostrazione **non si usa** il fatto che $C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$).

I sottoinsiemi di cardinalità k si possono dividere in due classi:

- 1) i sottoinsiemi C che contengono a_n
- 2) i sottoinsiemi D che non contengono a_n .

Quindi il numero C_k^n dei sottoinsiemi di cardinalità k si può esprimere come la somma del numero dei sottoinsiemi del primo tipo e del numero dei sottoinsiemi del secondo tipo.

D'altra parte

- I) gli insiemi C del primo tipo si possono esprimere come $C = C' \cup \{a_n\}$, con C' sottoinsieme di $\{a_1, a_2, \dots, a_{n-1}\}$ che ha cardinalità $k-1$, e quindi sono tanti quanti i sottoinsiemi di cardinalità $k-1$ di un insieme di cardinalità $n-1$, ovvero C_{k-1}^{n-1} .
- II) gli insiemi D del secondo tipo, sono sottoinsiemi di $\{a_1, a_2, \dots, a_{n-1}\}$ di cardinalità k e quindi sono esattamente C_k^{n-1} .

Come si contano le disposizioni-I

In questa nota vogliamo vedere come si può arrivare a contare sia le disposizioni che le combinazioni. Iniziamo con un esempio.

Consideriamo $n = 5$ e $k = 1, 2, 3$ e proviamo a scrivere tutte le disposizioni di 5 elementi di classe k . Indichiamo il nostro insieme di 5 elementi con $\{1, 2, 3, 4, 5\}$. Ovviamente tutte le disposizioni di classe 1 sono $D_1^5 = 5$:

(1)	(2)	(3)	(4)	(5)
-----	-----	-----	-----	-----

Per ottenere le disposizioni di classe 2 si può procedere mettendo insieme le disposizioni che iniziano per 1, quelle che iniziano per 2, etc. come segue

(1,2)	(2,1)	(3,1)	(4,1)	(5,1)
(1,3)	(2,3)	(3,2)	(4,2)	(5,2)
(1,4)	(2,4)	(3,4)	(4,3)	(5,3)
(1,5)	(2,5)	(3,5)	(4,5)	(5,4)

Vengono quindi 20 disposizioni, in quanto si tratta di 5 colonne ciascuna di lunghezza 4: infatti una volta scelto il primo elemento, rimangono $5 - 1 = 4$ scelte per il secondo elemento.

Per ottenere le disposizioni di classe 3 possiamo ancora dividere le disposizioni mettendo insieme le disposizioni che iniziano per una disposizione (i, j) di classe 2. Le disposizioni di classe 3 che iniziano per tali elementi sono del tipo (i, j, k) con k che varia nell'insieme $\{1, 2, 3, 4, 5\} \setminus \{i, j\}$ di cardinalità $5 - 2 = 3$. Otterremo quindi $60 = 20 \cdot 3$ disposizioni.

(1,2,3)	(2,1,3)	(3,1,2)	(4,1,2)	(5,1,2)
(1,2,4)	(2,1,4)	(3,1,4)	(4,1,3)	(5,1,3)
(1,2,5)	(2,1,5)	(3,1,5)	(4,1,5)	(5,1,4)
(1,3,2)	(2,3,1)	(3,2,1)	(4,2,1)	(5,2,1)
(1,3,4)	(2,3,4)	(3,2,4)	(4,2,3)	(5,2,3)
(1,3,5)	(2,3,5)	(3,2,5)	(4,2,5)	(5,2,4)
(1,4,2)	(2,4,1)	(3,4,1)	(4,3,1)	(5,3,1)
(1,4,3)	(2,4,3)	(3,4,2)	(4,3,2)	(5,3,2)
(1,4,5)	(2,4,5)	(3,4,5)	(4,3,5)	(5,3,4)
(1,5,2)	(2,5,1)	(3,5,1)	(4,5,1)	(5,4,1)
(1,5,3)	(2,5,3)	(3,5,2)	(4,5,2)	(5,4,2)
(1,5,4)	(2,5,4)	(3,5,4)	(4,5,3)	(5,4,3)

Come si contano le disposizioni-II

Quanto fatto nella nota di approfondimento precedente si generalizza al caso n e k (con $1 \leq k \leq n$) ottenendo

$$D_1^n = n, \quad \text{e la formula ricorsiva} \quad D_k^n = D_{k-1}^n (n - (k - 1))$$

da cui si ricava immediatamente che

$$D_1^n = n, \quad D_2^n = n(n-1), \quad D_3^n = n(n-1)(n-2), \quad D_k^n = n(n-1) \cdots (n - (k - 1)),$$

$$\text{ossia} \quad D_k^n = \frac{n!}{(n-k)!} \quad \text{e per le permutazioni} \quad P_n = D_n^n = n!$$

Relazioni tra disposizioni e combinazioni-I

Abbiamo visto che per trovare il numero C_k^n delle combinazioni di n elementi di classe k , basta osservare che

$$D_k^n = C_k^n \cdot P_k,$$

da cui immediatamente

$$C_k^n = \frac{D_k^n}{P_k} = \frac{n!}{k!(n-k)!}.$$

Iniziamo con il caso particolare $n = 5$ e $k = 3$. Riscriviamo tutte le 60 disposizioni mettendo in ogni riga tutte quelle che contengono gli stessi tre elementi, e poi permutiamo i tre elementi in tutti i modo possibili.

{1,2,3}	(1,2,3)	(1,3,2)	(2,1,3)	(2,3,1)	(3,1,2)	(3,2,1)
{1,2,4}	(1,2,4)	(1,4,2)	(2,1,4)	(2,4,1)	(4,1,2)	(4,2,1)
{1,2,5}	(1,2,5)	(1,5,2)	(2,1,5)	(2,5,1)	(5,1,2)	(5,2,1)
{1,3,4}	(1,3,4)	(1,4,3)	(3,1,4)	(3,4,1)	(4,1,3)	(4,3,1)
{1,3,5}	(1,3,5)	(1,5,3)	(3,1,5)	(3,5,1)	(5,1,3)	(5,3,1)
{1,4,5}	(1,4,5)	(1,5,4)	(4,1,5)	(4,5,1)	(5,1,4)	(5,4,1)
{2,3,4}	(2,3,4)	(2,4,3)	(3,2,4)	(3,4,2)	(4,2,3)	(4,3,2)
{2,3,5}	(2,3,5)	(2,5,3)	(3,2,5)	(3,5,2)	(5,2,3)	(5,3,2)
{2,4,5}	(2,4,5)	(2,5,4)	(4,2,5)	(4,5,2)	(5,2,4)	(5,4,2)
{3,4,5}	(3,4,5)	(3,5,4)	(3,1,5)	(3,5,1)	(3,1,4)	(5,4,3)

Abbiamo ottenuto quindi una tabella con 10 righe, ossia il numero dei sottoinsiemi di cardinalità 3 dell'insieme $\{1,2,3,4,5\}$ e 6 colonne, in quanto ogni riga ha 6 elementi, in quanto 6 sono le permutazioni di tre elementi. Questa tabella contiene effettivamente tutte le 60 disposizioni.

Relazioni tra disposizioni e combinazioni-II

In generale, per dimostrare che D_k^n è il prodotto di C_k^n per P_k si può procedere osservando che le disposizioni di n elementi di classe k si possono raggruppare in modo che ciascun gruppo è formato da disposizioni che contengono gli stessi elementi, ma differiscono solo per l'ordine: ognuno di tali gruppi individua quindi il sottoinsieme degli elementi comuni, ovvero una combinazione di classe k di n elementi.

Chiaramente ciascuno dei C_k^n gruppi è composto dallo stesso numero (P_k) di disposizioni: da ciascuna combinazione di classe k di n elementi si ottengono P_k disposizioni diverse, permutando tra loro i k elementi distinti che compongono la combinazione. Quest'ultima osservazione significa appunto che vale la relazione $D_k^n = C_k^n \cdot P_k$.

3.6 Esercizi di verifica

Esercizio 3.1. Le lettere AAMMM vengono ordinate a caso. Qual è la probabilità di ottenere la parola MAMMA?

Esercizio 3.2. Si fanno n lanci di una moneta perfetta. Per $1 \leq h \leq n$, qual è la probabilità di ottenere il risultato testa per la prima volta all' h -esimo lancio?

Esercizio 3.3. Da un'urna, che contiene 6 oggetti numerati da 1 a 6, si estraggono a caso tre oggetti contemporaneamente. Qual è la probabilità che il minimo numero estratto sia superiore a 2?

Esercizio 3.4. In una mano del gioco della roulette si punta su $\{pair\}$, $\{passe\}$, $\{16\}$. Qual è la probabilità di vincere almeno una di queste puntate?

Esercizio 3.5. Qual è la probabilità che il numero 16 esca almeno una volta su cinque mani del gioco della roulette?

Esercizio 3.6. Qual è la probabilità che esca il numero 16 in una delle cinque estrazioni su una ruota del lotto? (Si estrae senza reinserimento da un'urna contenente i numeri $\{1, 2, \dots, 90\}$).

Esercizio 3.7. Qual è la probabilità che esca la coppia di numeri 16 e 48 nelle cinque estrazioni su una ruota del lotto?

Esercizio 3.8. Qual è la probabilità che esca la terna di numeri 16, 48, 90 nelle cinque estrazioni su una ruota del lotto?

Esercizio 3.9. Vengono lanciati contemporaneamente 5 dadi perfetti.

Calcolate la probabilità degli eventi elencati qui di seguito:

- $\{tutti\ i\ dadi\ danno\ punteggi\ diversi\ fra\ loro\}$
- $\{due\ dadi\ danno\ punteggi\ uguali\ fra\ loro\ e\ gli\ altri\ tre\ danno\ punteggi\ tutti\ diversi\}$
("coppia")
- $\{tre\ dadi\ danno\ punteggi\ uguali\ fra\ loro\ e\ gli\ altri\ due\ danno\ due\ punteggi\ diversi\}$
("tris")
- $\{quattro\ dadi\ danno\ punteggi\ uguali\ fra\ loro\ e\ uno\ da\ un\ punteggio\ diverso\}$
("poker")
- $\{tutti\ i\ dadi\ danno\ lo\ stesso\ punteggio\}$

(“jazzi”)

f) {due diverse coppie di punteggi fra loro uguali e un punteggio diverso dagli altri due}

(“doppia coppia”)

g) {tre punteggi uguali fra loro e gli altri due uguali fra loro e diversi dal precedente}

(“full”).

Esercizio 3.9 bis. Riformulate da soli ove possibile, con gli opportuni cambiamenti, e poi risolvete l’analogo dell’esercizio precedente per il caso del lancio di soli tre dadi. (Questo esercizio si può saltare se non si sono trovate eccessive difficoltà a risolvere completamente l’esercizio precedente).

Esercizio 3.10. Un servizio da tè consiste di quattro tazzine e quattro piattini con due tazzine e due piattini di un colore e i rimanenti di un altro colore. Le tazzine sono poste a caso sopra i piattini. Calcolare le probabilità degli eventi:

{Nessuna tazzina è su un piattino dello stesso colore}

{Una sola tazzina è su un piattino dello stesso colore}

{Due sole tazzine sono su un piattino dello stesso colore}

Calcolare la probabilità dell’evento

{Nessuna tazzina su un piattino dello stesso colore}

se il servizio è composto di quattro tazzine e quattro piattini di quattro colori diversi.

4 Probabilità condizionate

In questa lezione verrà introdotta la definizione di probabilità condizionata e ne verranno illustrate alcune conseguenze immediate: la **Formula delle probabilità composte**, la **Formula delle probabilità totali** e la **Formula di Bayes**.

Abbiamo già visto che, in uno spazio di probabilità finito, la teoria della probabilità è in effetti già tutta contenuta nella formula (3), che mostra come si ottenga la probabilità di un evento composto, una volta assegnate le probabilità a ciascun evento elementare. Si vedrà comunque che le formule che verranno ottenute nel seguito (e la nozione di indipendenza stocastica che verrà illustrata a partire dalla prossima lezione) costituiscono spesso una guida al ragionamento probabilistico, che può rivelarsi complementare all'uso della formula (3). Tali nozioni infatti permettono, alcune volte, di assegnare probabilità ad eventi composti (oppure di calcolarle sulla base di probabilità assegnate ad altri eventi) in modo più diretto, senza necessariamente far intervenire tutta la collezione degli eventi semplici. Vedremo nelle successive lezioni, in particolare, come si possano risolvere, in modo alternativo, alcuni degli esercizi affrontati nella lezione precedente.

Prima di iniziare tale studio è opportuno **ricordare** il significato "logico" della nozione di **partizione di un insieme**.

Prima di tutto ricordiamo che una collezione di sottoinsiemi H_1, \dots, H_m dello spazio Ω ($H_l \in \mathcal{P}(\Omega), l = 1, \dots, m$), **costituisce** una partizione di Ω , **se e solo se è** tale che

$$\bigcup_{l=1}^m H_l = \Omega; \quad H_{l_1} \cap H_{l_2} = \emptyset, \text{ per } l_1 \neq l_2.$$

Interpretando H_1, \dots, H_m come eventi, abbiamo che essi sono a **due a due incompatibili** (cioè è certo che non se ne possono verificare due contemporaneamente) e, d'altra parte, essi sono **esaustivi** (è certo che se ne verifichi almeno uno); dunque: è certo che si verifichi uno ed uno soltanto degli eventi H_1, \dots, H_m (la nostra situazione di incertezza risiede nel fatto che non sappiamo quale di essi sia verificato).

Come si era visto quale immediata conseguenza degli assiomi della probabilità si ha che, se H_1, \dots, H_m costituisce una partizione di Ω , allora deve risultare

$$\sum_{l=1}^m P(H_l) = 1.$$

Inoltre osserviamo che, per qualunque evento $E \in \mathcal{P}(\Omega)$, possiamo scrivere

$$E = (E \cap H_1) \cup \dots \cup (E \cap H_m)$$

e dunque

$$P(E) = P(E \cap H_1) + \dots + P(E \cap H_m). \quad (23)$$

4.1 Definizione di probabilità condizionata

Cominciamo con un esempio

Esempio 4.1. *In un lancio di un dado a sei facce, quale probabilità dobbiamo assegnare all'evento $A \equiv \{X \text{ dispari}\}$, sapendo che si è verificato l'evento $B \equiv \{X \geq 2\}$?*

Soluzione. Tutti gli eventi elementari

$$\{X = 1\}, \quad \{X = 2\}, \dots, \quad \{X = 6\}$$

sono inizialmente ritenuti equiprobabili.

Sapere che si è verificato l'evento B equivale a sapere che si è verificato uno dei seguenti eventi elementari:

$$\{X = 2\}, \dots, \{X = 6\} \quad (\text{non sappiamo però "quale"}).$$

È naturale, a questo punto, **assumere** quanto segue:

l'informazione che si è verificato l'evento B non modifica la situazione di equiprobabilità fra gli eventi $\{X = 2\}, \dots, \{X = 6\}$.

A seguito di tale informazione, quindi, la probabilità di osservare l'evento A deve essere dunque valutata come la probabilità del verificarsi di uno fra 2 eventi elementari favorevoli su un totale di 5 eventi elementari possibili, equiprobabili fra loro; valuteremo quindi tale probabilità "condizionata" uguale a $\frac{2}{5}$.

La soluzione del precedente esempio mostra che, nel caso di un numero finito di eventi elementari equiprobabili, è naturale imporre che la probabilità da attribuire ad un evento A , quando si sappia per certo che si è verificato un evento B , sia data da

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|}{|\Omega|} \frac{|\Omega|}{|B|} = \frac{P(A \cap B)}{P(B)}.$$

Ciò suggerisce la seguente

Definizione 4.1 (Probabilità condizionata). *Siano E ed H due eventi, con $P(H) > 0$. Viene detta **probabilità condizionata di E dato H** , ed indicata con il simbolo $P(E|H)$, la quantità*

$$P(E|H) = \frac{P(E \cap H)}{P(H)}. \quad (24)$$

Osservazione 1 (di carattere euristico). All'interno di ciascuna delle interpretazioni della probabilità (classica, frequentista, soggettivista, ...) cui si è accennato in precedenza, il numero $P(E|H)$ definito nella (24) coincide effettivamente con la probabilità che, coerentemente con tale interpretazione, dovremmo attribuire al verificarsi di E , se sapessimo che si è verificato H .

Ciò costituisce la motivazione per definire "assiomaticamente" la nozione di probabilità condizionata attraverso la (24).

Esercizio proposto 4.1 (continuazione dell'Esempio 2.2, dado non equilibrato). *Un dado ha sei facce numerate da 1 a 6; esso è pesato in modo tale che ciascuna faccia abbia una probabilità di presentarsi (in un singolo lancio) proporzionale al suo valore. Siano*

$$A \equiv \{\text{Si presenta un numero pari}\}, \quad B \equiv \{\text{Si presenta un numero primo}\}.$$

Calcolare $P(B|A)$ e $P(A|B)$.

Vediamo ora le semplici, ma importanti, conseguenze della definizione di probabilità condizionata, già menzionate in precedenza.

4.2 Conseguenze immediate della definizione di probabilità condizionata

4.2.1 Formula delle probabilità composte

Dalla definizione di probabilità condizionata si ottiene immediatamente che, se $P(E_1) > 0$, allora

$$P(E_1 \cap E_2) = P(E_2|E_1)P(E_1). \quad (25)$$

Questa formula si può generalizzare.

Proposizione 1. (Formula delle probabilità composte) Consideriamo n eventi E_1, E_2, \dots, E_n , tali che

$$P(E_1 \cap E_2 \cap \dots \cap E_{n-1}) > 0. \quad (26)$$

Si ha

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \dots \cdot P(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1}). \quad (27)$$

Dimostrazione. Iniziamo con l'osservare che, essendo $E_1 \supseteq E_1 \cap E_2 \supseteq \dots \supseteq E_1 \cap E_2 \cap \dots \cap E_{n-1}$, per la proprietà di monotonia della probabilità si ha $P(E_1) \geq P(E_1 \cap E_2) \geq \dots \geq P(E_1 \cap E_2 \cap \dots \cap E_{n-1}) > 0$. Quindi la condizione (26) implica che $P(E_1) > 0$, $P(E_1 \cap E_2) > 0$, ..., $P(E_1 \cap E_2 \cap \dots \cap E_{n-1}) > 0$, per cui il prodotto a destra della (27) ha senso.

L'uguaglianza (27) segue immediatamente dalla definizione di probabilità condizionata: possiamo scrivere

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1})P(E_1 \cap E_2 \cap \dots \cap E_{n-1})$$

A sua volta $P(E_1 \cap E_2 \cap \dots \cap E_{n-1})$ può essere scritto come

$$P(E_{n-1}|E_1 \cap E_2 \cap \dots \cap E_{n-2})P(E_1 \cap E_2 \cap \dots \cap E_{n-2}).$$

La dimostrazione quindi si ottiene facilmente proseguendo così di seguito, fino a scrivere

$$P(E_1 \cap E_2) = P(E_2|E_1)P(E_1).$$

La precedente dimostrazione si può formalizzare utilizzando il principio di induzione. Il caso $n = 2$ corrisponde alla formula (25). Supposta vera l'affermazione della Proposizione per $m - 1$ eventi, ovvero

$$\begin{aligned} &P(E_1 \cap E_2 \cap \dots \cap E_{m-1}) = \\ &= P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \dots \cdot P(E_{m-1}|E_1 \cap E_2 \cap \dots \cap E_{m-2}). \end{aligned}$$

mostriamo ora che vale per m eventi, e infatti

$$\begin{aligned} &P(E_1 \cap E_2 \cap \dots \cap E_{m-1} \cap E_m) \\ &= P((E_1 \cap E_2 \cap \dots \cap E_{m-1}) \cap E_m) \\ &= P(E_1 \cap E_2 \cap \dots \cap E_{m-1}) \cdot P(E_m|E_1 \cap E_2 \cap \dots \cap E_{m-1}) \\ &= [P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \dots \cdot P(E_{m-1}|E_1 \cap E_2 \cap \dots \cap E_{m-2})] \\ &\quad \cdot P(E_m|E_1 \cap E_2 \cap \dots \cap E_{m-1}). \end{aligned}$$

□

La formula delle probabilità composte di solito viene usata per trovare la probabilità dell'intersezione di un numero finito di eventi, specialmente quando è più facile valutare le probabilità condizionate rispetto alle probabilità dell'intersezione. Questa idea viene illustrata nel seguente esempio (confrontare anche le osservazioni al successivo Esempio 5.2 della Lez. 5, e ancora l'Osservazione 1 della Lez. 6).

Esempio 4.2. *Un uomo ha un mazzo di n chiavi, una sola delle quali apre un porta. Egli prova le chiavi a caso ad una ad una, escludendo dal mazzo quelle già provate, finché non trova la chiave giusta. Vogliamo trovare la probabilità dell'evento*

$$E \equiv \{\text{chiave giusta al } k\text{-esimo tentativo}\}$$

Soluzione. Scriviamo E come intersezione di diversi eventi come segue:

$$E \equiv \{\text{chiave errata al primo tentativo}\} \cap \{\text{chiave errata al secondo tentativo}\} \cap \dots \cap \{\text{chiave giusta al } k\text{-esimo tentativo}\}.$$

Utilizzando la formula delle probabilità composte possiamo scrivere dunque

$$P(E) = P(\text{chiave errata al primo tentativo}) \cdot$$

$$\begin{aligned} & P(\{\text{chiave errata al secondo tentativo}\} | \{\text{chiave errata al primo tentativo}\}) \cdot \dots \cdot \\ & \dots \cdot P(\{\text{chiave giusta al } k\text{-esimo tentativo}\} | \{\text{chiave errata al primo tentativo}\} \cap \dots \cap \\ & \dots \cap \{\text{chiave errata al } (k-1)\text{-esimo tentativo}\}) = \\ & = \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \dots \cdot \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}. \end{aligned}$$

In simboli, ponendo $C_h = \{\text{chiave giusta al } h\text{-esimo tentativo}\}$

$$E = \overline{C}_1 \cap \overline{C}_2 \cap \dots \cap \overline{C}_{k-1} \cap C_k$$

e quindi

$$\begin{aligned} P(E) &= P(\overline{C}_1)P(\overline{C}_2|\overline{C}_1) \dots P(\overline{C}_{k-1}|\overline{C}_1 \cap \overline{C}_2 \cap \dots \cap \overline{C}_{k-2})P(C_k|\overline{C}_1 \cap \overline{C}_2 \cap \dots \cap \overline{C}_{k-1}) \\ &= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \dots \cdot \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}. \end{aligned}$$

4.2.2 Formula delle probabilità totali

Proposizione 4.1. *Sia $\{H_1, \dots, H_m\}$ una partizione di Ω , con $P(H_l) > 0$, $l = 1, \dots, m$. Per un qualunque evento $E \in \mathcal{P}(\Omega)$, risulta*

$$P(E) = P(E|H_1)P(H_1) + \dots + P(E|H_m)P(H_m),$$

o più brevemente

$$P(E) = \sum_{k=1}^m P(E|H_k)P(H_k).$$

Dimostrazione. Basta ricordare la precedente formula (23), ovvero $P(E) = \sum_{k=1}^m P(E \cap H_k)$, e tener conto che, essendo $P(H_l) > 0$, si ha, per $l = 1, \dots, m$, $P(E \cap H_l) = P(E|H_l)P(H_l)$. □

Come nel caso della formula delle probabilità composte, anche la formula delle probabilità totali si mostra particolarmente utile quando è più semplice valutare le probabilità condizionate $P(E|H_l)$ che quelle dell'intersezione $P(E \cap H_l)$.

Esempio 4.3 (Estrazione di un numero al lotto). *Qual è la probabilità che sia uguale a 16 il secondo estratto su una ruota del lotto?*

Soluzione. Indichiamo con X_1, X_2 i valori rispettivamente ottenuti nella prima e nella seconda estrazione e poniamo

$$E \equiv \{X_2 = 16\}.$$

Consideriamo la partizione $\{H, \bar{H}\}$, dove

$$H \equiv \{X_1 = 16\},$$

e applichiamo la formula delle probabilità totali; si ha così

$$P(E) = P(E|H)P(H) + P(E|\bar{H})P(\bar{H}),$$

da cui otteniamo facilmente

$$P(E) = 0 \times \frac{1}{90} + \frac{1}{89} \times \frac{89}{90} = \frac{1}{90}.$$

Esempio 4.4. *Una moneta perfetta viene lanciata n volte. Qual è la probabilità di ottenere un numero pari di risultati testa?*

Soluzione. Poniamo, per $k = 1, 2, \dots, n$

$$E_k \equiv \{\text{numero pari di risultati testa sui primi } k \text{ lanci}\}.$$

Si capisce subito che, se la moneta è perfetta, si avrà per motivi di simmetria,

$$P(E_n) = P(\bar{E}_n) = \frac{1}{2}.$$

È però utile anche ragionare come segue, adoperando la formula delle probabilità totali:

$$P(E_n) = P(E_{n-1}) \cdot P(E_n|E_{n-1}) + P(\bar{E}_{n-1}) \cdot P(E_n|\bar{E}_{n-1}), \quad (28)$$

e risulta

$$P(E_n|E_{n-1}) = P(\{\text{risultato croce all}'n\text{-esimo lancio}\}|E_{n-1})$$

$$P(E_n|\bar{E}_{n-1}) = P(\{\text{risultato testa all}'n\text{-esimo lancio}\}|\bar{E}_{n-1})$$

Il fatto che la moneta sia perfetta ci porterà ora a valutare:

$$P(E_n|E_{n-1}) = P(E_n|\bar{E}_{n-1}) = \frac{1}{2}.$$

Otteniamo dunque dalla (28)

$$P(E_n) = \frac{1}{2} (P(E_{n-1}) + P(\bar{E}_{n-1})) = \frac{1}{2}.$$

4.2.3 Formula di Bayes

Applicando la definizione di probabilità condizionata e poi la formula delle probabilità composte si ottiene, se $P(E) > 0$ e $P(H) > 0$,

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

che rappresenta la **forma elementare della formula di Bayes**. Quando $H = H_l$, dove $\{H_1, \dots, H_m\}$ è una partizione dell'evento certo questa formula si generalizza nel seguente modo.

Proposizione 3. (Formula di Bayes) Sia ancora $\{H_1, \dots, H_m\}$ una partizione di Ω , con $P(H_l) > 0$, $l = 1, \dots, m$. Per un qualunque evento $E \in \mathcal{P}(\Omega)$, risulta

$$P(H_l|E) = \frac{P(E|H_l) P(H_l)}{\sum_{r=1}^m P(E|H_r) P(H_r)}, \quad l = 1, \dots, m. \quad (29)$$

Dimostrazione.

Per la definizione di probabilità condizionata, si ha

$$P(H_l|E) = \frac{P(H_l \cap E)}{P(E)}.$$

Applicando la formula delle probabilità composte al numeratore del membro a destra otteniamo

$$P(H_l \cap E) = P(E|H_l) P(H_l);$$

applicando la formula delle probabilità totali al denominatore otteniamo

$$P(E) = \sum_{r=1}^m P(E|H_r) P(H_r).$$

Osservazione 2 (di carattere euristico). La formula di Bayes trova naturale applicazione nei problemi in cui si debba analizzare come un'“osservazione” porti a modificare lo stato di informazione sugli eventi di una partizione; spesso problemi di tale tipo sono originati da questioni di “inferenza statistica”.

Fissiamo l'attenzione su una partizione di Ω , $\mathcal{H} \equiv \{H_1, \dots, H_m\}$: sappiamo che è verificato uno ed uno soltanto degli eventi H_1, \dots, H_m , ma non sappiamo quale.

Attribuiamo, rispettivamente, probabilità $P(H_1), \dots, P(H_m)$ a ciascuno di tali eventi (possiamo pensare che tali probabilità esprimono il nostro stato di informazione “iniziale” su tale partizione).

Supponiamo poi di avere l'informazione che è verificato l'evento E e ci chiediamo come, in conseguenza di ciò, si debbano modificare le probabilità da attribuire agli eventi H_1, \dots, H_m (cioè come ciò modifichi il nostro stato di informazione “iniziale” su \mathcal{H}).

Tali “nuove” probabilità coincideranno con le probabilità condizionate $P(H_l|E)$ ($l = 1, \dots, m$), che vanno calcolate attraverso la formula (29). Dunque la formula di Bayes può essere vista come la regola secondo cui lo stato di informazione su \mathcal{H} si modifica sulla base dell'osservazione dell'evento E .

Più in generale, si può considerare la nuova probabilità

$$P_E : \mathcal{P}(\Omega) \mapsto [0, 1]; \quad A \mapsto P_E(A) := P(A|E).$$

È facile verificare che P_E così definita è una misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$. Si lascia al lettore la verifica, di queste proprietà.

Esercizio proposto 4.2 (P_E è una probabilità). *Controllare che P_E verifica gli assiomi i), ii) e iii) della Definizione 2.1, ovvero che*

i) $P_E(A) \in [0, 1]$

ii) $P_E(\Omega) = 1$,

iii) se A_1 e A_2 sono eventi incompatibili, cioè $A_1 \cap A_2 = \emptyset$, allora $P_E(A_1 \cup A_2) = P_E(A_1) + P_E(A_2)$.

Il seguente esempio costituisce un paradigma di tale uso della formula di Bayes; seguendo infatti la logica illustrata in tale esempio, la formula di Bayes può essere applicata in molti altri problemi, sostanzialmente analoghi, suggeriti in diversi campi di applicazione della teoria delle probabilità.

Esempio 4.5. *In un lotto di pezzi (che risultano, all'apparenza, simili) vi sono elementi di tipo A, B e C, rispettivamente nelle proporzioni del 50%, 30%, 20%.*

Quelli di tipo A hanno una probabilità del 10% di guastarsi durante il loro utilizzo. Le analoghe probabilità per quelli di tipo B e C sono rispettivamente del 15% e del 18%, rispettivamente. Viene scelto un pezzo a caso dal lotto. Quale probabilità si deve attribuire al fatto che esso sia di tipo C, se si osserva un suo guasto durante l'utilizzo?

Soluzione. Poniamo

$$E \equiv \{\text{si osserva un guasto del pezzo scelto durante l'utilizzo}\}$$

$$H_1 \equiv \{\text{il pezzo scelto è di tipo A}\}$$

$$H_2 \equiv \{\text{il pezzo scelto è di tipo B}\}$$

$$H_3 \equiv \{\text{il pezzo scelto è di tipo C}\}.$$

La condizione che il pezzo sia scelto "a caso" si traduce nella assegnazione di probabilità

$$P(H_1) = 0.5, \quad P(H_2) = 0.3, \quad P(H_3) = 0.2.$$

Gli altri dati del problema forniscono:

$$P(E|H_1) = 0.10, \quad P(E|H_2) = 0.15, \quad P(E|H_3) = 0.18.$$

Dalla formula delle probabilità totali otteniamo

$$P(E) = 0.5 \times 0.10 + 0.3 \times 0.15 + 0.2 \times 0.18 = 0.131$$

e la formula di Bayes fornisce:

$$P(H_1|E) = \frac{50}{131}, \quad P(H_2|E) = \frac{45}{131}, \quad P(H_3|E) = \frac{36}{131}.$$

Osservazione 3. La formula di Bayes (29) può essere più brevemente scritta nella forma

$$P(H_l|E) \propto P(H_l) \cdot P(E|H_l), \quad l = 1, \dots, m.$$

Notiamo che, essendo \mathcal{H} una partizione di Ω , sappiamo già a priori che deve risultare

$$\sum_{l=1}^m P(H_l|E) = 1.$$

Infatti considerando quanto detto nell'**Osservazione 2**,

$$\sum_{l=1}^m P(H_l|E) = \sum_{l=1}^m P_E(H_l) = 1.$$

La quantità $K = \frac{1}{P(E)} = \frac{1}{\sum_{r=1}^m P(E|H_r)P(H_r)}$ ha dunque il ruolo di **costante di normalizzazione**.

Esempio 4.6. (Un esempio medico) Siamo nel 1990, agli inizi del problema dell'HIV, e siamo nella seguente situazione: c'è un test (poco costoso) che permette di verificare la sieropositività per l'HIV, ma che non è infallibile, nel senso che può accadere di risultare sieropositivi, anche senza essere stati colpiti dal virus. È noto che un individuo colpito dal virus HIV risulta sieropositivo in 999 casi su 1000, mentre un individuo "sano" risulta sieropositivo in un caso su 100. Infine è noto che, nella popolazione di New York, la percentuale dei colpiti dal virus è dello 0,6%. John, che è stato scelto a caso tra gli abitanti di New York, si sottopone al test. Calcolare la probabilità che John risulti sieropositivo e, nel caso in cui sia risultato sieropositivo, la probabilità che sia stato colpito veramente dal virus HIV. Ripetere i calcoli considerando la sieronegatività.

Soluzione. Iniziamo esplicitando le probabilità sulla base dei dati del problema. Indicando con S^+ l'evento {John risulta sieropositivo}, con S^- l'evento complementare {John risulta sieronegativo} e con H l'evento {John ha il virus HIV}, abbiamo

$$P(H) = \frac{6}{1000}, \quad P(\bar{H}) = \frac{994}{1000},$$

$$P(S^+|H) = \frac{999}{1000}, \quad (\text{quindi } P(S^-|H) = \frac{1}{1000}), \quad P(S^+|\bar{H}) = \frac{1}{100}, \quad (\text{quindi } P(S^-|\bar{H}) = \frac{99}{100}).$$

Di conseguenza

$$P(S^+) = P(H)P(S^+|H) + P(\bar{H})P(S^+|\bar{H}) = \frac{6}{1000} \frac{999}{1000} + \frac{994}{1000} \frac{1}{100} = 0,015934,$$

e analogamente

$$P(S^-) = P(H)P(S^-|H) + P(\bar{H})P(S^-|\bar{H}) = \frac{6}{1000} \frac{1}{1000} + \frac{994}{1000} \frac{99}{100} = 0,984066 (= 1 - P(S^+)).$$

Dalla formula di Bayes si ha

$$\begin{aligned} P(H|S^+) &= \frac{P(H)P(S^+|H)}{P(H)P(S^+|H) + P(\bar{H})P(S^+|\bar{H})} = \frac{\frac{6}{1000} \frac{999}{1000}}{\frac{6}{1000} \frac{999}{1000} + \frac{994}{1000} \frac{1}{100}} = \frac{\frac{6}{1000} \frac{999}{1000}}{\frac{6}{1000} \frac{999}{1000} + \frac{994}{1000} \frac{10}{1000}} \\ &= \frac{6 \cdot 999}{6 \cdot 999 + 994 \cdot 10} \simeq 0,37617672 \end{aligned}$$

e analogamente

$$\begin{aligned} P(H|S^-) &= \frac{P(H)P(S^-|H)}{P(H)P(S^-|H) + P(\bar{H})P(S^-|\bar{H})} = \frac{\frac{6}{1000} \frac{1}{1000}}{\frac{6}{1000} \frac{1}{1000} + \frac{994}{1000} \frac{99}{100}} = \frac{\frac{6}{1000} \frac{1}{1000}}{\frac{6}{1000} \frac{1}{1000} + \frac{994}{1000} \frac{990}{1000}} \\ &= \frac{6 \cdot 1}{6 \cdot 1 + 994 \cdot 990} \simeq 0,00000609. \end{aligned}$$

La probabilità $P(H|S^+)$ ottenuta può sembrare sorprendentemente bassa: alcune osservazioni a questo proposito sono contenute nella seguente nota.

Alcune osservazioni sull'esempio medico 4.6

Come abbiamo detto, la probabilità $P(H|S^+)$ ottenuta può sembrare sorprendentemente bassa, ma (come è chiarito nel seguito di questa nota) il risultato è meno strano se si pone l'attenzione su due fatti:

1) John non è stato sottoposto al test perché si era esposto al rischio di infezione, ma è stato scelto a caso.

2) Il test non è simmetrico, nel senso che dà errore con frequenza diversa a seconda del caso in cui uno sia affetto dal virus o no: infatti la probabilità di errore nel caso di individuo sano, $P(S^-|\bar{H}) = \frac{1}{100}$, è dieci volte $P(S^-|H) = \frac{1}{1000}$, la probabilità di errore nel caso di individuo infetto.

Per quanto riguarda il punto 1), supponiamo invece che John sia stato a contatto con una o più persone malate di HIV. La probabilità (a priori) di essere stato colpito dal virus non sarebbe stata uguale a $6/1000$, ma molto più alta. A titolo di esempio supponiamo che sia $P(H) = P(\bar{H}) = 1/2$, allora l'essere risultato positivo al test avrebbe comportato il seguente calcolo:

$$\begin{aligned} P(H|S^+) &= \frac{P(H)P(S^+|H)}{P(H)P(S^+|H) + P(\bar{H})P(S^+|\bar{H})} \\ &= \frac{\frac{1}{2} \frac{999}{1000}}{\frac{1}{2} \frac{999}{1000} + \frac{1}{2} \frac{10}{1000}} = \frac{999}{999 + 10} \simeq 0,99. \end{aligned}$$

Sembra quindi che questo test sia abbastanza buono nel caso in cui si abbiano seri motivi per sospettare un contagio (ovviamente sarà bene in ogni caso fare altre indagini).

Tuttavia, il risultato ottenuto nel caso di un individuo scelto a caso suggerisce il dubbio se un tale test sia valido per fare uno screening per una popolazione, nel senso che sembra dare troppi falsi allarmi. Questo accade perché il test non è simmetrico, nel senso del punto 2). Se la asimmetria non fosse stata così grande il test sarebbe stato migliore per uno screening. A titolo di esempio, sempre nel caso in cui John fosse scelto a caso, cioè con $P(H) = \frac{6}{1000}$, $P(\bar{H}) = \frac{994}{1000}$, la precisione del test nel caso di infezione fosse sempre la stessa, cioè $P(S^-|H) = \frac{1}{1000}$, ma la precisione del test fosse migliore nel caso di non infezione, ad esempio se fosse $P(S^+|\bar{H}) = \frac{1,5}{1000}$, allora si avrebbe

$$\begin{aligned} P(H|S^+) &= \frac{P(H)P(S^+|H)}{P(H)P(S^+|H) + P(\bar{H})P(S^+|\bar{H})} \\ &= \frac{\frac{6}{1000} \frac{999}{1000}}{\frac{6}{1000} \frac{999}{1000} + \frac{994}{1000} \frac{1,5}{1000}} = \frac{6 \cdot 999}{6 \cdot 999 + 994 \cdot 1,5} \simeq 0,8008, \end{aligned}$$

che è una probabilità abbastanza alta (ovviamente anche in questo caso sarebbero opportune ulteriori indagini).

4.3 Esercizi di verifica

Esercizio 4.1. A e B sono due eventi tali che

$$P(A \cap B) = 0.3, \quad P(A \cap \bar{B}) = 0.2, \quad P(\bar{A} \cap B) = 0.1.$$

Calcolare $P(A \cup B)$, $P(A|B)$.

Esercizio 4.2. Vengono estratti a caso, senza reinserimento, due elementi dall'insieme $\{1, 2, \dots, 9\}$. Poniamo

$$A_i \equiv \{X_i \text{ pari}\}, \quad i = 1, 2.$$

Utilizzando la formula delle probabilità composte, calcolare le probabilità degli eventi $A_1 \cap A_2$, $A_1 \cap \bar{A}_2$, $\bar{A}_1 \cap \bar{A}_2$.

Esercizio 4.3. Nel lancio di due dadi, qual è la probabilità condizionata che nessuno dei due punteggi sia superiore a 4 sapendo che la somma dei due punteggi è uguale a 7?

Esercizio 4.4. Abbiamo due urne: l'urna U contiene una sola pallina gialla ed r palline rosse; l'urna V contiene una sola pallina rossa ed r palline gialle. Viene scelta a caso una fra queste due urne e ne estraiamo (ancora a caso) una pallina.

a) Calcolare la probabilità dell'evento

$$E \equiv \{\text{la pallina estratta è gialla}\}$$

b) Condizionatamente all'osservazione dell'evento E , qual è la probabilità di aver eseguito le estrazioni dall'urna U ?

Esercizio 4.5. Nel gioco della roulette, qual è la probabilità condizionata del risultato *pair*, dato che si è ottenuto il risultato *passee*? (Ricordiamo che il risultato $\{0\}$, non è *passee*, ne' *manque*, ne' *pair*, ne' *unpair*).

Esercizio 4.6. Un'urna contiene 3 palline rosse e 7 bianche; si esegue un'estrazione casuale e se ne reinserisce una pallina di colore opposto a quella estratta; si procede quindi ad una successiva estrazione casuale.

a) Qual è la probabilità di una pallina rossa alla seconda estrazione?

b) Sapendo che le palline estratte nelle due successive estrazioni sono dello stesso colore, qual è la probabilità che siano entrambe bianche?

Esercizio 4.7. Sto organizzando un appuntamento per una cena fra amici per questa sera. Non riesco a raggiungere Emilio per telefono e chiedo a Aldo e a Bruno di provare ad avvertirlo. Aldo e a Bruno proveranno separatamente ad avvertirlo, Aldo inviandogli un messaggio di posta elettronica e Bruno inviando un messaggio sul telefono cellulare. Dò le seguenti valutazioni di probabilità

$$P(\{\text{Emilio leggerà la sua posta elettronica}\}) = 0.7$$

$$P(\{\text{Emilio riceverà il messaggio sul suo cellulare}\}) = 0.8$$

$$P(\{\text{Emilio leggerà la posta elettronica e riceverà il messaggio sul cellulare}\}) = 0.56.$$

a) Come devo valutare la probabilità che Emilio venga all'appuntamento?

b) Dato che Emilio effettivamente si presenta all'appuntamento, come devo valutare la probabilità che egli abbia letto la sua posta elettronica?

Esercizio 4.8. Relativamente a due eventi A, B , suppongo di aver assegnato le probabilità

$$P(A), \quad P(B), \quad P(A|B).$$

Come devo valutare la probabilità che si sia verificato A , condizionatamente all'informazione che si è verificato almeno uno fra i due eventi A e B ?

5 Correlazione e indipendenza fra eventi

5.1 Il caso di due eventi: correlazione positiva, negativa e indipendenza

Riprendiamo l'Esempio 4.5 della precedente Lezione 4, sull'estrazione da un lotto di pezzi apparentemente identici. Come ci si poteva già aspettare intuitivamente prima di svolgere i calcoli, risulta

$$P(H_1|E) < P(H_1), \quad P(H_3|E) > P(H_3).$$

Da tale osservazione prendiamo spunto per formulare le seguenti definizioni.

Definizione 5.1. Siano A e B due eventi, con $P(A) > 0, P(B) > 0$. A e B si dicono **correlati positivamente** se risulta

$$P(A|B) > P(A).$$

Notiamo che tale condizione è equivalente a

$$P(A \cap B) > P(A) \cdot P(B)$$

e che, dunque, tale relazione è simmetrica.

Definizione 5.2. Due eventi A e B , con $P(A) > 0, P(B) > 0$, si dicono **correlati negativamente** se risulta

$$P(A|B) < P(A)$$

oppure

$$P(A \cap B) < P(A) \cdot P(B).$$

Definizione 5.3. Due eventi A e B si dicono **stocasticamente indipendenti** se risulta

$$P(A \cap B) = P(A) \cdot P(B).$$

Notiamo che non abbiamo richiesto necessariamente la condizione $P(A) > 0, P(B) > 0$. Se tale condizione è verificata e A e B sono indipendenti allora risulta

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

Esempio 5.1. Consideriamo l'esperimento relativo al lancio di due dadi. Imponiamo la condizione che ciascuno dei trentasei eventi elementari possibili abbia probabilità $\frac{1}{36}$. Consideriamo gli eventi composti $E_1 \equiv \{X_1 \text{ pari}\}$, $E_2 \equiv \{X_1 + X_2 \text{ pari}\}$, $E_3 \equiv \{X_1 + X_2 \leq 4\}$, $E_4 \equiv \{X_1 \leq 2\}$, $E_5 \equiv \{\max(X_1, X_2) > 3\}$.

È facile verificare che risulta:

E_1 ed E_2 sono stocasticamente indipendenti,

E_3 ed E_4 sono correlati positivamente,

E_3 ed E_5 sono correlati negativamente.

È importante a questo punto tener presente quanto segue.

Osservazione 1. Consideriamo ancora, a mò di esempio, l'esperimento del lancio di due dadi. Si verifica immediatamente che assegnare uguali probabilità $\frac{1}{36}$ a tutti gli eventi elementari implica che gli eventi (composti) del tipo $\{X_1 = i\}$, $\{X_2 = j\}$ sono tutti equiprobabili con probabilità uguale ad $\frac{1}{6}$ e $\{X_1 = i\}$, $\{X_2 = j\}$, per $1 \leq i \leq 6, 1 \leq j \leq 6$, costituiscono coppie di eventi indipendenti.

È d'altra parte immediato verificare anche il viceversa, cioè che la condizione

$$P(X_1 = i) = P(X_2 = j) = \frac{1}{6}$$

insieme all'indipendenza stocastica per tutte le coppie $\{X_1 = i\}, \{X_2 = j\}$ implica che tutti gli eventi elementari $\{X_1 = i\} \cap \{X_2 = j\}$ hanno probabilità uguale ad $\frac{1}{36}$.

Questa equivalenza prefigura un fatto piuttosto generale: *spesso l'assegnazione di probabilità non avviene imponendo le probabilità degli eventi elementari ma, piuttosto, imponendo che certi eventi composti abbiano probabilità assegnate ed imponendo l'indipendenza stocastica fra opportune coppie di eventi.*

Ciò può permettere di individuare quali debbano essere le corrispondenti probabilità per tutti gli eventi semplici (e quindi, attraverso la (3), per tutti gli eventi composti) oppure può permettere di individuare quali siano le probabilità almeno per certi eventi cui siamo effettivamente interessati.

Esempio 5.2. *Nel lancio di due dadi assumiamo che*

$$P(X_1 = i) = P(X_2 = j) = \frac{1}{6}, \quad 1 \leq i \leq 6, \quad 1 \leq j \leq 6$$

e che

$$P(\{X_1 \in I\} \cap \{X_2 \in J\}) = P(X_1 \in I) \cdot P(X_2 \in J),$$

essendo I, J una arbitraria coppia di sottoinsiemi di $\{1, 2, \dots, 6\}$.

Calcolare la probabilità dell'evento {almeno un punteggio ≥ 5 nei due lanci}.

Soluzione. Basta osservare che

$$\begin{aligned} & P(\{(X_1 \geq 5) \cup \{X_2 \geq 5\}\}) \\ &= P(X_1 \geq 5) + P(X_2 \geq 5) - P(\{X_1 \geq 5\} \cap \{X_2 \geq 5\}) \\ &= 2P(X_1 \geq 5) - [P(X_1 \geq 5)]^2 = \frac{2 \cdot 2}{6} - \frac{1}{9} = \frac{20}{36} = \frac{5}{9}. \end{aligned}$$

Osserviamo dunque che abbiamo calcolato tale probabilità imponendo i valori delle probabilità per alcuni eventi e di indipendenza stocastica fra certe coppie di eventi; non abbiamo fatto uso della formula (3) (che avrebbe ovviamente portato allo stesso risultato), cioè abbiamo trovato la soluzione dell'esercizio senza ricorrere ad un discorso di tipo combinatorio.

Possiamo osservare a tale proposito che un procedimento del tipo applicato qui è più sintetico, e ciò può costituire una caratteristica importante nei casi in cui $|\Omega|$ è un numero molto grande.

Esempio 5.3. *Tizio ha comprato due biglietti di ciascuna di due diverse lotterie. Sono stati emessi 700 biglietti per ciascuna lotteria e, in ogni lotteria, vengono estratti tre biglietti vincenti. Quale probabilità ha di vincere almeno un premio?*

Soluzione. Si sottointende che, per ciascuna lotteria, le estrazioni sono casuali e senza reinserimento; si sottointende inoltre che vi sia indipendenza stocastica fra quello che succede nelle due lotterie diverse.

Dunque, ponendo $E \equiv \{\text{Tizio vince almeno un premio}\}$ ed $E_i \equiv \{\text{Tizio vince almeno un premio nella lotteria } i\}$, con $i = 1, 2$, si ha

$$\begin{aligned} P(E) &= P(E_1 \cup E_2) \\ &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\ &= 2P(E_1) - [P(E_1)]^2. \end{aligned}$$

Per quanto riguarda il calcolo di $P(E_1)$, osserviamo che, applicando la formula delle probabilità composte, si ottiene

$$P(E_1) = 1 - P(\bar{E}_1) = 1 - \frac{697}{700} \cdot \frac{696}{699}.$$

Suggerimento per il calcolo della probabilità $P(\bar{E}_1)$

Per calcolare la probabilità di \bar{E}_1 possiamo ragionare in due modi:

(I MODO) Posto A_1^i l'evento all'estrazione i -sima non viene estratto nessuno dei due biglietti si ha

$$P(\bar{E}_1) = P(A_1^1 \cap A_1^2 \cap A_1^3) = P(A_1^1)P(A_1^2|A_1^1)P(A_1^3|A_1^1 \cap A_1^2) = \frac{698}{700} \frac{697}{699} \frac{696}{698} = \frac{697}{700} \cdot \frac{696}{699},$$

in quanto ci si riconduce a tre estrazioni senza reinserimento da un'urna che contiene 2 palline bianche (i due numeri dei due biglietti posseduti da Tizio, e 698 rosse, tutti gli altri: l'evento E_1 corrisponde allora all'evento nelle tre estrazioni escono solo palline rosse.

(II MODO) È interessante notare anche che a questo risultato si può arrivare anche pensando che invece si tratti di estrazioni in blocco, ossia di scegliere 2 biglietti tra i 700 di cui si sa che esattamente 3 sono vincenti, come sarebbe logico in un gratta e vinci, o una lotteria in cui si comprano biglietti su cui può essere scritta la frase *Non hai vinto ritenta* oppure *Hai vinto!*. Allora la situazione si riconduce a due estrazioni senza reinserimento da un'urna che contiene 3 palline verdi e 697 arancioni, e l'evento \bar{E}_1 diviene l'evento *{nelle due estrazioni si estraggono solo palline arancioni}*: la probabilità di non vincere diviene allora immediatamente

$$\frac{\binom{3}{0} \binom{697}{2}}{\binom{700}{2}} = \frac{697}{700} \cdot \frac{696}{699}.$$

In quanto segue approfondiremo alcuni aspetti critici della nozione di indipendenza stocastica fra eventi. D'ora in poi verrà utilizzato il simbolo $A \perp B$ per indicare l'indipendenza stocastica fra due eventi A e B .

5.2 Indipendenza fra partizioni e fra algebre di eventi

In uno spazio di probabilità, consideriamo due eventi A e B e le loro rispettive negazioni \bar{A} e \bar{B} .

È facile verificare che le seguenti relazioni sono fra di loro equivalenti:

$$A \perp B, \quad \bar{A} \perp B, \quad A \perp \bar{B}, \quad \bar{A} \perp \bar{B},$$

ad esempio, assumiamo $A \perp B$ e mostriamo che $\bar{A} \perp B$; infatti possiamo scrivere

$$\begin{aligned} P(\bar{A} \cap B) &= P(B) - P(A \cap B) = P(B) - P(A) \cdot P(B) \\ &= P(B) \cdot [1 - P(A)] = P(B) \cdot P(\bar{A}), \end{aligned}$$

e dunque $\bar{A} \perp B$.

Possiamo riassumere quanto sopra affermando che prendendo un qualunque evento della partizione $\{A, \bar{A}\}$ ed un qualunque evento della partizione $\{B, \bar{B}\}$ otteniamo una coppia di eventi indipendenti.

Ciò suggerisce la seguente definizione.

Definizione 5.4. Siano $\mathcal{A} \equiv \{A_1, A_2, \dots, A_n\}$ e $\mathcal{B} \equiv \{B_1, B_2, \dots, B_m\}$ due diverse partizioni (finite) di uno stesso spazio campione Ω . \mathcal{A} e \mathcal{B} sono due **partizioni indipendenti** se risulta

$$A_i \perp B_j, \quad \forall i = 1, \dots, n; j = 1, 2, \dots, m.$$

Esempio 5.4. Consideriamo di nuovo l'esperimento del lancio di due dadi e gli eventi

$$A_i \equiv \{X_1 = i\}, \quad i = 1, 2, \dots, 6; \quad B_j \equiv \{X_2 = j\}, \quad j = 1, 2, \dots, 6.$$

La condizione che tutti gli eventi elementari siano equiprobabili implica, come si è visto nei precedenti esempi, che le partizioni $\mathcal{A} \equiv \{A_1, A_2, \dots, A_6\}$ e $\mathcal{B} \equiv \{B_1, B_2, \dots, B_6\}$ sono indipendenti.

Osservazione 2. La nozione di indipendenza fra partizioni ha, nella teoria della probabilità, un importante significato concettuale, su cui ritorneremo in seguito. Per il momento ci limitiamo ad accennare che, in un certo senso, la nozione di indipendenza stocastica esprime una relazione che si addice ad una coppia di partizioni piuttosto che ad una coppia di eventi. In ogni caso vedremo presto che la nozione di indipendenza fra due partizioni ci servirà per definire in modo semplice e concettualmente efficiente la nozione di indipendenza fra due *variabili aleatorie* (Lez. 7 e 8). In tale prospettiva è utile presentare qui le seguenti nozioni.

Definizione 5.5 (Algebra). Una famiglia \mathcal{G} di eventi di Ω (dunque $\mathcal{G} \subseteq \mathcal{P}(\Omega)$) è un'**algebra**, se sono verificate le seguenti proprietà

- i) $\Omega \in \mathcal{G}$
- ii) $E \in \mathcal{G} \Rightarrow \bar{E} \in \mathcal{G}$
- iii) $E_1, E_2 \in \mathcal{G} \Rightarrow E_1 \cup E_2 \in \mathcal{G}$

È ovvio che se $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ è un'algebra allora si ha anche che $\emptyset \in \mathcal{G}$ e (per la *legge di De Morgan*) che $E_1, E_2 \in \mathcal{G} \Rightarrow E_1 \cap E_2 \in \mathcal{G}$.

Definizione 5.6 (Algebra generata da una famiglia di eventi). Sia $\mathcal{A} \equiv \{A_1, A_2, \dots, A_n\}$ una famiglia di eventi in uno spazio campione Ω . Si definisce **algebra generata da \mathcal{A}** , la famiglia $\mathcal{G}(\mathcal{A})$ di eventi di Ω caratterizzata dalle seguenti proprietà:

- * $\mathcal{G}(\mathcal{A})$ è un'algebra
- ** $\mathcal{A} \subseteq \mathcal{G}(\mathcal{A})$
- *** $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ algebra e $\mathcal{A} \subseteq \mathcal{C} \Rightarrow \mathcal{G}(\mathcal{A}) \subseteq \mathcal{C}$.

Possiamo dire cioè che $\mathcal{G}(\mathcal{A})$ è la più piccola famiglia di sottoinsiemi di Ω che abbia contemporaneamente le due proprietà di essere un'algebra e di contenere al suo interno tutti i sottoinsiemi della famiglia \mathcal{A} .

Esercizio proposto 5.1. Si dimostri che $\mathcal{P}(\Omega)$ è un'algebra.

Esercizio proposto 5.2 (algebra generata da una partizione). Si dimostri che se $\mathcal{A} \equiv \{A_1, A_2, \dots, A_n\}$ è una partizione (finita) dell'evento certo, allora $\mathcal{G}(\mathcal{A})$ è la famiglia degli insiemi $E \in \mathcal{P}(\Omega)$ che sono le unioni di eventi della partizione, ovvero

$$E = \bigcup_{i \in I} A_i = A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_k}, \quad \text{con } I = \{i_1, \dots, i_k\} \subseteq \{1, 2, \dots, n\},$$

con la convenzione che se $I = \emptyset$ allora $E = \bigcup_{i \in I} A_i = \emptyset$.

Vale in proposito la seguente

Proposizione 5.1. *Siano \mathcal{A} e \mathcal{B} due partizioni (finite) indipendenti di Ω . Allora, comunque scelti due eventi $E \in \mathcal{G}(\mathcal{A}), F \in \mathcal{G}(\mathcal{B})$ risulta $E \perp\!\!\!\perp F$.*

Per la verifica del precedente risultato è utile svolgere il seguente esercizio

Esercizio proposto 5.3. *Siano $\mathcal{A} \equiv \{A_1, A_2, \dots, A_n\}$ e $\mathcal{B} \equiv \{B_1, B_2, \dots, B_m\}$ due partizioni (finite) dell'evento certo. Mostrare che se $E = \cup_{r=1}^k A_{i_r} \in \mathcal{G}(\mathcal{A})$ ed $F = \cup_{s=1}^h B_{j_s} \in \mathcal{G}(\mathcal{B})$, allora*

$$E \cap F = \bigcup_{r=1}^k \bigcup_{s=1}^h A_{i_r} \cap B_{j_s}$$

$$P(E \cap F) = \sum_{r=1}^k \sum_{s=1}^h P(A_{i_r} \cap B_{j_s}).$$

Per la soluzione del precedente esercizio, si suggerisce di ricordare che

$$\sum_{r=1}^k \sum_{s=1}^h a_{i_r} b_{j_s} = \sum_{r=1}^k a_{i_r} \left(\sum_{s=1}^h b_{j_s} \right) = \left(\sum_{s=1}^h b_{j_s} \right) \left(\sum_{r=1}^k a_{i_r} \right)$$

Come applicazione della precedente **Proposizione 5.1** si suggerisce di risolvere il seguente esercizio.

Esercizio proposto 5.4. *Siano $\mathcal{A} \equiv \{A_1, A_2, \dots, A_6\}$ e $\mathcal{B} \equiv \{B_1, B_2, \dots, B_6\}$ le partizioni dell'Esempio 5.4, relativo al lancio di due dadi. Si verifichi che*

(a) $\mathcal{G}(\mathcal{A}) = \{\{X_1 \in I\}, \text{ per } I \subseteq \{1, 2, 3, 4, 5, 6\}\}$;

(b) $\mathcal{G}(\mathcal{B}) = \{\{X_2 \in J\}, \text{ per } J \subseteq \{1, 2, 3, 4, 5, 6\}\}$;

Come nell'Esempio 5.4 si assuma la sola condizione che tutti gli eventi elementari sono equiprobabili, in modo che le due partizioni \mathcal{A} e \mathcal{B} sono indipendenti. Utilizzando la **Proposizione 5.1** e i punti (a) e (b) precedenti si verifichi che

(c) qualunque siano $I, J \subseteq \{1, 2, 3, 4, 5, 6\}$, gli eventi $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti.

5.3 Indipendenza completa e prove bernoulliane

Dovrebbe essere abbastanza chiaro il seguente significato intuitivo della condizione di indipendenza stocastica fra due eventi A e B : A e B sono indipendenti se il sapere con certezza che si è verificato B , o anche il sapere con certezza che non si è verificato B , non modifica le aspettative circa il verificarsi, o meno, dell'evento A .

Ovviamente si tratta di un concetto limite, di una condizione ideale, che viene assunta quale ipotesi di lavoro per ottenere delle rilevanti semplificazioni nell'analisi di un problema reale (è utile, per fare un'analogia, pensare ad esempio al concetto di *punto materiale* in Meccanica: si tratta di una condizione limite, mai realizzata, ma che viene assunta ogni qualvolta sia accettabile entro una discreta approssimazione).

Come si è già detto, tale nozione è fondamentale nella costruzione di modelli probabilistici. Infatti, in pratica, nell'assegnare una misura di probabilità su uno spazio campionario, si parte sempre dall'individuazione di famiglie di eventi a ciascuno dei quali si impone uguale probabilità e a coppie di eventi entro cui si impone l'indipendenza stocastica. In base a tali posizioni si deduce quale debba essere la probabilità dei vari eventi composti, interessanti nel problema stesso (o almeno si deducono delle condizioni cui tali probabilità debbono soddisfare).

Per rimanere sul piano dell'esemplificazione più spicciola, assumiamo comunemente, ad esempio, che le estrazioni del lotto su ruote diverse in una stessa settimana siano fenomeni indipendenti fra di loro, i successivi lanci di una moneta perfetta siano indipendenti fra di loro, etc...

Vedremo comunque presto che vi sono delle naturali situazioni in cui la condizione di indipendenza è palesemente contraddetta; per ora accenniamo soltanto che ciò accade nei casi in cui una situazione di mancanza di informazione fa sì che ciascun evento osservato contiene un forte valore informativo, che si riflette sulle aspettative relative ad altri eventi connessi. Tale punto verrà sviluppato nella successiva Lezione 11.

Veniamo ora ad aspetti tecnici della nozione di indipendenza. Dobbiamo rilevare a questo proposito che la definizione precedentemente formulata, si rivela non adeguata ad esprimere compiutamente una condizione di indipendenza reciproca fra molti eventi diversi.

Ciò è efficacemente illustrato dal seguente semplice esempio.

Esempio 5.5. Riprendiamo ancora una volta il caso del lancio di due dadi e consideriamo gli eventi: $A \equiv \{X_1 \text{ pari}\}$, $B \equiv \{X_2 \text{ pari}\}$, $C \equiv \{X_1 + X_2 \text{ dispari}\}$. Imponendo la condizione di equiprobabilità fra gli eventi elementari abbiamo le relazioni: $A \perp B$, $A \perp C$, $B \perp C$. Notiamo però che ovviamente risulta

$$P(C|A \cap B) = 0.$$

Tale conclusione contrasta, naturalmente, con il significato di indipendenza e mostra l'esigenza di una definizione appropriata per il caso di più di due eventi.

Si dà allora la seguente definizione. Sia $\{E_1, E_2, \dots, E_n\}$ una famiglia di eventi in uno stesso spazio di probabilità e consideriamo le partizioni $\mathcal{P}_1 \equiv \{E_1, \bar{E}_1\}$, ..., $\mathcal{P}_n \equiv \{E_n, \bar{E}_n\}$.

Definizione 5.7. Gli eventi E_1, E_2, \dots, E_n sono una famiglia di eventi **completamente (o globalmente) indipendenti** se comunque estratti degli indici $\{j_1, \dots, j_m\}$ dall'insieme $\{1, 2, \dots, n\}$ ($2 \leq m \leq n$) e comunque scelti degli eventi $A_i \in \mathcal{P}_{j_i}$ (dunque $A_i = E_{j_i}$ oppure $A_i = \bar{E}_{j_i}$) risulta

$$P(A_1 \cap A_2 \cap \dots \cap A_m) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_m). \quad (30)$$

Chiaramente la precedente definizione implica le seguenti due condizioni: considerando solo il caso $A_i = E_{j_i}$ (e non il caso $A_i = \bar{E}_{j_i}$)

$$P(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = P(E_{j_1}) \cdot P(E_{j_2}) \cdot \dots \cdot P(E_{j_m}) \quad (31)$$

per ogni $\{j_1, \dots, j_m\} \subseteq \{1, 2, \dots, n\}$, con $2 \leq m \leq n$.

oppure, considerando solo il caso $m = n$,

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n), \quad \text{con } A_i = E_i \text{ oppure } A_i = \bar{E}_i \quad (32)$$

Osservazione 3 È importante notare che in alcuni testi la definizione di famiglia di eventi completamente indipendenti è data attraverso la relazione (31), mentre in altri è data attraverso

la (32). Ciò è dovuto al fatto che le relazioni (31) e (32) sono equivalenti, ed entrambe implicano la (30) (risultando quindi equivalenti alla (30)). Ciò è ovvio per $n = 2$, come detto all'inizio del paragrafo 5.2. Non diamo qui la dimostrazione di questa proprietà, cioè l'equivalenza tra le tre proprietà (31), (32) e (30), ma proponiamo al lettore il seguente esercizio.

Esercizio proposto 5.5. *Dimostrare l'equivalenza tra (31), (32) e (30) per $n = 3$.*

Suggerimento: *la soluzione è basata sull'osservazione che per tre eventi A, B, C si ha $A \cap B = (A \cap B \cap C) \cup (A \cap B \cap \bar{C})$, e che quindi $P(A \cap B) = P(A \cap B \cap C) + P(A \cap B \cap \bar{C})$, da cui anche $P(A \cap B \cap \bar{C}) = P(A \cap B) - P(A \cap B \cap C)$.*

Un caso assai particolare ma di notevole interesse è quello individuato dalla seguente definizione di **schema di Bernoulli** o delle **prove di Bernoulli** (detto anche delle **prove ripetute**).

Definizione 5.8 (Schema di Bernoulli, o prove bernoulliane). *Gli eventi E_1, E_2, \dots, E_n costituiscono delle prove bernoulliane se sono completamente indipendenti ed hanno tutti una stessa probabilità θ , con $0 < \theta < 1$.*

Se E_1, E_2, \dots, E_n costituiscono delle **prove bernoulliane** si ha dunque, in particolare, per $m \leq n$ e per qualunque $\{j_1, \dots, j_m\} \subseteq \{1, 2, \dots, n\}$,

$$P(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = \theta^m.$$

5.4 Indipendenza completa di partizioni

Abbiamo già osservato che la nozione di indipendenza completa tra n eventi $E_i, i = 1, \dots, n$, si può esprimere attraverso una proprietà che coinvolge gli elementi delle partizioni $\mathcal{P}_i, i = 1, \dots, n$, generate da tali eventi, ossia $\mathcal{P}_i = \{E_i, \bar{E}_i\}$.

Date n partizioni $\mathcal{P}_i = \{H_1^i, H_2^i, \dots, H_n^i\}, i = 1, \dots, n$, si può generalizzare immediatamente la definizione di indipendenza completa:

Definizione 5.9. *Le partizioni $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$ sono una famiglia di partizioni **completamente (o globalmente) indipendenti** se comunque estratti degli indici $\{j_1, \dots, j_m\}$ dall'insieme $\{1, 2, \dots, n\}$ (con la condizione che $2 \leq m \leq n$) e comunque scelti degli eventi $A_i \in \mathcal{P}_{j_i}$, per $i = 1, \dots, m$, risulta*

$$P(A_1 \cap A_2 \cap \dots \cap A_m) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_m). \quad (33)$$

Chiaramente la precedente definizione implica la seguente condizione: considerando solo il caso $m = n$, comunque scelti $A_i \in \mathcal{P}_i$, per $i = 1, \dots, n$,

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n). \quad (34)$$

Si può dimostrare che le due condizioni (33) e (34) precedenti sono equivalenti e, a loro volta, equivalgono alle stesse condizioni, ma con gli eventi che appartengono alle algebre generate dalle partizioni. Più precisamente vale la seguente generalizzazione della Proposizione 5.1.

Proposizione 5.2. *Siano $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_n$ una famiglia di partizioni completamente (o globalmente) indipendenti. Allora, comunque scelti n eventi $E_i \in \mathcal{G}(\mathcal{P}_i)$, per $i = 1, \dots, n$, risulta*

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \dots \cdot P(E_n). \quad (35)$$

Pur non dando la dimostrazione di questa affermazione facciamo notare che in questo caso è facile ottenere dalla (35) che vale la relazione: comunque scelto m (con $2 \leq m \leq n$), comunque estratti degli indici $\{j_1, \dots, j_m\}$ dall'insieme $\{1, 2, \dots, n\}$ e comunque scelti degli eventi $E_{j_i} \in \mathcal{G}(\mathcal{P}_{j_i})$, per $i = 1, \dots, m$, risulta

$$P(E_{j_1} \cap E_{j_2} \cap \dots \cap E_{j_m}) = P(E_{j_1}) \cdot P(E_{j_2}) \cdot \dots \cdot P(E_{j_m}).$$

Basterà infatti considerare $E_\ell = \Omega$ per $\ell \notin \{j_1, \dots, j_m\}$ e applicare la (35).

Esempio 5.6. Si lanciano tre dadi e sia X_i il risultato dell' i -esimo lancio. Siano

$$\mathcal{P}_i = \{\{X_i = 1\}, \{X_i = 2\}, \{X_i = 3\}, \{X_i = 4\}, \{X_i = 5\}, \{X_i = 6\}\}.$$

Chiaramente, comunque scelti j_1, j_2 , e j_3 si ha

$$P(X_1 = j_1, X_2 = j_2, X_3 = j_3) = \frac{1}{216} = P(X_1 = j_1) P(X_2 = j_2) P(X_3 = j_3)$$

e quindi, tenuto conto del fatto che la relazione (34) è verificata, e che è equivalente alla (31), le tre partizioni sono indipendenti. Inoltre, grazie alla Proposizione 5.2, risulta anche che, comunque scelti tre sottoinsiemi I_1, I_2 ed I_3 di $\{1, 2, 3, 4, 5, 6\}$, si ha

$$P(\{X_1 \in I_1\} \cap \{X_2 \in I_2\} \cap \{X_3 \in I_3\}) = P(\{X_1 \in I_1\}) P(\{X_2 \in I_2\}) P(\{X_3 \in I_3\}).$$

5.5 Esercizi di verifica

Esercizio 5.1. A e B sono due eventi tali che

$$P(A \cap B) = 0.3, \quad P(A \cap \bar{B}) = 0.2, \quad P(\bar{A} \cap B) = 0.1.$$

Verificare se A e B sono stocasticamente indipendenti.

Esercizio 5.2. Mostrare che se A e B sono due eventi indipendenti e $A \subseteq B$, allora si ha $P(A) = 0$, oppure $P(B) = 1$.

Esercizio 5.3. Siano A e B due eventi fra loro incompatibili. Mostrare che A e B risultano stocasticamente indipendenti se e solo se almeno uno di essi ha probabilità nulla.

Esercizio 5.4. X ed Y indicano i punteggi ottenuti nel lancio di due dadi a sei facce. Poniamo

$$A \equiv \{\max(X, Y) < 5\}, \quad B \equiv \{\min(X, Y) > 3\}$$

a) Calcolare $P(A \cap B)$, $P(A \cup B)$, $P(A|B)$, $P(B|A)$.

b) Gli eventi A e B sono indipendenti? A e B sono incompatibili?

Esercizio 5.5. Consideriamo i due risultati *pair* e *passee* nel gioco della roulette. Sono stocasticamente indipendenti, correlati positivamente o correlati negativamente?

Esercizio 5.6. Indichiamo con X un numero selezionato a caso nell'insieme dei primi 120 numeri naturale e consideriamo gli eventi

$$E \equiv \{X \text{ pari}\}, \quad F \equiv \{X \text{ divisibile per } 3\}.$$

Fra E ed F sussiste correlazione positiva, negativa o indipendenza stocastica?

Esercizio 5.7. a) Qual è la probabilità che il numero 16 venga estratto su una data ruota del lotto in una fissata giornata?

b) Qual è la probabilità che il numero 16 non venga mai estratto su una data ruota del lotto per n giornate consecutive?

c) Qual è la probabilità condizionata che il numero 16 venga estratto l' $(n+1)$ -esima giornata, dato che non è mai stato estratto nelle n giornate precedenti?

Esercizio 5.8. Non riesco ad avvertire Emilio per telefono dell'appuntamento per la cena di questa sera. Come al solito Aldo gli invierà allora un messaggio di posta elettronica, Bruno gli invierà un messaggio sul telefono cellulare, e interverrà anche Carla, cercando di avvertire di persona la sorella di Emilio. Essi avranno successo rispettivamente con probabilità $P(A) = 0.8$, $P(B) = 0.7$ e $P(C) = 0.6$; i tre eventi, inoltre, sono completamente indipendenti.

a) Trovare la probabilità che Emilio venga informato dell'appuntamento.

b) Dato che Emilio si presenta effettivamente all'appuntamento, come devo valutare la probabilità che egli abbia letto la sua posta elettronica?

Esercizio 5.9. Una moneta viene lanciata n volte. Per quanto riguarda il primo lancio imponiamo

$$P(T) = P(C) = \frac{1}{2}.$$

Per quanto riguarda i successivi $(n-1)$ lanci richiediamo soltanto che sia

$$P(T, T, \dots, T) = P(C, C, \dots, C) = 0.$$

a) Discutete la differenza che sussiste fra tale situazione e quella descritta nell'Esempio 4.4.

b)

Esercizio 5.10. Siano E_1 ed E_2 due eventi in uno spazio campione Ω . Elencate gli eventi appartenenti all'algebra generata da $\{E_1, E_2\}$.

Esercizio 5.11. Consideriamo di nuovo l'esperimento del lancio di due dadi e gli eventi

$$A_i \equiv \{X_1 = i\}, \quad i = 1, 2, \dots, 6; \quad B_j \equiv \{X_2 = j\}, \quad j = 1, 2, \dots, 6.$$

Trovate una misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$, diversa da quella uniforme considerata nell'Esempio 5.4 (cioè tale che gli eventi elementari non siano tutti fra di loro siano equiprobabili), per la quale le partizioni $\mathcal{A} \equiv \{A_1, A_2, \dots, A_6\}$ e $\mathcal{B} \equiv \{B_1, B_2, \dots, B_6\}$ siano comunque indipendenti.

Esercizio 5.12. Dimostrate la precedente **Proposizione 5.1**.

6 Probabilità binomiali e ipergeometriche; estrazioni casuali da urne

In questo paragrafo vogliamo discutere in modo più sistematico due particolari modelli probabilistici, che sono già comparsi in precedenti esempi e che portano alle probabilità binomiali e ipergeometriche.

6.1 Probabilità binomiali

Consideriamo, su uno spazio finito di probabilità, n prove bernoulliane E_1, \dots, E_n ; cioè, ricordando la Definizione 5.8, assumiamo che E_1, \dots, E_n sono completamente indipendenti ed equiprobabili: ponendo $P(E_i) = \theta$, per $i = 1, 2, \dots, n$ e ponendo

$$X_i = \begin{cases} 1 & \text{se si verifica } E_i \\ 0 & \text{se si verifica } \bar{E}_i \end{cases} \quad i = 1, 2, \dots, n, \quad (36)$$

si ha, per ogni n -upla $\mathbf{x} \equiv (x_1, \dots, x_n) \in \{0, 1\}^n$,

$$P(\{X_1 = x_1, \dots, X_n = x_n\}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}, \quad (37)$$

dove $\{X_1 = x_1, \dots, X_n = x_n\}$ è un modo rapido di scrivere l'evento $\{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}$.

Poniamo ora

$$S_n \equiv \sum_{i=1}^n X_i$$

e consideriamo, per $k = 0, 1, \dots, n$, la probabilità dell'evento composto $\{S_n = k\}$; osserviamo che potremo scrivere

$$\{S_n = k\} = \left\{ \sum_{i=1}^n X_i = k \right\} = \bigcup_{\mathbf{x} \in \{0,1\}^n: \sum_{i=1}^n x_i = k} \{X_1 = x_1, \dots, X_n = x_n\}.$$

Due eventi del tipo

$$\{X_1 = x'_1, \dots, X_n = x'_n\}, \quad \{X_1 = x''_1, \dots, X_n = x''_n\}$$

sono ovviamente incompatibili nel caso $\mathbf{x}' \equiv (x'_1, \dots, x'_n) \neq \mathbf{x}'' \equiv (x''_1, \dots, x''_n)$ e, nel caso in cui

$$\sum_{i=1}^n x'_i = \sum_{i=1}^n x''_i = k,$$

essi risultano equiprobabili, entrambi di probabilità $\theta^k (1 - \theta)^{n-k}$, in virtù dell'equazione (37).

E dunque, dal momento che la cardinalità dell'insieme

$$\{\mathbf{x} \in \{0, 1\}^n : \sum_{i=1}^n x_i = k\} \quad (38)$$

è uguale a $\binom{n}{k}$, potremo scrivere, per $k = 0, 1, \dots, n$,

$$\begin{aligned} P(\{S_n = k\}) &= \sum_{\mathbf{x} \in \{0,1\}^n: \sum_{i=1}^n x_i = k} P(\{X_1 = x_1, \dots, X_n = x_n\}) \\ &= \binom{n}{k} \theta^k (1 - \theta)^{n-k}. \end{aligned} \quad (39)$$

Probabilità del tipo in (39) sull'insieme $\{0, 1, \dots, n\}$, prendono il nome di **probabilità binomiali**.

Esempio 6.1. *Un dado viene lanciato 10 volte. Sia S il numero di volte in cui si ottiene il risultato asso. Calcolare $P(\{S = 5\})$.*

Soluzione. Si tratta di 10 prove bernoulliane, di probabilità $\frac{1}{6}$; dunque

$$P(\{S = 5\}) = \binom{10}{5} \cdot \left(\frac{5}{6}\right)^5 \cdot \left(\frac{1}{6}\right)^5 = \binom{10}{5} \cdot \frac{5^5}{6^{10}}.$$

Esempio 6.2. *Ciascun viaggiatore che occupa un posto in uno scompartimento “per fumatori” in un treno EuroStar è effettivamente un fumatore (o fumatrice) con probabilità uguale al 70%. Se lo scompartimento contiene 5 posti (oltre a quello da me prenotato), qual è la probabilità che io vi incontri meno di tre fumatori?*

Soluzione. Si sta sottointendendo che lo scompartimento venga riempito e che i viaggiatori si comportino, rispetto al fumo, in modo ciascuno indipendente dall'altro. Se S indica il numero dei fumatori nei 5 posti rimanenti, si avrà:

$$\begin{aligned} P(\{S < 3\}) &= P(\{S = 0\}) + P(\{S = 1\}) + P(\{S = 2\}) \\ &= \binom{5}{0} \left(\frac{7}{10}\right)^0 \left(\frac{3}{10}\right)^5 + \binom{5}{1} \left(\frac{7}{10}\right)^1 \left(\frac{3}{10}\right)^4 + \binom{5}{2} \left(\frac{7}{10}\right)^2 \left(\frac{3}{10}\right)^3. \end{aligned}$$

Esempio 6.3. *Un testo, contenente 20 errori di stampa, viene sottoposto a due diversi correttori di bozze. Ciascun errore contenuto nel testo viene individuato da ciascun correttore con probabilità $p = 0.6$ ed indipendentemente da quello che accade per gli altri errori. Inoltre i due correttori lavorano indipendentemente uno dall'altro.*

Trovare la probabilità che il numero degli errori individuati da almeno uno dei due correttori sia superiore a 15.

Soluzione. Ciascun errore viene individuato (da almeno uno dei due correttori) con probabilità¹⁰

$$\theta = 2p - p^2 = \frac{84}{100}$$

(non ci interessa se viene individuato da uno dei due correttori o dall'altro o, eventualmente, da entrambi; a noi interessa che almeno uno dei due individui l'errore).

Si tratta quindi di 20 prove bernoulliane, in ognuna delle quali vi è una probabilità di successo uguale a θ . Indicando dunque con S il numero complessivo degli errori individuati, si avrà

$$P(\{S > 15\}) = \sum_{k=16}^{20} \binom{20}{k} \left(\frac{84}{100}\right)^k \left(\frac{16}{100}\right)^{20-k}.$$

6.2 Estrazioni casuali da urne con reiserimento

Illustriamo ora una tipica situazione in cui si incontra uno schema di prove bernoulliane.

Pensiamo ad una popolazione composta di M oggetti, diciamo c_1, \dots, c_M , di cui m_1 di tipo A e i rimanenti $m_2 = M - m_1$ di un diverso tipo B . **Supponiamo, ad esempio, di aver numerato c_1, \dots, c_M in modo tale che**

$$c_1, \dots, c_{m_1} \text{ sono di tipo } A \text{ e } c_{m_1+1}, \dots, c_M \text{ di tipo } B. \quad (40)$$

¹⁰Il ragionamento per arrivare al calcolo di $\theta = 2p - p^2$ è simile a quello usato negli Esempi 5.2 e 5.3

Eseguiamo ora n estrazioni *casuali con reinserimento* da tale popolazione; con tale termine si vuole esprimere il fatto che si eseguono n estrazioni successive, reinserendo ogni volta nella popolazione l'oggetto estratto ed estraendo in modo tale che **ciascun** oggetto abbia, ogni volta, la **stessa probabilità** $\frac{1}{M}$ di essere estratto, sia esso di tipo A o di tipo B .

Lo spazio campione in tale esperimento (consistente nell'eseguire le n estrazioni) può essere identificato come l'insieme

$$\Omega \equiv \{1, \dots, M\}^n$$

costituito delle n -uple ordinate di elementi in $\{1, \dots, M\}$; esso ha dunque cardinalità M^n . In tale spazio campione, consideriamo, per $i = 1, \dots, n$, ora gli eventi del tipo:

$$E_i \equiv \{\text{l'oggetto estratto nella } i\text{-esima estrazione è di tipo } A\}.$$

Proposizione 1. Gli eventi E_1, \dots, E_n (definiti qui sopra) costituiscono delle prove bernoulliane con $P(E_i) = \frac{m_1}{M}$.

Dimostrazione. La condizione che le estrazioni sono casuali con reinserimento corrisponde all'assegnazione della stessa probabilità $\frac{1}{M^n}$ a ciascuno degli eventi elementari $(j_1, \dots, j_n) \in \Omega$. **Inoltre, in virtù della posizione (40) possiamo scrivere gli eventi E_i come segue**

$$\begin{aligned} E_i &\equiv \{(j_1, \dots, j_n) \in \Omega : 1 \leq j_i \leq m_1\} \\ &= \overbrace{\{1, \dots, M\} \times \dots \times \{1, \dots, M\}}^{i-1 \text{ volte}} \times \{1, 2, \dots, m_1\} \times \overbrace{\{1, \dots, M\} \times \dots \times \{1, \dots, M\}}^{n-i \text{ volte}}, \end{aligned}$$

ed in modo analogo

$$\begin{aligned} \bar{E}_i &\equiv \{(j_1, \dots, j_n) \in \Omega : m_1 + 1 \leq j_i \leq M\} \\ &= \overbrace{\{1, \dots, M\} \times \dots \times \{1, \dots, M\}}^{i-1 \text{ volte}} \times \{m_1 + 1, m_1 + 2, \dots, m_1 + m_2\} \times \overbrace{\{1, \dots, M\} \times \dots \times \{1, \dots, M\}}^{n-i \text{ volte}}. \end{aligned}$$

Dunque, considerando per $1 \leq i \leq n$, la quantità X_i come definita nella (36), si ha

$$\begin{aligned} P(\{X_i = 1\}) &= P(E_i) = \frac{|\{(j_1, \dots, j_n) \in \Omega : 1 \leq j_i \leq m_1\}|}{M^n} \\ &= \frac{m_1 \cdot M^{n-1}}{M^n} = \frac{m_1}{M}, \end{aligned}$$

e, analogamente

$$P(\{X_1 = x_1, \dots, X_n = x_n\}) = \frac{m_1^{\sum_{i=1}^n x_i} m_2^{n - \sum_{i=1}^n x_i}}{M^n} = \left(\frac{m_1}{M}\right)^{\sum_{i=1}^n x_i} \left(\frac{m_2}{M}\right)^{n - \sum_{i=1}^n x_i}.$$

Mettendo insieme le due relazioni dimostrate si ottiene che, per tutti gli eventi

$$A_i \in \mathcal{P}_i = \{E_i, \bar{E}_i\} = \left\{ \{X_i = 1\}, \{X_i = 0\} \right\}$$

si ha

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n),$$

che corrisponde alla relazione (32), che come osservato in Osservazione 3 della Lez. 5, è equivalente alla relazione (30) **che definisce l'indipendenza completa.** \square

Una precisazione sulla dimostrazione della precedente Proposizione 6.2

Al lettore più attento la dimostrazione potrebbe sembrare non soddisfacente, in quanto abbiamo dimostrato la condizione (32), senza aver dato la dimostrazione che essa è equivalente alla (30), che abbiamo preso come definizione dell'indipendenza completa. Tuttavia va osservato che, in modo assolutamente analogo, si potrebbe dimostrare direttamente la (30), ossia verificare che per ogni $2 \leq m \leq n$,

$$\begin{aligned} P(\{X_{j_1} = x_{j_1}, \dots, X_{j_m} = x_{j_m}\}) &= \frac{m_1^{\sum_{i=1}^m x_{j_i}} m_2^{m - \sum_{i=1}^m x_{j_i}} M^{n-m}}{M^n} \\ &= \left(\frac{m_1}{M}\right)^{\sum_{i=1}^m x_{j_i}} \left(\frac{m_2}{M}\right)^{m - \sum_{i=1}^m x_{j_i}} \\ &= \prod_{i=1}^m P(\{X_{j_i} = x_{j_i}\}). \end{aligned}$$

Consideriamo ora

$$S_n = \sum_{i=1}^n X_i.$$

S_n rappresenta dunque il numero di elementi di tipo A in un campionamento casuale *con reinserimento* di n oggetti da una popolazione complessivamente costituita da M elementi, di cui m_1 di tipo A e $m_2 = M - m_1$ di tipo B. Ricordando la formula (39) ed in virtù della **Proposizione 1**, possiamo concludere scrivendo

$$P(\{S_n = k\}) = \binom{n}{k} \left(\frac{m_1}{M}\right)^k \left(\frac{m_2}{M}\right)^{n-k}, \quad \text{per } k = 0, 1, \dots, n. \quad (41)$$

6.3 Estrazioni casuali da urne senza reinserimento e Probabilità ipergeometriche

Consideriamo ora la stessa situazione come descritta nel paragrafo precedente, ma con la differenza che le n estrazioni siano eseguite *senza reinserimento*. Poniamo di nuovo

$$E_i \equiv \{\text{oggetto di tipo A alla } i\text{-esima estrazione}\}, \quad i = 1, 2, \dots, n$$

$$X_i = \begin{cases} 1 & \text{se si verifica } E_i, \\ 0 & \text{se si verifica } \bar{E}_i, \end{cases} \quad i = 1, 2, \dots, n.$$

$$S_n = \sum_{i=1}^n X_i.$$

Nel caso di estrazioni senza reinserimento potremo considerare come spazio campione lo spazio costituito dalle n -uple di elementi di $\{1, \dots, M\}$, senza ripetizione:

$$\Omega \equiv \{(j_1, \dots, j_n) : 1 \leq j_i \leq M, j_1 \neq \dots \neq j_n\} = \{(j_1, \dots, j_n) : 1 \leq j_i \leq M, j_1, \dots, j_n \text{ tutti distinti}\}.$$

Dunque

$$|\Omega| = M(M-1)\dots(M-(n-1)) = M(M-1)\dots(M-n+1) = \frac{M!}{(M-n)!}.$$

Lo schema di estrazioni casuali senza reinserimento si traduce nella condizione che tutti gli elementi di Ω (eventi elementari) hanno uguale probabilità

$$\frac{1}{M(M-1)\dots(M-n+1)} = \frac{(M-n)!}{M!}.$$

Un qualunque evento composto, della forma $\{X_1 = x_1, \dots, X_n = x_n\}$, ha una cardinalità che dipende soltanto dal numero $k = \sum_{i=1}^n x_i$ ed, esattamente, è data da

$$\begin{aligned} & \overbrace{m_1 \cdot (m_1 - 1) \cdot \dots \cdot (m_1 - (k - 1))}^{k \text{ fattori}} \cdot \overbrace{m_2 \cdot (m_2 - 1) \cdot \dots \cdot (m_2 - (n - k - 1))}^{n-k \text{ fattori}} \\ &= m_1 \cdot (m_1 - 1) \cdot \dots \cdot (m_1 - k + 1) \cdot m_2 \cdot (m_2 - 1) \cdot \dots \cdot (m_2 - (n - k) + 1) \\ &= \frac{m_1!}{(m_1 - k)!} \cdot \frac{m_2!}{(m_2 - (n - k))!} \end{aligned}$$

e dunque

$$\begin{aligned} & P(\{X_1 = x_1, \dots, X_n = x_n\}) \\ &= \frac{m_1 \cdot (m_1 - 1) \cdot \dots \cdot (m_1 - k + 1) \cdot m_2 \cdot (m_2 - 1) \cdot \dots \cdot (m_2 - (n - k) + 1)}{M(M-1)\dots(M-n+1)} \\ &= \frac{m_1!}{(m_1 - k)!} \frac{m_2!}{(m_2 - (n - k))!} \frac{(M - n)!}{M!} \end{aligned}$$

da cui

$$\begin{aligned} P(\{S_n = k\}) &= \sum_{\mathbf{x} \in \{0,1\}^n; \sum_i x_i = k} P(\{X_1 = x_1, \dots, X_n = x_n\}) \\ &= \binom{n}{k} \frac{m_1 \cdot (m_1 - 1) \cdot \dots \cdot (m_1 - k + 1) \cdot m_2 \cdot (m_2 - 1) \cdot \dots \cdot (m_2 - (n - k) + 1)}{M(M-1)\dots(M-n+1)} \\ &= \binom{n}{k} \frac{m_1!}{(m_1 - k)!} \frac{m_2!}{(m_2 - (n - k))!} \frac{(M - n)!}{M!}; \end{aligned}$$

riscrivendo in forma più compatta tale ultima frazione, attraverso la notazione dei coefficienti binomiali, possiamo concludere: **per $k \in \{0, 1, \dots, n\}$**

$$P(\{S_n = k\}) = \frac{\binom{m_1}{k} \binom{m_2}{n-k}}{\binom{M}{n}} \quad \text{per } \max(0, n + m_1 - M) \leq k \leq \min(n, m_1), \quad (42)$$

$$P(\{S_n = k\}) = 0 \quad \text{altrimenti}$$

Probabilità del tipo (42) prendono il nome di *probabilità ipergeometriche*. In effetti già conosciamo questo risultato dall'Esempio 3.8 (si veda anche l'Esercizio proposto 3.2).

Dal momento che, per fissati M, m_1, n , la famiglia degli eventi $\{S_n = k\}$, per $\max(0, n + m_1 - M) \leq k \leq \min(n, m_1)$, costituisce una partizione dello spazio campione, deve ovviamente risultare

$$\sum_{k=0 \vee (n+m_1-M)}^{n \wedge m_1} P(\{S_n = k\}) = 1$$

cioè

$$\sum_{k=0 \vee (n+m_1-M)}^{n \wedge m_1} \frac{\binom{m_1}{k} \binom{m_2}{n-k}}{\binom{M}{n}} = 1.$$

In effetti quest'ultima identità coincide con la (20).

Osservazione 1. Si può pervenire al risultato (42) anche in un modo alternativo, applicando direttamente la formula delle probabilità composte agli eventi del tipo $\{X_1 = x_1, \dots, X_n = x_n\} = \{X_1 = x_1\} \cap \dots \cap \{X_n = x_n\}$:

$$P(\{X_1 = x_1, \dots, X_n = x_n\}) \\ = P(\{X_1 = x_1\})P(\{X_2 = x_2\}|\{X_1 = x_1\}) \cdots P(\{X_n = x_n\}|\{X_1 = x_1, \dots, X_{n-1} = x_{n-1}\}).$$

Possiamo infatti *imporre* direttamente, a partire dalla descrizione del problema, che le probabilità condizionate del tipo

$$P(\{X_{r+1} = 1\}|\{X_1 = x_1, \dots, X_r = x_r\}),$$

per $1 \leq r \leq n-1$ siano uguali a

$$\frac{m_1 - \sum_{i=1}^r x_i}{M - r + 1},$$

cioè uguali al rapporto fra il numero $m_1 - \sum_{i=1}^r x_i$ degli elementi di tipo A rimasti nella popolazione dopo le prime $(r-1)$ estrazioni ed il numero complessivo $M - (r-1) = M - r + 1$ degli elementi rimasti nella popolazione.

Notiamo dunque che tali probabilità condizionate non vengono **calcolate** utilizzando la loro definizione (data in Definizione 4.1 della Lezione 4), ma vengono direttamente **assegnate** a partire dalle condizioni del problema.

Ciò costituisce un esempio di quanto era stato già accennato più in generale in merito alla nozione di probabilità condizionata: si giunge cioè a calcolare delle probabilità di eventi composti non tramite un calcolo combinatorio, bensì assegnando delle probabilità condizionate e delle condizioni di simmetria fra eventi; tali considerazioni sono analoghe a quelle già svolte nell'Esempio 6.2.

Osservazione 2. Il problema qui affrontato riguarda il calcolo della probabilità di avere k elementi di tipo A in un campionamento casuale di n oggetti da una popolazione complessivamente costituita da M elementi, di cui m_1 di tipo A e $m_2 = (M - m_1)$ di tipo B. Tale calcolo si applica a problemi di diverso tipo (quali estrazioni da urne, sondaggio elettorale, analisi statistica di una popolazione, etc...) tutti, fra di loro, sostanzialmente isomorfi. È interessante confrontare fra loro le due formule (41) e (42). Entrambe risolvono il problema detto; la prima riguarda però estrazioni con reinserimento, mentre la seconda riguarda estrazioni senza reinserimento. Intuitivamente ci possiamo aspettare che le due formule tendano a coincidere nel caso in cui M , m_1 ed m_2 siano numeri molto grandi rispetto a n ; infatti in tal caso ci si può aspettare che non vi sia grande differenza fra estrazioni con o senza reinserimento. Ciò può essere formalizzato come segue: è possibile dimostrare che, per fissati valori di n , k (con $0 \leq k \leq n$) e θ (con $0 < \theta < 1$), **mandando M ed m_1 all'infinito in modo che $\frac{m_1}{M}$ tende a θ , allora risulta**

$$\lim_{\substack{M \rightarrow \infty, m_1 \rightarrow \infty \\ \frac{m_1}{M} \rightarrow \theta}} \frac{\binom{m_1}{k} \binom{M-m_1}{n-k}}{\binom{M}{n}} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Ad esempio, indicando con $[x]$ la *parte intera*¹¹ di un numero reale x , si può prendere $m_1 = [\theta M]$, e in effetti risulta

$$\lim_{M \rightarrow \infty} \frac{\binom{[\theta M]}{k} \binom{M-[\theta M]}{n-k}}{\binom{M}{n}} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

¹¹Ricordiamo che la parte intera $[x]$ di x è quel numero intero k tale che $k \leq x < k+1$

Convergenza delle probabilità ipergeometriche alle probabilità binomiali

Dimostriamo che

$$\lim_{\substack{M \rightarrow \infty, m_1 \rightarrow \infty \\ \frac{m_1}{M} \rightarrow \theta}} \frac{\binom{m_1}{k} \binom{M-m_1}{n-k}}{\binom{M}{n}} = \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Ricordiamo che, per arrivare alla (42), abbiamo mostrato che

$$\begin{aligned} & \binom{n}{k} \frac{m_1!}{(m_1-k)!} \frac{m_2!}{(m_2-(n-k))!} \frac{(M-n)!}{M!} \\ &= \frac{n!}{k!(n-k)!} \frac{m_1!}{(m_1-k)!} \frac{m_2!}{(m_2-(n-k))!} \frac{(M-n)!}{M!} \\ &= \frac{m_1!}{k!(m_1-k)!} \frac{m_2!}{(n-k)!(m_2-(n-k))!} \frac{(M-n)!n!}{M!} \\ &= \frac{\binom{m_1}{k} \binom{m_2}{n-k}}{\binom{M}{n}}. \end{aligned}$$

Utilizzando la precedente relazione con $m_2 = M - m_1$, basta quindi verificare che

$$\lim_{\substack{M \rightarrow \infty, m_1 \rightarrow \infty \\ \frac{m_1}{M} \rightarrow \theta}} \frac{m_1!}{(m_1-k)!} \frac{(M-m_1)!}{(M-m_1-n-k)!} \frac{(M-n)!}{M!} = \theta^k (1-\theta)^{n-k}.$$

Infatti si ha

$$\begin{aligned} & \frac{m_1!}{(m_1-k)!} \frac{(M-m_1)!}{(M-m_1-n-k)!} \frac{(M-n)!}{M!} \\ &= \frac{m_1}{M} \left(\frac{m_1}{M-1} - \frac{1}{M-1} \right) \cdots \left(\frac{m_1}{M-(k-1)} - \frac{k-1}{M-(k-1)} \right) \\ & \cdot \left(\frac{M}{M-k} - \frac{m_1}{M-k} \right) \left(\frac{M-1}{M-(k+1)} - \frac{m_1}{M-(k+1)} \right) \cdots \left(\frac{M-(n-k-1)}{M-(n-1)} - \frac{m_1}{M-(n-1)} \right) \end{aligned}$$

e la tesi si ottiene considerando che, per ogni i, j ,

$$\begin{aligned} \lim_{M \rightarrow \infty} \frac{j}{M-j} &= 0, \quad \lim_{M \rightarrow \infty} \frac{M-i}{M-j} = 1, \\ \lim_{\substack{M \rightarrow \infty, m_1 \rightarrow \infty \\ \frac{m_1}{M} \rightarrow \theta}} \frac{m_1}{M-j} &= \lim_{\substack{M \rightarrow \infty, m_1 \rightarrow \infty \\ \frac{m_1}{M} \rightarrow \theta}} \frac{m_1}{M} \frac{M}{M-j} = \theta. \end{aligned}$$

Nel caso in cui $m_1 = \lfloor \theta M \rfloor$ basta mandare M all'infinito in quanto chiaramente $m_1 = \lfloor \theta M \rfloor$ tende ad infinito se M tende all'infinito, e inoltre

$$\lim_{M \rightarrow \infty} \frac{\lfloor \theta M \rfloor}{M} = \theta.$$

L'ultimo limite si ottiene tenendo conto che

$$\lfloor \theta M \rfloor \leq \theta M < \lfloor \theta M \rfloor + 1$$

e che quindi

$$0 \leq \theta M - \lfloor \theta M \rfloor < 1, \quad \text{ovvero} \quad 0 \leq \frac{\theta M}{M} - \frac{\lfloor \theta M \rfloor}{M} < \frac{1}{M}.$$

6.4 Esercizi di verifica

Esercizio 6.1. Un candidato ad un'elezione ha bisogno di almeno 50 voti per essere eletto. Prepara allora una lettera per informare i potenziali elettori circa la sua candidatura, il suo programma elettorale, etc....

Egli valuta che ogni persona che riceve la lettera si recherà effettivamente a votare per lui con una probabilità del 40%, indipendentemente dal comportamento degli altri (e si sottintende che egli certamente non ottiene voti da coloro cui non ha inviato la lettera).

- Qual è la probabilità che egli riceva esattamente 51 voti se invia la lettera a 200 persone?
- Qual è la probabilità di essere eletto se invia la lettera a 100 persone?
- Qual è il numero minimo di persone cui deve inviare copia della lettera affinché la probabilità di essere eletto sia superiore all'80%?

Esercizio 6.2. Si prendono a caso $n = 5$ viti da una scatola contenente complessivamente $M = 26$ viti, di cui alcune nuove ed altre usurate.

- Supponendo che la scatola contiene $m_1 = 20$ viti nuove e $m_2 = 6$ viti usurate, calcolare la probabilità che almeno quattro delle cinque viti scelte siano nuove.
- Si supponga ora di non conoscere inizialmente il numero M_1 delle viti nuove nella scatola e si ponga

$$P(\{M_1 = h\}) = \binom{26}{h} \left(\frac{4}{5}\right)^h \left(\frac{1}{5}\right)^{26-h}, \quad h = 0, 1, 2, \dots, 26$$

Dopo aver verificato che tutte le cinque viti scelte sono nuove, come va calcolata la probabilità dell'ipotesi $\{M_1 = 26\}$?

Esercizio 6.3. In un lotto di 15 lampadine, ve ne sono 5 guaste. Se ne estraggono a caso 3. Calcolare la probabilità dei seguenti eventi:

- nessuna lampadina difettosa fra le tre estratte
- esattamente una lampadina difettosa fra le tre estratte
- almeno una lampadina difettosa fra le tre estratte.

Si considerino separatamente i due diversi casi in cui

- si estraggano le tre lampadine contemporaneamente
- le estrazioni sono con reimbussolamento

Esercizio 6.4. Si hanno m esemplari di un certo tipo di telecomando (TC) per televisore; ciascun TC ha bisogno di due batterie per il suo funzionamento. Si hanno a disposizione $2m$ batterie, di cui però h cariche e $(2m - h)$ scariche. Da tale gruppo di batterie vengono costituite in modo casuale m coppie, che vengono inserite negli m TC.

Calcolare la probabilità che un fissato TC abbia entrambe le batterie cariche.

Esercizio 6.5. Riottenere la formula (42) seguendo le indicazioni contenute nella precedente **Osservazione 1 di pagina 54.**

7 Variabili aleatorie e distribuzioni di probabilità

Varie questioni incontrate nelle precedenti lezioni trovano una adeguata formalizzazione tramite l'introduzione della nozione di **variabile aleatoria**. Nei precedenti esempi, infatti, ci siamo ripetutamente imbattuti in oggetti quali: somma dei punteggi nel lancio di due dadi, numero di votanti per uno schieramento in un sondaggio elettorale, numero di successi su n prove bernoulliane, massimo fra i cinque numeri risultanti da un'estrazione del lotto, etc....

A parte la diversa natura dei problemi considerati, notiamo che si è trattato in ogni caso di situazioni in cui possiamo elencare i valori che possono essere assunti da una certa grandezza X , ma sussiste una situazione di incertezza relativamente a quale sarà lo specifico valore che X effettivamente assume: il valore che essa assume sarà connesso (in qualche preciso modo) al risultato elementare di un qualche esperimento aleatorio; in base alla misura di probabilità assegnata sullo spazio campione in tale esperimento, potremo valutare la probabilità che si presentino i vari possibili valori per la grandezza.

Tali considerazioni motivano le definizioni seguenti.

Definizione 7.1 (provvisoria). *Sia Ω un insieme finito. Una applicazione $X : \Omega \rightarrow X(\Omega) \subseteq \mathbb{R}$ viene detta **variabile aleatoria** (definita su $(\Omega, \mathcal{P}(\Omega))$).*

Osservazione 1. Essendo Ω un insieme finito, l'immagine di X , ovvero $X(\Omega)$, sarà un insieme del tipo $\{x_1, \dots, x_n\}$, con $x_\ell \in \mathbb{R}$, per $\ell = 1, \dots, n$.

Consideriamo ora gli eventi

$$X^{-1}(\{x_\ell\}) \equiv \{\omega \in \Omega : X(\omega) = x_\ell\}, \quad \ell = 1, 2, \dots, n.$$

Tali eventi vengono anche indicati brevemente con i simboli

$$\{X = x_1\}, \dots, \{X = x_n\}.$$

È immediato verificare che la famiglia degli eventi

$$\{\{X = x_1\}, \dots, \{X = x_n\}\}$$

costituisce una partizione di Ω ; per tale motivo si ha ovviamente che, ponendo¹²

$$p_\ell \equiv P(X = x_\ell), \quad \ell = 1, \dots, n,$$

risulta

$$\sum_{\ell=1}^n p_\ell = 1, \quad p_\ell \geq 0, \quad \ell = 1, \dots, n.$$

¹²Per mettere in evidenza la dipendenza dalla variabile aleatoria X si usa anche la notazione

$$p_X(x_\ell) \equiv P(X = x_\ell), \quad \ell = 1, \dots, n.$$

La funzione $p_X : X(\Omega) \rightarrow \mathbb{R}; x \mapsto p_X(x) \equiv P(X = x)$, così definita viene anche detta **densità discreta** di X . Ovviamente vale

$$\sum_{\ell=1}^n p_X(x_\ell) = 1, \quad p_X(x_\ell) \geq 0, \quad \ell = 1, \dots, n.$$

Possiamo dunque considerare il nuovo spazio di probabilità

$$(X(\Omega), \mathcal{P}(X(\Omega)), P_X)$$

dove, per $A \in \mathcal{P}(X(\Omega))$

$$P_X(A) \equiv P(X^{-1}(A)) \quad \text{o equivalentemente} \quad P_X(A) \equiv \sum_{\ell: x_\ell \in A} p_\ell. \quad (43)$$

Ricordiamo che $X^{-1}(A)$ indica la controimmagine dell'insieme A tramite la funzione X , ovvero

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}.$$

Per comodità di notazione si scrive più brevemente $X^{-1}(A) = \{X \in A\}$.

Definizione 7.2 (provvisoria). *La misura di probabilità $P_X(\cdot)$ su $\mathcal{P}(X(\Omega))$ prende il nome di distribuzione di probabilità della variabile aleatoria X .*

Interpreteremo quindi $P_X(A)$ come $P(\{X \in A\})$. Tale interpretazione e l'equivalenza in (43) sono basate sul fatto che l'evento $X^{-1}(A) = \{X \in A\}$ si può scrivere come

$$\{X \in A\} = \bigcup_{\ell: x_\ell \in A} \{X = x_\ell\},$$

e di conseguenza

$$P(\{X \in A\}) = \sum_{\ell: x_\ell \in A} P(\{X = x_\ell\}) = \sum_{\ell: x_\ell \in A} p_X(x_\ell) = \sum_{\ell: x_\ell \in A} p_\ell$$

Osservazione 2. Ovviamente a qualunque variabile aleatoria definita su uno spazio di probabilità possiamo associare la sua distribuzione di probabilità. Per individuare la distribuzione di probabilità di una variabile aleatoria X basta specificare la sua immagine $X(\Omega)$, cioè l'insieme dei valori assumibili da X , ed i valori $p_\ell = P(X = x_\ell)$ per ogni $x_\ell \in X(\Omega)$. Due diverse variabili aleatorie, definite o meno su uno stesso spazio di probabilità, possono dar luogo ad una stessa distribuzione di probabilità (vedi i due successivi Esempi 7.1 e 7.2).

È opportuno innanzitutto richiamare l'attenzione sui due particolari tipi di variabili aleatorie: le variabili aleatorie **degeneri** e le variabili aleatorie **binarie**.

Definizione 7.3 (variabili aleatorie degeneri). *Diciamo che X , variabile aleatoria definita su Ω , è una variabile aleatoria **degenere** se esiste $\bar{x} \in \mathbb{R}$, tale che $X(\Omega) = \{\bar{x}\}$, cioè se X è costante su Ω .*

La distribuzione di una variabile aleatoria degenere è banale: $X(\Omega) = \{\bar{x}\}$, $\mathcal{P}(X(\Omega)) = \{\emptyset, \{\bar{x}\}\}$, con $P_X(\emptyset) = 0$ e $P_X(\{\bar{x}\}) = 1$.

Definizione 7.4 (variabili aleatorie binarie). *Diciamo che X , variabile aleatoria definita su Ω , è una variabile aleatoria **binaria** se $X(\Omega) = \{0, 1\}$.*

Definizione 7.5 (indicatore di un evento). *Sia $E \in \mathcal{P}(\Omega)$ un evento e sia \mathcal{X}_E la **funzione caratteristica**, (o anche **funzione indicatrice**) di E :*

$$\mathcal{X}_E(\omega) = \begin{cases} 1 & \text{per } \omega \in E \\ 0 & \text{per } \omega \notin E \end{cases}$$

\mathcal{X}_E è dunque una variabile aleatoria binaria, che viene indicata con il termine **indicatore** di E .

Ricordiamo infine che invece di \mathcal{X}_E si usa talvolta anche il simbolo $\mathbf{1}_E$.

Osservazione 3. Per qualunque v.a. binaria X esiste $E \subseteq \Omega = \{\omega_1, \dots, \omega_N\}$ tale che

$$X(\omega_i) = \mathcal{X}_E(\omega_i), \quad i = 1, \dots, N.$$

Poniamo infatti

$$E = \{\omega_i \in \Omega : X(\omega_i) = 1\}$$

Si avrà allora

$$\bar{E} = \{\omega_i \in \Omega : X(\omega_i) = 0\}$$

e dunque possiamo scrivere $X(\omega_i) = \mathcal{X}_E(\omega_i)$, $i = 1, \dots, N$.

La distribuzione di una variabile binaria \mathcal{X}_E , con $p = P(E)$, è individuata ovviamente dal fatto che $X(\Omega) = \{0, 1\}$ e da $p_0 = 1 - p$ e $p_1 = p$.

Proposizione 1. Una qualunque variabile aleatoria X si scrive come combinazione lineare di variabili aleatorie binarie.

Dimostrazione.

Sia $X(\Omega) = \{x_1, \dots, x_n\}$ l'insieme dei valori assumibili da X ; consideriamo gli eventi

$$H_\ell \equiv \{X = x_\ell\}, \quad \ell = 1, \dots, n$$

e le variabili aleatorie binarie X_1, \dots, X_n definite come indicatori di tali eventi, ovvero $X_\ell = \mathcal{X}_{H_\ell}$. È ovvio allora che possiamo scrivere

$$X(\omega_i) = \sum_{\ell=1}^n x_\ell \cdot X_\ell(\omega_i), \quad i = 1, \dots, N.$$

Infatti basta mostrare che la funzione $Y(\omega)$, definita su Ω da

$$Y(\omega) \equiv \sum_{\ell=1}^n x_\ell \cdot X_\ell(\omega)$$

coincide con $X(\omega)$ per ogni ω . A questo scopo basta osservare che la famiglia di eventi H_ℓ forma una partizione dell'evento certo Ω e di conseguenza basta mostrare che

$$Y(\omega) \equiv \sum_{\ell=1}^n x_\ell \cdot X_\ell(\omega) = X(\omega) \quad \text{per ogni } \omega \in H_\kappa, \quad \text{e per ogni } \kappa = 1, \dots, n.$$

E infatti per ogni $\omega \in H_\kappa$ ovviamente $X(\omega) = x_\kappa$, e anche $Y(\omega) = x_\kappa$, come si vede subito, tenendo conto che $\mathcal{X}_{H_\ell}(\omega) = 0$ se $\ell \neq \kappa$ e che ovviamente $\mathcal{X}_{H_\kappa}(\omega) = 1$:

$$Y(\omega) = x_\kappa \cdot \mathcal{X}_{H_\kappa}(\omega) + \sum_{\ell \neq \kappa} x_\ell \cdot \mathcal{X}_{H_\ell}(\omega) = x_\kappa \cdot 1 + 0 = x_\kappa.$$

Osservazione 4. Spesso si mira a determinare direttamente la distribuzione di probabilità di una variabile aleatoria, sulla base di considerazioni circa l'esperimento consistente nell'osservare il valore della variabile stessa. In tali casi non teniamo conto dello spazio di probabilità Ω su cui la variabile può essere definita, ne' come tale variabile vi possa essere definita, ne' quale sia la misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$.

Vediamo ora qualche esempio di distribuzione di probabilità.

Esempio 7.1. Consideriamo ancora una volta l'esperimento legato al lancio di due dadi, in cui

$$\Omega \equiv \{(h, k) : 1 \leq h \leq 6, 1 \leq k \leq 6\}.$$

$$P(\{(h, k)\}) = \frac{1}{36}, \quad 1 \leq h \leq 6, \quad 1 \leq k \leq 6$$

Su questo spazio possiamo definire diverse variabili aleatorie, ad esempio:

$X_1 : (h, k) \rightarrow h$ ("punteggio del primo dado"), $X_2 : (h, k) \rightarrow k$ ("punteggio del secondo dado"),
 $X : (h, k) \rightarrow h + k$ ("somma dei due punteggi"), $W : (h, k) \rightarrow \frac{h}{k}$, etc...

X_1 e X_2 hanno la stessa distribuzione di probabilità, data da :

$$P(\{X_1 = x\}) = P(\{X_2 = x\}) = \frac{1}{6}, \quad x = 1, 2, \dots, 6;$$

questa è la **distribuzione uniforme** su $\{1, 2, \dots, 6\}$.

La distribuzione di probabilità di $X = X_1 + X_2$ è invece data da

$$P(\{X = x\}) = \frac{x-1}{36}, \quad x = 2, 3, \dots, 7, \quad P(\{X = x\}) = \frac{13-x}{36}, \quad x = 8, 9, \dots, 12.$$

Esempio 7.2. Riprendiamo l'Esempio 4.2, considerando il caso $n = 6$. La variabile aleatoria

$$T \equiv \text{numero dei tentativi fino a trovare la chiave giusta}$$

può prendere i valori $1, 2, \dots, 6$ e risulta, grazie a quanto avevamo visto,

$$P(T = x) = \frac{1}{6}, \quad x = 1, 2, \dots, 6.$$

Confrontando tale risultato con quanto visto prima, troviamo l'esempio di due variabili aleatorie (cioè X_1 e T), definite su spazi diversi, che hanno la stessa distribuzione di probabilità.

Prima di considerare il successivo esempio è utile fare mente locale sulla seguente semplice

Osservazione 4. Siano E_1, \dots, E_n degli eventi in uno spazio $(\Omega, \mathcal{P}(\Omega))$ e indichiamo rispettivamente con X_1, \dots, X_n i loro indicatori. Ovviamente X_1, \dots, X_n sono delle variabili aleatorie definite su $(\Omega, \mathcal{P}(\Omega))$. Su $(\Omega, \mathcal{P}(\Omega))$ possiamo definire anche la variabile aleatoria $S_n \equiv \sum_{h=1}^n X_h$; poniamo cioè

$$S_n(\omega_i) \equiv \sum_{h=1}^n X_h(\omega_i) = \sum_{h=1}^n \mathbf{1}_{E_h}(\omega_i).$$

Ovviamente S_n ha il significato di **numero di successi fra gli eventi** E_1, \dots, E_n .

Prima di passare ad esaminare i due importanti Esempi 7.3 e 7.4, si noti che la variabile aleatoria S_n può assumere $n + 1$ valori, ossia $S_n(\Omega) = \{0, 1, \dots, n\}$, e che la famiglia degli n eventi E_h per $h = 1, \dots, n$ **non coincide** con la partizione $\{S_n = 0, S_n = 1, \dots, S_n = n\}$, che invece è costituita da $n + 1$ eventi. Questa osservazione mostra anche che la rappresentazione di una variabile aleatoria come combinazione lineare di variabili aleatorie binarie non è unica: posto $H_\ell = \{S_n = \ell\}$ per $\ell = 0, 1, \dots, n$ si ha che

$$S_n = \sum_{\ell=0}^n \ell \mathcal{X}_{H_\ell}$$

corrisponde alla rappresentazione usata nella dimostrazione della **Proposizione 1**, che è diversa dalla precedente rappresentazione

$$S_n = \sum_{h=1}^n \mathbf{1}_{E_h}.$$

Esempio 7.3. Consideriamo n prove bernoulliane, cioè n eventi completamente indipendenti, ciascuno di probabilità θ ($0 < \theta < 1$) e consideriamo la variabile aleatoria $S_n \equiv$ numero di successi sulle n prove. I valori possibili per tale variabile sono ovviamente $0, 1, \dots, n$ e, come abbiamo visto nella lezione precedente, si ha,

$$P(\{S_n = k\}) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \quad k = 0, 1, \dots, n.$$

Si dice che S_n segue una **distribuzione binomiale di parametri n e θ** ; ciò si indica con il simbolo $S_n \sim b(n, \theta)$.

Esempio 7.4. Vengono eseguite n estrazioni casuali senza reinserimento da una popolazione che contiene complessivamente M elementi, di cui m_1 elementi di tipo A e m_2 elementi di tipo B. Consideriamo la variabile aleatoria $S_n \equiv$ numero di elementi di tipo A fra gli n elementi estratti. Sappiamo che vale

$$P(\{S_n = k\}) = \frac{\binom{m_1}{k} \binom{m_2}{n-k}}{\binom{M}{n}}, \quad \max(0, n + m_1 - M) \leq k \leq \min(n, m_1).$$

Si dice che S_n segue una **distribuzione ipergeometrica di parametri M, m_1, n** e ciò si indica con il simbolo $S_n \sim Hyp(M, m_1, n)$.

Finora abbiamo quasi esclusivamente considerato variabili aleatorie **a valori interi** (cioè tali che $X(\Omega) \subseteq \mathbb{Z}$); ma questi non sono gli unici casi di possibile interesse; nel caso considerato nel precedente Esempio 7.3, è interessante considerare anche la variabile aleatoria Y_n ("frequenza dei successi") definita dalla relazione

$$Y_n = \frac{S_n}{n}.$$

A questo proposito è interessante più in generale, date n variabili aleatorie X_1, \dots, X_n , studiare il comportamento probabilistico della media aritmetica

$$Y_n = \frac{\sum_{h=1}^n X_h}{n}.$$

A tale tipo di variabile aleatoria, daremo particolare attenzione nel seguito, in particolare nella Lez. 10, dove verrà ottenuto uno specifico risultato, sotto particolari condizioni.

Nelle ultime lezioni ci occuperemo di variabili aleatorie che prendono valori in un intervallo continuo di numeri reali.

Tornando a variabili aleatorie a valori interi notiamo quanto segue.

Osservazione 6. Per una variabile aleatoria X , a valori interi, cioè $X(\Omega) \subseteq \mathbb{Z}$, può essere spesso conveniente calcolare la distribuzione di probabilità tenendo conto della relazione

$$P(X = k) = P(X \leq k) - P(X \leq k - 1), \quad \text{per } k \in \mathbb{Z}.$$

Altre volte può essere conveniente tenere conto invece che

$$\begin{aligned} P(X = k) &= P(X \geq k) - P(X \geq k + 1) \\ &= P(X > k - 1) - P(X > k), \quad \text{per } k \in \mathbb{Z}. \end{aligned}$$

La dimostrazione delle relazioni seguenti è basata sul fatto che

$$\{X \leq k\} = \{X = k\} \cup \{X \leq k - 1\} \quad \text{e} \quad \{X \geq k\} = \{X = k\} \cup \{X \geq k + 1\},$$

da cui

$$P(X \leq k) = P(X = k) + P(X \leq k - 1) \quad \text{e} \quad P(X \geq k) = P(X = k) + P(X \geq k + 1)$$

e infine che $\{X \geq h\} = \{X > h - 1\}$.

Esempio 7.5. Siano X_1 e X_2 i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria Z definita come il massimo dei due punteggi. Individuare i valori che può assumere Z e con quali probabilità.

Soluzione. I valori possibili per Z sono ovviamente $\{1, 2, \dots, 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} \equiv \{\{X_1 = i\}, i = 1, \dots, 6\}$ e $\mathcal{B} \equiv \{\{X_2 = j\}, j = 1, \dots, 6\}$ sono indipendenti¹³, (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti¹⁴), risulta

$$\begin{aligned} P(Z = x) &= P(Z \leq x) - P(Z \leq x - 1) \\ &= P(\{X_1 \leq x\} \cap \{X_2 \leq x\}) - P(\{X_1 \leq x - 1\} \cap \{X_2 \leq x - 1\}) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x) - P(X_1 \leq x - 1) \cdot P(X_2 \leq x - 1) \\ &= \left(\frac{x}{6}\right)^2 - \left(\frac{x-1}{6}\right)^2 = \frac{2x-1}{36}, \quad x = 1, 2, \dots, 6. \end{aligned}$$

Esempio 7.6. Siano X_1 e X_2 i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria W definita come il minimo dei due punteggi. Individuare i valori che può assumere W e con quali probabilità.

Soluzione. I valori possibili per W sono ovviamente $\{1, 2, \dots, 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} \equiv \{\{X_1 = i\}, i = 1, \dots, 6\}$ e $\mathcal{B} \equiv \{\{X_2 = j\}, j = 1, \dots, 6\}$ sono indipendenti (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti), risulta

$$\begin{aligned} P(W = x) &= P(W \geq x) - P(W \geq x + 1) \\ &= P(\{X_1 \geq x\} \cap \{X_2 \geq x\}) - P(\{X_1 \geq x + 1\} \cap \{X_2 \geq x + 1\}) \\ &= P(X_1 \geq x) \cdot P(X_2 \geq x) - P(X_1 \geq x + 1) \cdot P(X_2 \geq x + 1) \\ &= P(X_1 > x - 1) \cdot P(X_2 > x - 1) - P(X_1 > x) \cdot P(X_2 > x) \\ &= (1 - P(X_1 \leq x - 1)) \cdot (1 - P(X_2 \leq x - 1)) - (1 - P(X_1 \leq x)) \cdot (1 - P(X_2 \leq x)) \\ &= \left(1 - \frac{x-1}{6}\right)^2 - \left(1 - \frac{x}{6}\right)^2 = \frac{49 - 14x + x^2 - (36 - 12x + x^2)}{36} \\ &= \frac{13 - 2x}{36}, \quad x = 1, 2, \dots, 6. \end{aligned}$$

Esercizio proposto 7.1. Ripetere gli esempi precedenti nel caso in cui i due dadi sono truccati in modo che $P(X_l = i) = Ki$, per $i = 1, \dots, 6$, ed $l = 1, 2$, e si assuma l'indipendenza delle partizioni \mathcal{A} e \mathcal{B} .

¹³Questa proprietà sarà alla base della successiva definizione di indipendenza stocastica per le v.a. X_1 e X_2

¹⁴Si veda a questo proposito l'Esercizio proposto 5.4.

Osservazione 7. Il concetto di variabile aleatoria così come introdotto in questa lezione (vedi Definizione 7.1), può talvolta apparire un po' forzato o artificiale a chi sia all'inizio dello studio della teoria assiomatica della probabilità. Di fatto, invece, esso si rivela di importanza fondamentale sia nella formalizzazione rigorosa che nella comprensione di numerose questioni specifiche del calcolo delle probabilità.

In particolare esso permette di dare un chiaro significato alle operazioni fra variabili aleatorie, estendendo in modo diretto a questi oggetti (essendo funzioni a valori reali) le operazioni definite nel campo dei numeri reali.

Ad esempio, come abbiamo precedentemente visto, la somma $X = X_1 + X_2$ di due variabili aleatorie X_1, X_2 (definite su uno stesso spazio Ω) altro non è che la funzione definita dalla relazione

$$X(\omega) = X_1(\omega) + X_2(\omega), \text{ per } \omega \in \Omega.$$

Avevamo già avvertito comunque che la Definizione 7.1 di variabile aleatoria, così come è stata formulata, è provvisoria. Come si vedrà, essa va infatti adeguatamente modificata e completata quando si passi a trattare il caso in cui Ω non è un insieme finito.

7.1 Esercizi di verifica

Esercizio 7.1. Indichiamo con X_1, \dots, X_5 i 5 numeri estratti su una ruota del lotto (si estrae senza reinserimento da un'urna contenente i numeri $\{1, 2, \dots, 90\}$) e sia inoltre X il valore più alto fra X_1, \dots, X_5 .

Calcolate $P(X \leq k)$ e $P(X = k)$ per $k = 1, 2, \dots, 90$.

Esercizio 7.2. Supponiamo che X sia una variabile aleatoria a valori nell'insieme $\{0, 1, \dots, n\}$ e che, per una coppia di costanti positive A e ρ , risulti

$$P(X = k) = A \cdot \frac{\rho^k}{k!(n-k)!}, \quad k = 0, 1, \dots, n,$$

Dimostrate che X segue una distribuzione binomiale ed individuatene i parametri.

Esercizio 7.3. Individuate una distribuzione di probabilità (non degenera) per una variabile aleatoria X in modo tale che risulti degenera la distribuzione della variabile aleatoria $Y = X^2$.

Esercizio 7.4. Individuate una distribuzione di probabilità per una variabile aleatoria X (non binaria) in modo tale che $Y = X^2$ risulti una variabile aleatoria binaria.

Esercizio 7.5. Sia X una variabile aleatoria con distribuzione di probabilità binomiale $b(6, \frac{1}{3})$. Trovare qual è il valore più probabile per X .

Esercizio 7.6. Consideriamo una variabile aleatoria con distribuzione di probabilità ipergeometrica

$$X \propto Hyp(6, 3; 3).$$

Qual è il più probabile fra i due eventi $\{X \leq 1\}, \{X > 1\}$?

Esercizio 7.7. In una lotteria sono stati emessi 1000 biglietti e vengono distribuiti 2 primi premi del valore di 1000 Euro, 4 secondi premi del valore di 500 Euro e 20 terzi premi del valore di 100 Euro.

a) Trovate la distribuzione di probabilità della variabile aleatoria X che indica il valore della vincita associata ad un singolo biglietto.

b) Scrivete la distribuzione di probabilità della variabile aleatoria $2X$.

Un tizio ha acquistato 2 biglietti della lotteria ed indichiamo con Z la variabile aleatoria che indica il valore complessivo della sua vincita alla lotteria.

c) Trovate la distribuzione di probabilità di Z .

Esercizio 7.8. Sia S_n una variabile aleatoria binomiale di parametri n e θ .

a) Verificate che, per $k = 0, 1, \dots, n - 1$

$$P(S_n = k + 1) = \frac{n - k}{k + 1} \frac{\theta}{1 - \theta} P(S_n = k).$$

b) Utilizzando la proprietà precedente, verificate che esiste un \bar{k} tale che

$$\begin{aligned} P(S_n = k + 1) &\geq P(S_n = k) & k \leq \bar{k} \\ P(S_n = k + 1) &\leq P(S_n = k) & k > \bar{k} \end{aligned}$$

8 Distribuzioni congiunte di più variabili aleatorie

In questa lezione esaminiamo alcune definizioni relative al caso in cui si considerino contemporaneamente, su uno stesso spazio di probabilità, due variabili aleatorie.

I concetti che verranno introdotti si possono estendere senza difficoltà al caso di un numero di variabili maggiore di due.

Sia dunque $(\Omega, \mathcal{P}(\Omega), P)$ uno spazio di probabilità e X, Y una coppia di variabili aleatorie definite su di esso; si avrà, diciamo, $X : \Omega \rightarrow X(\Omega) \equiv \{x_1, \dots, x_n\}$, $Y : \Omega \rightarrow Y(\Omega) \equiv \{y_1, \dots, y_m\}$ e possiamo considerare, su Ω , la partizione costituita dagli eventi del tipo

$$\{X = x_i\} \cap \{Y = y_j\} \equiv \{\omega \in \Omega : X(\omega) = x_i, Y(\omega) = y_j\}, \quad i = 1, \dots, n; \quad j = 1, \dots, m.$$

Spesso, d'ora in poi, la scrittura $\{X = x_i\} \cap \{Y = y_j\}$ verrà più semplicemente sostituita da $\{X = x_i, Y = y_j\}$, come del resto già fatto nella Lez. precedente; porremo¹⁵

$$p_{ij} \equiv P(\{X = x_i, Y = y_j\}), \quad i = 1, \dots, n; \quad j = 1, \dots, m$$

e ovviamente risulterà

$$p_{ij} \geq 0, \quad i = 1, \dots, n; \quad j = 1, \dots, m; \quad \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1.$$

Possiamo dunque considerare, analogamente a quanto visto sopra per il caso di una singola variabile aleatoria, il nuovo spazio di probabilità, indotto da X, Y , definito come

$$(X(\Omega) \times Y(\Omega), \mathcal{P}(X(\Omega) \times Y(\Omega)), P_{X,Y}),$$

dove, per $E \subseteq X(\Omega) \times Y(\Omega)$, si pone

$$P_{X,Y}(E) \equiv P(\{(X, Y) \in E\})$$

o equivalentemente

$$P_{X,Y}(E) = \sum_{(i,j):(x_i,y_j) \in E} p_{ij};$$

la misura di probabilità $P_{X,Y}$ su $(X(\Omega) \times Y(\Omega), \mathcal{P}(X(\Omega) \times Y(\Omega)))$ prende il nome di **distribuzione di probabilità congiunta di X, Y** .

Analogamente al caso di una variabile aleatoria unidimensionale si usa la notazione abbreviata, ovvero

$$\{(X, Y) \in E\} = \{\omega \in \Omega : (X(\omega), Y(\omega)) \in E\}.$$

¹⁵In analogia con il caso di una sola variabile aleatoria utilizzeremo anche la notazione

$$p_{X,Y}(x_i, y_j) \equiv P(\{X = x_i, Y = y_j\}), \quad i = 1, \dots, n; \quad j = 1, \dots, m$$

per mettere in evidenza le variabili aleatorie X ed Y coinvolte, ed i valori $\{x_1, \dots, x_n\}$ e $\{y_1, \dots, y_m\}$ che possono assumere. La funzione $p_{X,Y}$ assume il nome di **densità discreta congiunta di X e Y** . Ovviamente, essendo $\{X = x_i, Y = y_j\}$, $i = 1, \dots, n; j = 1, \dots, m;$ una partizione, risulterà

$$p_{X,Y}(x_i, y_j) \geq 0, \quad i = 1, \dots, n; \quad j = 1, \dots, m; \quad \sum_{i=1}^n \sum_{j=1}^m p_{X,Y}(x_i, y_j) = 1.$$

Inoltre l'equivalenza tra le due definizioni di $P_{X,Y}$ è dovuta al fatto

$$\{(X, Y) \in E\} = \bigcup_{i,j:(x_i,y_j) \in E} \{(X, Y) = (x_i, y_j)\} = \bigcup_{i,j:(x_i,y_j) \in E} \{X = x_i, Y = y_j\},$$

da cui

$$P(\{(X, Y) \in E\}) = \sum_{i,j:(x_i,y_j) \in E} P(X = x_i, Y = y_j).$$

Siano ora date due variabili aleatorie X, Y i cui insiemi di valori possibili siano $\{x_1, \dots, x_n\}$, $\{y_1, \dots, y_m\}$ rispettivamente e sia la loro distribuzione di probabilità congiunta individuata dalle probabilità

$$p_{ij} \equiv P(\{X = x_i, Y = y_j\}), \quad i = 1, \dots, n; \quad j = 1, \dots, m.$$

La **distribuzione marginale** per la v.a. X è la distribuzione di probabilità concentrata sui valori $\{x_1, \dots, x_n\}$ e che loro attribuisce le probabilità

$$p'_i = \sum_{j=1}^m P(\{X = x_i, Y = y_j\}) = \sum_{j=1}^m p_{ij}, \quad i = 1, \dots, n, \quad (44)$$

come segue immediatamente dal fatto che

$$\{X = x_i\} = \bigcup_{j=1}^m \{X = x_i, Y = y_j\}.$$

Ovviamente p'_i definisce effettivamente una distribuzione di probabilità per una variabile aleatoria, in quanto risulta

$$p'_i \geq 0, \quad i = 1, \dots, n; \quad \sum_{i=1}^n p'_i = 1.$$

Analogamente la **distribuzione marginale** per la v.a. Y è la distribuzione di probabilità concentrata sui valori $\{y_1, \dots, y_m\}$ e che loro attribuisce le probabilità

$$p''_j = \sum_{i=1}^n P(\{X = x_i, Y = y_j\}) = \sum_{i=1}^n p_{ij}, \quad j = 1, \dots, m. \quad (45)$$

Esempio 8.1. Siano X ed Y sono due variabili aleatorie che possono rispettivamente assumere i valori $\{-1, 0, 1\}$ e $\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$, con probabilità congiunte indicate nella seguente tabella¹⁶

$Y \setminus X$	-1	0	1
1/4	0.1	0.2	0
1/2	0	0.12	0.05
3/4	0.05	0.1	0.04
1	0.1	0.04	0.2

¹⁶Si noti che tutti i calcoli che seguono non dipendono dallo spazio di probabilità $(\Omega, \mathcal{P}(\Omega), P)$ su cui le due variabili aleatorie X ed Y sono definite, ma possono essere effettuati utilizzando solamente la distribuzione congiunta.

Applichiamo la formula (44) per trovare la distribuzione di probabilità marginale della variabile X ; otteniamo allora

$$P(X = -1) = 0.1 + 0 + 0.05 + 0.1 = 0.25;$$

analogamente risulta

$$P(X = 0) = 0.46, \quad P(X = 1) = 0.29.$$

Abbiamo cioè ottenuto le probabilità della distribuzione marginale di X calcolando le somme degli elementi nelle diverse colonne della tabella. Analogamente, calcolando le somme degli elementi sulle righe, otteniamo la distribuzione marginale della variabile Y :

$$P\left(Y = \frac{1}{4}\right) = 0.3, \quad P\left(Y = \frac{1}{2}\right) = 0.17, \quad P\left(Y = \frac{3}{4}\right) = 0.19, \quad P(Y = 1) = 0.34.$$

Riportiamo allora tali distribuzioni marginali, inserendole in una riga ed in una colonna aggiunte rispettivamente nei margini in basso e a destra della tabella.

Cioè completiamo la tabella riportandovi anche le somme di riga e le somme di colonna; otteniamo dunque

$Y \setminus X$	-1	0	1	P_Y
1/4	0.1	0.2	0	0.30
1/2	0	0.12	0.05	0.17
3/4	0.05	0.1	0.04	0.19
1	0.1	0.04	0.2	0.34
P_X	0.25	0.46	0.29	1

Ogni volta che abbiamo una coppia di variabili aleatorie (indicate ad esempio con X, Y), risulta naturale descrivere la loro distribuzione di probabilità congiunta attraverso una *tabella a doppia entrata* (cioè, insomma, una matrice), come segue:

$Y \setminus X$	x_1	\dots	x_i	\dots	x_n	P_Y
y_1	$p_{1,1}$	\dots	$p_{i,1}$	\dots	$p_{n,1}$	p''_1
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
y_j	$p_{1,j}$	\dots	$p_{i,j} = P(X = x_i, Y = y_j)$	\dots	$p_{n,j}$	$p''_j = P(Y = y_j)$
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
y_m	$p_{1,m}$	\dots	$p_{i,m}$	\dots	$p_{n,m}$	p''_m
P_X	p'_1	\dots	$p'_i = P(X = x_i)$	\dots	p'_n	1

In questa tabella:

- la prima riga e la prima colonna rispettivamente indicano i valori possibili $\{x_i, i = 1, \dots, n\}$, $\{y_j, j = 1, \dots, m\}$ per le due variabili aleatorie;

- gli elementi nelle righe e colonne interne indicano le probabilità dei corrispondenti eventi, ovvero $p_{ij} = P(\{X = x_i, Y = y_j\})$;

- l'ultima riga e l'ultima colonna, dunque ai *margini* della tabella, vengono riportate rispettivamente le somme di colonna e le somme di riga, cioè le probabilità delle distribuzioni *marginali* (di qui il nome) delle due variabili, ovvero $p'_i = \sum_{j=1}^m p_{ij}$ e $p''_j = \sum_{i=1}^n p_{ij}$, rispettivamente.

In una situazione quale quella qui sopra descritta, con due variabili aleatorie X e Y definite su uno stesso spazio Ω , fissiamo ora $1 \leq j \leq m$ e consideriamo l'evento $\{Y = y_j\}$, che assumiamo avere probabilità strettamente positiva.

Supponiamo ora che sia stato osservato questo evento, mentre non è stato osservato il valore assunto dalla variabile X .

Ci possiamo allora domandare quale sia, data questa informazione, la distribuzione di probabilità che esprime lo stato di informazione parziale circa X .

A questo quesito, risulta naturale rispondere con la seguente:

Definizione 8.1. *La distribuzione di probabilità condizionata della variabile X , dato l'evento $\{Y = y_j\}$ è la distribuzione di probabilità che concentra, sui valori x_i ($1 \leq i \leq n$), le probabilità condizionate¹⁷ date da*

$$p'_{i|j} \equiv P(X = x_i | Y = y_j) = \frac{P(\{X = x_i, Y = y_j\})}{P(Y = y_j)}, \quad 1 \leq i \leq n,$$

dove $P(X = x_i | Y = y_j)$ è un modo tipograficamente rapido¹⁸ per scrivere $P(\{X = x_i\} | \{Y = y_j\})$.

È ovvio che le $p'_{i|j}$ si ricavano dalle probabilità congiunte p_{ij} tramite la formula¹⁹

$$p'_{i|j} = \frac{p_{ij}}{p''_j} = \frac{p_{ij}}{\sum_{i=1}^n p_{ij}}, \quad (46)$$

e che, qualunque sia $j = 1, \dots, m$

$$p'_{i|j} \geq 0, \quad i = 1, \dots, n, \quad \sum_{i=1}^n p'_{i|j} = 1.$$

Analogamente definiremo, per un fissato indice $1 \leq i \leq n$, la **distribuzione condizionata di Y , dato l'evento $\{X = x_i\}$** , come la distribuzione di probabilità che concentra, sui valori y_j ($1 \leq j \leq m$), le probabilità condizionate date da

$$p''_{j|i} \equiv P(Y = y_j | X = x_i) = \frac{p_{ij}}{p'_i} = \frac{p_{ij}}{\sum_{j=1}^m p_{ij}}. \quad (47)$$

Anche in questo caso, qualunque sia $i = 1, \dots, n$

$$p''_{j|i} \geq 0, \quad j = 1, \dots, m, \quad \sum_{j=1}^m p''_{j|i} = 1.$$

¹⁷Useremo anche le seguenti notazioni

$$p_X(x_i | Y = y_j) \equiv P(X = x_i | Y = y_j) \quad \text{oppure} \quad p_{X|Y}(x_i | y_j) \equiv P(X = x_i | Y = y_j)$$

al posto di $p'_{i|j}$ per mettere in evidenza quali sono le variabili aleatorie X ed Y ed i valori x_i ed y_j coinvolti.

¹⁸In modo analogo scriveremo anche $P(X = x_i)$ o $P(X \in I)$ invece di $P(\{X = x_i\})$ o $P(\{X \in I\})$.

¹⁹Le formule (46) e (47) si possono scrivere rispettivamente anche nel seguente modo:

$$p_{X|Y}(x_i | y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{i=1}^n p_{X,Y}(x_i, y_j)},$$

$$p_{Y|X}(y_j | x_i) \equiv P(Y = y_j | X = x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{j=1}^m p_{X,Y}(x_i, y_j)}.$$

Osservazione 1. Consideriamo una coppia di variabili aleatorie X ed Y , per le quali gli insiemi di valori possibili siano $X(\Omega) \equiv \{x_1, \dots, x_n\}$ e $Y(\Omega) \equiv \{y_1, \dots, y_m\}$. La distribuzione di probabilità congiunta è allora individuata dall'insieme delle probabilità congiunte

$$\{p_{ij}; i = 1, \dots, n; j = 1, \dots, m\}.$$

In base a $\{p_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$, attraverso le formule (44), (45), (46), (47), si determinano univocamente

- le probabilità marginali

$$\{p'_i; i = 1, \dots, n\}, \quad \{p''_j; j = 1, \dots, m\}$$

- e le probabilità condizionate

$$\{p'_{i|j}; j = 1, \dots, m, i = 1, \dots, n\}, \quad \{p''_{j|i}; i = 1, \dots, n, j = 1, \dots, m\}.$$

Supponiamo ora di assegnare la coppia $\{p'_i; i = 1, \dots, n\}, \{p''_j; j = 1, \dots, m\}$, (con $p'_i \geq 0$ e $p''_j \geq 0$); si ricordi che vanno rispettati i vincoli dati dalle condizioni di normalizzazione

$$\sum_{i=1}^n p'_i = 1; \quad \sum_{j=1}^m p''_j = 1.$$

Vi sono ovviamente diverse distribuzioni congiunte che ammettono $\{p'_i; i = 1, \dots, n\}, \{p''_j; j = 1, \dots, m\}$ come probabilità marginali; guardiamo infatti il sistema lineare

$$\begin{cases} \sum_{j=1}^m p_{ij} = p'_i, & i = 1, \dots, n \\ \sum_{i=1}^n p_{ij} = p''_j, & j = 1, \dots, m \\ \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1 \end{cases}$$

nelle variabili p_{ij} : si tratta di un sistema di $(n + m + 1)$ equazioni in $n \cdot m$ variabili che risulta indeterminato²⁰, nei casi $n \geq 2, m \geq 2$.

²⁰In realtà, date $\{p'_i; i = 1, \dots, n\}$ e $\{p''_j; j = 1, \dots, m\}$, con le condizioni di normalizzazione

$$\sum_{i=1}^n p'_i = 1; \quad \sum_{j=1}^m p''_j = 1,$$

si tratta del sistema nelle $n \cdot m$ incognite $p_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, m$,

$$\begin{cases} \sum_{j=1}^m p_{ij} = p'_i, & i = 1, \dots, n \\ \sum_{i=1}^n p_{ij} = p''_j, & j = 1, \dots, m \\ p_{ij} \geq 0, & i = 1, \dots, n \quad j = 1, \dots, m. \end{cases}$$

La condizione $\sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1$ è infatti automaticamente soddisfatta, grazie alle condizioni di normalizzazione per $\{p'_i; i = 1, \dots, n\}$ e $\{p''_j; j = 1, \dots, m\}$. I gradi di libertà del sistema sono $(n-1) \cdot (m-1)$. Per convincersene si cominci con il caso $n = m = 2$. Chiaramente in questo caso basta fissare, ad esempio p_{11} , con $0 \leq p_{11} \leq p'_1$ e $p_{11} \leq p''_1$, ovvero $p_{11} \leq \min(p'_1, p''_1)$, per ottenere automaticamente i valori $p_{12} = p'_1 - p_{11}$ e $p_{21} = p''_1 - p_{11}$. Infine $p_{22} = p'_2 - p_{21} = p''_2 - p_{12} = 1 - (p_{11} + p_{12} + p_{21})$. Il caso generale è analogo, ma, ad esempio, si possono fissare i valori di $p_{i,j}$ per $i = 1, \dots, n-1$ e $j = 1, \dots, m-1$, in modo che

$$p_{i,j} \geq 0, \quad i = 1, \dots, n-1, \quad j = 1, \dots, m-1 \quad \sum_{j=1}^{m-1} p_{ij} \leq p'_i, \quad i = 1, \dots, n-1 \quad e \quad \sum_{i=1}^{n-1} p_{ij} \leq p''_j, \quad j = 1, \dots, m-1.$$

I rimanenti $n-1+m-1+1 = n+m-1$ valori di $p_{n,j}$ per $j = 1, \dots, m-1$, $p_{i,m}$, per $i = 1, \dots, n-1$, e $p_{n,m}$ sono automaticamente ricavati dalle equazioni del sistema.

Invece $\{p_{ij}; i = 1, \dots, n; j = 1, \dots, m\}$ risultano univocamente determinate quando si impongano ad esempio sia le probabilità marginali $\{p'_i; i = 1, \dots, n\}$ che le probabilità condizionate $\{p''_{j|i}; i = 1, \dots, n, j = 1, \dots, m\}$; infatti si deve avere, proprio per la definizione di distribuzione condizionata,

$$p_{ij} = p'_i \cdot p''_{j|i}, \quad i = 1, \dots, n, \quad j = 1, \dots, m.$$

Ciò mostra anche che la conoscenza di $\{p'_i; i = 1, \dots, n\}$ e $\{p''_{j|i}; i = 1, \dots, n, j = 1, \dots, m\}$ determina la coppia $\{p''_j; j = 1, \dots, m\}$ $\{p'_{i|j}; i = 1, \dots, n, j = 1, \dots, m\}$.

Notiamo anche che la relazione che lega fra loro tali probabilità non è nient'altro che la Formula di Bayes

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i)P(Y = y_j | X = x_i)}{P(Y = y_j)},$$

che, riadattata al simbolismo qui introdotto, può essere riscritta nella forma:

$$p'_{i|j} = \frac{p'_i \cdot p''_{j|i}}{p''_j},$$

oppure, equivalentemente,

$$p''_{j|i} = \frac{p''_j \cdot p'_{i|j}}{p'_i}.$$

Esercizio proposto 8.1. Consideriamo i due punteggi X_1 e X_2 derivanti dal lancio di due dadi e definiamo le variabili aleatorie

$$X = X_1 + X_2, \quad Y = \max(X_1, X_2).$$

Per tali variabili sono già state calcolate (nell'Esempio 7.1 e nell'Esempio 7.5 della precedente lezione) le distribuzioni marginali. Costruite ora la tabella delle probabilità congiunte e calcolate la distribuzione condizionata di Y , dato l'evento $\{X = 9\}$.

Esercizio proposto 8.2. Consideriamo i due punteggi X_1, X_2, X_3 e X_4 derivanti dal lancio di quattro dadi e definiamo le variabili aleatorie

$$X = X_1 + X_2, \quad Y = \max(X_1, X_2).$$

$$X' = X_3 + X_4, \quad Y' = Y = \max(X_1, X_2).$$

Dopo aver osservato che le variabili aleatorie X ed X' hanno la stessa distribuzione marginale, e che lo stesso vale ovviamente per Y ed Y' , costruite tabella delle probabilità congiunte di X' e Y' e calcolate la distribuzione condizionata di X , dato l'evento $\{Y = 5\}$, e quella di X' dato l'evento $\{Y' = 5\}$.

8.1 Indipendenza stocastica fra variabili aleatorie.

Vogliamo ora definire il concetto di indipendenza stocastica fra due variabili aleatorie X, Y .

Ricordando quanto discusso nel caso di due eventi potremo dire, dal punto di vista euristico, che due variabili aleatorie sono indipendenti se, qualunque informazione raccolta su una delle due variabili, ad esempio X , non porta a modificare lo stato di informazione su Y . Arriveremo subito, in effetti, a formulare una definizione rigorosa proprio partendo da tali considerazioni.

Cominciamo con il seguente, semplice, ma **fondamentale**

Esercizio proposto 8.3. Siano X ed Y due variabili aleatorie. Verificare²¹ che, le seguenti condizioni sono fra di loro equivalenti:

(i) le probabilità condizionate $p''_{j|i}$ non dipendono dall'indice i ;

(ii) $\forall 1 \leq i \leq n, 1 \leq j \leq m$ risulta

$$p''_{j|i} = p''_j;$$

(iii) $\forall 1 \leq i \leq n, 1 \leq j \leq m$ risulta

$$p_{ij} = p'_i \cdot p''_j.$$

Potremo dire allora che X ed Y sono stocasticamente indipendenti qualora si verifichi una (e quindi tutte) delle condizioni (i), (ii), o (iii) del precedente esercizio.

A partire, in particolare, da (iii) e, ricordando la Definizione 4 della Lez. 5 di partizioni indipendenti, potremo allora giungere alla seguente

Definizione 8.2. Le variabili aleatorie X ed Y si dicono **stocasticamente indipendenti** se sono fra loro indipendenti le due partizioni

$$\mathcal{A} \equiv \{\{X = x_1\}, \dots, \{X = x_n\}\} \quad e \quad \mathcal{B} \equiv \{\{Y = y_1\}, \dots, \{Y = y_m\}\}.$$

In altre parole le variabili aleatorie X ed Y sono stocasticamente indipendenti se e solo se, vale²²

$$p_{ij} = p'_i \cdot p''_j, \quad \forall 1 \leq i \leq n, \quad 1 \leq j \leq m,$$

ovvero

$$P(\{X = x_i\} \cap \{Y = y_j\}) = P(\{X = x_i\}) \cdot P(\{Y = y_j\}), \quad \forall 1 \leq i \leq n, \quad 1 \leq j \leq m$$

Grazie alla **Proposizione 1** della Lezione 5 possiamo affermare che

Proposizione 1 Se X ed Y sono indipendenti secondo la precedente Definizione 8.2, allora si ha

$$P(X \in I, Y \in J) = P(X \in I)P(Y \in J), \quad \text{qualunque siano } I \subseteq X(\Omega) \text{ e } J \subseteq Y(\Omega) \quad (48)$$

²¹Le implicazioni (iii) \Rightarrow (ii), (ii) \Rightarrow (i) sono ovvie. Basta a questo punto dimostrare l'implicazione (i) \Rightarrow (iii): si osservi che, se vale (i) e se si pone $q_j := p''_{j|1} \equiv p''_{j|i}$, allora

$$p_{ij} = p'_i \cdot p''_{j|i} = p'_i \cdot q_j.$$

Da ciò, sommando sugli indici $i = 1, \dots, n$ si ottiene che

$$p''_j = \sum_{i=1}^n p_{ij} = \sum_{i=1}^n p'_i \cdot q_j = q_j \cdot \sum_{i=1}^n p'_i = q_j \cdot 1 = q_j.$$

Confrontando tra loro le ultime due relazioni si ottiene immediatamente la (iii).

²²Si noti che si tratta della condizione (iii) dell'Esercizio proposto 8.3.

o anche, posto $\mathcal{G}(X) = \mathcal{G}(\mathcal{A})$, l'algebra generata dalla partizione $\mathcal{A} \equiv \{\{X = x_1\}, \dots, \{X = x_n\}\}$ e $\mathcal{G}(Y) = \mathcal{G}(\mathcal{B})$, l'algebra generata dalla partizione $\mathcal{B} \equiv \{\{Y = y_1\}, \dots, \{Y = y_m\}\}$, si ha

$$P(A \cap B) = P(A)P(B), \quad \text{qualunque siano } A \in \mathcal{G}(X) \text{ e } B \in \mathcal{G}(Y).$$

Inoltre vale anche il viceversa, ovvero se vale (48), allora X ed Y sono indipendenti secondo la precedente Definizione 8.2.

Dimostrazione. In realtà, come già detto, basta solo applicare la **Proposizione 1** della Lezione 5. Tuttavia ci sembra utile riportare la dimostrazione diretta: gli eventi $\{X \in I\}$ ed $\{Y \in J\}$ si possono scrivere rispettivamente come

$$\{X \in I\} = \bigcup_{i: x_i \in I} \{X = x_i\}, \quad \{Y \in J\} = \bigcup_{j: y_j \in J} \{Y = y_j\},$$

e quindi

$$\{X \in I\} \cap \{Y \in J\} = \left(\bigcup_{i: x_i \in I} \{X = x_i\} \right) \cap \left(\bigcup_{j: y_j \in J} \{Y = y_j\} \right) = \bigcup_{i: x_i \in I} \bigcup_{j: y_j \in J} \{X = x_i\} \cap \{Y = y_j\}.$$

Di conseguenza, se X ed Y sono indipendenti secondo la precedente Definizione 8.2, allora²³

$$\begin{aligned} P(\{X \in I\} \cap \{Y \in J\}) &= \sum_{i: x_i \in I} \sum_{j: y_j \in J} P(\{X = x_i\} \cap \{Y = y_j\}) \\ &= \sum_{i: x_i \in I} \sum_{j: y_j \in J} P(\{X = x_i\}) \cdot P(\{Y = y_j\}) \\ &= \sum_{i: x_i \in I} P(\{X = x_i\}) \cdot \sum_{j: y_j \in J} P(\{Y = y_j\}) \\ &= \sum_{i: x_i \in I} P(\{X = x_i\}) \cdot P(\{Y \in J\}) \\ &= P(X \in I)P(Y \in J). \end{aligned}$$

Il viceversa è ovvio, basta prendere $I = \{x_i\}$ e $J = \{y_j\}$.

²³Si confrontino i passaggi effettuati con gli Esercizi proposti 5.2 e 5.3.

8.2 Esercizi di verifica

Esercizio 8.1. Siano X ed Y variabili aleatorie, entrambe a valori nell'insieme $\{-1, 0, 1\}$ e con distribuzione congiunta data dalla tabella

$Y \setminus X$	-1	0	1
-1	0.1	0.1	0
0	0	0.1	0.3
1	0.1	0.15	θ

essendo θ un opportuno valore $0 < \theta < 1$.

- Determinare il valore di θ .
- Determinare la distribuzione di probabilità congiunta della coppia (X, Z) , dove $Z \equiv X \cdot Y$
- Determinare la distribuzione di probabilità marginale della variabile Z
- Determinare la distribuzione di probabilità condizionata di X , dato $\{Z = 1\}$.

Esercizio 8.2. Verificare se sono stocasticamente indipendenti le variabili aleatorie X ed Y , la cui distribuzione di probabilità congiunta è stata considerata nel precedente Esempio 8.1.

Esercizio 8.3. Verificare se sono stocasticamente indipendenti le variabili aleatorie X ed Y (rispettivamente somma e massimo dei punteggi risultanti dal lancio di due dadi), considerate nell'Esercizio proposto 1 di questa lezione.

Esercizio 8.4. Siano X ed Y variabili aleatorie stocasticamente indipendenti con $X(\Omega) \equiv \{x_1, \dots, x_n\}$ e $Y(\Omega) \equiv \{y_1, \dots, y_m\}$ e siano $\{p'_i; i = 1, \dots, n\}$ e $\{p''_j; j = 1, \dots, m\}$ le relative probabilità marginali. Trovare la distribuzione di probabilità della variabile $Z \equiv X + Y$.

Esercizio 8.5. Siano X ed Y variabili aleatorie stocasticamente indipendenti, con distribuzioni binomiali di parametri (r, θ) e (s, θ) , rispettivamente.

- Determinare la distribuzione di probabilità di $Z \equiv X + Y$.
- Determinare la distribuzione di probabilità condizionata di X , dato $\{Z = k\}$, con $0 \leq k \leq r + s$.

8.2.1 Soluzione di alcuni esercizi importanti

In questa sezione segnaliamo al lettore e risolviamo in modo rapido gli Esercizi 8.4 e 8.5.

Soluzione dell'Esercizio 8.4. Ricordiamo che X e Y sono due variabili aleatorie e che $Z = X + Y$.

Iniziamo notando immediatamente che $Z(\Omega) = \{z \in \mathbb{R} : \exists x_k \in X(\Omega), y_h \in Y(\Omega) : x_k + y_h = z\}$, e

$$\begin{aligned} \{Z = z\} &= \{X + Y = z\} = \bigcup_{x_k, y_h : x_k + y_h = z} \{X = x_k, Y = y_h\} \\ &= \bigcup_k \{X = x_k, Y = z - x_k\} = \bigcup_h \{X = z - y_h, Y = y_h\} \end{aligned}$$

dove gli eventi $\{X = x_k, Y = y_h\}$ sono disgiunti a due a due, e lo stesso vale per gli eventi $\{X = x_k, Y = z - x_k\}$ (o per gli eventi $\{X = z - y_h, Y = y_h\}$), e dove alcuni degli eventi $\{X = x_k, Y = z - x_k\}$ (o $\{X = z - y_h, Y = y_h\}$) possono essere vuoti.

Quindi, **anche senza l'ipotesi di indipendenza** tra X ed Y , si ha

$$\begin{aligned} P(\{Z = z\}) &= P(\{X + Y = z\}) = \sum_{x_k, y_h: x_k + y_h = z} P(\{X = x_k, Y = y_h\}) \\ &= \sum_k P(\{X = x_k, Y = z - x_k\}) = \sum_h P(\{X = z - y_h, Y = y_h\}). \end{aligned}$$

Inoltre, **se si suppone l'indipendenza** tra X ed Y , allora

$$\begin{aligned} P(\{Z = z\}) &= P(X + Y = z) = \sum_{x_k, y_h: x_k + y_h = z} P(X = x_k)P(Y = y_h) \\ &= \sum_k P(X = x_k)P(Y = z - x_k) = \sum_h P(X = z - y_h)P(Y = y_h). \end{aligned}$$

Soluzione dell'Esercizio 8.5. Ricordiamo che $X \sim b(r, \theta)$ ed $Y \sim b(s, \theta)$ sono due variabili aleatorie indipendenti e che $Z = X + Y$.

Soluzione dell'Esercizio 8.5. Punto i) Utilizzando il risultato dell'Esercizio 8.4 precedente, essendo $X \sim b(r, \theta)$ ed $Y \sim b(s, \theta)$, indipendenti, si ha che

$$Z(\Omega) = 0, 1, 2, \dots, r + s,$$

$$\begin{aligned} P(Z = \ell) &= \sum_{\substack{k+h=\ell \\ 0 \leq k \leq r, 0 \leq h \leq s}} \binom{r}{k} \theta^k (1 - \theta)^{r-k} \binom{s}{h} \theta^h (1 - \theta)^{s-h} \\ &= \sum_{\substack{k+h=\ell \\ 0 \leq k \leq r, 0 \leq h \leq s}} \theta^\ell (1 - \theta)^{r+s-\ell} \binom{r}{k} \binom{s}{h} \\ &= \theta^\ell (1 - \theta)^{r+s-\ell} \sum_{0 \leq k \leq r, 0 \leq \ell-k \leq s} \binom{r}{k} \binom{s}{\ell-k}, \end{aligned}$$

da cui, ricordando la (20),

$$P(Z = \ell) = \theta^\ell (1 - \theta)^{r+s-\ell} \binom{r+s}{\ell} \quad \ell = 0, 1, \dots, r + s.$$

Soluzione dell'Esercizio 8.5. Punto ii) Utilizzando le definizioni di probabilità condizionata si ha:

$$P(X = i | Z = k) = \frac{P(X = i, Z = k)}{P(Z = k)}$$

considerando che $\{X = i, Z = k\} = \{X = i, Y = k - i\}$, e che $\{X = i, Y = k - i\} \neq \emptyset$, solo se $0 \leq i \leq r$ e $0 \leq k - i \leq s$

$$\begin{aligned} P(X = i | Z = k) &= \frac{P(X = i, Y = k - i)}{P(Z = k)} = \frac{P(X = i)P(Y = k - i)}{P(Z = k)} \\ &= \frac{\binom{r}{i} \theta^i (1 - \theta)^{r-i} \binom{s}{k-i} \theta^{k-i} (1 - \theta)^{s-(k-i)}}{\binom{r+s}{k} \theta^k (1 - \theta)^{r+s-k}} = \frac{\binom{r}{i} \binom{s}{k-i}}{\binom{r+s}{k}} \quad 0 \leq i \leq r \text{ e } 0 \leq k - i \leq s, \end{aligned}$$

cioè la distribuzione condizionata di X data $Z = k$ è una $Hyp(r + s, r; k)$.

Osservazione Questi ultimi due risultati non sono sorprendenti, infatti si potrebbe anche ragionare come segue:

siano E_j , per $j = 1, 2, \dots, r + s$, eventi globalmente indipendenti di probabilità θ , cioè tale da formare uno schema di Bernoulli. Consideriamo le variabili $X' = \sum_{j=1}^r \mathbf{1}_{E_j}$ e $Y' = \sum_{j=r+1}^{r+s} \mathbf{1}_{E_j}$. Per tali variabili aleatorie valgono le seguenti proprietà:

- a) X' ha la stessa distribuzione di X , cioè $b(r, \theta)$,
- b) Y' ha la stessa distribuzione di Y , cioè $b(s, \theta)$,
- c) X' ed Y' sono indipendenti²⁴,

e quindi (X', Y') **ha la stessa distribuzione congiunta di** (X, Y) . Di conseguenza,

- da una parte $Z' := X' + Y' = \sum_{i=1}^{r+s} \mathbf{1}_{E_j}$ ha la stessa distribuzione di $Z = X + Y$,
- mentre dall'altra parte Z' ha chiaramente distribuzione binomiale $b(r + s, \theta)$.

E con ciò si ottiene che anche Z ha distribuzione binomiale $b(r + s, \theta)$.

Inoltre anche la distribuzione condizionata di X data $Z = k$ coincide con la distribuzione di X' data $Z' = k$.

Il fatto di sapere che $Z' = k$, permette di affermare che si sono verificate solo le sequenze di

$$(E_1^* \cap E_2^* \cap \dots \cap E_r^*) \cap (E_{r+1}^* \cap E_{r+2}^* \cap \dots \cap E_{r+s}^*)$$

nelle quali per esattamente k tra gli E_ℓ^* si ha $E_\ell^* = E_\ell$, mentre per i restanti $r + s - k$ si ha $E_\ell^* = \bar{E}_\ell$.

Per individuare uno di questi eventi basta specificare quali sono i k indici ℓ per cui $E_\ell^* = E_\ell$, ovvero basta specificare un sottoinsieme $K \subseteq \{1, 2, \dots, r + s\}$ di cardinalità k . Ciascuno degli $\binom{r+s}{k}$ eventi di questo tipo ha la stessa probabilità (ed esattamente $\theta^k \cdot (1 - \theta)^{r+s-k}$).

L'evento $X' = i$, per $i = 0, 1, \dots, r$ corrisponde al caso in cui la cardinalità di $K_A := K \cap \{1, 2, \dots, r\}$ è uguale ad i (e quindi la cardinalità di $K_B := K \cap \{r + 1, r + 2, \dots, r + s\}$ è uguale ad $k - i$).

Tenendo conto di ciò ci si può convincere facilmente che la distribuzione condizionata di X' data $Z' = k$ è

$$P(X' = i | Z' = k) = \frac{\binom{r}{i} \binom{s}{k-i}}{\binom{r+s}{k}} \quad \text{per } 0 \leq i \leq r, \quad 0 \leq k - i \leq s,$$

cioè è una $\text{Hyp}(r + s, r; k)$.

²⁴Per dimostrare la proprietà c) bisogna utilizzare la **Proposizione 1** della Lezione 5, considerando che i) la partizione \mathcal{A} generata dagli eventi $\{E_1, E_2, \dots, E_r\}$ e la partizione \mathcal{B} generata dagli eventi $\{E_{r+1}, E_{r+2}, \dots, E_{r+s}\}$ sono indipendenti:

infatti un evento $A \in \mathcal{A}$ se e solo se $A = E_1^* \cap E_2^* \cap \dots \cap E_r^*$ dove $E_i^* = E_i$ oppure $E_i^* = \bar{E}_i$, e analogamente un evento $B \in \mathcal{B}$ se e solo se $B = E_{r+1}^* \cap E_{r+2}^* \cap \dots \cap E_{r+s}^*$ dove $E_j^* = E_j$ oppure $E_j^* = \bar{E}_j$, e quindi

$$P((E_1^* \cap E_2^* \cap \dots \cap E_r^*) \cap (E_{r+1}^* \cap E_{r+2}^* \cap \dots \cap E_{r+s}^*)) = P(E_1^*) \cdot P(E_2^*) \cdot \dots \cdot P(E_r^*) \cdot P(E_{r+1}^*) \cdot P(E_{r+2}^*) \cdot \dots \cdot P(E_{r+s}^*)$$

↓

$$P(A \cap B) = P(A)P(B)$$

ii) gli eventi del tipo $E = \{X' = k\}$ e gli eventi del tipo $F = \{Y' = k\}$ appartengono rispettivamente a $\mathcal{G}(\mathcal{A})$ e a $\mathcal{G}(\mathcal{B})$, le algebre generate dalle partizioni \mathcal{A} e \mathcal{B} .

9 Valore atteso di una variabile aleatoria e relative proprietà

In questa lezione verrà introdotta la nozione di “valore atteso” di una variabile aleatoria e ne verranno messe in evidenza proprietà ed aspetti fondamentali.

In molti problemi di tipo probabilistico, data una variabile aleatoria X , sorge la necessità di individuare una quantità deterministica che, in qualche senso e entro certi fini, sia *equivalente* ad X .

Ad esempio se X rappresenta il valore (aleatorio) del ricavo derivante da una operazione finanziaria²⁵, nasce spesso l'esigenza di valutare una cifra deterministica, che risulti equa quale importo da pagare per avere il diritto di godere di tale ricavo²⁶. Similmente, in un gioco d'azzardo vi è la necessità di verificare se il gioco sia, o meno, equo, e così via...

Come si comincerà a vedere qui di seguito, il concetto di valore atteso ha un ruolo fondamentale in tali problematiche e risulta anche ugualmente importante sia sul piano teorico, sia in altre applicazioni, ad esempio di tipo fisico, di tipo statistico, etc...

Cominciamo intanto con una definizione rigorosa di tale concetto.

Definizione 9.1 (valore atteso). Sia $X : \Omega \equiv \{\omega_1, \dots, \omega_N\} \rightarrow X(\Omega) \subseteq \mathbb{R}$ una variabile aleatoria definita su $(\Omega, \mathcal{P}(\Omega))$. Si definisce **valore atteso**²⁷ di X il numero

$$\mathbb{E}(X) \equiv \sum_{i=1}^N p(\omega_i) X(\omega_i),$$

dove, come al solito, $p(\omega_i) = P(\{\omega_i\})$.

Prima di presentare alcuni esempi illustrativi è bene elencare alcune proprietà immediate di tale definizione; in quanto segue X, Y, Z, \dots sono variabili aleatorie definite sullo stesso spazio $\Omega = \{\omega_1, \dots, \omega_N\}$.

La prima proprietà è immediata²⁸ (ricordando che deve essere, ovviamente, $\sum_{i=1}^N p(\omega_i) = 1$):

Proposizione 1. Se X è una variabile aleatoria degenera, cioè tale che, per un valore $\hat{x} \in \mathbb{R}$, vale $X(\omega_i) = \hat{x}, i = 1, \dots, N$, allora si ha

$$\mathbb{E}(X) = \hat{x}.$$

Anche le dimostrazioni delle seguenti due proprietà seguono banalmente dalla definizione.

Proposizione 2. Sia $E \subseteq \Omega$ un evento e $X = \mathcal{X}_E$ la variabile aleatoria indicatore di E . Si ha

$$\mathbb{E}(X) = P(E).$$

²⁵Ad esempio nel caso di un'*opzione call* l'operazione consiste nel comprare la possibilità, ma non l'obbligo, di pagare una azione ad un prezzo prefissato K ad un istante prefissato T , invece che al prezzo (aleatorio) di mercato dell'azione. Ovviamente tale possibilità viene esercitata se il prezzo (aleatorio) al tempo T è maggiore di K , altrimenti non viene esercitata: non è conveniente pagare K quello che si può ottenere sul mercato ad un prezzo minore.

²⁶Più in generale quando ci si assicura contro eventuali danni o furti, quando si scommette o si gioca in borsa, si deve pagare una quantità di denaro certa (il premio di un'assicurazione, l'ammontare della scommessa) in cambio di una quantità aleatoria (il denaro che si potrebbe ottenere in caso di danno o furto, l'importo della scommessa, in caso di vincita).

È interessante considerare anche il caso del gioco in borsa, in cui la quantità certa è, ad esempio, il prezzo di un'*opzione call*, ovvero il prezzo per ottenere la possibilità di pagare una azione ad un prezzo prefissato K , nell'istante T , mentre la quantità aleatoria è il ricavo tra il prezzo dell'azione al tempo T e K , se questa differenza è positiva.

²⁷Oltre a valore atteso si usano anche i termini *valore medio*, *aspettazione*, *speranza matematica*, e a volte anche *media*, che però è meglio evitare, perché spesso quando si parla di media ci si riferisce alla media aritmetica.

²⁸E del resto, pensando all'interpretazione di $\mathbb{E}(X)$ come quantità certa che si è disposti a scambiare con X , si capisce che se X è a sua volta una quantità certa \hat{x} , ci si aspetta che $\mathbb{E}(X)$ sia uguale a \hat{x} .

Proposizione 3. Siano $a \leq b$ due numeri reali tali che $a \leq X(\omega_i) \leq b, \forall 1 \leq i \leq N$. Allora

$$a \leq \mathbb{E}(X) \leq b.$$

Più in generale vale la **proprietà di monotonia**: se X ed Y sono due variabili aleatorie con la proprietà che $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$, allora

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

Una proprietà assolutamente fondamentale del valore atteso è la **linearità**, che verrà enunciata e verificata qui di seguito.

Proposizione 4. Siano a, b due numeri reali arbitrari e X, Y due variabili aleatorie definite su Ω . Consideriamo la variabile aleatoria Z , combinazione lineare di X, Y , con coefficienti a, b , cioè poniamo

$$Z(\omega_i) \equiv a \cdot X(\omega_i) + b \cdot Y(\omega_i), \quad \forall 1 \leq i \leq N.$$

Si ha

$$\mathbb{E}(Z) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y),$$

ovvero

$$\mathbb{E}(aX + bY) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).$$

Dimostrazione.

Si tratta in effetti di una semplice verifica. In virtù della definizione, abbiamo

$$\begin{aligned} \mathbb{E}(Z) &= \sum_{i=1}^N p(\omega_i) [a \cdot X(\omega_i) + b \cdot Y(\omega_i)] = \\ &= a \cdot \sum_{i=1}^N p(\omega_i) X(\omega_i) + b \cdot \sum_{i=1}^N p(\omega_i) Y(\omega_i) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y). \end{aligned}$$

Come immediati corollari di quanto sopra, otteniamo

Proposizione 5. Sia a un numero reale fissato e poniamo $Z = a \cdot X$. Allora risulta

$$\mathbb{E}(Z) = a \cdot \mathbb{E}(X).$$

(Basta prendere $b = 0$)

Proposizione 6. Sia b un numero reale fissato e poniamo $Z = X + b$. Allora risulta

$$\mathbb{E}(Z) = \mathbb{E}(X) + b.$$

(Basta prendere $a = 1$ ed Y la variabile aleatoria degenera con $Y(\omega_i) = 1$ per ogni $\omega_i \in \Omega$.)

Infine va ricordato che la proprietà della **Proposizione 4** si estende immediatamente al caso di n variabili aleatorie:

$$\mathbb{E}\left(\sum_{k=1}^n a_k X_k\right) = \sum_{k=1}^n a_k \mathbb{E}(X_k)$$

Di fatto, per calcolare il valore atteso $\mathbb{E}(X)$ di una variabile aleatoria X , occorre conoscere la distribuzione di probabilità di X ma non è necessariamente richiesto di conoscere quale sia lo spazio di probabilità Ω su cui X è definita, ne' come X vi possa essere definita, ne' quale sia la misura di probabilità su $(\Omega, \mathcal{P}(\Omega))$.

Il valore atteso $\mathbb{E}(X)$ dipende infatti soltanto dalla distribuzione di probabilità di X , come mostrato nel seguente risultato.

Proposizione 7. Supponiamo che X sia una variabile aleatoria definita su un arbitrario spazio finito Ω e tale che $X(\Omega) \equiv \{x_1, \dots, x_n\}$. Risulta allora

$$\mathbb{E}(X) = \sum_{j=1}^n x_j \cdot P(\{X = x_j\}). \quad (49)$$

Dimostrazione.

I modo

Dal momento che $\{\{X = x_1\}, \dots, \{X = x_n\}\}$ costituisce una partizione di Ω , potremo scrivere

$$\mathbb{E}(X) = \sum_{i=1}^N p(\omega_i) \cdot X(\omega_i) = \sum_{j=1}^n \left[\sum_{i: X(\omega_i)=x_j} p(\omega_i) \cdot X(\omega_i) \right].$$

D'altra parte, per ciascun $1 \leq j \leq n$ fissato, si ha ovviamente:

$$\begin{aligned} \sum_{i: X(\omega_i)=x_j} p(\omega_i) \cdot X(\omega_i) &= \sum_{i: X(\omega_i)=x_j} p(\omega_i) \cdot x_j = \\ &= x_j \left(\sum_{i: X(\omega_i)=x_j} p(\omega_i) \right) = x_j \cdot P(\{X = x_j\}). \end{aligned}$$

La dimostrazione è quindi completata.

II modo

La precedente dimostrazione è autocontenuta. Tuttavia è interessante notare che c'è una dimostrazione alternativa: posto come nella **Proposizione 1** della Lezione 7, $E_j = \{X = x_j\}$ e $X_j = \mathcal{X}_{E_j} = \mathbf{1}_{E_j}$, si ha (come dimostrato appunto in tale Proposizione)

$$X = \sum_{j=1}^n x_j X_j$$

e quindi per linearità

$$\mathbb{E}(X) = \sum_{j=1}^n x_j \mathbb{E}(X_j) = \sum_{j=1}^n x_j P(E_j) = \sum_{j=1}^n x_j P(\{X = x_j\}).$$

Questa dimostrazione si estende immediatamente anche al caso in cui si voglia calcolare $\mathbb{E}(h(X))$. Si veda a questo proposito la **Proposizione 10**.

Esempio 9.1. Negli Esempi 7.1 e 7.2 della Lezione 7 abbiamo visto due variabili aleatorie, X_1 e T , definite su spazi diversi, che hanno la stessa distribuzione di probabilità, uniforme su $\{1, 2, \dots, 6\}$. Il loro comune valore atteso è dato da

$$\mathbb{E}(X_1) = \sum_{j=1}^6 \frac{j}{6} = 3.5 (= \mathbb{E}(T)).$$

Questo esempio si generalizza immediatamente al caso in cui X è una variabile aleatoria uniforme su $\{1, 2, \dots, n\}$, ovvero con $P(X = j) = \frac{1}{n}$, per $j = 1, 2, \dots, n$ per cui²⁹

$$\mathbb{E}(X) = \sum_{j=1}^n j \frac{1}{n} = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Nel caso specifico di una variabile aleatoria X a valori in $\{0, 1, 2, \dots, n\}$ o in $\{1, 2, \dots, n\}$, il calcolo del valore atteso può anche essere convenientemente eseguito in termini della funzione

$$\bar{F}(j) \equiv P(\{X > j\});$$

vale infatti la seguente

Proposizione 8. Sia $X(\Omega) \subseteq \{0, 1, \dots, n\}$. Allora si ha

$$\mathbb{E}(X) = \sum_{j=0}^{n-1} P(\{X > j\}) = \sum_{j=0}^{n-1} \bar{F}(j)$$

In particolare se $X(\Omega) = \{1, \dots, n\}$, allora si ha

$$\mathbb{E}(X) = 1 + \sum_{j=1}^{n-1} P(\{X > j\})$$

Dimostrazione.

Come già notato nel corso della Lezione 7, si può scrivere, per $1 \leq j \leq n$

$$P(\{X = j\}) = P(\{X > j-1\}) - P(\{X > j\}),$$

e, tenendo conto che $P(\{X > n\}) = 0$,

$$P(\{X = n\}) = P(\{X > n-1\}).$$

Dunque

$$\begin{aligned} \mathbb{E}(X) &= \sum_{j=0}^n j \cdot P(\{X = j\}) = \sum_{j=1}^n j \cdot P(\{X = j\}) = \\ &= \sum_{j=1}^n j \cdot [P(\{X > j-1\}) - P(\{X > j\})] \\ &= \sum_{j=1}^n j \cdot P(\{X > j-1\}) - \sum_{j=1}^n j \cdot P(\{X > j\}) \end{aligned}$$

²⁹Si ricordi che

$$\sum_{j=1}^n j = \frac{n(n+1)}{2}.$$

ponendo $h = j - 1$, e tenendo conto che $P(\{X > n\}) = 0$,

$$\begin{aligned}
 &= \sum_{h=0}^{n-1} (h+1) \cdot P(\{X > h\}) - \sum_{j=1}^{n-1} j \cdot P(\{X > j\}) \\
 &= 1 \cdot P(\{X > 0\}) + 2 \cdot P(\{X > 1\}) + 3 \cdot P(\{X > 2\}) + \dots + nP(\{X > n-1\}) \\
 &\quad - 1 \cdot P(\{X > 1\}) - 2 \cdot P(\{X > 2\}) - \dots - (n-1)P(\{X > n-1\}) \\
 &= P(\{X > 0\}) + (2-1)P(\{X > 1\}) + \dots + (n-(n-1))P(\{X > n-1\}) \\
 &= P(\{X > 0\}) + \sum_{h=1}^{n-1} ((h+1) - h) \cdot P(\{X > h\}) \\
 &= \sum_{h=0}^{n-1} P(\{X > h\}),
 \end{aligned}$$

il che prova la prima affermazione. La seconda affermazione dipende solo dal fatto che se $X(\Omega) = \{1, \dots, n\}$, allora $P(\{X > 0\}) = 1$.

Esempio 9.2. Siano X e Y i punteggi ottenuti nel lancio di due dadi e consideriamo la variabile aleatoria definita da $Z = X \vee Y$ (cioè $Z = \max(X, Y)$). Calcolare $\mathbb{E}(Z)$.

Soluzione. Avevamo visto nell'Esempio 7.5 della precedente Lezione 7 che $P(\{Z \leq x\}) = (\frac{x}{6})^2$, $x = 1, 2, \dots, 6$. Dunque $P(\{Z > x\}) = 1 - P(\{Z \leq x\}) = 1 - (\frac{x}{6})^2$. Da cui

$$\mathbb{E}(Z) = 1 + \sum_{j=1}^{6-1} P(\{Z > j\}) = 1 + 5 - \frac{1+4+9+16+25}{36} = 5 - \frac{19}{36} \cong 4.472$$

Anche questo esempio si generalizza immediatamente al caso in cui X ed Y sono variabili aleatorie indipendenti, ciascuna uniforme in $\{1, 2, \dots, n\}$:

$$\begin{aligned}
 \mathbb{E}(X \vee Y) &= \sum_{j=0}^{n-1} P(\{X \vee Y > j\}) = \sum_{j=0}^{n-1} (1 - P(\{X \vee Y \leq j\})) \\
 &= \sum_{j=0}^{n-1} 1 - \sum_{j=0}^{n-1} P(\{X \leq j, Y \leq j\}) = n - \sum_{j=0}^{n-1} P(\{X \leq j, Y \leq j\}).
 \end{aligned}$$

Si noti che finora si è usato solo il fatto che le due variabili aleatorie sono a valori in $\{1, 2, \dots, n\} \subseteq \{0, 1, 2, \dots, n\}$.

A questo punto per l'indipendenza delle due variabili aleatorie X ed Y si ha³⁰

$$P(\{X \leq j, Y \leq j\}) = P(\{X \leq j\})P(\{Y \leq j\}),$$

da cui

$$\mathbb{E}(X \vee Y) = n - \sum_{j=0}^{n-1} P(\{X \leq j\})P(\{Y \leq j\}). \tag{50}$$

³⁰Si noti che la formula (50) vale in generale per variabili aleatorie indipendenti X e Y , a valori in $\{0, 1, 2, \dots, n\}$.

Essendo sia X che Y uniformi in $\{1, 2, \dots, n\}$ si ha³¹

$$P(\{X \leq j\})P(\{Y \leq j\}) = \left(\frac{j}{n}\right)^2, \quad \text{per } j = 0, 1, \dots, n,$$

e quindi

$$\begin{aligned} \mathbb{E}(X \vee Y) &= n - \sum_{j=0}^{n-1} \left(\frac{j}{n}\right)^2 = n - \frac{1}{n^2} \frac{(n-1)(n-1+\frac{1}{2})(n-1+1)}{3} \\ &= n - \frac{(n-1)(n-\frac{1}{2})}{3n} = \frac{6n^2 - (n-1)(2n-1)}{6n} = \frac{4n^2 + 3n - 1}{6n}. \end{aligned}$$

Esercizio proposto 9.1. Si ripeta il procedimento del precedente Esempio 9.2 considerando il minimo al posto del massimo. Si ripeta il procedimento del precedente Esempio 9.2, sia per il minimo che per il massimo, considerando X ed Y sempre indipendenti, ma uniformi in $\{0, 1, \dots, n\}$ invece che uniformi in $\{1, 2, \dots, n\}$.

C'è da notare comunque che, **in molti casi, il valore atteso di una variabile aleatoria può essere ottenuto in modo semplice, senza neanche calcolare la distribuzione di probabilità** (o, nel caso di variabili a valori interi positivi, la funzione $\bar{F}(j)$).

Piuttosto si tratta di sfruttare adeguatamente la proprietà di linearità. Il ruolo di tale proprietà verrà in parte illustrato in quanto segue.

Esempio 9.3. Calcolare il valore atteso di una variabile aleatoria X con una distribuzione binomiale di parametri n, θ .

Soluzione. Specializzando al nostro caso la formula (49), potremo scrivere:

$$\mathbb{E}(X) = \sum_{k=0}^n k \binom{n}{k} \theta^k (1-\theta)^{n-k},$$

e fare i conti relativi³². Ma possiamo anche ottenere il valore di $\mathbb{E}(X)$ senza calcoli, ricordando l'Esempio 7.3 della precedente Lezione 7 e ragionando come segue.

Consideriamo n variabili aleatorie binarie X_1, \dots, X_n , con $P(\{X_j = 1\}) = \theta$; per la proprietà di linearità del valore atteso si ha dunque

$$\mathbb{E}(X_j) = \theta, \quad \mathbb{E}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \mathbb{E}(X_j) = n\theta.$$

³¹Si tenga conto del fatto che

$$\sum_{k=1}^n k^2 = \frac{n(n+\frac{1}{2})(n+1)}{3}.$$

³²Si ottiene che $\mathbb{E}(X) = n\theta$, infatti

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \binom{n}{k} \theta^k (1-\theta)^{n-k} = \sum_{k=1}^n k \frac{n!}{k!(n-k)!} \theta^k (1-\theta)^{n-k} \\ &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} \theta^k (1-\theta)^{n-k} \end{aligned}$$

Sappiamo d'altra parte che, nel caso particolare in cui X_1, \dots, X_n sono indicatori di eventi completamente indipendenti, la variabile aleatorie $S := \sum_{j=1}^n X_j$ segue appunto una distribuzione binomiale di parametri n, θ , la stessa di X . E dunque, visto che il valore atteso di una variabile aleatoria dipende soltanto dalla sua distribuzione di probabilità, abbiamo che, anche per la nostra variabile aleatoria X , risulta $\mathbb{E}(X) = \mathbb{E}(S) = n\theta$.

Esercizio proposto 9.2. 25 studenti sostengono una prova di esonero. Supponendo che, per ognuno di loro, la probabilità di successo è uguale a 0.80, qual è il valore atteso del numero di studenti che passano la prova?

Naturalmente possiamo avere distribuzioni di probabilità diverse che danno luogo allo stesso valore atteso, e ne vedremo ora diversi esempi. In particolare due distribuzioni binomiali $b(n', \theta')$ e $b(n'', \theta'')$ danno luogo allo stesso valore atteso qualora risulti $n' \cdot \theta' = n'' \cdot \theta''$.

Riprendiamo ora il caso visto nell'Esempio 3.4 della Lezione 3

Esempio 3.4 rivisitato (“Paradosso del Cavalier De Méré”). Il numero di volte in cui si ottiene risultato “asso” in quattro lanci di un dado è una variabile aleatoria con distribuzione $b(4, \frac{1}{6})$; dunque il suo valore atteso è dato da $\frac{1}{6} \times 4 = \frac{2}{3}$. Tale valore coincide anche con il valore atteso del numero di volte in cui si presenta il doppio asso in ventiquattro lanci di una coppia di dadi.

Possiamo concludere quindi che in entrambi i tipi di gioco d'azzardo si ha uguale valore atteso del numero dei successi (mentre sono diverse le probabilità di ottenere almeno un successo, come avevamo visto).

Osservazione 1. Dati n eventi, A_1, A_2, \dots, A_n , consideriamo la variabile S_n che conta il numero dei successi su tali eventi, ovvero

$$S_n = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \dots + \mathbf{1}_{A_n} = X_1 + X_2 + \dots + X_n,$$

con $X_i = \mathbf{1}_{A_i}$, $i = 1, 2, \dots, n$.

Riprendendo quanto visto nel corso della soluzione del precedente Esempio 9.3, possiamo notare che, nel caso in cui ciascuno di tali eventi abbia probabilità θ , il valore atteso $\mathbb{E}(S_n)$ è uguale, in forza della proprietà di linearità, al prodotto $n \cdot \theta$.

Per giungere a tale conclusione **non abbiamo fatto alcuna ipotesi circa l'indipendenza stocastica, o meno, fra tali eventi**. Naturalmente il tipo di correlazione fra tali eventi avrà influenza sulla distribuzione di probabilità di S_n , ma non sul suo valore atteso che resta in ogni caso uguale a $n \cdot \theta$:

$$\mathbb{E}(S_n) = n\theta.$$

ponendo $h = k - 1$, e tenendo conto che $1 \leq k \leq n \Leftrightarrow 0 \leq h \leq n - 1$, che $n - k = n - 1 - (k - 1) = n - 1 - h$, $n! = n \cdot (n - 1)!$, $k = h + 1$, si ha

$$\begin{aligned} &= \sum_{h=0}^{n-1} n \frac{(n-1)!}{h!(n-1-h)!} \theta^{h+1} (1-\theta)^{n-1-h} \\ &= n\theta \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} \theta^h (1-\theta)^{n-1-h} = n\theta \sum_{h=0}^{n-1} \binom{n-1}{h} \theta^h (1-\theta)^{n-1-h} \end{aligned}$$

tenendo conto dello sviluppo della potenza del binomio

$$= n\theta (\theta + (1-\theta))^{n-1} = n\theta \cdot 1.$$

La variabile aleatoria

$$Y_n \equiv \frac{S_n}{n}$$

può essere interpretata come la **frequenza relativa dei successi** sugli n eventi. Ancora per la proprietà di linearità, avremo, in ogni caso,

$$\mathbb{E}(Y_n) = \theta.$$

Esempio 9.4 (valore atteso di una ipergeometrica). Consideriamo in particolare **una distribuzione ipergeometrica di parametri** M, m_1, n . Come sappiamo questa è la distribuzione di probabilità di una variabile aleatoria S che conta il numero di successi su n eventi (non indipendenti) ciascuno di probabilità $\frac{m_1}{M}$; e quindi si ha che **il suo valore atteso è dato da** $n \cdot \frac{m_1}{M}$.

Estendiamo ora la discussione su un punto, già introdotto nella precedente **Osservazione 1**, che risulta di notevole importanza nella teoria della probabilità.

Osservazione 2 (valore atteso di una media aritmetica). Consideriamo n variabili aleatorie X_1, \dots, X_n definite su uno stesso spazio $(\Omega, \mathcal{P}(\Omega))$ e che abbiano tutte lo stesso valore atteso; cioè, usando il simbolo μ per indicare brevemente $\mathbb{E}(X_j)$,

$$\mathbb{E}(X_1) = \dots = \mathbb{E}(X_n) = \mu.$$

Consideriamo ora, sullo stesso spazio $(\Omega, \mathcal{P}(\Omega))$, la variabile aleatoria definita come **media aritmetica** delle X_1, \dots, X_n ,

$$Y_n(\omega_i) = \frac{\sum_{j=1}^n X_j(\omega_i)}{n}, \quad i = 1, \dots, N;$$

si ha ovviamente, in virtù della proprietà di linearità del valore atteso,

$$\mathbb{E}(Y_n) = \mu,$$

ottenendo quindi una generalizzazione del caso considerato nella precedente **Osservazione 1**.

A parità di valore atteso possono però sussistere situazioni assai diverse per quanto riguarda la distribuzione di probabilità di Y_n , a seconda delle proprietà di **correlazione** fra le variabili X_1, \dots, X_n . Questo è quanto vedremo meglio nella prossima lezione, introducendo i concetti di **varianza** di una variabile aleatoria e di **covarianza** fra due variabili aleatorie.

Un esempio, in un certo senso *estremo*, di comportamento di una media aritmetica è il seguente; questo mette anche in luce alcuni aspetti basilari del concetto di valore atteso.

Esempio 9.5. Consideriamo una lotteria in cui vengono venduti n biglietti, numerati progressivamente. Supponiamo che tale lotteria distribuisca un totale di r ($r < n$) premi, di cui, ad esempio,

r_1 “primi premi” di entità c_1 ,

r_2 “secondi premi” di entità c_2 e

r_3 “terzi premi” di entità c_3

(dunque $r = r_1 + r_2 + r_3$ e $c_1 > c_2 > c_3 > 0$).

Indichiamo con X_1, \dots, X_n le vincite rispettivamente associate al biglietto 1, al biglietto 2, ..., al biglietto n .

Ovviamente X_1, \dots, X_n sono delle variabili aleatorie i cui valori possibili costituiscono l'insieme $\{0, c_1, c_2, c_3\}$. X_1, \dots, X_n hanno tutte, la stessa distribuzione di probabilità, data da³³

$$P(\{X_j = c_1\}) = \frac{r_1}{n}, \quad P(\{X_j = c_2\}) = \frac{r_2}{n}, \quad P(\{X_j = c_3\}) = \frac{r_3}{n},$$

mentre

$$P(\{X_j = 0\}) = \frac{n-r}{n} (= 1 - P(\{X_j = c_1\}) - P(\{X_j = c_2\}) - P(\{X_j = c_3\}));$$

e dunque risulta

$$\mathbb{E}(X_j) = \frac{r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3}{n}.$$

Si può pervenire anche più semplicemente a quest'ultima conclusione, sfruttando di nuovo la proprietà di linearità del valore atteso e ragionando come segue: per motivi di simmetria, è ovvio che X_1, \dots, X_n abbiano tutte uno stesso valore atteso $\mathbb{E}(X_j) = \mu$. Andiamo a considerare la variabile aleatoria

$$S_n = \sum_{j=1}^n X_j;$$

S_n , avendo il significato di ammontare complessivo dei premi distribuiti dalla lotteria, deve essere una variabile aleatoria degenera, di valore $r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3$, ovvero

$$S_n(\omega) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3, \quad \text{qualunque sia } \omega \in \Omega$$

quindi

$$\mathbb{E}(S_n) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3.$$

Per la proprietà di linearità, d'altra parte, si deve avere

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{j=1}^n X_j\right) = n \cdot \mu$$

e dunque deve essere

$$\mu = \frac{\mathbb{E}(S_n)}{n} = \frac{r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3}{n}. \quad (51)$$

Ovviamente anche il valore atteso della variabile aleatoria $Y_n \equiv \frac{S_n}{n}$ (Y_n media aritmetica di X_1, \dots, X_n) è uguale alla quantità μ , come sappiamo deve essere; ma in questo speciale caso Y_n è una variabile aleatoria con distribuzione degenera, tutta concentrata sul valore μ .

Dal confronto cioè fra la distribuzione di X_j e quella di Y_n , vediamo che entrambe hanno valore atteso μ , ma la X_j non ha una distribuzione degenera, come invece accade per Y_n .

³³Infatti si può pensare che vengano effettivamente messe in un'urna n palline con i numeri di tutti i possibili biglietti 1, 2, ..., n e che vengano effettuate $r_1 + r_2 + r_3$ estrazioni senza reinserimento. L'evento "vincita della cifra c_1 " viene allora espresso come l'evento "viene estratto il numero j , nelle prime r_1 estrazioni". Analogamente l'evento "vincita della cifra c_2 " viene espresso come l'evento "viene estratto il numero j , in una delle estrazioni tra la $(r_1 + 1)$ -sima e la $(r_1 + r_2)$ -sima estrazione". Infine l'evento "vincita della cifra c_3 " viene espresso come l'evento "viene estratto il numero j , in una delle estrazioni tra la $(r_1 + r_2 + 1)$ -sima e l'ultima estrazione, ovvero la $(r_1 + r_2 + r_3)$ -sima". Basta poi notare che la probabilità che il numero j sia estratto alla k -sima estrazione vale $\frac{1}{n}$, qualunque sia k per ottenere la distribuzione di X_j .

È interessante anche notare che si ottiene la stessa distribuzione anche nel caso in cui invece la lotteria sia del tipo **gratta e vinci**. Ovvero ci sono n biglietti e si prende un biglietto a caso tra n in cui r_i danno diritto al premio c_i , per $i = 1, 2, 3$.

Esempio 9.6. La rilevanza della proprietà di linearità del valore atteso può essere illustrata dal seguente esempio. Supponiamo di possedere un biglietto di una lotteria A ed un biglietto della lotteria B .

Supponiamo per semplicità che A distribuisca

r'_1 premi di entità c'_1 e r'_2 premi di entità c'_2 , su un totale di n' biglietti;

mentre B distribuisce

r''_1 premi di entità c''_1 e r''_2 premi di entità c''_2 , su un totale di n'' biglietti.

Indicando con X la variabile aleatoria che indica la vincita complessivamente derivante dai due biglietti, calcolare $\mathbb{E}(X)$.

Soluzione. Indichiamo rispettivamente con X' e X'' le vincite relative ai due singoli biglietti, cosicché $X = X' + X''$.

X' è una variabile aleatoria che può assumere i valori $\{0, c'_1, c'_2\}$ e X'' è una variabile aleatoria che può assumere i valori $\{0, c''_1, c''_2\}$ e risulta, procedendo come nel precedente Esempio 9.5,

$$\mathbb{E}(X') = \frac{c'_1 r'_1 + c'_2 r'_2}{n'}, \quad \mathbb{E}(X'') = \frac{c''_1 r''_1 + c''_2 r''_2}{n''}.$$

È importante sottolineare che, sempre grazie alla proprietà di linearità del valore atteso, **per calcolare il valore atteso $\mathbb{E}(X)$ non c'è bisogno di calcolare interamente la distribuzione di probabilità di X** ; possiamo infatti ottenere immediatamente, in virtù della proprietà di linearità,

$$\mathbb{E}(X) = \mathbb{E}(X') + \mathbb{E}(X'') = \frac{c'_1 r'_1 + c'_2 r'_2}{n'} + \frac{c''_1 r''_1 + c''_2 r''_2}{n''}.$$

Osservazione 3. Riprendiamo il caso di una singola lotteria, considerata nel precedente Esempio 9.5, ed indichiamo con c il costo di un singolo biglietto.

L'acquirente del biglietto j paga dunque il prezzo certo c ed ottiene in cambio il guadagno aleatorio X_j , il cui valore atteso è $\mathbb{E}(X_j) = \mu$, dove μ è ottenuto come in (51) dell'Esempio 9.5. Si ha invece che l'organizzatore della lotteria ottiene un ricavo R pari a

$$R = n \cdot (c - \mu);$$

tale ricavo (positivo, negativo o nullo a seconda che sia μ minore, maggiore o uguale a c) è certo, nel senso che non dipende dal risultato aleatorio della lotteria (cioè da quali saranno i biglietti estratti). Su tale base, si può interpretare il valore atteso $\mu = \mathbb{E}(X_j)$ come il prezzo "equo" per l'acquisto di un singolo biglietto.

Tale interpretazione della nozione di valore atteso è fondamentale nel contesto della finanza matematica, in relazione al concetto di "non arbitraggio"³⁴.

Il ruolo della nozione di valore atteso nella comprensione del concetto di gioco equo è anche illustrata nel seguente

³⁴Si dice che in una operazione finanziaria c'è opportunità di arbitraggio se l'operazione finanziaria di sicuro non comporta perdite, e comporta un guadagno strettamente positivo con probabilità strettamente positiva. Per questo tipo di problemi si vedano anche le note della Prof. Nappo del corso "Metodi probabilistici per l'economia e la finanza". Gli appunti, nella versione del 19-01-2004, si possono trovare in rete all'indirizzo <http://www.mat.uniroma1.it/people/nappo/nappo.html#MPEF2003-04>

Esempio 9.7. Un giocatore, in possesso di un capitale iniziale di 31 Euro, gioca **al raddoppio** in una serie di puntate in ciascuna delle quali può vincere o perdere la cifra puntata: inizialmente punta 1 Euro, se vince ottiene 1 Euro e si ferma; se perde, raddoppia la puntata e continua così di seguito finché non vince la prima volta o finché non esaurisce il suo capitale iniziale.

Supponiamo che, in ciascuna puntata, il giocatore vince o perde con probabilità rispettivamente date da θ o $1 - \theta$.

Indicando con X la variabile aleatoria che rappresenta il capitale del giocatore al termine del gioco, vogliamo calcolare $\mathbb{E}(X)$.

Soluzione. Osserviamo innanzitutto che X è una variabile aleatoria che prende solo i due valori dell'insieme $\{0, 32\}$. Infatti il giocatore continua a giocare al raddoppio fino a che non raggiunge il capitale di 32 Euro, a meno che non perda per 5 volte consecutive, nel qual caso esaurisce appunto tutto il suo capitale iniziale di $31 = 1 + 2 + 4 + 8 + 16$ Euro. Ne segue quindi

$$P(\{X = 0\}) = (1 - \theta)^5, \quad P(\{X = 32\}) = 1 - (1 - \theta)^5,$$

da cui

$$\mathbb{E}(X) = 0 \cdot (1 - \theta)^5 + 32 \left(1 - (1 - \theta)^5\right) = 32 - 32 \cdot (1 - \theta)^5.$$

Vediamo quindi che, così facendo, il giocatore decide di scambiare il suo capitale certo di 31 Euro con un capitale aleatorio X di valore atteso $\mathbb{E}(X)$.

Osserviamo a tale proposito che risulta $\mathbb{E}(X) < 31$, $\mathbb{E}(X) = 31$, $\mathbb{E}(X) > 31$, a seconda che sia $\theta < \frac{1}{2}$, $\theta = \frac{1}{2}$, oppure $\theta > \frac{1}{2}$, cioè a seconda che il gioco sia sfavorevole, equo, oppure favorevole per il giocatore stesso.

Osservazione 4. Supponiamo di distribuire su una retta delle masse p_1, p_2, \dots, p_n , ponendole rispettivamente sui punti di ascissa x_1, x_2, \dots, x_n . Osserviamo allora che la quantità $\sum_{j=1}^n p_j \cdot x_j$ equivale al baricentro³⁵ di tale distribuzione di masse.

Così come il baricentro costituisce un valore che, a certi fini, risulta “riassuntivo” di tutta la distribuzione di masse, così il valore atteso di una variabile aleatoria costituisce un valore che, entro certi fini, riassume la conoscenza completa della distribuzione di probabilità.

Una proprietà di fondamentale importanza è data dalla seguente

Proposizione 9. Siano X, Y due variabili aleatorie indipendenti definite su uno stesso spazio di probabilità. Allora

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Dimostrazione. Siano rispettivamente

$$X(\Omega) = \{x_1, \dots, x_n\} \quad \text{e} \quad Y(\Omega) = \{y_1, \dots, y_m\}.$$

Possiamo allora considerare la partizione di Ω costituita dagli eventi del tipo

$$E_{h,k} = \{X = x_h\} \cap \{Y = y_k\} \equiv \{\omega_i \in \Omega : X(\omega_i) = x_h, Y(\omega_i) = y_k\},$$

³⁵Infatti in generale se le masse sono m_1, m_2, \dots, m_n e se sono poste rispettivamente sui punti di ascissa x_1, x_2, \dots, x_n , allora il baricentro è

$$\hat{x} = \frac{x_1 \cdot m_1 + x_2 \cdot m_2 + \dots + x_n \cdot m_n}{m_1 + m_2 + \dots + m_n}.$$

Nel caso in cui $m_i = p_i$, con p_i densità discreta, la precedente espressione diviene

$$\hat{x} = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n}{p_1 + p_2 + \dots + p_n} = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n = \mathbb{E}(X).$$

in quanto il denominatore $p_1 + p_2 + \dots + p_n$ vale 1.

per $h = 1, 2, \dots, n$, e $k = 1, 2, \dots, m$, e scrivere

$$\begin{aligned}\mathbb{E}(X \cdot Y) &= \sum_{i=1}^N p(\omega_i) [X(\omega_i) \cdot Y(\omega_i)] = \sum_{h=1}^n \sum_{k=1}^m \left[\sum_{i: \omega_i \in \{X=x_h\} \cap \{Y=y_k\}} p(\omega_i) [X(\omega_i) \cdot Y(\omega_i)] \right] \\ &= \sum_{h=1}^n \sum_{k=1}^m \left[\sum_{i: \omega_i \in \{X=x_h\} \cap \{Y=y_k\}} p(\omega_i) x_h \cdot y_k \right] = \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k \left[\sum_{i: \omega_i \in \{X=x_h\} \cap \{Y=y_k\}} p(\omega_i) \right] \\ &= \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k P(\{X = x_h\} \cap \{Y = y_k\}).\end{aligned}$$

Si noti che a questo risultato si poteva arrivare anche considerando che

$$X \cdot Y = \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k \mathbf{1}_{E_{h,k}}$$

da cui immediatamente, per la proprietà di linearità del valore atteso si ha

$$\mathbb{E}(X \cdot Y) = \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k \mathbb{E}(\mathbf{1}_{E_{h,k}}) = \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k P(\{X = x_h\} \cap \{Y = y_k\}). \quad (52)$$

Si noti inoltre che fino a questo punto non si è usata l'ipotesi di indipendenza, mentre ora, tenendo conto del fatto che, per l'indipendenza $P(\{X = x_h\} \cap \{Y = y_k\}) = P(\{X = x_h\}) \cdot P(\{Y = y_k\})$, si ottiene³⁶

$$\begin{aligned}\mathbb{E}(X \cdot Y) &= \sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k \cdot P(X = x_h) \cdot P(Y = y_k) \\ &= \sum_{h=1}^n x_h \cdot P(X = x_h) \cdot \sum_{k=1}^m y_k \cdot P(Y = y_k) = \mathbb{E}(X) \cdot \mathbb{E}(Y),\end{aligned}$$

e la dimostrazione è terminata.

In questa ultima parte estendiamo i risultati ottenuti in **Proposizione 7** e **Proposizione 9**, al caso di trasformazioni di variabili aleatorie.

Proposizione 10. Sia X una variabile aleatoria definita su uno spazio di probabilità finito, e sia data la funzione $h : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto h(x)$. Allora

$$\mathbb{E}(h(X)) = \sum_{j=1}^n h(x_j) \cdot P(\{X = x_j\}). \quad (53)$$

³⁶Se i passaggi che seguono risultassero difficili, si noti che

$$\begin{aligned}\sum_{h=1}^n \sum_{k=1}^m x_h \cdot y_k \cdot P(X = x_h) \cdot P(Y = y_k) &= \sum_{h=1}^n \left(\sum_{k=1}^m x_h \cdot y_k \cdot P(X = x_h) \cdot P(Y = y_k) \right) \\ &= \sum_{h=1}^n x_h \cdot P(X = x_h) \left(\sum_{k=1}^m y_k \cdot P(Y = y_k) \right) = \sum_{h=1}^n x_h \cdot P(X = x_h) \mathbb{E}(Y) = \mathbb{E}(Y) \sum_{h=1}^n x_h \cdot P(X = x_h) = \mathbb{E}(Y) \mathbb{E}(X)\end{aligned}$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilità finito, e sia data la funzione $g: \mathbb{R}^2 \rightarrow \mathbb{R}; (x, y) \mapsto g(x, y)$. Allora

$$\mathbb{E}(g(X, Y)) = \sum_{h=1}^n \sum_{k=1}^m g(x_h, y_k) P(\{X = x_h\} \cap \{Y = y_k\}) \quad (54)$$

Prima di dare la dimostrazione di questo risultato osserviamo che la precedente **Proposizione 10** assicura che **per calcolare il valore atteso della variabile aleatoria di una trasformazione di variabili aleatorie (ovvero di $W = h(X)$ o $Z = g(X, Y)$) non è necessario calcolare la sua distribuzione di probabilità**, ma basta applicare la (53) o la (54) rispettivamente.

Dimostrazione. La dimostrazione di (53) non è altro che l'immediata estensione della dimostrazione della **Proposizione 7**: infatti, con le stesse notazioni usate nel (*II modo*) di tale dimostrazione, basta notare³⁷ che

$$h(X) = \sum_{j=1}^n h(x_j) X_j = \sum_{j=1}^n h(x_j) \mathbf{1}_{E_j}$$

e quindi per linearità

$$\mathbb{E}(h(X)) = \sum_{j=1}^n h(x_j) \mathbb{E}(X_j) = \sum_{j=1}^n h(x_j) P(E_j) = \sum_{j=1}^n h(x_j) P(\{X = x_j\}).$$

La dimostrazione di (54) è simile:

con le stesse notazioni usate nella dimostrazione della **Proposizione 9**

$$g(X, Y) = \sum_{h=1}^n \sum_{k=1}^m g(x_h, y_k) \mathbf{1}_{E_{h,k}}$$

da cui immediatamente, per la proprietà di linearità del valore atteso si ha

$$\mathbb{E}(g(X, Y)) = \sum_{h=1}^n \sum_{k=1}^m g(x_h, y_k) \mathbb{E}(\mathbf{1}_{E_{h,k}}) = \sum_{h=1}^n \sum_{k=1}^m g(x_h, y_k) P(\{X = x_h\} \cap \{Y = y_k\}). \quad (55)$$

Infine va notato che nel caso di indipendenza tra X e Y si ha anche

Proposizione 11. Siano X, Y due variabili aleatorie **indipendenti** definite su uno stesso spazio di probabilità, e siano $h_i: \mathbb{R} \rightarrow \mathbb{R}; x \mapsto h_i(x)$, per $i = 1, 2$ due funzioni reali, allora

$$\mathbb{E}(h_1(X) \cdot h_2(Y)) = \mathbb{E}(h_1(X)) \cdot \mathbb{E}(h_2(Y)).$$

Dimostrazione. Basta prendere $g(x, y) = h_1(x) \cdot h_2(y)$ in (55) ottenendo così

$$\mathbb{E}(h_1(X) \cdot h_2(Y)) = \sum_{h=1}^n \sum_{k=1}^m h_1(x_h) \cdot h_2(y_k) \cdot P(\{X = x_h\} \cap \{Y = y_k\})$$

³⁷Ovviamente se $\omega \in E_j = \{X = x_j\}$, ovvero è tale che $X(\omega) = x_j$, allora $W(\omega) = h(X(\omega)) = h(x_j)$.

e sfruttando l'indipendenza tra X ed Y

$$= \sum_{h=1}^n \sum_{k=1}^m h_1(x_h) \cdot h_2(y_k) \cdot P(\{X = x_h\}) \cdot P(\{Y = y_k\}).$$

A questo punto la dimostrazione è identica alla dimostrazione della **Proposizione 9**, pur di sostituire $x_h \cdot y_k$ con $h_1(x_h) \cdot h_2(y_k)$.

Come già osservato, per calcolare il valore atteso della variabile aleatoria di una trasformazione di variabili aleatorie (ovvero di $W = h(X)$ o $Z = g(X, Y)$) non è necessario calcolare la sua distribuzione di probabilità. Tuttavia questo problema è un problema interessante, e abbiamo già calcolato alcune distribuzioni di trasformazioni di variabili aleatorie: si vedano i precedenti esempi ed esercizi che riguardano la somma di variabili aleatorie, il massimo o il minimo, etc., in particolare si ricordino Esempio 7.5, Esempio 7.6, Esercizio proposto 7.1, Esercizio proposto 8.1, Esercizio 8.1 (punti b) e c), Esercizio 8.4, Esercizio 8.5.

La prossima proposizione riassume il metodo generale per ottenere la distribuzione di trasformazioni di variabili aleatorie.

Proposizione 12 Sia X una variabile aleatoria definita su uno spazio di probabilità finito, e sia data la funzione $h : \mathbb{R} \rightarrow \mathbb{R}; x \mapsto h(x)$. Allora, posto $W = h(X)$, e $W(\Omega) = \{w_1, w_2, \dots, w_\ell\}$ si ha

$$\mathbb{P}(W = w_k) = \mathbb{P}(h(X) = w_k) = \sum_{j: h(x_j)=w_k} P(\{X = x_j\}), \quad k = 1, 2, \dots, \ell \quad (56)$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilità finito, e sia data la funzione $g : \mathbb{R}^2 \rightarrow \mathbb{R}; (x, y) \mapsto g(x, y)$. Allora, posto $Z = g(X, Y)$ e $Z(\Omega) = \{z_1, z_2, \dots, z_r\}$ si ha

$$\mathbb{P}(g(X, Y) = z_j) = \sum_{h,k: g(x_h, y_k)=z_j}^{1 \leq h \leq n, 1 \leq k \leq m} P(\{X = x_h\} \cap \{Y = y_k\}), \quad j = 1, 2, \dots, r. \quad (57)$$

9.1 Valore atteso condizionato e valore atteso

ATTENZIONE Sezione NUOVA

Definizione 9.2 (Valore atteso condizionato ad un evento di una variabile aleatoria discreta). *Sia A un evento con $P(A) > 0$ e sia X una variabile aleatoria (discreta), a valori in $X(\Omega) = \{x_1, \dots, x_m\}$. Allora possiamo si definisce*

$$E(X|A) = \sum_{k=1}^m x_k P(X = x_k|A)$$

Dalla precedente definizione e dalla formula delle probabilità totali si ottiene immediatamente il seguente risultato.

Proposizione 9.1. *Se $\{H_1, \dots, H_n\}$ formano una partizione e si ha $P(H_i) > 0$ per ogni $i = 1, \dots, n$ allora si ha*

$$E(X) = \sum_{j=1}^n E(X|H_j) P(H_j). \quad (58)$$

Dimostrazione. Sappiamo che, per ogni $j = 1, \dots, n$

$$E(X|H_j) = \sum_{k=1}^m x_k P(X = x_k|H_j)$$

per cui

$$\sum_{j=1}^n E(X|H_j) P(H_j) = \sum_{j=1}^n \sum_{k=1}^m x_k P(X = x_k|H_j) P(H_j).$$

Essendo una somma finita, possiamo scambiare l'ordine di sommatoria e quindi ottenere la tesi osservando che

$$\begin{aligned} \sum_{j=1}^n E(X|H_j) P(H_j) &= \sum_{k=1}^m \sum_{j=1}^n x_k P(X = x_k|H_j) P(H_j) = \sum_{k=1}^m x_k \sum_{j=1}^n P(X = x_k|H_j) P(H_j) \\ &= \sum_{k=1}^m x_k \sum_{j=1}^n P(X = x_k), \end{aligned}$$

dove l'ultima uguaglianza deriva della formula delle probabilità totali applicata all'evento $\{X = x_k\}$ e alla partizione $\{H_1, \dots, H_n\}$. \square

Un caso particolarmente interessante è il caso in cui la partizione è data dalla partizione generata da una variabile aleatoria (discreta) Y con valori in $Y(\Omega) = \{y_1, \dots, y_n\}$

$$H_j = \{Y = y_j\}, \quad j = 1, \dots, n.$$

È interessante allora applicare il risultato precedente al caso precedente, specialmente se la distribuzione (discreta) congiunta di X e di Y è data attraverso la distribuzione marginale di Y , o attraverso la sua densità discreta

$$p_Y(y_j) = P(Y = y_j), \quad j = 1, \dots, n.$$

e la distribuzione condizionata di X data Y , ovvero attraverso

$$p_{X|Y}(x_k|y_j) = P(X = x_k|\{Y = y_j\}), \quad k = 1, \dots, m$$

per ogni $j = 1, \dots, m$.

In questo caso la formula (59) diviene

$$E(X) = \sum_{j=1}^n E(X|\{Y = y_j\}) P(Y = y_j). \quad (59)$$

Vediamo come si può applicare questa formula.

Esempio 9.8. Sia Y una variabile aleatoria uniforme nell'insieme $\{1, \dots, 10\}$, e supponiamo di sapere che, condizionatamente a $\{Y = j\}$, X ha distribuzione binomiale di parametri $n = j$ e $p = \frac{1}{4}$, cioè, per $j = 1, \dots, 10$,

$$P(X = k|Y = j) = \binom{j}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{j-k}, \quad k = 0, 1, \dots, j.$$

Allora, considerando che chiaramente $E(X|\{Y = j\}) = j \frac{1}{4}$,

$$E(X) = \sum_{k=1}^{10} E(X|Y = j) P(Y = j) = \sum_{k=1}^{10} j \frac{1}{4} \frac{1}{10} = \frac{1}{4} \frac{10 \cdot 11}{2} \frac{1}{10} = \frac{11}{8}.$$

Più in generale se Y una variabile aleatoria a valori nell'insieme $\{1, \dots, n\}$ e, condizionatamente a $\{Y = j\}$, X ha distribuzione binom(j, p), si ha

$$E(X) = \sum_{j=1}^n E(X|Y = j) P(Y = j) = \sum_{j=1}^n j p P(Y = j) = E(Y p) = p E(Y).$$

9.2 Esercizi di verifica

Esercizio 9.1. Sia X una variabile aleatoria tale che

$$P(X = -1) = q, \quad P(X = 1) = p$$

con $0 < q = 1 - p < 1$. Calcolate $\mathbb{E}(X)$.

Esercizio 9.2. Sia X una variabile aleatoria a valori nell'insieme $\{1, 2, \dots, n\}$ con distribuzione di probabilità data dalla posizione

$$P(\{X = k\}) \propto k, \quad k = 1, \dots, n.$$

Calcolate³⁸ $P(\{X = k\})$, per $k = 1, \dots, n$, e $\mathbb{E}(X)$.

Esercizio 9.3. Calcolate il valore atteso di una variabile aleatoria X a valori nell'insieme

$$\{0, 1, 2, \dots, n\},$$

sapendo che risulta

$$P(\{X = k\}) \propto \frac{\rho^k}{k!(n-k)!}, \quad k = 0, 1, \dots, n.$$

con ρ costante positiva assegnata.

Esercizio 9.4. Siano g_1, \dots, g_n dei numeri positivi assegnati e $x_1 < \dots < x_n$ assegnati numeri reali. Sia X una variabile aleatoria a valori nell'insieme $\{x_1, \dots, x_n\}$ e tale che

$$P(\{X = x_k\}) \propto g_k, \quad k = 1, \dots, n.$$

Ottenete la formula per calcolare il valore atteso di X .

Esercizio 9.5. Una persona possiede un biglietto di una lotteria in cui si distribuisce un premio da 5000 euro e tre premi da 2000 Euro ciascuno; possiede inoltre due biglietti di una lotteria che distribuisce un solo premio di 10000 Euro. Per entrambe le lotterie sono stati emessi 1000 biglietti. Indichiamo con X la variabile aleatoria che rappresenta l'importo complessivo della vincita di questa persona.

(a) Calcolate la probabilità di vincere almeno un premio, cioè $P(X > 0)$.

(b) Calcolate il valore atteso di X .

Esercizio 9.6. Tizio punta sul lancio di due dadi; relativamente al primo dado vince 1 Euro se si presenta il punto "3" o il punto "4" e vince 2 Euro se si presenta il punto "5" o il punto "6"; relativamente al secondo dado vince 1 Euro se si presenta il punto "3" e vince 2 Euro se si presenta un qualunque punto pari. In tutti gli altri casi non riceve alcuna vincita.

a) Determinate il valore atteso della vincita complessiva.

b) Calcolate la probabilità che Tizio vinca su entrambi i dadi.

Esercizio 9.7. Renato e Stefano lanciano una moneta perfetta n volte ciascuno; indichiamo con X_r ed X_s il numero di risultati testa da loro rispettivamente ottenuti.

Calcolate

$$\mathbb{E}(X_r^2 - X_r \cdot X_s).$$

³⁸Si ricordi che

$$\sum_{k=1}^n k = \frac{n(n+1)}{2} \quad \sum_{k=1}^n k^2 = \frac{1}{3}n(n + \frac{1}{2})(n+1)$$

Esercizio 9.8. Renato e Stefano lanciano una moneta perfetta n volte ciascuno ed il vincitore è quello fra i due che realizza il maggior numero di risultati testa.

Indichiamo con X il punteggio del vincitore e con Y il punteggio del perdente (se Renato e Stefano pareggiano, cioè se, utilizzando le notazioni dell'esercizio precedente, $X_r = X_s$ allora $X = Y = X_r = X_s$).

Trovare $\mathbb{E}(X - Y)$.

Esercizio 9.9. Verificate che risulta, per $n = 2, 3, \dots$,

$$2^n - 1 = \sum_{k=0}^{n-1} 2^k$$

e riformulate più in generale il testo e la soluzione dell'Esempio 9.7 (sostituendo un valore n generico al posto di $n = 5$).

Soluzione dell'Esercizio 9.8

In generale, per ogni a e b numeri reali, $\max(a, b) - \min(a, b) = |a - b|$, e quindi anche per le due variabili aleatorie X_r e X_s si ha che $\max(X_r, X_s) - \min(X_r, X_s) = |X_r - X_s|$. Poiché X_r e X_s sono variabili aleatorie indipendenti, entrambe con distribuzione $Bin(n, 1/2)$, le variabili aleatorie X_r e $n - X_s$, hanno la stessa distribuzione congiunta di X_r e X_s , e quindi

$$E[|X_r - X_s|] = E[|X_r - (n - X_s)|] = E[|X_r + X_s - n|].$$

Come sappiamo la somma di due variabili aleatorie indipendenti con distribuzione binomiale $Bin(n, p)$ e $Bin(m, p)$, rispettivamente, ha ancora distribuzione binomiale di parametri $n + m$ e p . Quindi $Z = X_r + X_s$ ha distribuzione $Bin(2n, p)$ e si tratta di calcolare

$$\begin{aligned} E[|Z - n|] &= \sum_{k=0}^{2n} |k - n| P(Z = k) = \sum_{k=0}^{2n} |k - n| \binom{2n}{k} \frac{1}{2^{2n}} \\ &= \sum_{k=0}^n (n - k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{k=n}^{2n} (k - n) \binom{2n}{k} \frac{1}{2^{2n}}, \end{aligned}$$

(si noti che il caso $k = n$ è ripetuto due volte, ma dà un contributo nullo alla somma)

$$= \sum_{k=0}^n (n - k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{k=n}^{2n} (k - n) \binom{2n}{2n - k} \frac{1}{2^{2n}}$$

ponendo $h = 2n - k$ di modo che $k - n = n - h$

$$\begin{aligned} &= \sum_{k=0}^n (n - k) \binom{2n}{k} \frac{1}{2^{2n}} + \sum_{h=0}^n (n - h) \binom{2n}{h} \frac{1}{2^{2n}} \\ &= 2 \sum_{k=0}^n (n - k) \binom{2n}{k} \frac{1}{2^{2n}} = 2 \frac{1}{2^{2n}} \left[n \sum_{k=0}^n \binom{2n}{k} - \sum_{k=0}^n k \binom{2n}{k} \right] \\ &= \frac{n}{2^{2n-1}} \left[n \sum_{k=0}^n \binom{2n}{k} - \sum_{k=1}^n 2n \binom{2n-1}{k-1} \right] \end{aligned}$$

Bisogna ora ricordare che

$$\binom{2n}{k} = \binom{2n-1}{k} + \binom{2n-1}{k-1}, \quad k \geq 1,$$

da cui

$$\begin{aligned} E[|Z - n|] &= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^n \binom{2n-1}{k} + \sum_{k=1}^n \binom{2n-1}{k-1} - 2 \sum_{k=1}^n \binom{2n-1}{k-1} \right] \\ &= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^n \binom{2n-1}{k} - \sum_{k=1}^n \binom{2n-1}{k-1} \right] \\ &= \frac{n}{2^{2n-1}} \left[\sum_{k=0}^n \binom{2n-1}{k} - \sum_{h=0}^{n-1} \binom{2n-1}{h} \right] \\ &= \frac{n}{2^{2n-1}} \binom{2n-1}{n}. \end{aligned}$$

10 Varianza, Covarianza e comportamento delle medie aritmetiche di variabili aleatorie

Nella precedente lezione abbiamo visto che è spesso interessante instaurare un confronto fra due variabili aleatorie che ammettano uguale valore atteso. Una nozione utile a tale riguardo è quella di **varianza**.

In questa lezione introdurremo dunque il concetto di varianza di una variabile aleatoria e ne vedremo alcune proprietà fondamentali; verrà introdotto, a tale proposito, anche il concetto di covarianza di una coppia di variabili aleatorie.

Come si vedrà, tali due nozioni forniscono, fra l'altro, utili informazioni relativamente al comportamento probabilistico di medie aritmetiche di una collezione di variabili aleatorie.

Cominciamo con la seguente

Definizione 10.1 (Varianza). Sia $X : \Omega \equiv \{\omega_1, \dots, \omega_N\} \rightarrow X(\Omega) \equiv \{x_1, \dots, x_n\}$ una variabile aleatoria ed indichiamo brevemente con μ il valore atteso di X . La **varianza** di X viene definita come il valore atteso di $(X - \mu)^2$ e si indica brevemente con il simbolo $Var(X)$, cioè

$$Var(X) = \mathbb{E} \left((X - \mathbb{E}(X))^2 \right) \equiv \sum_{\omega_i \in \Omega} p(\omega_i) (X(\omega_i) - \mu)^2.$$

Le proprietà fondamentali di tale definizione sono elencate qui di seguito.

Proposizione 1. La varianza di una variabile aleatoria X è sempre non negativa, ovvero $Var(X) \geq 0$. Inoltre, se $p(\omega_i) > 0$ per ogni $\omega_i \in \Omega$, si ha $Var(X) = 0$ se e solo se X è una variabile aleatoria degenera³⁹.

La dimostrazione è immediata e viene lasciata quale esercizio.

Proposizione 2. Per ogni variabile aleatoria X

$$Var(X) = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad (60)$$

Dimostrazione.

Applicando la proprietà di linearità del valore atteso, si ha immediatamente

$$Var(X) = \mathbb{E}(X^2 - 2\mu X + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mu^2.$$

³⁹Nel caso in cui Ω è un insieme finito, l'ipotesi che $p(\omega_i) > 0$ appare del tutto naturale. Tuttavia vale la pena di esaminare il caso generale in cui possa accadere che $p(\omega_i) = 0$ per qualche $\omega_i \in \Omega$. In tale caso non è più vero che $Var(X) = 0$ sia equivalente all'esistenza di un valore \hat{x} , tale che $X(\omega_i) = \hat{x}$ per ogni $\omega_i \in \Omega$. Si può invece affermare che $Var(X) = 0$ se e solo se esiste un $\hat{x} \in X(\Omega)$ tale che $P(X = \hat{x}) = 1$. In entrambi i casi, \hat{x} coincide con il valore atteso $\mu := \mathbb{E}(X)$.

Infatti, posto $\mathbb{E}(X) = \mu$, si ha

$$Var(X) = \sum_{i=1}^N p(\omega_i) (X(\omega_i) - \mu)^2 = 0$$

se e solo se

$$p(\omega_i) (X(\omega_i) - \mu)^2 = 0, \quad \text{per ogni } i = 1, \dots, N$$

in quanto una somma di termini non negativi è nulla se e solo se tutti gli addendi sono nulli. Di conseguenza almeno uno tra $p(\omega_i)$ e $X(\omega_i) - \mu$ deve essere nullo, cioè

$$\begin{cases} \text{o } X(\omega_i) - \mu = 0 & \text{e allora può essere sia } p(\omega_i) > 0, \text{ sia } p(\omega_i) = 0, \\ \text{o } X(\omega_i) - \mu \neq 0 & \text{e allora necessariamente } p(\omega_i) = 0. \end{cases}$$

Basta poi osservare che $P(\{\omega : X(\omega) \neq \mu\}) = \sum_{i: X(\omega_i) - \mu \neq 0} p(\omega_i) = 0$, e che ciò equivale a $P(\{X(\omega_i) = \mu\}) = 1$.

Un modo equivalente di esprimere la precedente proprietà è il seguente: $Var(X) = 0$ se e solo se X coincide con la variabile aleatoria degenera identicamente uguale a $\mu := \mathbb{E}(X)$ con probabilità 1.

Anche per le seguenti due proprietà le dimostrazioni sono immediate, tenendo conto ad esempio di (60), e vengono lasciate per esercizio.

Proposizione 3. Sia $Y = X + b$, essendo $b \in \mathbb{R}$. Allora

$$\text{Var}(Y) = \text{Var}(X).$$

Proposizione 4. Sia $Y = a \cdot X$, essendo $a \in \mathbb{R}$. Allora

$$\text{Var}(Y) = a^2 \cdot \text{Var}(X).$$

Proposizione 5. Siano X, Y due variabili aleatorie definite su $(\Omega, \mathcal{P}(\Omega))$. Allora

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2[\mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)]. \quad (61)$$

Dimostrazione.

Anche in questo caso la dimostrazione è immediata; ricordando la (60) e applicando la proprietà di linearità del valore atteso, si ha infatti

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y)^2] - [\mathbb{E}(X + Y)]^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - [\mathbb{E}(X)^2 + 2\mathbb{E}(X) \cdot \mathbb{E}(Y) + \mathbb{E}(Y)^2] \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 + 2[\mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)] \\ &= \text{Var}(X) + \text{Var}(Y) + 2[\mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)]. \end{aligned}$$

Osservazione 1. Come il valore atteso $\mathbb{E}(X)$, anche la varianza $\text{Var}(X)$ dipende soltanto dalla distribuzione di probabilità di X . Si veda a questo proposito anche la **Proposizione 7** della precedente Lezione 9.

Esempio 10.1. Sia X una variabile aleatoria binaria, ad esempio l'indicatore di un evento E , con $P(E) = p$.

Allora si ha, ricordando anche⁴⁰ la **Proposizione 2** della precedente Lezione 9

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X) - \mathbb{E}(X)^2 = \mathbb{E}(X)(1 - \mathbb{E}(X)) = p(1 - p).$$

Esempio 10.2. Consideriamo una variabile aleatoria X tale

$$P(X = -1) = q, \quad P(X = 1) = p$$

con $0 < q = 1 - p < 1$. Allora si ha $\mathbb{E}(X) = 2p - 1$ e, osservando che X^2 è una variabile aleatoria con distribuzione degenerata su 1,

$$\text{Var}(X) = \mathbb{E}[X^2] - (2p - 1)^2 = 4p(1 - p).$$

⁴⁰Si tenga anche conto che se $X = \mathcal{X}_E$, allora $X^2 = X$, in quanto ovviamente $0^2 = 0$ e $1^2 = 1$, e che

$$\mathbb{E}(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p.$$

Consideriamo ora, per un dato $a > 0$, la variabile $W = a \cdot X$:

$$P(W = -a) = q, \quad P(W = a) = p;$$

allora W^2 è una variabile aleatoria con distribuzione degenere su a^2 e si ha

$$\mathbb{E}(W) = a\mathbb{E}(X) = a(2p - 1);$$

$$\text{Var}(W) = a^2 - a^2(2p - 1)^2 = a^2 \cdot 4p(1 - p) = a^2 \cdot \text{Var}(X),$$

come del resto era ovvio, tenendo conto della **Proposizione 4**.

Definizione 10.2 (Covarianza). Siano date, su uno stesso spazio di probabilità, due variabile aleatorie X ed Y e, per comodità, indichiamo brevemente con μ_X e μ_Y i loro rispettivi valori attesi. Si definisce **covarianza** fra X, Y la quantità

$$\text{Cov}(X, Y) \equiv \mathbb{E}[(X - \mu_X) \cdot (Y - \mu_Y)] \left(= \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))] \right). \quad (62)$$

Osservazione 2. Si osservi che se $Y = X$, allora $\text{Cov}(X, Y) = \text{Cov}(X, X) = \text{Var}(X)$. Inoltre, svolgendo il prodotto nel secondo membro della (62) ed applicando la proprietà di linearità del valore atteso, si ottiene immediatamente

$$\text{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mu_X \cdot \mu_Y \left(= \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y) \right). \quad (63)$$

Dunque, ricordando (61), possiamo scrivere

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \quad (64)$$

Esercizio proposto 10.1. Verificare⁴¹ che, date tre variabili aleatorie X, Y, Z e due costanti α, β risulta

$$\text{Cov}(X, \alpha Y + \beta Z) = \alpha \text{Cov}(X, Y) + \beta \text{Cov}(X, Z).$$

Come corollario della **Proposizione 9** della precedente Lezione 9, ed in virtù della (64), otteniamo immediatamente

Proposizione 6. Siano date X ed Y due variabili aleatorie definite su uno stesso spazio di probabilità. Se X ed Y sono stocasticamente indipendenti, allora

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Dimostrazione. Basta ricordare la (63) e che, per la **Proposizione 9** della Lezione 9, se X ed Y sono indipendenti allora $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

⁴¹Per la simmetria della covarianza, la proprietà che

$$\text{Cov}(X, \alpha Y_1 + \beta Y_2) = \alpha \text{Cov}(X, Y_1) + \beta \text{Cov}(X, Y_2),$$

implica ovviamente anche

$$\text{Cov}(aX_1 + bX_2, Y) = a \text{Cov}(X_1, Y) + b \text{Cov}(X_2, Y).$$

Le due precedenti proprietà insieme sono note come la proprietà di bilinearità della covarianza o anche più brevemente:

$$\text{Cov}(aX_1 + bX_2, \alpha Y_1 + \beta Y_2) = a \cdot \alpha \text{Cov}(X_1, Y_1) + a \cdot \beta \text{Cov}(X_1, Y_2) + b \cdot \alpha \text{Cov}(X_2, Y_1) + b \cdot \beta \text{Cov}(X_2, Y_2).$$

Osservazione 3. Siano X ed Y due variabili aleatorie binarie definite su uno stesso spazio di probabilità. Osserviamo che anche il loro prodotto⁴² $X \cdot Y$ è una variabile binaria, con

$$\mathbb{E}(X \cdot Y) = P(\{X \cdot Y = 1\}) = P(\{X = 1\} \cap \{Y = 1\})$$

e dunque

$$Cov(X, Y) = P(\{X = 1\} \cap \{Y = 1\}) - P(\{X = 1\}) \cdot P(\{Y = 1\}).$$

Notiamo che la condizione $Cov(X, Y) = 0$, ovvero

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{65}$$

implica in tale caso che X ed Y sono stocasticamente indipendenti⁴³ in quanto

$$\begin{aligned} P(\{X = 1\} \cap \{Y = 1\}) &= \mathbb{E}(X \cdot Y) \\ &= \mathbb{E}(X) \cdot \mathbb{E}(Y) = P(\{X = 1\}) \cdot P(\{Y = 1\}). \end{aligned}$$

Osservazione 4. In generale, cioè per coppie di variabili non entrambe binarie, **la condizione (65) (ovvero $Cov(X, Y) = 0$) è soltanto necessaria, ma non sufficiente per l'indipendenza stocastica**, come mostra infatti il seguente **controesempio**.

Sia X una variabile aleatoria tale che

$$P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$$

e poniamo $Y \equiv X^2$, cosicché $X \cdot Y = X^3$ e ovviamente⁴⁴ X, Y non possono essere stocasticamente indipendenti; notiamo che, in questo caso, X^3 ha la stessa distribuzione di probabilità di X e $\mathbb{E}(X) = \mathbb{E}(X^3) = 0$; dunque risulta

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X^3) = 0 = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

La precedente osservazione ci porta naturalmente a dare la seguente definizione

Definizione 10.3 (Variabili aleatorie non correlate). *Diremo che due variabili aleatorie X ed Y , definite su uno stesso spazio di probabilità, sono **non correlate** (o anche **scorrelate**) se risulta verificata la condizione (65).*

⁴²A questo proposito è utile osservare che, se $X = \mathbf{1}_A$ ed $Y = \mathbf{1}_B$ allora $X \cdot Y = \mathbf{1}_A \cdot \mathbf{1}_B = \mathbf{1}_{A \cap B}$, come si vede subito, considerando che

$\mathbf{1}_A \cdot \mathbf{1}_B(\omega) = 1$ se e solo se $\omega \in A$ e simultaneamente $\omega \in B$, ovvero $\omega \in A \cap B$

e anche

$\mathbf{1}_{A \cap B}(\omega) = 1$ se e solo se $\omega \in A \cap B$.

⁴³Si osservi che, se come prima $X = \mathbf{1}_A$ ed $Y = \mathbf{1}_B$ allora la distribuzione congiunta di X e di Y è individuata da

$$\begin{aligned} P(\{X = 0\} \cap \{Y = 0\}) &= P(\overline{A} \cap \overline{B}) \\ P(\{X = 0\} \cap \{Y = 1\}) &= P(\overline{A} \cap B) \\ P(\{X = 1\} \cap \{Y = 0\}) &= P(A \cap \overline{B}) \\ P(\{X = 1\} \cap \{Y = 1\}) &= P(A \cap B), \end{aligned}$$

e che quindi l'indipendenza degli eventi $A = \{X = 1\}$ e $B = \{Y = 1\}$, è sufficiente per ottenere l'indipendenza delle variabili aleatorie X ed Y .

⁴⁴Il fatto che X ed $Y = X^2$ non siano indipendenti si vede immediatamente: ad esempio $P(Y = 0) = P(X^2 = 0) = P(X = 0) = 1/3 > 0$, e quindi $P(Y = 0) \cdot P(X = 1) = (1/3) \cdot (1/3) > 0$, mentre $P(\{Y = 0\} \cap \{X = 1\}) = P(\emptyset) = 0$.

Esercizio proposto 10.2. Generalizzare il controesempio dell'Osservazione 4, mostrando che le variabili aleatorie X ed $Y = X^2$, sono scorrelate, ma non indipendenti, *ogni volta che la variabile aleatoria X è simmetrica rispetto all'origine, ossia con $X(\Omega) = \{\pm x_1, \pm x_2, \dots, \pm x_r\} \cup \{0\}$, per un $r \geq 1$, (oppure $X(\Omega) = \{\pm x_1, \pm x_2, \dots, \pm x_r\}$, per un $r \geq 2$), e con*

$$P(X = x_i) = P(X = -x_i)$$

per ogni $i = 1, \dots, r$.

Perché il caso $X(\Omega) = \{-x_1, x_1\}$ non fornisce un controesempio?

Prima di proseguire è utile generalizzare la relazione (64) che permette di calcolare la varianza della somma di due variabili aleatorie al caso della somma di un numero finito X_1, X_2, \dots, X_n di variabili aleatorie:

$$\text{Var} \left(\sum_{k=1}^n X_k \right) = \sum_{h=1}^n \sum_{k=1}^n \text{Cov}(X_h, X_k) \quad (66)$$

$$= \sum_{k=1}^n \text{Var}(X_k) + \sum_{\substack{1 \leq h, k \leq n \\ h \neq k}} \text{Cov}(X_h, X_k) \quad (67)$$

$$= \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{h=1}^{n-1} \sum_{k=h+1}^n \text{Cov}(X_h, X_k). \quad (68)$$

Basterà mostrare (66), in quanto le (67) e (68), sono solo forme differenti della prima espressione. Tuttavia rimandiamo la dimostrazione della (66), nell'Appendice alla fine di questa Lezione, e preferiamo illustrare prima le sue applicazioni.

Esempio 10.3. Sia X una variabile aleatoria con distribuzione binomiale $b(n, \theta)$. Quanto vale $\text{Var}(X)$?

Soluzione. Conviene ragionare lungo la linea già svolta nella soluzione dell'Esempio 9.3 della precedente lezione, cioè *riguardiamo*⁴⁵ X come la somma di n variabili aleatorie binarie indipendenti X_1, \dots, X_n , ciascuna di valore atteso θ e di varianza $\theta \cdot (1 - \theta)$; dunque $\text{Var}(X) = \text{Var}(\sum_{i=1}^n X_i) = n\theta \cdot (1 - \theta)$, per la (67), in quanto $\text{Cov}(X_h, X_k) = 0$, per $h \neq k$.

Esempio 10.4. Vogliamo ora calcolare la varianza di una variabile aleatoria X con distribuzione ipergeometrica di parametri M, m_1, n . Come nel precedente esempio, possiamo riguardare X come la somma di n variabili aleatorie binarie X_1, \dots, X_n , ciascuna di valore atteso $\frac{m_1}{M}$ e dunque di varianza $\frac{m_1}{M} \cdot (1 - \frac{m_1}{M})$.

Nel presente caso però X_1, \dots, X_n non sono indipendenti; è chiaro che esse, prese a due a due, esse hanno una stessa covarianza, cioè $\text{Cov}(X_h, X_k) = \text{Cov}(X_1, X_2)$ e quindi possiamo scrivere

$$\text{Var}(X) = n \cdot \frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right) + n \cdot (n - 1) \cdot \text{Cov}(X_1, X_2).$$

⁴⁵Riguardare X come la somma di n variabili aleatorie binarie indipendenti (e quindi non correlate: $\text{Cov}(X_i, X_j) = 0$ per $i \neq j$) X_1, \dots, X_n , ciascuna di valore atteso θ e di varianza $\theta \cdot (1 - \theta)$, equivale a considerare che X ha la stessa distribuzione di $S := \sum_{i=1}^n X_i$, e quindi

$$\text{Var}(X) = \text{Var}(S) = \text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{\substack{1 \leq h, k \leq n \\ h \neq k}} \text{Cov}(X_h, X_k) = \sum_{i=1}^n \text{Var}(X_i) = n\theta(1 - \theta).$$

Ora, essendo X_1, X_2 variabili aleatorie binarie, si ha

$$\begin{aligned} \text{Cov}(X_1, X_2) &= P(\{X_1 = 1\} \cap \{X_2 = 1\}) - P(\{X_1 = 1\}) \cdot P(\{X_2 = 1\}) \\ &= P(\{X_1 = 1\}) \cdot P(\{X_2 = 1\} | \{X_1 = 1\}) - \left(\frac{m_1}{M}\right)^2 \\ &= \frac{m_1}{M} \cdot \frac{m_1 - 1}{M - 1} - \left(\frac{m_1}{M}\right)^2 = -\frac{m_1}{M} \cdot \frac{M - m_1}{M(M - 1)} \end{aligned}$$

e possiamo concludere scrivendo

$$\text{Var}(X) = n \cdot \frac{m_1}{M} \cdot \frac{M - m_1}{M} \left(1 - \frac{n - 1}{M - 1}\right) = np(1 - p) \left(1 - \frac{n - 1}{M - 1}\right),$$

dove si è posto $p = \frac{m_1}{M}$, la percentuale di palline bianche presenti nell'urna⁴⁶.

Esempio 10.5. Consideriamo le vincite X_h, X_k associate a due diversi biglietti nella lotteria considerata nell'Esempio 9.5 della precedente Lezione 9. Volendo calcolare $\text{Cov}(X_h, X_k)$, potremmo procedere semplicemente come segue. Innanzitutto, come nel precedente esempio, possiamo osservare che $\text{Cov}(X_h, X_k)$ non dipende dalla coppia di indici h, k (purché $h \neq k$) e da ciò segue

$$\text{Var}(S_n) = n \cdot \text{Var}(X_1) + n(n - 1)\text{Cov}(X_1, X_2);$$

si ha d'altra parte che la distribuzione di probabilità di S_n è degenere; ne segue

$$\text{Var}(S_n) = 0.$$

Possiamo dunque concludere

$$\text{Cov}(X_1, X_2) = -\frac{\text{Var}(X_1)}{n - 1}.$$

Consideriamo ora n variabili aleatorie X_1, \dots, X_n (non necessariamente binarie) e, per semplicità, le assumiamo tali che

$$\mathbb{E}(X_1) = \dots = \mathbb{E}(X_n) = \mu, \quad (69)$$

$$\text{Var}(X_1) = \dots = \text{Var}(X_n) = \sigma^2 \quad (70)$$

$$\text{Cov}(X_h, X_k) = \varphi, \quad 1 \leq h \neq k \leq n. \quad (71)$$

Che cosa possiamo dire circa la loro **media aritmetica**

$$Y_n \equiv \frac{1}{n} \sum_{h=1}^n X_h? \quad (72)$$

A tale proposito è immediato verificare la seguente proposizione (basterà tener presente sia la linearità del valore atteso che le regole viste qui sopra circa il calcolo della varianza).

Proposizione 7. Siano date n variabili aleatorie X_1, \dots, X_n , che verificano (69), (70) e (71). Si ha

$$\mathbb{E}(Y_n) = \mu$$

⁴⁶Si noti il fatto che quindi se il numero totale M delle palline presenti nell'urna è molto grande rispetto al numero n delle estrazioni, allora la varianza del numero di palline estratte (**senza reinserimento**) dall'urna è molto vicina alla varianza del numero delle palline estratte nel caso in cui le estrazioni siano **con reinserimento**.

e

$$\text{Var}(Y_n) = \frac{1}{n} [\sigma^2 + (n-1)\varphi],$$

dove Y_n è definito come sopra in (72).

Osservazione 5. La varianza $\text{Var}(X)$ di una variabile aleatoria X è un indice del grado di dispersione della distribuzione di probabilità rispetto al valore atteso; è chiaro dalla definizione di varianza che, a parità di valore atteso μ , scarti grandi (in modulo) di X rispetto a μ sono tanto più probabili quanto più risulta grande $\text{Var}(X)$.

A tale proposito è utile osservare quanto segue: così come $\mathbb{E}(X)$ corrisponde al concetto di baricentro quando le probabilità $p_j = P(X = x_j)$ vengono interpretate come delle masse concentrate sui diversi punti x_1, \dots, x_n , analogamente $\text{Var}(X)$, nella stessa interpretazione, corrisponde al concetto di **momento di inerzia** della distribuzione stessa.

In collegamento con quanto appena detto, possiamo affermare che la conoscenza del valore atteso e della varianza di una variabile aleatoria fornisce un'idea indicativa, seppur riassuntiva, della sua distribuzione di probabilità e se ne possono ricavare alcune utili disuguaglianze; in particolare si ha la seguente⁴⁷

Proposizione 8. (Diseguaglianza di Chebyshev). Sia X una variabile aleatoria, con valore atteso uguale a μ e varianza uguale a σ^2 . Allora, $\forall \varepsilon > 0$,

$$P(\{|X - \mu| > \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2},$$

o equivalentemente

$$P(\{|X - \mathbb{E}[X]| > \varepsilon\}) \leq \frac{\text{Var}(X)}{\varepsilon^2}, \quad \forall \varepsilon > 0.$$

Dimostrazione.

La dimostrazione segue immediatamente dalla definizione stessa di varianza, infatti possiamo scrivere

$$\begin{aligned} \sigma^2 &= \sum_{\omega_i \in \Omega} p(\omega_i) \cdot (X(\omega_i) - \mu)^2 \geq \sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot (X(\omega_i) - \mu)^2 \geq \\ &\geq \sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot \varepsilon^2 = \varepsilon^2 \cdot \sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) = \varepsilon^2 \cdot P(\{|X - \mu| > \varepsilon\}). \end{aligned}$$

⁴⁷Si osservi che ovviamente nella diseguaglianza di Chebyshev si prende $\varepsilon > 0$ in quanto, per $\varepsilon < 0$ si avrebbe $P(\{|X - \mu| > \varepsilon\}) = 1$, mentre per $\varepsilon = 0$ non avrebbe senso il secondo membro della diseguaglianza. Infine va osservato che la diseguaglianza di Chebyshev è interessante solo se $\frac{\sigma^2}{\varepsilon^2} < 1$: in caso contrario si ottiene solo una banalità (ovvero che $P(\{|X - \mu| > \varepsilon\})$ è minore o uguale di un numero strettamente maggiore di 1, il che è ovvio in quanto $P(\{|X - \mu| > \varepsilon\})$ è un numero minore o uguale ad 1).

Il lettore più attento può notare che la dimostrazione rimane invariata se si considera la probabilità dell'evento $\{|X - \mu| \geq \varepsilon\}$, per cui vale anche

$$P(\{|X - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

La dimostrazione è quindi terminata.

Tuttavia si osservi che nel passaggio $\sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot (X(\omega_i) - \mu)^2 \geq \sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot \varepsilon^2$ c'è il segno di minore uguale e non di minore stretto, in quanto è possibile che $\sum_{\omega_i: |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) = 0$, nel qual caso si ha che i membri della disuguaglianza sono entrambi uguali a zero.

Proposizione 9. Siano date n variabili aleatorie X_1, \dots, X_n ed indichiamo con Y_n la loro media aritmetica. Se X_1, \dots, X_n verificano (69), (70) e (71), ovvero sono tali che

$$\begin{aligned}\mathbb{E}(X_1) &= \dots = \mathbb{E}(X_n) = \mu, \\ \text{Var}(X_1) &= \dots = \text{Var}(X_n) = \sigma^2 \\ \text{Cov}(X_h, X_k) &= 0, \quad 1 \leq h \neq k \leq n,\end{aligned}$$

allora, qualunque sia $\varepsilon > 0$

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2}.$$

Dimostrazione.

La dimostrazione segue immediatamente ricordando le precedenti **Proposizioni 7** ed **8**. Infatti si ha prima di tutto per la disuguaglianza di Chebyshev (**Proposizione 8**) si ha

$$P(\{|Y_n - \mu_{Y_n}| > \varepsilon\}) \leq \frac{\text{Var}(Y_n)}{\varepsilon^2}.$$

Inoltre per la **Proposizione 7** si ha che $\mu_{Y_n} = \mu$ e che $\text{Var}(Y_n) = \frac{1}{n} \sigma^2$.

Osservazione 6 La tesi della **Proposizione 9** si può anche riscrivere come

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2},$$

o equivalentemente

$$P(\{\mu - \varepsilon \leq Y_n \leq \mu + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2}.$$

In altre parole si può dire che l'evento "la media aritmetica Y_n di X_1, X_2, \dots, X_n , differisce dal valore atteso μ (comune a tutte le v.a. X_i) meno di ε " ha probabilità maggiore o uguale a $1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2}$. Se n è "molto grande", in modo che $1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2}$ sia "vicino" ad 1, la tesi si può parafrasare anche dicendo che, se n è "molto grande", media aritmetica e valore atteso differiscono tra loro meno di ε , con probabilità "vicina" ad 1.

Più interessante ancora, dal punto di vista applicativo, è il fatto che siamo in grado di rispondere **alla domanda**:

Quante prove si devono effettuare, ovvero quanto si deve prendere grande n , affinché, con probabilità almeno $1 - \delta$, la media aritmetica differisca dal valore atteso μ meno di ε ?

La risposta alla precedente domanda è molto semplice: **è sufficiente prendere**

$$n \geq \frac{\sigma^2}{\varepsilon^2 \delta},$$

infatti in tale caso $\frac{\sigma^2}{\varepsilon^2 n} \leq \delta$ e quindi $1 - \frac{\sigma^2}{\varepsilon^2 n} \geq 1 - \delta$:

$$P(\{\mu - \varepsilon \leq Y_n \leq \mu + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \geq 1 - \frac{\sigma^2}{\varepsilon^2 n} \geq 1 - \delta,$$

$$\Updownarrow$$

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \leq \delta.$$

Esempio 10.6. Una coppia di dadi perfetti a sei facce viene lanciata n volte ed indichiamo con S_n il numero dei lanci in cui il maggiore fra i due punteggi risulta maggiore o uguale a 5.

Calcolare il minimo valore di n per il quale, in base alla disuguaglianza di Chebyshev, si possa scrivere

$$P\left(\left\{\left|\frac{S_n}{n} - \frac{5}{9}\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{10}$$

Soluzione Per iniziare osserviamo che S_n è la somma di n variabili aleatorie binarie X_i indipendenti ed ugualmente distribuite. Posto Z il valore del primo dado e W il valore del secondo dado (entrambi al primo lancio) si ha

$$\begin{aligned} P(X_1 = 1) &= P(\max(Z, W) \geq 5) = 1 - P(\max(Z, W) < 5) = 1 - P(Z < 5, W < 5) \\ &= 1 - P(Z < 5)P(W < 5) = 1 - P(Z \leq 4)P(W \leq 4) = 1 - \left(\frac{4}{6}\right)^2 = \frac{20}{36} = \frac{5}{9} = \theta. \end{aligned}$$

Quindi in questo caso $\mu = \theta = \frac{5}{9}$, mentre $\sigma^2 = \theta(1 - \theta) = \frac{5}{9} \cdot \frac{4}{9}$, e infine $\varepsilon = \frac{1}{30}$, di conseguenza

$$P\left(\left\{\left|\frac{S_n}{n} - \frac{5}{9}\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} = \frac{1}{n} \frac{\theta(1 - \theta)}{\varepsilon^2} = \frac{1}{n} \frac{\frac{5}{9} \cdot \frac{4}{9}}{\left(\frac{1}{30}\right)^2} = \frac{1}{n} \frac{5 \cdot 4 \cdot 3^2 \cdot 10^2}{9 \cdot 9} = \frac{1}{n} \frac{2000}{9} \leq \frac{1}{10}$$

$$\Updownarrow$$

$$n \geq \frac{20000}{9} \simeq 2222,22 \quad \Leftrightarrow \quad n \geq 2223$$

Definizione 10.4 (Variabili aleatorie standard). Una variabile aleatoria Z si dice **standard** quando

$$\mathbb{E}(Z) = 0, \quad \text{Var}(Z) = 1.$$

Come applicazione della disuguaglianza di Chebyshev, si ha che, se Z è una variabile aleatoria standard allora⁴⁸, $\forall \varepsilon > 0$,

$$P(\{|Z| > \varepsilon\}) \leq \frac{1}{\varepsilon^2}. \tag{73}$$

Data una variabile aleatoria X , con valore atteso μ e varianza σ^2 , possiamo costruire una variabile aleatoria standard Z , funzione⁴⁹ di X , ponendo

$$Z = \frac{X - \mu}{\sigma} \left(= \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \right)$$

⁴⁸Ovviamente la (73) ha interesse solo se $\frac{1}{\varepsilon^2} < 1$, cioè solo se $\varepsilon > 1$.

⁴⁹Si tratta di una funzione affine di X , ovvero $Z = aX + b$, con $(a, b) = \left(\frac{1}{\sigma}, -\frac{\mu}{\sigma}\right)$. Si noti che $\left(\frac{1}{\sigma}, -\frac{\mu}{\sigma}\right)$ è l'unica coppia di valori (a, b) per cui $aX + b$ è una variabile aleatoria standard.

Z viene detta **standardizzata** di X ; con σ si indica la determinazione positiva della radice quadrata di σ^2 , che prende il nome di **scarto standard**. In altre parole si può dire che una variabile aleatoria X , con valore atteso μ e varianza σ^2 si può sempre scrivere nella forma

$$X = \sigma Z + \mu,$$

essendo Z una variabile aleatoria standard; e inoltre la (73) diviene

$$P\left(\left\{\left|\frac{X - \mu}{\sigma}\right| > \varepsilon\right\}\right) \leq \frac{1}{\varepsilon^2}.$$

Consideriamo ora n prove bernoulliane E_1, \dots, E_n di probabilità θ .

Indichiamo con X_1, \dots, X_n gli indicatori di E_1, \dots, E_n e con $S_n = \sum_{j=1}^n X_j$ il numero di successi sulle n prove, e quindi la loro media aritmetica

$$Y_n = \frac{S_n}{n},$$

è la variabile aleatoria con il significato di **frequenza relativa dei successi sulle n prove**.

Relativamente alla loro media aritmetica $Y_n = \frac{S_n}{n}$, abbiamo

Proposizione 10. Per ogni $\varepsilon > 0$

$$P\left(\left\{\left|\sqrt{n} \frac{Y_n - \theta}{\sqrt{\theta(1-\theta)}}\right| > \varepsilon\right\}\right) \leq \frac{1}{\varepsilon^2}.$$

Dimostrazione.

Basta ricordare che $\mathbb{E}(S_n) = n\theta$ e (Esempio 10.3) che $Var(S_n) = n \cdot \theta(1-\theta)$ e dunque

$$\mathbb{E}(Y_n) = \theta, \quad Var(Y_n) = \frac{\theta(1-\theta)}{n}$$

e quindi applicare la (73) alla variabile aleatoria standardizzata di Y_n

$$\sqrt{n} \frac{Y_n - \theta}{\sqrt{\theta(1-\theta)}}.$$

Chiudiamo questo paragrafo notando che l'interesse della precedente **Proposizione 10** risiede nel fatto che, per la variabile standardizzata $\sqrt{n} \frac{Y_n - \theta}{\sqrt{\theta(1-\theta)}}$ della media aritmetica, le probabilità di differire dal valore atteso (cioè zero) più di ε non si può rendere piccola (nemmeno prendendo n grande), come invece accade per la variabile Y_n . Questa proprietà è connessa con il Teorema Centrale del Limite, che verrà discusso più avanti.

10.1 Disuguaglianza di Cauchy e coefficiente di correlazione

Siano date due variabili aleatorie X ed Y , vale allora la seguente disuguaglianza

$$|Cov(X, Y)| \leq \sqrt{Var(X)} \cdot \sqrt{Var(Y)}. \quad (74)$$

Tale diseguaglianza è nota come **diseguaglianza di Cauchy** e generalizza la diseguaglianza di Cauchy-Schwartz per vettori, ovvero, se \mathbf{u}, \mathbf{v} sono vettori di \mathbb{R}^n , allora

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| \left(\sqrt{u_1^2 + u_2^2 + \dots + u_n^2} \cdot \sqrt{v_1^2 + v_2^2 + \dots + v_n^2} \right),$$

dove con $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^n u_i \cdot v_i$ si indica il prodotto scalare tra i vettori \mathbf{u} e \mathbf{v} , e con $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \dots + u_n^2}$ si indica il modulo del vettore \mathbf{u} .

Infatti si consideri il caso particolare in cui $X(\Omega) = \{x_1, x_2, \dots, x_n\}$ ed $Y(\Omega) = \{y_1, y_2, \dots, y_n\}$, e si abbia

$$\begin{aligned} P(X = x_i, Y = y_i) &= \frac{1}{n}, \quad i = 1, 2, \dots, n \\ P(X = x_i, Y = y_j) &= 0, \quad i \neq j, i, j = 1, 2, \dots, n. \end{aligned}$$

Allora, posto $\mu = \mathbb{E}(X)$ e $\nu = \mathbb{E}(Y)$, la diseguaglianza (74) diviene

$$\left| \sum_{i=1}^n (x_i - \mu)(y_i - \nu) \frac{1}{n} \right| \leq \sqrt{\sum_{i=1}^n (x_i - \mu)^2 \frac{1}{n}} \sqrt{\sum_{i=1}^n (y_i - \nu)^2 \frac{1}{n}}$$

che è esattamente la diseguaglianza di Cauchy-Schwartz per i vettori \mathbf{u} e \mathbf{v} con $u_i = x_i - \mu$ e $v_i = y_i - \nu$, per $i = 1, 2, \dots, n$, a parte per il fattore $\frac{1}{n}$.

La dimostrazione della diseguaglianza (74) di Cauchy è basata sull'osservazione che la funzione

$$\varphi(x) := \mathbb{E}[(X - \mu) - x(Y - \nu)]^2$$

gode di due proprietà:

$\varphi(x) \geq 0$ per ogni x , in quanto valore atteso di una variabile aleatoria non negativa,

$\varphi(x) = \text{Var}(X) - 2x \text{Cov}(X, Y) + x^2 \text{Var}(Y)$, come si vede subito per la linearità del valore atteso e considerando che

$$((X - \mu) - x(Y - \nu))^2 = (X - \mu)^2 - 2x(X - \mu)(Y - \nu) + x^2(Y - \nu)^2.$$

Di conseguenza $\varphi(x) = ax^2 + bx + c$, con $a = \text{Var}(Y)$, $b = -2\text{Cov}(X, Y)$ e $c = \text{Var}(X)$, ed il discriminante $b^2 - 4ac = 4(\text{Cov}(X, Y))^2 - 4\text{Var}(X)\text{Var}(Y)$ è minore o uguale a zero, ovvero

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

Per ottenere la (74) basta estrarre la radice quadrata.

Come conseguenza il rapporto

$$\rho_{X,Y} := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

che è detto **coefficiente di correlazione** tra X ed Y è sempre minore o uguale ad 1 in valore assoluto.

Seguendo l'analogia con il caso vettoriale, in qualche senso il coefficiente di correlazione $\rho_{X,Y}$ generalizza il coseno tra due vettori, infatti è noto che, se $\widehat{\mathbf{u}\mathbf{v}}$ indica l'angolo formato fra due vettori,

$$\cos(\widehat{\mathbf{u}\mathbf{v}}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}.$$

Per continuare l'analogia notiamo che se $Y = \alpha X$, allora $\rho_{X,Y} = \pm 1$ a seconda del segno di α , come del resto accade che $\cos(\widehat{\mathbf{u}\mathbf{v}}) = \pm 1$, nel caso in cui $\mathbf{v} = \alpha \mathbf{u}$.

10.2 Appendice: Covarianza della somma di n variabili aleatorie

Se X_1, X_2, \dots, X_n sono n variabili aleatorie, definite sullo stesso spazio di probabilità, allora

$$\begin{aligned} \text{Var} \left(\sum_{k=1}^n X_k \right) &= \sum_{h=1}^n \sum_{k=1}^n \text{Cov}(X_h, X_k) \\ &= \sum_{k=1}^n \text{Var}(X_k) + \sum_{\substack{1 \leq h, k \leq n \\ h \neq k}} \text{Cov}(X_h, X_k) \\ &= \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{h=1}^{n-1} \sum_{k=h+1}^n \text{Cov}(X_h, X_k). \end{aligned}$$

Come già osservato basta mostrare la prima uguaglianza, in quanto le altre due sono solo forme differenti della prima espressione.

Per iniziare si ponga per semplicità $\mathbb{E}(X_i) = \mu_i$, così $\mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mu_i$. Si tratta quindi di calcolare il valore atteso di

$$\begin{aligned} \left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \right)^2 &= \left(\sum_{i=1}^n (X_i - \mu_i) \right)^2 = \left(\sum_{h=1}^n (X_h - \mu_h) \right) \cdot \left(\sum_{k=1}^n (X_k - \mu_k) \right) \\ &= \sum_{h=1}^n \sum_{k=1}^n (X_h - \mu_h)(X_k - \mu_k). \end{aligned}$$

dove l'ultima uguaglianza dipende dal fatto che, come è facile convincersi, in generale

$$\begin{aligned} (a_1 + a_2 + \dots + a_n) \cdot (a_1 + a_2 + \dots + a_n) &= a_1 \cdot a_1 + a_1 \cdot a_2 + \dots + a_1 \cdot a_n \\ &\quad + a_2 \cdot a_1 + a_2 \cdot a_2 + \dots + a_2 \cdot a_n \\ &\quad + \dots + \dots + \dots + \dots \\ &\quad + a_n \cdot a_1 + a_n \cdot a_2 + \dots + a_n \cdot a_n \\ &= \sum_{h=1}^n \sum_{k=1}^n a_h \cdot a_k. \end{aligned}$$

A questo punto basta passare al valore atteso e sfruttarne la proprietà di linearità:

$$\begin{aligned} \text{Var} \left(\sum_{i=1}^n X_i \right) &= \mathbb{E} \left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i \right)^2 \right] \\ &= \mathbb{E} \left[\sum_{h=1}^n \sum_{k=1}^n (X_h - \mu_h)(X_k - \mu_k) \right] \\ &= \sum_{h=1}^n \sum_{k=1}^n \mathbb{E} [(X_h - \mu_h)(X_k - \mu_k)] \\ &= \sum_{h=1}^n \sum_{k=1}^n \text{Cov}(X_h, X_k). \end{aligned}$$

10.3 Esercizi di verifica

Esercizio 10.1. Sia X il punteggio ottenuto nel lancio di un dado a sei facce. Calcolare $Var(X)$.

Esercizio 10.2. Sia X la vincita associata ad un biglietto di una lotteria che, su un totale di 10000 biglietti distribuisce 10 premi da 200 Euro e 20 premi da 100 Euro. Calcolare $Var(X)$.

Esercizio 10.3. Siano Y_1 ed Y_2 i primi due numeri estratti su una ruota del lotto e poniamo $E_1 \equiv \{Y_1 > 45\}$, $E_2 \equiv \{Y_2 < 45\}$. Calcolare la covarianza fra gli indicatori X_1 ed X_2 degli eventi E_1, E_2 .

Esercizio 10.4. Sia S_{100} il numero di elettori per lo schieramento A in un campione casuale (senza reinserimento) di 100 elettori estratti da una popolazione di 1000 elettori di cui m votano per A e $(1000 - m)$ votano per B . Che cosa si ottiene applicando la disuguaglianza di Chebyshev alla variabile aleatoria S_{100} ?

Esercizio 10.5. Il Dipartimento di Matematica acquista 20 copie di un software; ciascuna copia ha probabilità $\frac{1}{100}$ di dare degli errori di funzionamento, indipendentemente dal comportamento delle altre. Indichiamo con S , la variabile aleatoria che conta il numero di copie che danno errori. Scrivete la disuguaglianza che si ottiene applicando ad S la *disuguaglianza di Chebyshev*.

Esercizio 10.6. Siano X, Y due variabili aleatorie standardizzate e consideriamo, $\forall t \in \mathbb{R}$, la variabile aleatoria

$$T_t \equiv (X - tY)^2.$$

a) Calcolare $\mathbb{E}(T_t)$

b) Tenendo conto che deve risultare $\mathbb{E}(T_t) \geq 0, \forall t \in \mathbb{R}$, dimostrare che risulta

$$|Cov(X, Y)| \leq 1.$$

Esercizio 10.7. Sia X una variabile aleatoria con valore atteso μ e varianza σ^2 . Definiamo, per $t \in \mathbb{R}$, la funzione

$$f(t) := \mathbb{E}(X - t)^2.$$

a) Calcolare esplicitamente $f(t)$.

b) Mostrare che μ è il punto di minimo di f e che

$$\sigma^2 = \min_{t \in \mathbb{R}} f(t).$$

In altre parole mostrare che, per ogni $t \in \mathbb{R}$,

$$\mathbb{E}(X - \mu)^2 \leq \mathbb{E}(X - t)^2.$$

11 Campionamento da popolazioni con composizione incognita; indipendenza condizionata

In gran parte di questa lezione concentreremo essenzialmente l'attenzione sulla distribuzione congiunta di una coppia di variabili aleatorie R, S (a valori interi non negativi), nei casi in cui la distribuzione condizionata di S dato R sia binomiale oppure ipergeometrica.

Si tratta quindi di sviluppare un'analisi di casi piuttosto particolari. Tale analisi può risultare interessante per varie applicazioni e per le connessioni con problematiche di tipo statistico; essa permetterà inoltre di illustrare ulteriormente varie nozioni viste nelle precedenti lezioni.

Pur se con qualche modifica nella notazione, verranno innanzitutto richiamati e sviluppati alcuni aspetti ed alcuni esempi, cui si è già accennato in precedenza.

Consideriamo una **popolazione** costituita da un totale di M **elementi**, di cui alcuni di **tipo A** ed altri di **tipo B**.

Qui analizziamo il caso in cui **il numero complessivo di elementi di tipo A** (e quindi anche di tipo B) sia **non noto** e viene visto come una variabile aleatoria, **che indicheremo con il simbolo R** ; ovviamente R sarà dunque una variabile aleatoria a valori in $\{0, 1, \dots, M\}$.

Lo stato di informazione su R viene descritto attraverso la distribuzione di probabilità:

$$p_R(r) \equiv P(\{R = r\}), \quad r = 0, 1, \dots, M. \quad (75)$$

Eseguiamo ora n estrazioni dalla popolazione, registrando il tipo di elemento (A o B) che, man mano, viene estratto ed **indichiamo con S il numero di elementi di tipo A estratti**, o, meglio, risultanti nel **campione** estratto.

Ovviamente S è, in generale, una variabile aleatoria; la distribuzione di probabilità di S , condizionata al valore assunto da R , è data dalle probabilità condizionate

$$p_S(s|\{R = r\}) \equiv P(\{S = s\}|\{R = r\}), \quad s = 0, 1, \dots, n. \quad (76)$$

È chiaro che tale distribuzione condizionata sarà determinata in base alle modalità con cui vengono effettuate le n estrazioni.

Una volta che siano state assegnate la distribuzione marginale della variabile R e le distribuzioni condizionate di S data R , ne risulta univocamente determinata la distribuzione di probabilità congiunta della coppia (R, S) , attraverso la formula⁵⁰

$$p_{R,S}(r, s) \equiv P(\{R = r, S = s\}) = p_R(r) \cdot p_S(s|R = r), \quad 0 \leq s \leq n, \quad 0 \leq r \leq M. \quad (77)$$

A partire da tale distribuzione congiunta possiamo ottenere la distribuzione marginale di S attraverso la formula

$$p_S(s) = \sum_{r=0}^M p_{r,s} = \sum_{r=0}^M p_R(r) \cdot p_S(s|R = r), \quad s = 0, 1, \dots, n. \quad (78)$$

Attraverso l'uso della Formula di Bayes possiamo ora anche ottenere la distribuzione condizionata di R data l'osservazione di un valore s per S :

possiamo scrivere, ponendo $p_R(r|S = s) = P(\{R = r\}|\{S = s\})$,

$$p_R(r|S = s) = \frac{p_R(r) \cdot p_S(s|R = r)}{p_S(s)}, \quad r = 0, 1, \dots, M. \quad (79)$$

⁵⁰Si tratta della solita formula delle probabilità composta:

$$P(\{R = r, S = s\}) = P(\{R = r\})P(\{S = s\}|\{R = r\}).$$

Osservazione 1 (di carattere euristico). Il problema risolto dalla formula (79), cioè quello di calcolare la distribuzione di probabilità condizionata della variabile R (numero complessivo di elementi di tipo A fra gli M elementi della popolazione) data l'osservazione di un valore s per la variabile S (numero di elementi di tipo A fra tutti gli n elementi esaminati) si riallaccia chiaramente ad una problematica di tipo statistico.

Tale problema è legato infatti all'esigenza di ricavare, in merito al numero di elementi di un tipo fissato presenti all'interno di una popolazione, dell'informazione rilevante senza scrutinare tutta la popolazione, ma bensì scrutinandone soltanto una parte. Problemi di tale genere si presentano frequentemente in molti campi applicativi, ad esempio nel controllo di qualità che si deve effettuare su pezzi di una produzione industriale o nelle proiezioni di un risultato elettorale, etc....

Vi sono, nella pratica, vari metodi per formalizzare ed affrontare tali problemi. Il metodo qui esaminato (che in un certo senso è quello più puramente probabilistico) si può riassumere come segue:

* assegnando una distribuzione di probabilità (marginale) ad R , si esprime **lo stato di informazione**, circa il valore che può assumere tale variabile, di cui si dispone **prima di fare il "campionamento"** (cioè prima delle estrazioni degli n elementi da esaminare)

* quindi si assegnano le distribuzioni condizionate di S date le possibili ipotesi sul valore assunto da R ; tali distribuzioni condizionate riflettono **le modalità con cui vengono effettuate le estrazioni** degli n elementi

* in base ai due ingredienti fin qui descritti si ottiene, applicando la Formula di Bayes, la distribuzione condizionata di R dato il valore s osservato per la variabile S ; tale distribuzione condizionata si interpreta come quella distribuzione di probabilità che rappresenta **lo stato di informazione su R , a cui si perviene dopo aver osservato l'evento $\{S = s\}$** . Si suggerisce a tale proposito di tornare all'**Osservazione 2** della Lezione 4.

Le considerazioni svolte nella precedente **Osservazione 1** mettono anche in luce l'interesse di calcolare

$$\mathbb{E}(R|S = s) := \sum_{i=1}^N R(\omega_i) P(\{\omega_i\} | \{S = s\}),$$

cioè il **valore atteso condizionato** di R data l'osservazione di un valore s per S .

Si può dimostrare che

$$\mathbb{E}(R|S = s) = \sum_{r=0}^M r \cdot P(\{R = r\} | \{S = s\}) = \sum_{r=0}^M r \cdot p_R(r | \{S = s\}).$$

Tale quantità non è altro che il valore atteso calcolato rispetto alla distribuzione espressa nella (79) e cioè

$$\mathbb{E}(R|S = s) = \frac{\sum_{r=0}^M r \cdot p_R(r) \cdot p_S(s|R = r)}{p_S(s)}. \quad (80)$$

Consideriamo ora in dettaglio due particolari modalità di estrazioni dalla popolazione:

(i) estrazioni casuali con reinserimento

ed

(ii) estrazioni casuali senza reinserimento.

Questi casi danno luogo a due specifici modelli per quanto riguarda le distribuzioni condizionate per la variabile S dati i valori della variabile R .

Come sappiamo, nel caso **(i)** delle **estrazioni casuali con reinserimento** risulta (si veda a questo proposito l'Esempio 7.3 nella Lezione 7)

$$p_S(s|R = r) = \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}, \quad s = 0, 1, \dots, n. \quad (81)$$

cioè la distribuzione condizionata di S dato ($R = r$) è una distribuzione binomiale di parametri n ed $\frac{r}{M}$.

Nel caso **(ii)** delle **estrazioni casuali senza reinserimento** risulta (si veda ora l'Esempio 7.4 nella Lezione 7)

$$p_S(s|R=r) = \frac{\binom{r}{s} \binom{M-r}{n-s}}{\binom{M}{n}}, \quad \max(0, r+n-M) \leq s \leq \min(r, n), \quad (82)$$

cioè la distribuzione condizionata è una distribuzione ipergeometrica $Hyp(M, r, n)$.

Imponendo rispettivamente tali due condizioni, le precedenti formule (78), (79) e (80) diventano dunque:

(i) estrazioni casuali con reinserimento

$$p_S(s) = \frac{\binom{n}{s}}{M^n} \sum_{r=0}^M p_R(r) \cdot r^s \cdot (M-r)^{n-s} \quad (83)$$

$$p_R(r|S=s) = \frac{p_R(r) \cdot \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}, \quad r = 0, 1, \dots, M \quad (84)$$

e di conseguenza

$$\mathbb{E}(R|S=s) = \frac{\sum_{r=0}^M r \cdot p_R(r) \cdot \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}. \quad (85)$$

(ii) estrazioni casuali senza reinserimento

$$p_S(s) = \sum_{r=s}^{M-n+s} p_R(r) \cdot \frac{\binom{r}{s} \binom{M-r}{n-s}}{\binom{M}{n}} \quad (86)$$

$$p_R(r|S=s) = \frac{p_R(r) \cdot \binom{r}{s} \binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}}, \quad r = s, s+1, \dots, M-n+s, \quad (87)$$

e di conseguenza⁵¹

$$\mathbb{E}(R|S=s) = \frac{\sum_{r=s}^{M-n+s} r \cdot p_R(r) \cdot \binom{r}{s} \binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}}. \quad (88)$$

A questo punto vediamo che $p_S(s)$, $p_R(r|S=s)$ e $\mathbb{E}(R|S=s)$ sono completamente determinati, in modi diversi a seconda che ci si trovi nel caso **(i)** o nel caso **(ii)**, una volta specificata la distribuzione marginale di R .

È interessante ora analizzare in particolare il caso in cui R segue una distribuzione binomiale.

⁵¹Si noti che la condizione che $r = s, s+1, \dots, M-n+s$ in (87), deriva immediatamente dalle condizioni su s in (82), che a loro volta derivano da

$$\begin{cases} 0 \leq s \leq r \\ 0 \leq n-s \leq M-r. \end{cases}$$

Queste ultime, viste come condizioni su r , per s fissato, divengono

$$\begin{cases} r \geq s \\ r \leq M - (n-s) \end{cases}$$

Del resto, nel caso **estrazioni casuali senza reinserimento**, sapere che $\{S = s\}$ equivale ad aver osservato s elementi di tipo A ed $n-s$ elementi di tipo B, e quindi equivale a sapere che il numero totale di elementi di tipo A della popolazione è almeno s , e che il numero totale di elementi di tipo B è almeno $n-s$, e quest'ultima condizione implica che gli elementi di tipo A non possono essere più di $M - (n-s) = M - n + s$.

11.1 Caso estrazioni casuali senza reinserimento e con distribuzione binomiale per R .

Ci sono alcuni aspetti da notare nel caso particolare in cui R segue una distribuzione binomiale e ci si ponga nel caso **(ii)**.

Siccome R può prendere valori nell'insieme $\{0, 1, \dots, M\}$ allora, se R segue una distribuzione binomiale, sarà $R \sim \text{bin}(M, \theta)$ per un qualche valore $0 < \theta < 1$:

$$p_R(r) = \binom{M}{r} \theta^r (1 - \theta)^{M-r}, \quad r = 0, 1, \dots, M.$$

Sostituendo tale espressione nelle (86) e (87) otteniamo quanto segue

$$\begin{aligned} p_S(s) &= \sum_{r=s}^{M-n+s} \binom{M}{r} \theta^r (1 - \theta)^{M-r} \cdot \frac{\binom{r}{s} \binom{M-r}{n-s}}{\binom{M}{n}} \\ &= \sum_{k=0}^{M-n} \binom{M}{s+k} \theta^{s+k} (1 - \theta)^{M-(s+k)} \cdot \frac{\binom{s+k}{s} \binom{M-(s+k)}{n-s}}{\binom{M}{n}} \\ &= \theta^s (1 - \theta)^{n-s} \sum_{k=0}^{M-n} \binom{M}{s+k} \theta^k (1 - \theta)^{M-n-k} \cdot \frac{(k+s)! n! (M-n)! (M-k-s)!}{s! k! M! (n-s)! (M-n-k)!} \\ &= \binom{n}{s} \theta^s (1 - \theta)^{n-s} \sum_{k=0}^{M-n} \binom{M}{s+k} \binom{M-n}{k} \theta^k (1 - \theta)^{M-n-k} \cdot \frac{(k+s)! (M-k-s)!}{M!} \\ &= \binom{n}{s} \theta^s (1 - \theta)^{n-s} \sum_{k=0}^{M-n} \binom{M-n}{k} \theta^k (1 - \theta)^{M-n-k} \\ &= \binom{n}{s} \theta^s (1 - \theta)^{n-s}; \end{aligned}$$

dunque **anche la distribuzione marginale di S è binomiale**, più precisamente $S \sim \text{bin}(n, \theta)$.

Per quanto riguarda la distribuzione condizionata di R data S , abbiamo

$$\begin{aligned} p_R(r|S=s) &= \frac{\binom{M}{r} \theta^r (1 - \theta)^{M-r} \cdot \binom{r}{s} \binom{M-r}{n-s}}{\binom{n}{s} \theta^s (1 - \theta)^{n-s} \cdot \binom{M}{n}} & r = s, s+1, \dots, M - (n-s) \\ &= \theta^{r-s} (1 - \theta)^{M-n-(r-s)} \frac{M! r! (M-r)! s! (n-s)! n! (M-n)!}{r! (M-r)! s! (r-s)! (n-s)! [(M-r) - (n-s)]! n! M!} \\ &= \theta^{r-s} (1 - \theta)^{M-n-(r-s)} \frac{(M-n)!}{(r-s)! [(M-r) - (n-s)]!} \end{aligned}$$

Possiamo concludere dunque scrivendo **la distribuzione condizionata di R data S è individuata da**

$$p_R(r|S=s) = \binom{M-n}{r-s} \cdot \theta^{r-s} (1 - \theta)^{M-n-(r-s)}, \quad r = s, s+1, \dots, M - (n-s). \quad (89)$$

Consideriamo ora la variabile aleatoria

$$T \equiv R - S,$$

che rappresenta **il numero di elementi di tipo A fra gli $(M-n)$ non estratti**.

Ragionando analogamente a quanto si è fatto per S possiamo dedurre che anche T deve avere una distribuzione binomiale, esattamente di parametri $M - n$ e θ . Dalla (89) possiamo poi dedurre⁵² **la distribuzione di probabilità condizionata di T , dato $\{S = s\}$,**

$$\begin{aligned} p_T(t|S = s) &= P(\{R - S = t\}|\{S = s\}) = P(\{R = s + t\}|\{S = s\}) \\ &= \binom{M - n}{t} \cdot \theta^t (1 - \theta)^{M - n - t}, \quad t = 0, 1, \dots, M - n. \end{aligned}$$

Cioè la distribuzione di probabilità condizionata di T , dato $\{S = s\}$, qualunque sia il valore s , è uguale alla distribuzione di probabilità marginale di T e dunque⁵³ S e T sono variabili aleatorie stocasticamente indipendenti!

Osservazione 2 (di carattere euristico). È utile soffermarsi ad illustrare lo specifico significato intuitivo che è possibile rintracciare, riguardando a posteriori quanto abbiamo qui ottenuto. **Si tratta di vedere come si sarebbe potuto arrivare alle stesse conclusioni anche sulla base di un ragionamento intuitivo.**

L'assegnazione della distribuzione binomiale $\text{bin}(M, \theta)$ alla variabile R traduce la seguente condizione: ogni elemento nella popolazione ha probabilità θ di essere di tipo A e probabilità $(1 - \theta)$ di essere di tipo B ; inoltre ogni elemento si comporta in modo indipendente dagli altri.

Estraiamo ora a caso n elementi dalla popolazione; la circostanza che l'estrazione sia casuale permette di asserire che anche ciascuno degli elementi estratti ha probabilità θ di essere di tipo A e che si comporta in modo indipendente dagli altri.

Questa osservazione ci permette subito di concludere (senza fare troppi calcoli) che la distribuzione di S deve essere $\text{bin}(n, \theta)$.

Guardiamo ora alla distribuzione condizionata di R data S . Decomponiamo $R = S + T$, come la somma delle due variabili aleatorie S (numero di elementi di tipo A fra gli n estratti) e $T \equiv R - S$ (numero di elementi di tipo A fra gli $M - n$ non estratti). Per quanto sopra osservato circa il significato della posizione $R \sim \text{bin}(M, \theta)$, possiamo vedere intuitivamente che anche T deve avere una distribuzione binomiale $\text{bin}(M - n, \theta)$ e che T, S debbono essere stocasticamente indipendenti; dunque possiamo scrivere

$$\begin{aligned} p_R(r|S = s) &= P(R = r|S = s) = P(R - S = r - s|S = s) \\ &= P(T = r - s|S = s) = P(T = r - s) \\ &= \binom{M - n}{r - s} \cdot \theta^{r - s} (1 - \theta)^{M - n - (r - s)} \quad r = s, s + 1, \dots, M - (n - s) \end{aligned}$$

e cioè ritrovare la (89).

Consideriamo ora il valore atteso di R dato $\{S = s\}$.

⁵²Il fatto che $P(\{R - S = t\}|\{S = s\}) = P(\{R = s + t\}|\{S = s\})$ è intuitivo, ma può essere dedotto facilmente ragionando come segue:

$$P(\{R - S = t\}|\{S = s\}) = \frac{P(\{R - S = t\} \cap \{S = s\})}{P(\{S = s\})},$$

inoltre $\{R - S = t\} \cap \{S = s\} = \{R - s = t\} \cap \{S = s\}$, e quindi

$$P(\{R - S = t\}|\{S = s\}) = \frac{P(\{R - s = t\} \cap \{S = s\})}{P(\{S = s\})} = P(\{R = s + t\}|\{S = s\})$$

⁵³Il fatto che la distribuzione di probabilità condizionata di T , dato $\{S = s\}$, sia la stessa qualunque sia il valore s , implica l'indipendenza delle variabili aleatorie T ed S , si veda a questo proposito l'Esercizio proposto 8.3, ed in particolare l'equivalenza tra le condizioni (i) e (iii), ivi indicate.

Possiamo scrivere, in virtù dell'indipendenza fra S e T ,

$$\mathbb{E}(T|S = s) = \mathbb{E}(T) = \theta(M - n)$$

e possiamo facilmente concludere, sfruttando la proprietà di linearità del valore atteso,

$$\mathbb{E}(R|S = s) = \mathbb{E}(T|S = s) + \mathbb{E}(S|S = s) = \theta(M - n) + s.$$

Osservazione 3 (ancora di carattere euristico). Abbiamo dunque notato che, se R segue una distribuzione binomiale, allora T, S sono stocasticamente indipendenti. Bisogna tener bene presente che **in generale ciò non accade, se si attribuisce ad R un tipo di distribuzione di probabilità diverso dalla binomiale.**

L'indipendenza stocastica fra S e T esprime il fatto che l'osservazione di S non apporta dell'informazione circa il valore di T . Possiamo dunque concludere che **l'assunzione che la distribuzione marginale di R sia binomiale non è molto realistica nel problema del campionamento**, fin qui illustrato. In altre parole potremo dire che, se si assume che R sia binomiale, allora non è molto utile eseguire un campionamento allo scopo di trarre dell'informazione rilevante, circa il comportamento degli elementi non scrutinati, sulla base del comportamento degli elementi già scrutinati.

Ciò può risultare abbastanza evidente, ad esempio, nel problema del sondaggio elettorale: supponiamo di assumere che ogni elettore (in un gruppo di M elettori) abbia una probabilità fissata, p , di votare per lo schieramento A, indipendentemente dal comportamento degli altri elettori (in tal caso il numero complessivo R di elettori per A ha una distribuzione $\text{bin}(M, p)$). È chiaro intuitivamente che in un tal caso è poco utile eseguire un sondaggio, in quanto la risposta di un elettore non fornisce indicazioni circa il comportamento degli altri.

La situazione più frequente, comunque, è quella in cui non sussiste indipendenza stocastica fra i vari elettori.

A questo punto intervengono aspetti piuttosto delicati circa la condizione di indipendenza stocastica nel caso di estrazioni da una popolazione con caratteristiche non note. Anche se qui non è ne' possibile ne' particolarmente opportuno chiarire completamente tali aspetti, essenzialmente connessi a problematiche di tipo statistico, sarà però utile avviare in proposito un discorso circa il ruolo della nozione di **indipendenza condizionata**. Ciò sarà l'argomento del prossimo sottoparagrafo.

11.2 Esempi

Esaminiamo intanto un esempio per illustrare un caso in cui la distribuzione di R non è binomiale (e quindi in cui non sussiste indipendenza stocastica fra S e T).

Esempio 11.1. *Un'urna contiene 5 oggetti, alcuni di colore Arancio ed altri di colore Blu.*

Il numero di oggetti di colore Arancio non è a noi noto, bensì è una variabile aleatoria R , con distribuzione data da

$$p_0 = \frac{1}{20}, \quad p_1 = \frac{2}{10}, \quad p_2 = \frac{1}{20}, \quad p_3 = 0, \quad p_4 = \frac{5}{10}, \quad p_5 = \frac{2}{10}.$$

Eseguiamo 3 estrazioni casuali senza reinserimento da quest'urna ed indichiamo con S il numero di oggetti di colore Arancio ottenuti in tali estrazioni.

Qual è la distribuzione di probabilità di S ?

Soluzione. Ovviamente risulta:

$$\begin{aligned} P(\{S = 0\}) &= \sum_{m=0}^5 p_m P(\{S = 0\}|\{R = m\}) \\ &= \frac{1}{20} \cdot 1 + \frac{2}{10} \cdot \frac{2}{5} + \frac{1}{20} \cdot \frac{1}{10} = \frac{27}{200}; \end{aligned}$$

analogamente

$$\begin{aligned} P(\{S = 1\}) &= \frac{2}{10} \cdot \frac{6}{10} + \frac{1}{20} \cdot \frac{6}{10} = \frac{30}{200}; \\ P(\{S = 2\}) &= \frac{1}{20} \cdot \frac{3}{10} + \frac{5}{10} \cdot \frac{6}{10} = \frac{63}{200}; \\ P(\{S = 3\}) &= \frac{5}{10} \cdot \frac{4}{10} + \frac{2}{10} \cdot 1 = \frac{80}{200}. \end{aligned}$$

Esempio 11.2. Nelle condizioni del precedente *Esempio 11.1*, qual è la distribuzione condizionata di R data l'osservazione $\{S = 2\}$?

Soluzione. Risulta ovviamente

$$P(\{R = 0\}|\{S = 2\}) = P(\{R = 1\}|\{S = 2\}) = P(\{R = 5\}|\{S = 2\}) = 0;$$

inoltre, essendo $P(\{R = 3\}) = 0$ si ha⁵⁴

$$P(\{R = 3\}|\{S = 2\}) = 0;$$

infine

$$\begin{aligned} P(\{R = 2\}|\{S = 2\}) &= \frac{\frac{1}{20} \cdot \frac{3}{10}}{\frac{63}{200}} = \frac{3}{63}, \\ P(\{R = 4\}|\{S = 2\}) &= \frac{\frac{5}{10} \cdot \frac{6}{10}}{\frac{63}{200}} = \frac{60}{63}. \end{aligned}$$

È facile verificare che, in questo caso, le due variabili S e $T = R - S$ non possono essere stocasticamente indipendenti.

Esempio 11.3. Continuando sempre a considerare l'urna dell'*Esempio 11.1*, esaminiamo ora il caso in cui **le estrazioni siano con reinserimento**.

Qual è la distribuzione di probabilità di S ?

Soluzione. Si ha

$$P(\{S = k\}) = \sum_{m=0}^5 p_m \binom{3}{k} \left(\frac{m}{5}\right)^k \left(\frac{5-m}{5}\right)^{3-k}.$$

⁵⁴Si osservi che $P(\{R = 3\}|\{S = 2\}) = 0$, in quanto per ipotesi $P(\{R = 3\}) = 0$ e quindi

$$0 \leq P(\{R = 3\} \cap \{S = 2\}) \leq P(\{R = 3\}) = 0.$$

Di conseguenza

$$P(\{R = 3\}|\{S = 2\}) = \frac{P(\{R = 3\} \cap \{S = 2\})}{P(\{S = 2\})} = 0.$$

In particolare⁵⁵

$$P(\{S = 2\}) = \frac{1}{125} \left(0 + \frac{2}{10} \cdot 12 + \frac{1}{20} \cdot 36 + \frac{5}{10} \cdot 48 \right)$$

Esempio 11.4. Nelle stesse condizioni del precedente Esempio 11.4 qual è in questo caso la distribuzione condizionata di R data l'osservazione $\{S = 2\}$?

Soluzione. Si debbono ovviamente ancora escludere i casi $\{R = 0\}$, $\{R = 5\}$, nel senso che le rispettive probabilità condizionate sono nulle, e si ha⁵⁶

$$P(\{R = 1\}|\{S = 2\}) = \frac{P(\{R = 1\})P(\{S = 2\}|\{R = 1\})}{P(\{S = 2\})} = \frac{4}{47}$$

analogamente

$$P(\{R = 2\}|\{S = 2\}) = \frac{3}{47}, \quad P(\{R = 4\}|\{S = 2\}) = \frac{40}{47}$$

Di nuovo $P(\{R = 3\}|\{S = 2\}) = 0$, in quanto per ipotesi $P(\{R = 3\}) = 0$.

11.3 Indipendenza condizionata

Cominciamo questo sottoparagrafo insistendo su nozioni che dovrebbero essere ormai chiare, per passare subito dopo a sottolineare aspetti critici, relativi al caso di estrazioni casuali da una popolazione con composizione aleatoria.

Consideriamo allora di nuovo il caso di n estrazioni casuali da una popolazione che contiene oggetti di due tipi, ad esempio A e B.

⁵⁵Si consideri che per $k = 2$, si ha $3 - k = 1$ e quindi, per $m = 0$ si ha $\binom{0}{5} = 0$, che, per $m = 5$ si ha $\binom{5-5}{5} = 0$, e che $p_3 = 0$. Di conseguenza

$$\begin{aligned} P(\{S = k\}) &= p_1 \binom{3}{2} \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^1 + p_2 \binom{3}{2} \left(\frac{2}{5}\right)^2 \left(\frac{3}{5}\right)^1 + p_4 \binom{3}{2} \left(\frac{4}{5}\right)^2 \left(\frac{1}{5}\right)^1 \\ &= \frac{3}{125} \left(\frac{2}{10} \cdot 4 + \frac{1}{20} \cdot 2^2 \cdot 3 + \frac{5}{10} \cdot 4^2 \right) = \frac{1}{125} \left(\frac{2}{10} \cdot 12 + \frac{1}{20} \cdot 36 + \frac{5}{10} \cdot 48 \right) \\ &= \frac{3}{125} \left(\frac{1}{5} \cdot 4 + \frac{1}{5} \cdot 3 + \frac{5}{5} \cdot 8 \right) = \frac{3}{625} (4 + 3 + 40) \\ &= \frac{3}{625} \cdot 47. \end{aligned}$$

⁵⁶Infatti

$$\begin{aligned} P(\{R = 1\}|\{S = 2\}) &= \frac{P(\{R = 1\})P(\{S = 2\}|\{R = 1\})}{P(\{S = 2\})} \\ &= \frac{\frac{3}{625} \cdot 4}{\frac{3}{625} \cdot (4 + 3 + 40)} = \frac{4}{47}; \\ P(\{R = 2\}|\{S = 2\}) &= \frac{P(\{R = 2\})P(\{S = 2\}|\{R = 1\})}{P(\{S = 2\})} \\ &= \frac{\frac{3}{625} \cdot 3}{\frac{3}{625} \cdot (4 + 3 + 40)} = \frac{3}{47}; \\ P(\{R = 4\}|\{S = 2\}) &= \frac{P(\{R = 4\})P(\{S = 2\}|\{R = 1\})}{P(\{S = 2\})} \\ &= \frac{\frac{3}{625} \cdot 40}{\frac{3}{625} \cdot (4 + 3 + 40)} = \frac{40}{47}; \end{aligned}$$

Poniamo

$$E_j \equiv \{\text{oggetto di tipo A alla } j\text{-esima estrazione}\}, \quad j = 1, 2, \dots, n;$$

poniamo anche

$$S \equiv \sum_{j=1}^n \mathbf{1}_{E_j}$$

Supponiamo di sapere che la popolazione contiene r elementi di tipo A e $(M - r)$ elementi di tipo B.

In tale caso, **sotto la condizione che le estrazioni siano senza reinserimento**, c'è **dependenza stocastica** fra E_1, \dots, E_n . Inoltre, S ha **distribuzione ipergeometrica** $Hyp(M, r, n)$.

Sotto la condizione che le estrazioni siano con reinserimento, E_1, \dots, E_n sono **eventi stocasticamente indipendenti**, di probabilità $\frac{r}{M}$, e S ha invece **distribuzione binomiale** $bin(n, \frac{r}{M})$.

Ora dobbiamo sottolineare quanto segue.

A proposito di questo ultimo caso (di estrazioni con reinserimento), si deve fare attenzione al fatto che l'indipendenza stocastica fra E_1, \dots, E_n sussiste in virtù della concomitanza fra due diverse circostanze:

(a) le **estrazioni** (casuali) sono con reinserimento

(b) **conosciamo la composizione della popolazione** da cui si effettua il campionamento (cioè sono note le proporzioni $\frac{r}{M}$ e $\frac{M-r}{M}$ degli elementi di tipo rispettivamente A e B).

Vedremo infatti qui di seguito, iniziando con un semplice esempio, che, fermo restando la condizione (a) di estrazioni con reinserimento, **non vi può essere in generale indipendenza stocastica fra** E_1, \dots, E_n **se viene a mancare la condizione (b)**.

Notiamo d'altra parte che, come abbiamo visto prima, la motivazione per effettuare un campionamento è data proprio dall'esigenza di ricavare informazioni circa la composizione di una popolazione; è dunque realistico pensare che, se si effettua il campionamento, il numero di elementi del tipo, ad esempio, A sia una variabile aleatoria R , piuttosto che un valore noto r ; ed in tale caso, ripetiamo, pur se le estrazioni casuali sono effettuate con reinserimento, non vi è in generale indipendenza stocastica fra E_1, \dots, E_n .

Prima di proseguire vediamo infatti il seguente esempio illustrativo a cui avevamo accennato poco fa.

Esempio 11.5. *Due urne, U_1 e U_2 , contengono 10 palline ciascuna. U_1 contiene 3 palline verdi e 7 blu, mentre U_2 ne contiene 3 blu e 7 verdi. Ci viene data a caso una delle due urne (non sappiamo quale) e da tale urna eseguiamo due successive estrazioni (ciò significa che stiamo facendo delle estrazioni da una popolazione in cui il numero delle palline verdi è una variabile aleatoria R che può assumere, con uguale probabilità, il valore 3 oppure il valore 7). Poniamo*

$$E_j \equiv \{\text{pallina verde alla } j\text{-esima estrazione}\}, \quad j = 1, 2$$

Vogliamo calcolare $P(E_1)$, $P(E_2)$ e $P(E_2|E_1)$, confrontandoli tra loro. A tale scopo osserviamo che vi sono due ipotesi alternative:

$$H_1 \equiv \{\text{abbiamo eseguito le estrazioni da } U_1\}$$

$$H_2 \equiv \{\text{abbiamo eseguito le estrazioni da } U_2\};$$

certamente una di queste due ipotesi è vera, ma non sappiamo quale ed attribuiamo le probabilità (visto che l'urna è stata scelta "a caso")

$$P(H_1) = P(H_2) = \frac{1}{2}.$$

Abbiamo intanto, applicando la definizione di probabilità condizionata e poi la formula delle probabilità totali:

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{P(H_1)P(E_1 \cap E_2|H_1) + P(H_2)P(E_1 \cap E_2|H_2)}{P(H_1)P(E_1|H_1) + P(H_2)P(E_1|H_2)}$$

e quindi in questo caso specifico

$$= \frac{\frac{1}{2}P(E_1 \cap E_2|H_1) + \frac{1}{2}P(E_1 \cap E_2|H_2)}{\frac{1}{2}P(E_1|H_1) + \frac{1}{2}P(E_1|H_2)} = \frac{P(E_1 \cap E_2|H_1) + P(E_1 \cap E_2|H_2)}{P(E_1|H_1) + P(E_1|H_2)}.$$

Come già accennato vogliamo analizzare specificamente il caso in cui le estrazioni siano casuali e con reinserimento.

In tal caso si ha che, sotto l'ipotesi di eseguire le estrazioni da H_1 , la probabilità di ottenere una pallina verde in una singola estrazione è uguale a $\frac{3}{10}$ e, visto che le estrazioni avvengono con reinserimento da un'urna contenente 3 palline verdi e 7 blu, la probabilità di ottenere due volte pallina verde in due successive estrazione è uguale a $(\frac{3}{10})^2$; possiamo scrivere in formule

$$P(E_1|H_1) = P(E_2|H_1) = \frac{3}{10}, \quad P(E_1 \cap E_2|H_1) = P(E_1|H_1)P(E_2|H_1) = \left(\frac{3}{10}\right)^2.$$

Analogamente, per quanto riguarda il condizionamento all'ipotesi H_2 , possiamo scrivere

$$P(E_1|H_2) = P(E_2|H_2) = \frac{7}{10}, \quad P(E_1 \cap E_2|H_2) = P(E_1|H_2)P(E_2|H_2) = \left(\frac{7}{10}\right)^2.$$

Possiamo concludere quindi

$$P(E_1) = P(H_1)P(E_1|H_1) + P(H_2)P(E_1|H_2) = \frac{1}{2} \left(\frac{3}{10} + \frac{7}{10} \right) = \frac{1}{2}$$

e ovviamente

$$P(E_2) \left(= P(H_1)P(E_2|H_1) + P(H_2)P(E_2|H_2) \right) = P(E_1) = \frac{1}{2},$$

mentre

$$P(E_2|E_1) = \frac{\left(\frac{3}{10}\right)^2 + \left(\frac{7}{10}\right)^2}{\frac{3}{10} + \frac{7}{10}} = \frac{58}{100}.$$

Dunque

$$P(E_2|E_1) > P(E_2)$$

da cui vediamo che E_1 , E_2 **non sono stocasticamente indipendenti**, bensì **positivamente correlati**.

Esercizio proposto 11.1. Calcolare la probabilità condizionata $P(E_1|E_2)$ nel caso di estrazioni casuali dall'urna considerata nel precedente Esempio 11.5.

Esercizio proposto 11.2. Calcolare la probabilità condizionata $P(E_2|E_1)$ e $P(E_1|E_2)$ nel caso di estrazioni casuali dall'urna considerata nel precedente Esempio 11.3, in cui R è aleatorio

Generalizziamo ora quanto visto nell'Esempio 11.5 e nei precedenti Esercizi proposti 11.1 e 11.2.

Una popolazione contiene M oggetti, alcuni di tipo A ed altri di tipo B. Supponiamo che il numero complessivo di quelli di tipo A sia una variabile aleatoria R , con distribuzione di probabilità data da

$$p_0 = P(\{R = 0\}), \quad p_1 = P(\{R = 1\}), \quad \dots, \quad p_M = P(\{R = M\})$$

con $\sum_{r=0}^M p_r = 1$. Eseguiamo delle estrazioni casuali con reinserimento dalla popolazione e poniamo

$$E_j \equiv \{\text{oggetto di tipo A alla } j\text{-esima estrazione}\}, \quad j = 1, 2, \dots$$

Vogliamo calcolare $P(E_2|E_1)$. Estendendo quanto svolto nel precedente esempio, abbiamo

$$\begin{aligned} P(E_1) &= P(E_2) = \sum_{r=0}^M P(E_i|R=r) \cdot p_r = \sum_{r=0}^M \frac{r}{M} \cdot p_r = \frac{\mathbb{E}(R)}{M} \\ P(E_2|E_1) &= \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{r=0}^M P(E_1 \cap E_2|R=r) \cdot p_r}{\sum_{r=0}^M P(E_1|R=r) \cdot p_r} \\ &= \frac{\sum_{r=0}^M \left(\frac{r}{M}\right)^2 \cdot p_r}{\sum_{r=0}^M \frac{r}{M} \cdot p_r} = \frac{1}{M} \frac{\sum_{r=0}^M r^2 \cdot p_r}{\sum_{r=0}^M r \cdot p_r} = \frac{1}{M} \frac{\mathbb{E}(R^2)}{\mathbb{E}(R)}. \end{aligned}$$

È facile verificare che E_1, E_2 sono stocasticamente indipendenti se e solo se R è una variabile degenerata⁵⁷ con $R(\omega_i) = \hat{r} \in \{0, 1, \dots, M\}$, per ogni $\omega_i \in \Omega$, nel qual caso $\text{Var}(R) = 0$ e, quindi, ricordando la **Proposizione 1** della Lezione 10, $\mathbb{E}(R^2) = (\mathbb{E}(R))^2 = \hat{r}^2$.

Quanto svolto fin qui suggerisce le seguenti definizioni

Definizione 11.1 (indipendenza condizionata rispetto ad un evento e rispetto ad una partizione). Siano E_1, E_2, H tre eventi. Diremo che E_1 ed E_2 **sono condizionatamente indipendenti dato l'evento H** se risulta

$$P(E_1 \cap E_2|H) = P(E_1|H) \cdot P(E_2|H).$$

Sia ora $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ una partizione dell'evento certo. Diremo che E_1 ed E_2 **sono condizionatamente indipendenti data la partizione \mathcal{H}** se risulta

$$P(E_1 \cap E_2|H_k) = P(E_1|H_k) \cdot P(E_2|H_k), \quad \text{per ogni } k = 1, 2, \dots, m.$$

Definizione 11.2 (indipendenza condizionata rispetto ad una variabile aleatoria). Siano E_1, E_2 due eventi ed Z una variabile aleatoria. Diremo che E_1 ed E_2 **sono condizionatamente indipendenti data Z** se risulta

$$P(E_1 \cap E_2|\{Z = z\}) = P(E_1|\{Z = z\}) \cdot P(E_2|\{Z = z\}), \quad \text{per ogni } z \in Z(\Omega) = \{z_1, z_2, \dots, z_\ell\}.$$

Si osservi che la precedente definizione equivale all'indipendenza condizionata di E_1 ed E_2 data la partizione $\mathcal{H} = \{\{Z = z_k\}, k = 1, 2, \dots, \ell\}$.

⁵⁷Infatti $P(E_2) = P(E_2|E_1)$ se e solo se $\frac{\mathbb{E}(R)}{M} = \frac{1}{M} \frac{\mathbb{E}(R^2)}{\mathbb{E}(R)}$, cioè se e solo se $(\mathbb{E}(R))^2 = \mathbb{E}(R^2)$, ovvero se e solo se $\text{Var}(R) = 0$.

Definizione 11.3 (indipendenza condizionata di due v.a. rispetto ad una v.a.). *Siano X, Y, Z tre variabili aleatorie. Diremo che X ed Y sono condizionatamente indipendenti data Z se risulta*

$$P(\{X = x\} \cap \{Y = y\} | \{Z = z\}) = P(X = x | Z = z) \cdot P(Y = y | Z = z),$$

per ogni $x \in X(\Omega) = \{x_i, i = 1, 2, \dots, n\}$, per ogni $y \in Y(\Omega) = \{y_j, j = 1, 2, \dots, m\}$, per ogni $z \in Z(\Omega) = \{z_k, k = 1, 2, \dots, \ell\}$.

Coppie di eventi condizionatamente indipendenti e coppie di variabili aleatorie condizionatamente indipendenti si incontrano comunemente in varie problematiche, in particolare nelle situazioni di tipo statistico. Ciò accade anche in situazioni al di fuori degli schemi di estrazioni casuali (con reinserimento) da popolazioni con composizione aleatoria. Guardiamo in proposito il seguente

Esempio 11.6. *Vi sono a disposizione tre diversi canali di comunicazione, C_1, C_2 e C_3 , per spedire dei messaggi. Ogni messaggio può essere spedito da ciascuno dei tre canali. La probabilità di trasmettere il messaggio correttamente tramite C_1 è uguale a $p^{(1)} = 0.9$. Le analoghe probabilità per C_2 e C_3 sono rispettivamente date da $p^{(2)} = 0.6$ e $p^{(3)} = 0.3$. Supponiamo ora che il canale venga scelto a caso da un meccanismo e **non sia noto** all'operatore. Questi spedisce il messaggio due volte consecutive (sempre sullo stesso canale, che gli è stato riservato per quel messaggio) al fine di aumentare l'affidabilità della trasmissione.*

a) *Trovare la probabilità che il messaggio sia trasmesso correttamente in almeno una delle due volte.*

b) *Si supponga di sapere in seguito che il messaggio è stato trasmesso correttamente la seconda volta, ma non la prima volta. Condizionatamente a questa osservazione, come bisogna valutare le probabilità che sia stato utilizzato il canale C_1, C_2 e C_3 , rispettivamente?*

Soluzione. Poniamo

$$\begin{aligned} E_i &\equiv \{\text{messaggio trasmesso correttamente nell}'i\text{-esimo tentativo}\}, & i &= 1, 2 \\ H_j &\equiv \{\text{è stato assegnato il canale } C_j\}, & j &= 1, 2, 3. \end{aligned}$$

a) Dobbiamo calcolare $P(E_1 \cup E_2)$. Applicando la formula delle probabilità totali possiamo scrivere

$$P(E_1 \cup E_2) = \sum_{j=1}^3 P(E_1 \cup E_2 | H_j) \cdot P(H_j)$$

Visto che il canale si assume scelto a caso possiamo porre

$$P(H_j) = \frac{1}{3}, \quad j = 1, 2, 3.$$

Ora possiamo osservare che gli eventi E_1, E_2 **non sono indipendenti** bensì **sono condizionatamente indipendenti**, dati gli eventi della partizione $\{H_1, H_2, H_3\}$; quindi

$$\begin{aligned} P(E_1 \cup E_2 | H_j) &= P(E_1 | H_j) + P(E_2 | H_j) - P(E_1 \cap E_2 | H_j) = \\ &= P(E_1 | H_j) + P(E_2 | H_j) - P(E_1 | H_j) \cdot P(E_2 | H_j) = p^{(j)} (2 - p^{(j)}). \end{aligned}$$

Dunque

$$P(E_1 \cup E_2) = \frac{1}{3} [0.9 \times 1.1 + 0.6 \times 1.4 + 0.3 \times 1.7]$$

b) Dobbiamo calcolare $P(H_j|\bar{E}_1 \cap E_2)$. Si ha, ancora in virtù della condizione di indipendenza condizionata

$$\begin{aligned} P(\bar{E}_1 \cap E_2) &= \sum_{j=1}^3 P(\bar{E}_1 \cap E_2|H_j) \cdot P(H_j) = \sum_{j=1}^3 P(\bar{E}_1|H_j) \cdot P(E_2|H_j)P(H_j) \\ &= (0.1 \times 0.91 + 0.4 \times 0.6 + 0.3 \times 0.7) \frac{1}{3} = \frac{1}{3} 0.54 = 0.18; \end{aligned}$$

utilizzando la Formula di Bayes abbiamo dunque

$$P(H_j|\bar{E}_1 \cap E_2) = \frac{P(H_j)P(\bar{E}_1 \cap E_2|H_j)}{P(\bar{E}_1 \cap E_2)} = \frac{\frac{1}{3}P(\bar{E}_1|H_j) \cdot P(E_2|H_j)}{0.18} = \frac{(1-p^{(j)})p^{(j)}}{0.54}.$$

e quindi

$$P(H_1|\bar{E}_1 \cap E_2) = \frac{9}{54} = \frac{3}{18}, \quad P(H_2|\bar{E}_1 \cap E_2) = \frac{24}{54} = \frac{8}{18}, \quad P(H_3|\bar{E}_1 \cap E_2) = \frac{21}{54} = \frac{7}{18}.$$

Una proprietà fondamentale della nozione di indipendenza condizionata è mostrata dalla seguente **Proposizione 1**; la dimostrazione è lasciata al lettore per esercizio⁵⁸.

Proposizione 1. Siano E_1, E_2 due eventi condizionatamente indipendenti data una variabile aleatoria Z con $Z(\Omega) = \{z_1, z_2, \dots, z_n\}$. Allora

$$P(E_2|E_1 \cap \{Z = z_j\}) = P(E_2|\{Z = z_j\})$$

$$P(E_2|E_1) = \sum_{j=1}^n P(E_2|\{Z = z_j\})P(\{Z = z_j\}|E_1).$$

⁵⁸Diamo come suggerimento per la dimostrazione i seguenti elementi:

$$P(E_2|E_1 \cap \{Z = z_j\}) = \frac{P(E_2 \cap E_1 \cap \{Z = z_j\})}{P(E_1 \cap \{Z = z_j\})} = \frac{P(E_2 \cap E_1|\{Z = z_j\})P(\{Z = z_j\})}{P(E_1|\{Z = z_j\})P(\{Z = z_j\})}$$

per l'ipotesi di indipendenza condizionata

$$= \frac{P(E_2|\{Z = z_j\})P(E_1|\{Z = z_j\})P(\{Z = z_j\})}{P(E_1|\{Z = z_j\})P(\{Z = z_j\})}$$

Inoltre

$$\begin{aligned} P(E_2|E_1) &= \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{j=1}^n P(E_1|\{Z = z_j\})P(E_2|\{Z = z_j\})P(\{Z = z_j\})}{\sum_{i=1}^n P(E_1|\{Z = z_i\})P(\{Z = z_i\})} \\ &= \sum_{j=1}^n P(E_2|\{Z = z_j\}) \frac{P(E_1|\{Z = z_j\})P(\{Z = z_j\})}{\sum_{i=1}^n P(E_1|\{Z = z_i\})P(\{Z = z_i\})}, \end{aligned}$$

e infine

$$P(\{Z = z_j\}|E_1) = \frac{P(E_1|\{Z = z_j\})P(\{Z = z_j\})}{\sum_{i=1}^n P(E_1|\{Z = z_i\})P(\{Z = z_i\})}.$$

11.4 Esercizi di verifica

Esercizio 11.1. Due urne, U_1 e U_2 , contengono 10 palline ciascuna. U_1 contiene 10 palline blu, mentre U_2 ne contiene 5 blu e 5 verdi. Viene scelta a caso una delle due urne (non sappiamo quale) e da tale urna eseguiamo due successive estrazioni casuali con reinserimento, ottenendo ciascuna volta pallina blu.

Condizionatamente a tale evento, qual è la probabilità che sia stata scelta l'urna U_1 ?

Esercizio 11.2. Una pianta produce R semi, dove R è una variabile aleatoria binomiale con parametri n e p . Supponiamo inoltre che ciascun seme, fra gli R prodotti, germogli con probabilità θ , indipendentemente dagli altri. Sia S il numero dei germogli risultanti.

(a) Calcolare $P(S = j | R = i)$ e $P(S = j, R = i)$.

(b) Calcolare $P(S = j)$.

(c) Calcolare $P(R = i | S = j)$.

(d) Calcolare $\mathbb{E}(R | S = j)$.

Esercizio 11.3. Il testo di esame scritto consiste di quattro esercizi. Ogni esercizio può contenere un errore con probabilità 0.1, indipendentemente dagli altri. Supponiamo che, dopo aver redatto il testo, ciascun esercizio venga ricontrollato e che la presenza di un eventuale errore sia rilevata con probabilità 0.8. Gli errori inizialmente presenti e poi rilevati vengono corretti.

(a) Determinare la probabilità che, dopo il controllo, non vi siano più esercizi contenenti errori

(b) Condizionatamente al fatto che non vi siano più esercizi contenenti errori dopo il controllo, qual è la probabilità che non vi fossero errori neanche prima del controllo?

Esercizio 11.4. Indichiamo con R il numero di pezzi difettosi un lotto di 20 pezzi e facciamo la seguente valutazione di probabilità:

$$P(R = 0) = 0.5, \quad P(R = 10) = 0.4, \quad P(R = 20) = 0.1.$$

Qual è la probabilità di avere entrambi i pezzi difettosi, scegliendo a caso (senza reinserimento) due pezzi dal lotto?

Esercizio 11.5. Si hanno m esemplari di un certo tipo di telecomando (TC) per televisore; ciascun TC ha bisogno di due batterie per il suo funzionamento. Si hanno a disposizione $2m$ batterie, di cui alcune possono essere scariche. Da tale gruppo di batterie vengono costituite in modo casuale m coppie, che vengono inserite negli m TC.

Indichiamo con R il numero complessivo delle batterie cariche; supponiamo che ciascuna batteria sia carica con probabilità $\frac{1}{2}$ ed indipendentemente da quello che accade alle altre.

(a) Calcolare la probabilità che un fissato TC abbia entrambe le batterie cariche.

(b) Sia S il numero complessivo dei TC con entrambe le batterie cariche. Calcolare $\mathbb{E}(S)$.

Esercizio 11.6. Abbiamo un dado di cui non siamo certi se è regolare oppure se è truccato. Truccato qui significa che esso fornisce il risultato “sei” con probabilità doppia rispetto a quella di tutti gli altri risultati. Attribuiamo probabilità 0.9 all'ipotesi che esso sia regolare e probabilità 0.1 all'ipotesi che invece sia truccato. Viene eseguito un lancio del dado e si ottiene il risultato “sei”.

(a) Condizionatamente a questo risultato, qual è la probabilità che il dado sia truccato?

(b) Condizionatamente a questo risultato, qual è la probabilità che si ottenga il risultato “sei” anche in un secondo lancio?

Esercizio 11.7. Una popolazione, composta da 100 elementi, contiene R elementi di tipo A e $100 - R$ elementi di tipo B, R essendo una variabile con una distribuzione binomiale $\text{bin}(100, \frac{3}{4})$.

Da tale popolazione si eseguono cinquanta estrazioni casuali senza reinserimento ed indichiamo con

S il numero di elementi di tipo A fra quelli estratti.

(a) Trovare la distribuzione di probabilità di S

(b) Trovare la distribuzione di probabilità condizionata di R dato $\{S = 40\}$

(c) Calcolare il valore atteso condizionato di R dato $\{S = 40\}$

(d) Calcolare la varianza della distribuzione condizionata di R dato $\{S = 40\}$ e, pensando ad R come la somma di $S + (R - S)$ calcolare la minorazione fornita dalla disuguaglianza di Chebychev per $P(\{45 \leq R \leq 65\}|\{S = 40\})$

Esercizio 11.8. Una popolazione, composta da 9 elementi, contiene R elementi di tipo A e $9 - R$ elementi di tipo B, R essendo una variabile con una distribuzione uniforme su $\{0, 1, 2, \dots, 9\}$:

$$P(\{R = r\}) = \frac{1}{10}, \quad r = 0, 1, \dots, 9.$$

Da tale popolazione si eseguono quattro estrazioni casuali senza reinserimento ed indichiamo con S il numero di elementi di tipo A fra i quattro estratti.

(a) Trovare la distribuzione di probabilità di S

(b) Trovare la distribuzione di probabilità condizionata di R dato $\{S = 1\}$

(c) Calcolare il valore atteso condizionato di R dato $\{S = 1\}$.

12 Modelli di occupazione e schemi di estrazioni da urne

Consideriamo l'insieme $A_{n,r}$ ($n, r \in \mathbb{N}$) costituito dalle n -uple ordinate $\mathbf{x} \equiv (x_1, \dots, x_n)$, con $x_j \geq 0$, interi, e tali che $\sum_{j=1}^n x_j = r$, in simboli

$$A_{n,r} = \left\{ \mathbf{x} \equiv (x_1, \dots, x_n) : \sum_{j=1}^n x_j = r, x_i \in \mathbb{N} \cup \{0\}, i = 1, 2, \dots, n \right\}.$$

Ad esempio per $n = 3$ ed $r = 4$,

$$A_{3,4} = \left\{ (4, 0, 0), (0, 4, 0), (0, 0, 4), \right. \\ (3, 1, 0), (3, 0, 1), (1, 3, 0), (1, 0, 3), (0, 3, 1), (0, 1, 3), \\ (2, 1, 1), (1, 2, 1), (1, 1, 2), \\ \left. (2, 2, 0), (2, 0, 2), (0, 2, 2) \right\}$$

Qui vogliamo illustrare il significato delle distribuzioni di probabilità su $A_{n,r}$ ed esaminare alcuni casi notevoli.

Vediamo subito una naturale interpretazione di una distribuzione di probabilità su $A_{n,r}$: siano dati r **soggetti** O_1, \dots, O_r ed n **diversi siti** $1, \dots, n$.

Supponiamo che gli r soggetti si dispongano negli n siti in modo aleatorio⁵⁹ e consideriamo per $i = 1, 2, \dots, r$ e $j = 1, 2, \dots, n$ le variabili aleatorie binarie:

$$B_i^{(j)} = \begin{cases} 1 & \text{se il soggetto } O_i \text{ cade nel sito } j \\ 0 & \text{altrimenti} \end{cases}.$$

Consideriamo ora le variabili aleatorie definite come segue:

$$X_j \equiv \sum_{i=1}^r B_i^{(j)}, \quad j = 1, \dots, n.$$

La variabile aleatoria X_j indica dunque il **numero complessivo dei soggetti nel sito** j .

Il vettore aleatorio

$$\mathbf{X} \equiv (X_1, \dots, X_n)$$

è quindi un vettore aleatorio a valori in $A_{n,r}$ e ciascuna distribuzione di probabilità su $A_{n,r}$ può essere vista come un modello probabilistico di scelta dei siti da parte dei soggetti.

Le variabili aleatorie X_1, \dots, X_n vengono detti **numeri di occupazione** e le distribuzioni di probabilità su $A_{n,r}$ sono indicate con il termine **modelli di occupazione**.

Prima di illustrare i casi più notevoli di modelli di occupazione, è opportuno premettere le seguenti osservazioni.

Osservazione 1. L'insieme $A_{n,r}$ ha cardinalità data da $\binom{n+r-1}{n-1}$.

⁵⁹La frase **gli r soggetti si dispongono negli n siti** non tragga in inganno: nelle applicazioni potrebbe benissimo trattarsi di r **oggetti che vengono disposti in n siti**. Si veda a questo proposito il successivo Esempio 12.1, in cui soggetto viene interpretato come prova, mentre sito come risultato o esito. Nel successivo Esempio 12.4, invece i soggetti sono persone e i siti sono piani di un edificio.

Infatti il generico elemento di $A_{n,r}$ può essere ad esempio rappresentato attraverso il seguente tipo di “disegno”

$$\begin{array}{c} \overbrace{** \dots **}^{x_1} \mid \overbrace{** \dots **}^{x_2} \mid \dots \mid \overbrace{** \dots **}^{x_n} \\ 1^0 \text{ sito} \quad 2^0 \text{ sito} \quad \dots \quad n^{\text{esimo}} \text{ sito} \end{array} \tag{90}$$

dove, guardando da sinistra verso destra, x_1 è il numero dei simboli * a sinistra della prima barretta |, x_2 è il numero dei simboli * compresi fra la prima e la seconda barretta ... e così via; x_n è il numero dei simboli * a destra dell'ultima barretta (se $x_i = 0$, la $(i - 1)$ -esima e la i -esima barretta sono contigue, senza simboli * in mezzo).

Notiamo d'altra parte che ad ognuno di tali disegni corrisponde un'unica n -upla $x \in A_{n,r}$. Possiamo stabilire dunque una corrispondenza biunivoca fra l'insieme costituito dai disegni stessi e l'insieme $A_{n,r}$.

Ad esempio per $r = 4$ ed $n = 3$, il disegno $** \mid ** \mid$ o più esplicitamente,

$$\begin{array}{c} \overbrace{**}^2 \mid \overbrace{**}^2 \mid \overbrace{}^0 \\ 1^0 \text{ sito} \quad 2^0 \text{ sito} \quad 3^0 \text{ sito} \end{array}$$

corrisponde alla tripletta dei numeri di occupazione $(2, 2, 0)$.

Analogamente, la tripletta dei numeri di occupazione $(1, 0, 3)$ corrisponde al disegno $* \mid \mid **$, o più esplicitamente a

$$\begin{array}{c} \overbrace{*}^1 \mid \overbrace{}^0 \mid \overbrace{***}^3 \\ 1^0 \text{ sito} \quad 2^0 \text{ sito} \quad 3^0 \text{ sito} \end{array}$$

La cardinalità dell'insieme $A_{n,r}$ è quindi la stessa dell'insieme di tutti i disegni del tipo (90)

Ora, ciascun disegno contiene in totale $r + n - 1$ simboli, di cui r simboli sono asterischi * ed $(n - 1)$ simboli sono barrette |. Inoltre ciascun disegno corrisponde ad un modo di disporre le $(n - 1)$ barrette sul totale degli $(r + n - 1)$ posti.

Dunque vi sono $\binom{n+r-1}{n-1}$ diversi possibili disegni e tale è anche la cardinalità dell'insieme $A_{n,r}$.

Osservazione 2. Consideriamo qui il caso in cui gli r soggetti siano fra di loro distinguibili. Pensiamo cioè che ciascuno dei simboli * abbia un suo indice distintivo ed avremo dunque gli r simboli distinti $*_1, \dots, *_r$.

In tal caso il generico risultato elementare dell'esperimento aleatorio in questione (consistente nel disporre i soggetti nei siti) si può rappresentare come una “configurazione” data da un disegno del tipo in (90) in cui vengono indicati però anche gli indici distintivi dei soggetti; avremo dunque **una configurazione** del tipo

$$\begin{array}{c} \overbrace{*_{i_1} *_{i_2} \dots *_{i_{x_1-1}} *_{i_{x_1}}}^{x_1} \mid \overbrace{*_{i_{x_1+1}} *_{i_{x_1+2}} \dots *_{i_{x_1+x_2}}}^{x_2} \mid \dots \mid \overbrace{*_{i_{x_1+\dots+x_{n-1}+1}} \dots *_{i_r}}^{x_n} \\ 1^0 \text{ sito} \quad 2^0 \text{ sito} \quad \dots \quad n^{\text{esimo}} \text{ sito} \end{array} \tag{91}$$

In questo caso si considerano diverse due configurazioni anche se danno luogo alla stessa n -upla per i numeri di occupazione (cioè allo stesso disegno), purché vi sia almeno un soggetto che, nelle due configurazioni, cade in due siti diversi. Una tale **configurazione** si può anche vedere come una funzione che associa a ciascuno dei soggetti (distinguibili) il suo sito, cioè come **una applicazione dall'insieme** $\{O_1, O_2, \dots, O_r\}$ **all'insieme** $\{1, 2, \dots, n\}$.

Il numero complessivo di tali configurazioni è dunque dato da n^r .

È chiaro che i disegni sono delle classi di equivalenza nell'insieme delle configurazioni⁶⁰: vi sono in generale più configurazioni che danno luogo ad una stessa n -upla $x \in A_{n,r}$ ed il loro numero è dato dal **coefficiente multinomiale**

$$\binom{r}{x_1 x_2 \dots x_n} \equiv \frac{r!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!}. \quad (92)$$

Tale quantità infatti esprime il numero dei possibili modi⁶¹ in cui un insieme di r elementi può essere suddiviso in n sottoinsiemi, di rispettive cardinalità x_1, x_2, \dots, x_n .

⁶⁰Un modo di ragionare per arrivare a calcolare il coefficiente multinomiale è il seguente: una configurazione potrebbe essere data da una permutazione di $\{O_1, O_2, \dots, O_r\}$ che dà luogo ad un vettore di numeri di occupazione $x \in A_{n,r}$ secondo la seguente regola: i primi x_1 elementi della permutazione vengono messi nel sito 1, i successivi x_2 elementi sono messi nel sito 2, e così via fino agli ultimi x_n elementi che vengono messi nel sito n (ovviamente se uno degli x_i è nullo nessun elemento viene messo nel sito i). È chiaro che permutazioni che hanno gli stessi elementi ai primi x_1 posti, gli stessi elementi ai successivi x_2 posti, e così via fino agli ultimi x_n posti, sono associati allo stesso vettore x dei numeri di occupazione. Le classi di equivalenza sono dunque

$$\frac{r!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!}.$$

Per rendere più concreto quanto detto ripetiamo nell'esempio in cui $n = 3$ ed $r = 4$. Ad esempio, le permutazioni (O_1, O_2, O_3, O_4) (O_1, O_2, O_4, O_3) (O_2, O_1, O_3, O_4) (O_2, O_1, O_4, O_3) , nel caso di $x = (2, 2, 0)$ danno tutte luogo alla stessa situazione in cui gli elementi $\{O_1, O_2\}$ vengono messi nel sito 1, gli elementi $\{O_3, O_4\}$ vengono messi nel sito 2 e nessuno nel sito 3. Un ragionamento analogo vale per le permutazioni (O_1, O_3, O_2, O_4) (O_1, O_3, O_4, O_2) (O_3, O_1, O_2, O_4) (O_3, O_1, O_4, O_2) .

⁶¹Per arrivare a tale coefficiente multinomiale si può anche ragionare come segue: Prima di tutto va scelto un insieme B_1 di cardinalità $\binom{r}{x_1}$. Per ciascuna scelta di B_1 si passa poi a scegliere un sottoinsieme B_2 di cardinalità x_2 , tra gli $r - x_1$ elementi rimasti. Ciò può essere fatto in $\binom{r-x_1}{x_2}$ modi diversi. Si prosegue scegliendo il sottoinsieme B_3 di cardinalità x_3 tra gli $r - x_1 - x_2 = r - (x_1 + x_2)$ elementi rimasti, e così via fino a che non rimangono $r - (x_1 + x_2 + \dots + x_{n-1}) = x_n$ elementi, che rappresentano il sottoinsieme B_n . Si ottengono quindi

$$\begin{aligned} & \binom{r}{x_1} \cdot \binom{r-x_1}{x_2} \cdot \dots \cdot \binom{r-(x_1+\dots+x_{n-2})}{x_{n-1}} \cdot \binom{r-(x_1+\dots+x_{n-1})}{x_n} \\ &= \frac{r!}{x_1!(r-x_1)!} \frac{(r-x_1)!}{x_2!(r-x_1-x_2)!} \dots \frac{(r-x_1-\dots-x_{n-2})!}{x_{n-1}!(r-x_1-\dots-x_{n-1})!} \\ &= \frac{r!}{x_1!x_2! \dots x_{n-1}!(r-x_1-\dots-x_{n-1})!} = \frac{r!}{x_1!x_2! \dots x_{n-1}!x_n!} \end{aligned}$$

La configurazione ottenuta va interpretata nel senso che gli elementi di B_i vengono messi nel sito i , per $i = 1, \dots, n$.

Ad esempio per $r = 4$ ed $n = 3$, nel caso in cui i 4 soggetti siano distinguibili, si consideri la seguente configurazione che coincide con la seguente funzione da $\{O_1, O_2, O_3, O_4\}$ in $\{1, 2, 3\}$

$$\begin{aligned} O_1 &\mapsto 1 \\ O_2 &\mapsto 1 \\ O_3 &\mapsto 2 \\ O_4 &\mapsto 2 \end{aligned}$$

Questa funzione è una delle possibili fra quelle che danno luogo al vettore dei numeri di occupazione $(2, 2, 0)$. La famiglia completa di tali funzioni è data da

$$\begin{array}{cccccc} O_1 \mapsto 1 & O_1 \mapsto 1 & O_1 \mapsto 1 & O_1 \mapsto 2 & O_1 \mapsto 2 & O_1 \mapsto 2 \\ O_2 \mapsto 1 & O_2 \mapsto 2 & O_2 \mapsto 2 & O_2 \mapsto 1 & O_2 \mapsto 1 & O_2 \mapsto 2 \\ O_3 \mapsto 2 & O_3 \mapsto 1 & O_3 \mapsto 2 & O_3 \mapsto 1 & O_3 \mapsto 2 & O_3 \mapsto 1 \\ O_4 \mapsto 2 & O_4 \mapsto 2 & O_4 \mapsto 1 & O_4 \mapsto 2 & O_4 \mapsto 1 & O_4 \mapsto 1 \end{array}$$

e corrispondono, rispettivamente, alle partizioni “ordinate” di $\{O_1, O_2, O_3, O_4\}$

$$\begin{aligned} B_1 &= \{O_1, O_2\}, B_2 = \{O_3, O_4\}, B_3 = \emptyset, \\ B_1 &= \{O_1, O_3\}, B_2 = \{O_3, O_4\}, B_3 = \emptyset, \\ B_1 &= \{O_1, O_4\}, B_2 = \{O_2, O_3\}, B_3 = \emptyset, \\ B_1 &= \{O_2, O_3\}, B_2 = \{O_1, O_4\}, B_3 = \emptyset, \\ B_1 &= \{O_2, O_4\}, B_2 = \{O_1, O_3\}, B_3 = \emptyset, \\ B_1 &= \{O_3, O_4\}, B_2 = \{O_1, O_2\}, B_3 = \emptyset. \end{aligned}$$

Parliamo di partizioni “ordinate” nel senso che, ad esempio distinguiamo la prima e l’ultima partizione, che pur essendo composte dagli stessi insiemi, differiscono in quanto il primo elemento della partizione corrisponde alla specificazione degli elementi che si trovano nel sito 1, mentre la seconda a quelli che si trovano nel sito 2 ed infine la terza a quelli che si trovano nel sito 3.

Tutte le precedenti partizioni concorrono al disegno $** | ** |$ e sono esattamente $6 = \frac{4!}{2!2!0!}$: si possono calcolare pensando che prima di tutto si deve scegliere l’insieme B_1 in $\binom{4}{2}$ modi, per ciascuno di questi modi si ha poi la possibilità di scegliere l’insieme B_2 in $\binom{4-2}{2}$ modi ed infine B_3 è obbligatoriamente determinato da $B_3 = \{O_1, O_2, O_3, O_4\} \setminus (B_1 \cup B_2)$, ottenendo un totale di

$$\binom{4}{2} \cdot \binom{4-2}{2} = \frac{4!}{2!(4-2)!} \cdot \frac{(4-2)!}{2!(4-2-2)!} = \frac{4!}{2!2!0!}$$

configurazioni per il vettore dei numeri di occupazione $(2, 2, 0)$.

12.1 Modello di Maxwell-Boltzmann

Si suppone che gli r soggetti siano **distinguibili** e che vengano distribuiti negli n siti in modo tale da assegnare uguale probabilità $\frac{1}{n^r}$ a ciascuna configurazione del tipo (91).

Per $\mathbf{x} \equiv (x_1, x_2, \dots, x_n) \in A_{n,r}$ si ha dunque⁶²

$$\begin{aligned} P(\{\mathbf{X} = \mathbf{x}\}) &= P(\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}) \\ &= \frac{\binom{r}{x_1 x_2 \dots x_n}}{n^r} = \frac{r!}{x_1! x_2! \cdots x_{n-1}! x_n! n^r}, \quad \text{per ogni } \mathbf{x} \in A_{n,r}. \end{aligned}$$

⁶²Si noti che sarebbe più evocativo mettere al posto del simbolo P un simbolo del tipo P_{MB} per mettere in evidenza che si tratta delle probabilità relative al modello di Maxwell-Boltzmann. Tuttavia lasciamo come al solito la scrittura P per non appesantire le notazioni.

Osservazione 3. Da quanto sopra, segue

$$\sum_{\mathbf{x} \in A_{n,r}} \binom{r}{x_1 x_2 \dots x_n} = n^r.$$

ovvero

$$\sum_{\mathbf{x} \in A_{n,r}} \frac{r!}{x_1! x_2! \dots x_{n-1}! x_n!} = n^r.$$

12.2 Modello di Bose-Einstein

Si suppone che gli r soggetti siano **indistinguibili** e che vengano distribuiti negli n siti in modo da assegnare uguale probabilità a ciascun vettore di numeri di occupazione, cioè a ciascuna n -upla $\mathbf{x} \in A_{n,r}$.

Si pone dunque⁶³

$$P(\{\mathbf{X} = \mathbf{x}\}) = P(\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}) = \frac{1}{\binom{n+r-1}{n-1}}, \quad \text{per ogni } \mathbf{x} \in A_{n,r}.$$

12.3 Modello di Fermi-Dirac

Supponiamo $r \leq n$ e poniamo

$$\hat{A}_{n,r} \equiv \{\mathbf{x} \in A_{n,r} : x_i = 0, \text{ oppure } x_i = 1, i = 1, 2, \dots, n\}$$

La cardinalità di $\hat{A}_{n,r}$ è ovviamente uguale a $\binom{n}{r}$.

Ora si suppone che gli r soggetti siano distribuiti nei siti in modo da assegnare uguale probabilità a ciascuno dei disegni in $\hat{A}_{n,r}$ (ciò in particolare implica che in ciascun sito non può cadere più di un oggetto). Si ha allora⁶⁴

$$P(\{\mathbf{X} = \mathbf{x}\}) = P(\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}) = \begin{cases} \frac{1}{\binom{n}{r}} & \text{per } \mathbf{x} \in \hat{A}_{n,r} \\ = 0 & \text{per } \mathbf{x} \notin \hat{A}_{n,r} \end{cases}$$

12.4 Schemi di estrazioni da urne

Verrà illustrato qui di seguito che ai modelli di occupazione si può anche dare, equivalentemente, un'interpretazione in termini di schemi di estrazioni casuali da urne.

Supponiamo di avere inizialmente n oggetti di n diversi tipi⁶⁵ in un'urna e di eseguire in modo aleatorio r estrazioni dall'urna.

⁶³Si noti che sarebbe più evocativo mettere al posto del simbolo P un simbolo del tipo P_{BE} per mettere in evidenza che si tratta delle probabilità relative al modello di Bose-Einstein. Tuttavia lasciamo come al solito la scrittura P per non appesantire le notazioni.

⁶⁴Si noti che sarebbe più evocativo mettere al posto del simbolo P un simbolo del tipo P_{FD} per mettere in evidenza che si tratta delle probabilità relative al modello di Fermi-Dirac. Anche in questo caso lasciamo la scrittura P per non appesantire le notazioni.

⁶⁵Gli n diversi tipi sono denotati come tipo 1, tipo 2, ..., tipo n .

Tale esperimento ha come risultati elementari le n^r r -uple (i_1, i_2, \dots, i_r) con $i_j \in \{1, 2, \dots, n\}$ e poniamo, per $i = 1, \dots, n$,

$\tilde{X}_i \equiv$ numero di volte in cui si presenta un oggetto di tipo i nelle r estrazioni

Anche qui dunque $\tilde{\mathbf{X}} \equiv (\tilde{X}_1, \dots, \tilde{X}_n)$ è un vettore aleatorio a valori in $A_{n,r}$.

Come ora vedremo, diversi tipi di estrazioni casuali corrispondono ai diversi modelli di occupazione visti sopra e ciò permette agevolmente di spiegare questi ultimi in termini del meccanismo casuale con cui ogni soggetto sceglie un sito.

L'interpretazione nel linguaggio della sezione precedente di n siti ed r soggetti è il seguente: per $k = 1, 2, \dots, r$, il soggetto O_k viene posto nel sito i_k , che corrisponde al tipo dell'oggetto estratto nella k -sima estrazione.

Ricordiamo che il termine "estrazioni casuali" si riferisce all'ipotesi che ciascuno degli oggetti, presenti nell'urna al momento di una estrazione, abbia la stessa probabilità di presentarsi, indipendentemente dal suo tipo.

Le differenze fra i diversi modelli di estrazione risiedono nelle modalità della composizione dell'urna nelle successive estrazioni.

Estrazioni casuali con reinserimento e modello di Maxwell-Boltzmann

Dopo ogni estrazione ***l'oggetto estratto viene reinserito nell'urna***. Quindi ad ogni estrazione vi sono n oggetti di diversi tipi e ciascun tipo ha la stessa probabilità di presentarsi.

Ciascuno degli risultati elementari possibili ha la stessa probabilità $\frac{1}{n^r}$.

Ciò corrisponde dunque, per il vettore $\tilde{\mathbf{X}}$, al ***modello di Maxwell-Boltzmann***.

Vale forse la pena di ricordare l'interpretazione nel linguaggio precedente di n siti ed r soggetti: il soggetto O_k viene posto nel sito i_k estratto nella k -sima estrazione (con reinserimento).

Estrazioni casuali senza reinserimento e modello di Fermi-Dirac

L'oggetto estratto non viene più reinserito nell'urna; deve dunque essere $r \leq n$.

Alla j -esima estrazione vi sono $n - (j - 1) = (n - j + 1)$ oggetti nell'urna.

Ciascun tipo può presentarsi al più una sola volta nel complesso delle estrazioni.

Tutte le r -uple (i_1, i_2, \dots, i_r) , con $i_1 \neq i_2 \neq \dots \neq i_r$ hanno la stessa probabilità

$$\frac{1}{n(n-1)\dots(n-r+1)}.$$

Tutte le n -uple $x \in \hat{A}_{n,r}$ sono equiprobabili per $\tilde{\mathbf{X}}$. Ciò corrisponde dunque al ***modello di Fermi-Dirac***.

Dal punto di vista dei numeri di occupazione tuttavia ognuna delle $r!$ permutazioni (i_1, i_2, \dots, i_r) che dia luogo allo stesso insieme di cardinalità k , corrisponde allo stesso vettore di numeri di

occupazione, ovvero alla stessa n -pla $\mathbf{x} \in \widehat{A}_{n,r}$. Questo è il motivo per cui si ottiene esattamente che

$$P(\widetilde{\mathbf{X}} = \mathbf{x}) = \frac{r!}{n(n-1)\dots(n-r+1)} = \frac{1}{\binom{n}{r}}, \quad \text{per } \mathbf{x} \in \widehat{A}_{n,r},$$

$$P(\widetilde{\mathbf{X}} = \mathbf{x}) = 0 \quad \text{per } \mathbf{x} \in A_{n,r} \setminus \widehat{A}_{n,r}$$

Anche qui vale forse la pena di ricordare l'interpretazione nel linguaggio precedente di n siti ed r soggetti: il soggetto O_k viene posto nel sito i_k estratto nella k -sima estrazione (senza reinserimento).

Estrazioni casuali con doppio reinserimento e modello di Bose-Einstein

Dopo ciascuna estrazione, **viene inserito nell'urna, insieme all'oggetto estratto, anche un altro oggetto dello stesso tipo**, cosicché alla j -esima estrazione vi sono $(n+j-1)$ individui nell'urna.

Vogliamo ora calcolare $P(\{\mathbf{X} = \mathbf{x}\})$ per $\mathbf{x} \equiv (x_1, x_2, \dots, x_n) \in A_{n,r}$ (ricordiamo che risulta $\sum_{h=1}^n x_h = r$); cominciamo a considerare la probabilità del risultato elementare (i_1, i_2, \dots, i_r) definito da

$$\begin{aligned} i_j &= 1, & 1 \leq j \leq x_1 \\ i_j &= 2, & x_1 + 1 \leq j \leq x_1 + x_2 \\ &\vdots \\ i_j &= n, & \sum_{h=1}^{n-1} x_h + 1 \leq j \leq \sum_{h=1}^n x_h = r \end{aligned}$$

(cioè: tutti i primi x_1 elementi estratti sono di tipo 1, poi segue l'estrazione di x_2 elementi tutti di tipo 2 e così via). Tale probabilità sarà data da

$$\begin{aligned} &\frac{1}{n} \cdot \frac{2}{n+1} \cdot \frac{3}{n+2} \cdots \frac{x_1}{n+x_1-1} \\ &\cdot \frac{1}{n+x_1} \cdot \frac{2}{n+x_1+1} \cdots \frac{x_2}{n+x_1+x_2-1} \\ &\cdot \frac{1}{n+x_1+x_2} \cdots \\ &\vdots \\ &\cdot \frac{1}{n+x_1+x_2+\dots+x_{n-1}} \cdot \frac{2}{n+x_1+x_2+\dots+x_{n-1}+1} \cdots \frac{x_n-1}{n+\sum_{h=1}^n x_h-2} \cdot \frac{x_n}{n+\sum_{h=1}^n x_h-1} \\ &= \frac{x_1!x_2!\dots x_n!}{\frac{(n+r-1)!}{(n-1)!}}. \end{aligned}$$

Possiamo ora osservare che qualunque altro risultato elementare (i_1, i_2, \dots, i_r) , che ugualmente abbia x_1 coordinate uguali ad 1, x_2 coordinate uguali a 2, ..., x_n coordinate uguali a n , seppure in un ordine diverso, ha ancora la stessa probabilità dell' r -upla precedentemente considerata.

Tale probabilità è data cioè ancora da

$$\frac{x_1!x_2!\dots x_n!}{\frac{(n+r-1)!}{(n-1)!}}.$$

Dunque

$$\begin{aligned} P(\{X = x\}) &= \binom{r}{x_1 x_2 \dots x_n} \cdot \frac{x_1!x_2!\dots x_n!}{\frac{(n+r-1)!}{(n-1)!}} \\ &= \binom{r}{x_1 x_2 \dots x_n} \cdot \frac{x_1!x_2!\dots x_n!}{r! \binom{n+r-1}{n-1}} = \frac{1}{\binom{n+r-1}{n-1}}. \end{aligned}$$

Ciò corrisponde quindi al **modello di Bose-Einstein**.

Anche qui, nell'interpretazione nel linguaggio precedente di n siti ed r soggetti, il soggetto O_k viene posto nel sito i_k estratto nella k -sima estrazione (con doppio reinserimento).

12.5 Alcuni esempi

Abbiamo dunque illustrato fin qui due linguaggi diversi ma equivalenti, per illustrare i modelli di occupazione.

Bisogna però tenere presente che, con opportune modifiche di linguaggio, gli stessi schemi si presentano in moltissimi tipi di applicazioni diverse; per tale motivo la conoscenza degli aspetti basilari sui modelli di occupazione si rivela fruttuosa nella soluzione di moltissimi tipi di problemi di probabilità nel discreto.

In particolare il modello di Maxwell-Boltzmann si ripresenta in moltissime situazioni diverse, ed è interessante in particolare capire la differenza che sussiste fra tale modello e quelli di Bose-Einstein o di Fermi-Dirac.

Ora presentiamo alcuni esempi di problemi in cui si ritrovano dei modelli di occupazione; successivamente verrà illustrato un modello (la distribuzione multinomiale) utilissimo in tutte le applicazioni e che deriva da una naturale generalizzazione del modello di Maxwell-Boltzmann.

Esempio 12.1. *Un esperimento può dar luogo, con uguali probabilità $p = \frac{1}{3}$, a tre diversi risultati, che per semplicità indichiamo con 1, 2, 3.*

Supponiamo che l'esperimento venga condotto per 10 successive volte, sempre con le stesse modalità ed in modo indipendente una volta dall'altra e si indichi con X_1, X_2, X_3 il numero di volte in cui, rispettivamente, si verifica il risultato 1, oppure 2, oppure 3.

Calcolare le probabilità dell'evento $\{X_1 = 4, X_2 = 3, X_3 = 3\}$.

Soluzione. Notiamo che ovviamente risulta

$$P(\{X_1 + X_2 + X_3 = 10\}) = 1$$

e anche, ovviamente

$$P(\{X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_1 + X_2 + X_3 = 10\}) = 1$$

Tale problema può essere risolto guardando a X_1, X_2, X_3 come a dei numeri di occupazione, con $r = 10$ ed $n = 3$. (ogni prova è vista come un soggetto ed ogni possibile risultato come un sito dentro cui si inserisce ciascuna prova/soggetto).

Infatti le terne di valori possibili per il vettore aleatorio (X_1, X_2, X_3) costituiscono l'insieme $A_{3,10}$. Per

l'indipendenza e l'equiprobabilità nel comportamento delle diverse prove si ha che la distribuzione di probabilità congiunta di (X_1, X_2, X_3) coincide con il corrispondente modello di Maxwell-Boltzmann. E dunque

$$P(\{X_1 = 4, X_2 = 3, X_3 = 3\}) = \frac{\binom{10}{4\ 3\ 3}}{3^{10}} = \frac{10!}{4!(3!)^2} \frac{1}{3^{10}}.$$

Esempio 12.2. *Un dado viene lanciato 12 volte ed indichiamo con X_i ($i = 1, \dots, 6$) il numero dei lanci in cui si presenta il punteggio i .*

Calcolare

$$P(\{X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3\}).$$

Soluzione. Si tratta nuovamente di un modello di Maxwell-Boltzmann, questa volta con $n = 6$, $r = 12$.

L'evento $\{X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3\}$ implica ovviamente $X_5 = 0, X_6 = 0$. La probabilità cercata è data da

$$\begin{aligned} & P(\{X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3\}) \\ &= P(\{X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 3, X_5 = 0, X_6 = 0\}) \\ &= \frac{12!}{(3!)^4 (0!)^2 6^{12}} = \frac{12!}{6^{16}}. \end{aligned}$$

Esercizio proposto 12.1 (Il problema del compleanno, come modello di occupazione). *Si consideri la situazione del problema del compleanno (Esempio 3.3), in cui ci sono r persone. Si numerino da 1 a 365 i giorni dell'anno (1 corrisponde al primo gennaio, 32 al primo febbraio, etc. fino a 365, che corrisponde al 31 dicembre). Posto X_i la variabile aleatoria che conta il numero di persone nate nel giorno i , per $i = 1, 2, \dots, 365$.*

(a) *A quale tipo di modello di occupazione ci si riferisce?*

(b) *Scrivere, in termini delle variabili aleatorie X_i , l'evento*

$$\{\text{tutte le persone sono nate in giorni diversi}\}.$$

Esercizio proposto 12.2 (Estrazioni di assi). *Sia dato un mazzo di carte italiane (40 carte, 4 semi, carte numerate da 1 a 10 per ciascun seme, con 1 che corrisponde all'asso, 8 al fante, 9 al cavallo e 10 al re) ben mescolate. Si estraggano ad una ad una le carte dal mazzo fino ad esaurire tutte le carte. Sia X_1 il numero di carte estratte (diverse da un asso) prima del primo asso. Siano inoltre X_2 il numero di carte estratte (diverse da un asso) dopo il primo asso e prima del secondo asso, cioè tra il primo ed il secondo asso, X_3 il numero di carte uscite (diverse da un asso) dopo il secondo asso e prima del terzo asso, X_4 , le carte (diverse da un asso) uscite tra il terzo ed il quarto asso, ed infine X_5 le carte (diverse da un asso) uscite dopo il quarto asso. Ad esempio⁶⁶ se i quattro assi sono usciti come seconda, terza, quindicesima e trentaseiesima carta, allora $X_1 = 1, X_2 = 0, X_3 = 11, X_4 = 20$ e $X_5 = 4$. Ovviamente*

$$X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_4 \geq 0, X_5 \geq 0, \quad e \quad X_1 + X_2 + X_3 + X_4 + X_5 = 36$$

⁶⁶Più in generale se i quattro assi escono come k_1 - *sima*, k_2 - *sima*, k_3 - *sima* e k_4 - *sima* carta, allora $X_1 = k_1 - 1$, $X_2 = k_2 - k_1 - 1$, $X_3 = k_3 - k_2 - 1$, $X_4 = k_4 - k_3 - 1$ ed infine $X_5 = 40 - k_4$.

in quanto 36 è il numero totale di carte diverse da un asso.

Si dimostri che $(X_1, X_2, X_3, X_4, X_5)$ corrisponde al modello di occupazione di Bose-Einstein con $n = 5$ ed $r = 36$.

Esempio 12.3. Un testo contiene 20 caratteri. Supponiamo di sapere che esso contiene 5 errori (cioè 5 caratteri errati) e di valutare che ciascun carattere abbia uguale probabilità di essere errato rispetto a ciascun altro. Qual è la probabilità che gli errori si trovino nel $3^0, 5^0, 10^0, 11^0, 12^0$ carattere?

Questo è un esempio di modello di Fermi-Dirac, con $n = 20$ ed $r = 5$, in quanto in ogni posto ci può stare al più un errore. E dunque la probabilità cercata è data da $\frac{1}{\binom{20}{5}}$.

Esempio 12.4. 5 persone sono in attesa dell'ascensore nella hall di un albergo di 4 piani. Poniamo, per $i = 1, 2, 3, 4$,

$$X_i := \text{numero persone che scendono al piano } i$$

Se le persone, una volta entrate nell'ascensore, si distribuiscono a caso fra i piani, a quanto è uguale

$$P(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\})?$$

Soluzione. Si tratta di un modello di occupazione con $n = 4$ ed $r = 5$. La dizione “le persone si distribuiscono a caso fra i piani” è equivoca: potrebbe trattarsi di modello di Maxwell-Boltzmann oppure di Bose-Einstein, a seconda che le persone vengano considerate o meno **indipendenti**⁶⁷ fra di loro (o diciamo, **distinguibili**).

Nei due diversi casi si avrebbe, indicando rispettivamente con P_{MB} e con P_{BE} le probabilità nei due modelli,

$$P_{MB}(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{5!}{(0!)^3 4! 4^5} = \frac{5!}{4! 4^5} = \frac{5}{4^5}$$

oppure

$$P_{BE}(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{1}{\binom{8}{3}} = \frac{3!5!}{8!} = \frac{3 \cdot 2}{8 \cdot 7 \cdot 6} = \frac{1}{56}.$$

Riguardo al modello di Maxwell-Boltzmann in particolare, osserviamo che esso si rivela, per questo caso, piuttosto irrealistico: è poco ragionevole dare per scontato che vi sia indipendenza stocastica completa fra le diverse persone ma soprattutto che tutti i piani siano equiprobabili fra di loro!

In merito a quest'ultima condizione di equiprobabilità, vediamo che è opportuno estendere il modello di Maxwell-Boltzmann al caso in cui vi sia ancora indipendenza stocastica completa fra i diversi oggetti, ma non vi sia equiprobabilità fra i diversi siti. Tale estensione porta alla definizione di distribuzione multinomiale.

12.6 Distribuzione multinomiali

Consideriamo ancora un modello di occupazione con r soggetti ed n siti. Supponiamo ancora che vi sia indipendenza stocastica completa circa la scelta dei siti da parte dei diversi soggetti, ma non vi sia equiprobabilità fra i diversi siti: supponiamo che ciascun soggetto scelga il sito j con fissata probabilità p_j ($j = 1, \dots, n; p_j \geq 0, \sum_{j=1}^n p_j = 1$).

⁶⁷La frase “le persone vengano considerate o meno indipendenti”, va intesa nel senso che eventi relativi a persone diverse (sempre riguardo la scelta del piano al quale scendere) sono eventi completamente indipendenti.

Generalizzando quanto arguito per il modello di Maxwell-Boltzmann, possiamo concludere che, per i numeri di occupazione X_1, \dots, X_n , risulta:

$$P(\{X_1 = x_1, \dots, X_n = x_n\}) = \binom{r}{x_1 \ x_2 \ \dots \ x_n} (p_1)^{x_1} \cdot (p_2)^{x_2} \cdot \dots \cdot (p_n)^{x_n}. \quad (93)$$

$$= \frac{r!}{x_1! \ x_2! \ \dots \ x_n!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_n^{x_n}, \quad \text{per } (x_1, x_2, \dots, x_n) \in A_{n,r}. \quad (94)$$

Diciamo in tal caso che la distribuzione congiunta di X_1, \dots, X_n è una *distribuzione multinomiale* di parametri $(r, n; p_1, \dots, p_n)$.

È chiaro che un modello di Maxwell-Boltzmann costituisce un caso particolare di distribuzione multinomiale con $p_j = \frac{1}{n}$.

Osserviamo anche che una distribuzione binomiale $\text{bin}(r, p)$ è connessa ad una distribuzione multinomiale di parametri $(r, 2; p, 1 - p)$; in quale modo? (Vedere l'Esercizio di verifica 12.9). Per la comprensione della formula (93) in realtà è più comodo pensare alla distribuzione multinomiale proprio come a una generalizzazione della binomiale⁶⁸.

Osserviamo infine che essendo ovviamente $P(\mathbf{X} \in A_{n,r}) = 1$, si ha che

$$\sum_{(x_1, x_2, \dots, x_n) \in A_{n,r}} \frac{r!}{x_1! \ x_2! \ \dots \ x_n!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_n^{x_n} = 1,$$

che non è altro che un caso particolare della formula della potenza del multinomio, ovvero

$$\sum_{(x_1, x_2, \dots, x_n) \in A_{n,r}} \frac{r!}{x_1! \ x_2! \ \dots \ x_n!} a_1^{x_1} \cdot a_2^{x_2} \cdot \dots \cdot a_n^{x_n} = (a_1 + a_2 + \dots + a_n)^r, \quad (95)$$

che a sua volta è una generalizzazione della formula della potenza del binomio.

Esempio 12.5. *Un gioco fra quattro persone viene ripetuto 5 volte, sempre con gli stessi giocatori A, B, C, D. Ogni volta c'è un singolo vincitore: A, B e C hanno probabilità di vincere del 20% e D del 40%.*

Con X_A, X_B, X_C, X_D si indichino, rispettivamente, il numero delle vittorie di A, B, C, D sulle 5 volte. Calcolare le probabilità dell'evento

$$\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}.$$

⁶⁸L'idea è che ci sono r prove ad n esiti possibili, ovvero

$$E_k^{(j)}, \quad \text{per } j = 1, 2, \dots, n, \ k = 1, 2, \dots, r,$$

dove il verificarsi di $E_k^{(j)}$, significa il verificarsi dell'esito di tipo j nella k -sima prova ed esclude il verificarsi, nella prova k -sima degli altri esiti. In altre parole $\{E_k^{(j)}, \text{ per } j = 1, 2, \dots, n, \}$ sono partizioni per ogni $k = 1, 2, \dots, r$. Inoltre si suppone che, qualunque sia $k = 1, 2, \dots, r$, la probabilità $P(E_k^{(j)}) = p_j$.

Infine si suppone l'indipendenza delle prove, cioè che, qualunque sia $(j_1, j_2, \dots, j_r) \in \{1, 2, \dots, n\}^r$,

$$P(E_1^{(j_1)} \cap E_2^{(j_2)} \cap E_3^{(j_3)} \cap \dots \cap E_r^{(j_r)}) = P(E_1^{(j_1)}) \cdot P(E_2^{(j_2)}) \cdot P(E_3^{(j_3)}) \cdot \dots \cdot P(E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \dots \cdot p_{j_n}.$$

Il caso della binomiale $\text{bin}(r, \theta)$ e dello schema di Bernoulli per r prove indipendenti, con probabilità di successo θ rientra in questo modello, con $n = 2$, $E_k^{(1)} = E_k$, $E_k^{(2)} = \bar{E}_k$, $p_1 = \theta$, $p_2 = 1 - \theta$.

L'evento $\{X_1 = x_1, \dots, X_n = x_n\}$, con $\mathbf{x} \in A_{n,r}$ si verifica se e solo se si verificano eventi del tipo $E_1^{(j_1)} \cap E_2^{(j_2)} \cap \dots \cap E_r^{(j_r)}$, con x_1 indici di tipo 1, x_2 indici di tipo 2, ..., x_n indici di tipo n . Per calcolare la probabilità dell'evento $\{X_1 = x_1, \dots, X_n = x_n\}$, basta allora considerare che gli eventi di questo tipo hanno tutti la stessa probabilità:

$$P(E_1^{(j_1)} \cap E_2^{(j_2)} \cap \dots \cap E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \dots \cdot p_{j_r} = p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_n^{x_n},$$

e che inoltre gli eventi di questo tipo sono esattamente $\frac{r!}{x_1! \ x_2! \ \dots \ x_n!}$.

Soluzione. Le variabili aleatorie X_A, X_B, X_C, X_D hanno una distribuzione congiunta multinomiale di parametri $(5, 4; \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})$ e risulta

$$P(\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}) = \frac{5!}{(1!)^3 2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{5!}{2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{48}{5^4} = \frac{48}{625}.$$

Esempio 12.6. In una giornata del campionato di calcio si attribuisce probabilità 0.5 alla vittoria della squadra che gioca in casa (risultato 1), probabilità 0.2 alla sconfitta della squadra che gioca in casa (risultato 2), probabilità 0.3 al pareggio (risultato x), e il risultato di ciascuna partita è giudicato essere indipendente dai risultati delle altre. Si consideri la colonna “vincente” della schedina del totocalcio⁶⁹; si ponga

$$Z_1 \equiv \text{numero di risultati 1 sulle tredici partite}$$

e si dia analogo significato alle variabili Z_x, Z_2 .

Le variabili aleatorie Z_1, Z_2, Z_x hanno una distribuzione congiunta multinomiale di parametri $(13, 3; \frac{5}{10}, \frac{2}{10}, \frac{3}{10})$.

Esercizio proposto 12.3 (Il problema del compleanno: mesi di nascita). Si consideri la situazione del problema del compleanno (Esempio 3.3), in cui ci sono r persone. Si numerino da 1 a 12 i mesi dell'anno (1 corrisponde a gennaio, 2 a febbraio, etc.. fino a 12, che corrisponde a dicembre). Posto X_i la variabile aleatoria che conta il numero di persone nate nel mese i , per $i = 1, 2, \dots, 12$.

Chiaramente si tratta di un modello multinomiale. (a) Calcolare p_i per $i = 1, 2, \dots, 12$.

(b) Scrivere, in termini delle variabili aleatorie X_i , l'evento

$$\{\text{tutte le persone sono nate nel mese di giugno}\}$$

e calcolarne la probabilità.

(c) Scrivere, in termini delle variabili aleatorie X_i , l'evento

$$\{\text{tutte le persone sono nate nello stesso mese}\}$$

e calcolarne la probabilità.

12.7 Distribuzioni marginali e condizionate nei modelli di occupazione

Sia fissato ora un generico modello di occupazione, con n siti ed r soggetti. Vogliamo ricavare la distribuzione di probabilità marginale della variabile aleatoria X_1 .

Si avrà, per definizione di distribuzione marginale,

$$P(\{X_1 = x\}) = \sum_{(z_2, \dots, z_n) \in A_{n-1, r-x}} P(\{X_1 = x, X_2 = z_2, \dots, X_n = z_n\}) \quad (96)$$

Un'analogha formula vale per le distribuzioni marginali di X_2 , di X_3 , ... e di X_n .

⁶⁹Una schedina del totocalcio è composta da una colonna costituita dai risultati di 13 partite fissate. I risultati possibili sono 1, 2 e x, per ogni elemento della colonna. in totale ci sono quindi 3^{13} possibili colonne, ovvero possibili schedine da giocare.

Osservazione Per motivi di simmetria, nel caso dei modelli di Maxwell-Boltzmann, Bose-Einstein e Fermi-Dirac, si ha che le distribuzioni marginali di X_i coincidono tutte con la distribuzione marginale di X_1 . Infatti, ad esempio,

$$P(\{X_n = x\}) = \sum_{(z_1, \dots, z_{n-1}) \in A_{n-1, r-x}} P(\{X_1 = z_1, X_2 = z_2, \dots, X_{n-1} = z_{n-1}, X_n = x\})$$

ed inoltre, qualunque siano i valori di

$$P(\{X_1 = x, X_2 = z_2, \dots, X_{n-1} = z_{n-1}, X_n = z_1\}) = P(\{X_1 = z_1, X_2 = z_2, \dots, X_{n-1} = z_{n-1}, X_n = x\})$$

Consideriamo ora la distribuzione di probabilità della variabile aleatoria

$$Y = \sum_{j=2}^n X_j$$

Osserviamo che risulta

$$\{Y = y\} = \{X_1 = r - y\}$$

da cui

$$P(\{Y = y\}) = P(\{X_1 = r - y\})$$

e quindi, dalla (96),

$$P(\{Y = y\}) = \sum_{(z_2, \dots, z_n) \in A_{n-1, y}} P(\{X_1 = r - y, X_2 = z_2, \dots, X_n = z_n\}).$$

Supponiamo ora di voler calcolare la **distribuzione condizionata di X_2 data X_1** . Osserviamo allora che vale

$$P(\{X_1 = x_1, X_2 = x_2\}) = \sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} P(\{X_1 = x_1, X_2 = x_2, X_3 = z_3, \dots, X_n = z_n\}),$$

da cui

$$P(\{X_2 = x_2\} | \{X_1 = x_1\}) = \frac{\sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} P(\{X_1 = x_1, X_2 = x_2, X_3 = z_3, \dots, X_n = z_n\})}{\sum_{(\xi, \zeta_3, \dots, \zeta_n) \in A_{n-1, r-x_1}} P(\{X_1 = x_1, X_2 = \xi, X_3 = \zeta_3, \dots, X_n = \zeta_n\})}.$$

Esempio 12.7. Consideriamo il modello di Maxwell-Boltzmann con r soggetti ed n siti. Allora, per $x_1 = 0, 1, \dots, r$, avremo (ricordando che $P(\{\mathbf{X} = \mathbf{x}\}) = \frac{\binom{r}{x_1, x_2, \dots, x_n}}{n^r}$)

$$\begin{aligned} P(\{X_1 = x\}) &= \sum_{(z_2, \dots, z_n) \in A_{n-1, r-x}} P(\{X_1 = x, X_2 = z_2, \dots, X_n = z_n\}) \\ &= \frac{1}{n^r} \sum_{(z_2, \dots, z_n) \in A_{n-1, r-x}} \frac{r!}{x! (r-x)!} \frac{(r-x)!}{z_2! \dots z_n!} \\ &= \frac{1}{n^r} \frac{r!}{x! (r-x)!} \sum_{(z_2, \dots, z_n) \in A_{n-1, r-x}} \frac{(r-x)!}{z_2! \dots z_n!}, \end{aligned}$$

da cui, ricordando l'**Osservazione 3**, o equivalentemente la formula della potenza del multinomio (95)

$$\begin{aligned} P(\{X_1 = x\}) &= \binom{r}{x} \frac{1}{n^r} (n-1)^{r-x} = \\ &= \binom{r}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{r-x} \quad \text{per } x = 0, 1, \dots, r. \end{aligned}$$

Cioè, come ci si doveva immaginare⁷⁰, la distribuzione marginale di X_1 è binomiale di parametri r e $\frac{1}{n}$.

Per quanto riguarda la distribuzione condizionata di X_2 data X_1 si ha innanzitutto che

$$\begin{aligned} P(\{X_1 = x_1, X_2 = x_2\}) &= \sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} \frac{\binom{r}{x_1 \ x_2 \ z_3 \ \dots \ z_n}}{n^r} \\ &= \frac{1}{n^r} \frac{r!}{x_1! x_2! (r-x_1-x_2)!} \sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} \frac{(r-x_1-x_2)!}{z_3! \dots z_n!} \\ &= \binom{r}{x_1 \ x_2 \ (r-x_1-x_2)} \left(\frac{1}{n}\right)^{x_1} \left(\frac{1}{n}\right)^{x_2} \left(1 - \frac{2}{n}\right)^{r-x_1-x_2}, \end{aligned}$$

per $x_1 \geq 0$, $x_2 \geq 0$ e $x_1 + x_2 \leq r$.

La distribuzione condizionata di X_2 dato il valore x_1 per X_1 (con ovviamente $0 \leq x_1 \leq r$) è data da

$$\begin{aligned} P(\{X_2 = x_2\} | \{X_1 = x_1\}) &= \frac{P(\{X_1 = x_1, X_2 = x_2\})}{P(\{X_1 = x_1\})} \\ &= \binom{r-x_1}{x_2} \left(\frac{1}{n-1}\right)^{x_2} \left(1 - \frac{2}{n-1}\right)^{r-x_1-x_2}; \end{aligned}$$

per $x_2 \geq 0$, $x_1 + x_2 \leq r$, ovvero per $0 \leq x_2 \leq r - x_1$. Si tratta cioè, come ci si doveva aspettare, di una distribuzione binomiale di parametri $r - x_1$ e $\frac{1}{n-1}$.

Analoghe formule si possono facilmente ottenere per quanto riguarda le distribuzioni marginali e condizionate di più di due fra le n variabili X_1, \dots, X_n , per le distribuzioni di probabilità delle loro somme parziali etc... ; è anche interessante vedere come tali formule si specializzino per i vari modelli di occupazione notevoli elencati in precedenza.

Non vale qui la pena di scrivere sistematicamente tali formule e tali risultati specifici, che possono invece costituire utili esercizi per il lettore.

12.8 Distribuzioni marginali e condizionate per la distribuzione multinomiale

Si supponga che $\mathbf{X} = (X_1, X_2, \dots, X_n)$ segua una distribuzione multinomiale di parametri $(r, n; p_1, p_2, \dots, p_n)$.

È facile convincersi che la distribuzione marginale di X_i è binomiale di parametri r e p_i , così come la distribuzione di $X_i + X_j$ è $\text{bin}(r, p_i + p_j)$, e così via.

⁷⁰Il modello di Maxwell-Boltzmann è un caso particolare del modello multinomiale, con $p_j = \frac{1}{n}$, per ogni $j = 1, 2, \dots, n$, che a sua volta deriva dal modello di prove ripetute ad n esiti. Interessarsi di X_1 significa controllare ad ogni prova solo se si è verificato l'esito di tipo 1, o no. Ovvero ci si riconduce al caso binomiale: numero di successi in r prove ripetute in cui solo due esiti sono possibili. Risulta evidente che, nel caso della distribuzione multinomiale, la distribuzione di X_1 sia allora binomiale di parametri (r, p_1) .

Infatti piuttosto che mettersi a fare i calcoli, basta pensare che X_i conta il numero di successi in r prove indipendenti in cui “successo” alla k -sima prova significa esito di tipo i alla k -sima prova (e “insuccesso” alla k -sima prova significa esito di tipo ℓ , per un $\ell \neq i$ alla k -sima prova). Analogamente $X_i + X_j$ conta il numero di successi in r prove indipendenti in cui però stavolta “successo” alla k -sima prova significa esito di tipo i oppure j alla k -sima prova (e “insuccesso” alla k -sima prova significa esito di tipo ℓ , per un $\ell \notin \{i, j\}$ alla k -sima prova).

Ancora se vogliamo calcolare invece la distribuzione condizionata di X_2 dato X_1 si procede in modo simile a quanto fatto per il modello di Maxwell-Boltzmann: si ha innanzitutto che

$$P(\{X_1 = x_1, X_2 = x_2\}) = \frac{r!}{x_1! x_2! (r - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2}, \quad (97)$$

per $x_1 \geq 0, x_2 \geq 0$ e $x_1 + x_2 \leq r$.

Il risultato non è sorprendente: infatti basta pensare che se si considerano solo gli esiti di tipo 1, di tipo 2 e di tipo “ne’ 1 ne’ 2” e si indica con $W = X_3 + \dots + X_n$ la variabile aleatoria che conta gli esiti di tipo “ne’ 1 ne’ 2”, allora immediatamente si è nel caso di una distribuzione multinomiale di parametri $(r, 3; p_1, p_2, 1 - (p_1 + p_2))$.

$$\begin{aligned} P(\{X_1 = x_1, X_2 = x_2\}) &= P(\{X_1 = x_1, X_2 = x_2, W = r - (x_1 + x_2)\}) \\ &= \frac{r!}{x_1! x_2! (r - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2}. \end{aligned}$$

Tuttavia anche senza questo discorso euristico, per dimostrare la (97) basta osservare che

$$\begin{aligned} &P(\{X_1 = x_1, X_2 = x_2\}) \\ &= \sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} \binom{r}{x_1 \ x_2 \ z_3 \ \dots \ z_n} p_1^{x_1} p_2^{x_2} p_3^{z_3} \dots p_n^{z_n} \\ &= \frac{r!}{x_1! x_2! (r - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} \sum_{(z_3, \dots, z_n) \in A_{n-2, r-x_1-x_2}} \frac{(r - x_1 - x_2)!}{z_3! \dots z_n!} p_3^{z_3} \dots p_n^{z_n} \\ &= \binom{r}{x_1 \ x_2 \ (r - x_1 - x_2)} p_1^{x_1} p_2^{x_2} (p_3 + \dots + p_n)^{r-x_1-x_2} \\ &= \frac{r!}{x_1! x_2! (r - x_1 - x_2)!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2}, \end{aligned}$$

per $x_1 \geq 0, x_2 \geq 0$ e $x_1 + x_2 \leq r$.

La distribuzione condizionata di X_2 dato il valore x_1 per X_1 (con ovviamente $0 \leq x_1 \leq r$) è data da

$$P(\{X_2 = x_2\} | \{X_1 = x_1\}) = \binom{r - x_1}{x_2} \left(\frac{p_2}{1 - p_1} \right)^{x_2} \left(1 - \frac{p_2}{1 - p_1} \right)^{r-x_1-x_2} \quad (98)$$

per $0 \leq x_2 \leq r - x_1$. Si tratta cioè, come ci si poteva aspettare, di una distribuzione binomiale di parametri $r - x_1$ e $\frac{p_2}{1 - p_1}$. Infatti le prove continuano ad essere indipendenti, ma oramai solo due tipi di esiti sono possibili l’esito di tipo 2 oppure non due. Inoltre si devono considerare solo le $r - x_1$ in cui non si è avuto esito di tipo 1. Infine la probabilità di esito 2 va valutata condizionatamente a sapere che su ciascuna di tali prove non è verificato un esito di tipo 1: in fondo questa è l’interpretazione di $\frac{p_2}{1 - p_1}$.

Tuttavia anche in questo caso non è necessario questo discorso euristico, e per dimostrare la (98) basta osservare che

$$\begin{aligned}
 P(\{X_2 = x_2\}|\{X_1 = x_1\}) &= \frac{P(\{X_1 = x_1, X_2 = x_2\})}{P(\{X_1 = x_1\})} \\
 &= \frac{\binom{r}{x_1 x_2 (r-x_1-x_2)} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2}}{\binom{r}{x_1} p_1^{x_1} (1 - p_1)^{r-x_1}} \\
 &= \frac{\frac{r!}{x_1! x_2! (r-x_1-x_2)!} p_1^{x_1} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2}}{\frac{r!}{x_1! (r-x_1)!} p_1^{x_1} (1 - p_1)^{r-x_1}} \\
 &= \frac{1}{x_2! (r-x_1-x_2)!} p_2^{x_2} (1 - (p_1 + p_2))^{r-x_1-x_2} \\
 &\quad \frac{1}{(r-x_1)! (1 - p_1)^{r-x_1}} \\
 &= \binom{r-x_1}{x_2} \left(\frac{p_2}{1-p_1}\right)^{x_2} \left(1 - \frac{p_2}{1-p_1}\right)^{r-x_1-x_2}
 \end{aligned}$$

per $x_2 \geq 0$, $x_1 + x_2 \leq r$, ovvero per $0 \leq x_2 \leq r - x_1$.

12.9 Esercizi di verifica

Esercizio 12.1. Sia $r = 6, n = 4$. Rappresentate la quaterna (1,2,2,1) sotto forma di un disegno del tipo in (90).

Esercizio 12.2. Calcolate la probabilità del risultato

$$* | * | * |$$

assumendo rispettivamente che valga il modello di Maxwell-Boltzmann, o di Bose-Einstein o di Fermi-Dirac.

Esercizio 12.3. Consideriamo i 120 studenti del primo anno ed i 3 esami del primo semestre (chiamiamoli A, B, C). Facciamo una statistica per rilevare qual è l'esame che è stato superato per primo da ciascuno studente (supponiamo che tutti abbiano superato almeno un esame) e poniamo $X_A =$ numero degli studenti del primo anno che hanno superato l'esame A come primo (o unico) esame; analogamente si definiscano X_B e X_C .

È ragionevole assumere un modello di Maxwell-Boltzmann per X_A, X_B, X_C ? Ed un modello multinomiale?

Esercizio 12.4. I 40 membri di un dipartimento devono votare per eleggere il direttore. Vi sono i 4 candidati A, B, C e D. Ogni elettore deve esprimere un solo voto e tutti votano (non vi sono schede bianche, per tradizione). X_A, X_B, X_C e X_D sono i voti riportati dai vari candidati.

Calcolare $P(\{X_A = 20, X_B = 5, X_C = 14, X_D = 1\})$ sotto l'ipotesi che si tratti di uno schema di Maxwell-Boltzmann.

Esercizio 12.5. Un gioco viene ripetuto 5 volte, fra i giocatori A, B, C e D, dove A, B e C hanno probabilità di vincere del 20% e D del 40%.

Indicando con X_A, X_B, X_C, X_D , rispettivamente, il numero delle vittorie di A, B, C, D sulle 5 volte, calcolare le probabilità degli eventi:

(a) $\{X_A = 2\}$

(b) $\{X_A + X_D = 5\}$

(c) $\{X_A + X_B = 3, X_C = 1, X_D = 1\}$.

Esercizio 12.6. Si consideri la colonna “vincente” della schedina del Totocalcio, nel caso in cui si attribuisce probabilità 0.5 al risultato 1, probabilità 0.2 al risultato 2 e probabilità 0.3 al risultato x, e il risultato di ciascuna partita è giudicato indipendente dai risultati delle altre (si veda il precedente Esempio 12.6).

- (a) Qual è la colonna più probabile? Quanto vale la sua probabilità?
- (b) Quanto vale la probabilità del risultato (1, 1, x, 2, 1, 1, 1, 1, x, x, 1, 2, x)?
- (c) Qual è la probabilità che vi siano 7 risultati 1, 2 risultati 2 e 4 risultati x?

Esercizio 12.7. Un esperimento, che può dar luogo, con uguali probabilità $p = \frac{1}{3}$ a tre diversi risultati, viene condotto per 10 successive volte, in modo indipendente una volta dall'altra e si indica con X_1, X_2, X_3 il numero di volte in cui, rispettivamente, si verifica il risultato 1, oppure 2, oppure 3.

- (a) Calcolare le probabilità dell'evento $\{X_2 + X_3 = 6\}$
- (b) Calcolare la probabilità condizionata $P(\{X_2 = 3\}|\{X_1 = 4\})$.
- (c) Calcolare la distribuzione di probabilità condizionata di X_2 dato l'evento $\{X_1 = 4\}$.

Esercizio 12.8. Calcolare la distribuzione marginale di X_1 e la distribuzione condizionata di X_2 data X_1 in un modello di Bose-Einstein con r oggetti e n siti ed in un modello di Fermi-Dirac con r oggetti e n siti ($r \leq n$).

Esercizio 12.9. Indichiamo con S il numero di successi in n prove bernoulliane di probabilità p e poniamo $T = n - S$.

Che cosa si ottiene come distribuzione congiunta di S e T ?

13 Spazi di probabilità e variabili aleatorie in casi più generali

13.0.1 Definizione generale di spazio di probabilità

Fin qui abbiamo trattato spazi di probabilità con un numero finito di eventi elementari e variabili aleatorie che possono assumere soltanto un numero finito di valori possibili.

Tale limitazione non consente però di trattare tutti i casi di possibile interesse e dobbiamo quindi estendere tali definizioni a casi più generali; in effetti il lettore era stato avvertito che, in attesa di riformulazioni più generali e definitive, i concetti di spazio di probabilità e di variabile aleatoria venivano introdotti in forma provvisoria.

Oramai, nelle precedenti lezioni, abbiamo familiarizzato con tali concetti e possiamo dunque passare ora a trattare tali casi più generali e a capire come e perchè le definizioni stesse vadano parzialmente modificate.

Per cominciare il discorso, rimaniamo però ancora nel caso di spazi di probabilità con un numero finito di eventi elementari e consideriamo quanto segue:

vi sono situazioni in cui non possiamo o comunque non siamo interessati ad assegnare la probabilità a tutti i sottoinsiemi dello spazio campione, in un esperimento aleatorio.

Esempio 13.1. *m lanci di un dado costituiscono un esperimento in cui Ω coincide con $\{1, 2, \dots, 6\}^m$. Pensiamo ora ad un gioco, basato su tali m lanci, in cui un giocatore vince o perde sulla base del valore della somma degli m successivi punteggi. Allora ci interesserà soltanto assegnare le probabilità a quei sottoinsiemi di Ω definiti in termini di tale somma.*

Esempio 13.2. *Pensiamo all'esperimento che consiste nel disporre, in modo aleatorio, r oggetti in n siti. Come si era visto nella precedente Lezione 12, possiamo schematizzare come spazio campione Ω l'insieme $\{1, 2, \dots, n\}^r$ costituito dalle applicazioni $\varphi : \{1, 2, \dots, r\} \rightarrow \{1, 2, \dots, n\}$. Supponiamo ora che gli oggetti siano indistinguibili: non vogliamo o non possiamo distinguere fra loro due diversi eventi elementari che prevedano un ugual numero di oggetti per ciascuno dei siti (che prevedano cioè uguali valori per i numeri di occupazione X_1, \dots, X_n).*

In questo caso ciascun evento, che possa essere effettivamente osservato, sarà una unione di eventi del tipo $\{X_1 = x_1, \dots, X_n = x_n\}$ (con $\sum_j x_j = r$).

Esempio 13.3. *Supponiamo di lanciare un dado sul quale abbiamo applicato dei bollini rossi, sopra i numeri pari, e blu, sopra i numeri dispari, in modo che sia possibile capire se è uscito un numero pari o dispari, ma non sia possibile sapere precisamente quale numero sia uscito, senza togliere il bollino.*

*È naturale prendere come spazio campione $\Omega = \{1, 2, 3, 4, 5, 6\}$. Tuttavia, **se non ci è permesso togliere i bollini**, gli unici sottoinsiemi per i quali possiamo affermare se si sono verificati oppure no sono*

$$\{\text{esce un numero pari}\} \quad \text{oppure} \quad \{\text{esce un numero dispari}\}$$

oltre ovviamente all'evento impossibile \emptyset e all'evento certo Ω .

In una situazione del tipo descritto nei due precedenti esempi, indichiamo con \mathcal{F} la famiglia dei sottoinsiemi di Ω a cui viene attribuita una probabilità. **Soltanto gli elementi di \mathcal{F} verranno chiamati eventi: non tutti i sottoinsiemi di Ω sono dunque degli eventi, bensì soltanto quelli cui viene attribuita una probabilità;** nella trattazione che abbiamo esposto all'inizio si aveva $\mathcal{F} = \mathcal{P}(\Omega)$, mentre ora poniamo $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, prevedendo anche situazioni in cui la famiglia \mathcal{F} sia

strettamente contenuta in $\mathcal{P}(\Omega)$.

Oltre ad essere $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, quali altre proprietà dovrà avere \mathcal{F} ?

È ragionevole richiedere che, dati due arbitrari eventi A e B (cioè due sottoinsiemi di Ω appartenenti a \mathcal{F}), anche $A \cup B$, $A \cap B$, \overline{A} , \overline{B} siano degli eventi (cioè sottoinsiemi appartenenti a \mathcal{F}).

Si richiederà inoltre che i sottoinsiemi banali di Ω , cioè \emptyset (evento impossibile) e Ω (evento certo), facciano parte di \mathcal{F} .

Assumeremo dunque che \mathcal{F} sia un'algebra (secondo la Definizione 5.5 della Lezione 5).

Passiamo ora a considerare il caso in cui lo spazio campione Ω sia un insieme infinito.

Situazioni di tale tipo si presentano necessariamente quando si debbano considerare infiniti eventi diversi, o variabili aleatorie che prendano valore in un insieme infinito, o successioni di variabili aleatorie fra loro diverse.

Estendendo quanto svolto per il caso con Ω finito e tenendo conto delle precedenti considerazioni, anche nel caso generale uno spazio di probabilità è definito come una terna (Ω, \mathcal{F}, P) dove Ω è un arbitrario spazio di punti, \mathcal{F} è un'algebra di sottoinsiemi di Ω (gli eventi) e $P : \mathcal{F} \rightarrow [0, 1]$ è una funzione di insieme con la proprietà di additività⁷¹ e tale che $P(\Omega) = 1$.

Vi è però ora qualcosa da precisare in merito a degli aspetti che si presentano soltanto nel caso di Ω infinito.

Supponiamo che E_1, E_2, \dots sia una successione di eventi, cioè di elementi di \mathcal{F} ; sappiamo che l'unione finita di un numero arbitrario di tali eventi dovrà ancora essere un evento (avendo assunto che \mathcal{F} sia un'algebra). Ciò non garantisce però che anche l'unione numerabile $\bigcup_{j=1}^{\infty} E_j$ sia un elemento di \mathcal{F} . Se \mathcal{F} , oltre ad essere un'algebra, soddisfa anche tale condizione allora diremo che \mathcal{F} è una σ -algebra. Diamo cioè la seguente

Definizione 13.1 (σ -algebra). *Una famiglia \mathcal{F} di sottoinsiemi di Ω è detta σ -algebra se sono verificate le condizioni seguenti*

i) $\Omega \in \mathcal{F}$

ii) $E \in \mathcal{F} \Rightarrow \overline{E} \in \mathcal{F}$

iii) $E_1, E_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{j=1}^{\infty} E_j \in \mathcal{F}$

Ovviamente $\mathcal{P}(\Omega)$ è sempre una σ -algebra.

⁷¹Per comodità del lettore diamo una definizione formale di spazio di probabilità, nel caso finitamente additivo, in un modo che permette di evidenziare la connessione con la successiva Definizione 13.3.

Uno spazio di probabilità (finitamente additiva) è una terna (Ω, \mathcal{F}, P) con \mathcal{F} un'algebra, cioè tale che

i) $\Omega \in \mathcal{F}$

ii) se $A \in \mathcal{F}$, allora $\overline{A} \in \mathcal{F}$

iii) se $A_k \in \mathcal{F}$, per $k = 1, \dots, n$, allora $\bigcup_{k=1}^n A_k \in \mathcal{F}$

e $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ una funzione tale che

i) per ogni $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$

ii) $\mathbb{P}(\Omega) = 1$

iii) se $A_k \in \mathcal{F}$, per $k = 1, \dots, n$, e $A_h \cap A_k = \emptyset$, allora

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \mathbb{P}(A_k).$$

Inoltre se $E_1, E_2, \dots \in \mathcal{F}$ è una successione di eventi allora $\bigcap_{j=1}^{\infty} E_j \in \mathcal{F}$, come si vede subito applicando la formula di de Morgan

$$\bigcap_{j=1}^{\infty} E_j = \overline{\bigcup_{j=1}^{\infty} \overline{E_j}} \quad (99)$$

e le proprietà *ii*) e *iii*).

Consideriamo allora una successione di eventi E_1, E_2, \dots e assumiamo che \mathcal{F} sia una σ -algebra; dunque $\bigcup_{j=1}^{\infty} E_j \in \mathcal{F}$ e possiamo considerare la probabilità

$$P\left(\bigcup_{j=1}^{\infty} E_j\right).$$

Supponiamo ora che gli eventi E_1, E_2, \dots siano a due a due disgiunti e consideriamo la serie

$$\sum_{j=1}^{\infty} P(E_j);$$

possiamo chiederci allora se debba o meno valere l'identità

$$P\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} P(E_j); \quad (100)$$

osserviamo infatti che essa non è implicata dalla proprietà di additività finita.

Possiamo assumere la (100) come un ulteriore assioma imposto alla P , cioè che questa risulti numerabilmente additiva secondo la seguente

Definizione 13.2. Sia $P : \mathcal{F} \rightarrow [0, 1]$ una funzione di insieme. Diremo che P è **numerabilmente additiva** (o **σ -additiva**) se per qualunque successione E_1, E_2, \dots , con $E_j \in \mathcal{F}$, incompatibili a due a due, cioè $E_i \cap E_j = \emptyset$ per ogni $i \neq j$, risulta verificata la (100).

A seconda che si imponga o meno tale condizione sulla P , verremo dunque a costruire due diverse teorie della probabilità: una più "forte" (cioè in cui è possibile ricavare un numero maggiore di risultati) in cui vale la σ -additività della P ed una più "debole", ma più generale, in cui si considerano soltanto quei risultati che si possono ottenere imponendo che P sia finitamente additiva.

Gran parte della attuale letteratura sulla teoria e le applicazioni della probabilità danno per scontato l'assioma dell'additività numerabile.

Anche in questi appunti il termine "misura di probabilità" verrà d'ora in poi utilizzato, salvo avviso in contrario, per designare esclusivamente il caso σ -additivo; e giungiamo così alla seguente definizione (questa volta definitiva) di spazio di probabilità.

Definizione 13.3 (spazio di probabilità). *Uno spazio di probabilità è una terna (Ω, \mathcal{F}, P) dove*

1. Ω è un arbitrario spazio di punti,
2. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è una σ -algebra di sottoinsiemi di Ω
3. $P : \mathcal{F} \rightarrow [0, 1]$ è una misura di probabilità, cioè una funzione di insieme σ -additiva tale che $P(\Omega) = 1$.

Osservazione 1. È possibile trovare dei controesempi che mostrano quanto segue: esistono degli insiemi Ω tali che non è possibile costruire alcuna misura di probabilità (cioè σ -additiva) che risulti ben definita su tutto $\mathcal{P}(\Omega)$. Sulla stessa retta reale, la misura che associa agli intervalli la loro lunghezza non può essere coerentemente estesa a tutti i sottoinsiemi sotto il vincolo che valga la σ -additività (si veda ad esempio D. Billingsley “Probability and Measure”, Ed. Wiley and Sons, 1995).

Possiamo dunque dire in altre parole che, nel caso in cui si assume la σ -additività, il fatto di limitare il dominio della probabilità ad una sotto σ -algebra $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ (invece che a tutto $\mathcal{P}(\Omega)$) può diventare in molti casi una necessità piuttosto che una questione di scelta, come invece avviene nel caso finito.

Osservazione 2. (Spazi di probabilità numerabili) Sia (Ω, \mathcal{F}, P) uno spazio di probabilità con $\Omega = \{\omega_i; i = 1, 2, \dots\}$ numerabile e $\mathcal{F} = \mathcal{P}(\Omega)$ ed assumiamo che P sia σ -additiva. Notiamo che, analogamente a quanto accade nel caso di uno spazio di probabilità finito, $P(E)$ risulta univocamente determinata $\forall E \in \mathcal{F}$, una volta fissate le probabilità degli eventi elementari $P(\{\omega_i\})$, $i = 1, 2, \dots$; si avrà infatti

$$P(E) = \sum_{i: \omega_i \in E} P(\{\omega_i\}).$$

Ovviamente è necessaria anche la condizione che

$$P(\{\omega_i\}) \geq 0 \quad \forall i, \quad e \quad \sum_{i=1}^{\infty} P(\{\omega_i\}) = 1.$$

Tale circostanza cessa invece di valere nel caso in cui P sia finitamente additiva ma non σ -additiva. È anche interessante a tale proposito considerare il seguente

Esempio 13.4 (Controesempio). Consideriamo il caso in cui lo spazio Ω coincida con l'insieme dei numeri naturali: $\Omega \equiv \{1, 2, \dots\}$. Non è possibile definire alcuna probabilità P σ -additiva su $(\Omega, \mathcal{P}(\Omega))$ in modo che risulti

$$P(\{i\}) = c, \quad i = 1, 2, \dots$$

essendo c una costante indipendente da i . Infatti, se così fosse, dovrebbe risultare

$$1 = P(\Omega) = \sum_{i=1}^{\infty} c$$

Ma una serie a termini costanti $\sum_{i=1}^{\infty} c$ può convergere se e solo se si pone $c = 0$; e tale posizione implicherebbe $P(\Omega) = 0 \neq 1$.

Nella Lezione 2 avevamo verificato, nel caso di spazi di probabilità finiti, la validità di alcune proprietà fondamentali della probabilità quali conseguenze immediate degli assiomi; fra tali proprietà, troviamo in particolare la proprietà di monotonia: per $A, B \in \mathcal{P}(\Omega)$

$$A \subseteq B \quad \implies \quad P(A) \leq P(B).$$

Nel caso di spazi di probabilità infiniti l'aggiunta dell'assioma della σ -additività oltre a quello dell'additività finita garantisce che continuino a valere tutte le proprietà elencate⁷² nella Lezione 2.

⁷²Tutte queste proprietà continuano a valere anche negli spazi di probabilità più generali, in quanto si basano sulla proprietà di additività. L'unica accortezza consiste nello specificare sempre che si richiede che gli insiemi in considerazione appartengano ad \mathcal{F} : ad esempio nella precedente proprietà di monotonia va specificato che $A, B \in \mathcal{F}$, in modo che abbia senso calcolare $P(A)$ e $P(B)$.

In verità non occorre “aggiungere” - basta bensì sostituire - l’assioma della σ -additività a quello relativo all’additività finita, in quanto la σ -additività rende automaticamente verificata anche la additività finita, come verrà mostrato nella successiva Proposizione 2.

L’assioma della σ -additività, inoltre, dà luogo ad un ulteriore importante conseguenza: la proprietà di “continuità” della probabilità. Si ha cioè la seguente

Proposizione 1 In uno spazio di probabilità (Ω, \mathcal{F}, P) siano dati, per $n = 1, 2, \dots$, $A_n \in \mathcal{F}, B_n \in \mathcal{F}$ tali che

$$A_n \subseteq A_{n+1}; \quad B_n \supseteq B_{n+1}$$

e poniamo

$$A \equiv \bigcup_{n=1}^{\infty} A_n; \quad B \equiv \bigcap_{n=1}^{\infty} B_n.$$

Se P è σ -additiva allora risulta

$$P(A) = \lim_{n \rightarrow \infty} P(A_n); \quad P(B) = \lim_{n \rightarrow \infty} P(B_n)$$

Dimostrazione. Innanzitutto notiamo quanto segue:

In virtù della proprietà di monotonia della P , $\{P(A_n)\}_{n=1,2,\dots}$ è una successione non decrescente ed è limitata superiormente (in quanto $P(A_n) \leq 1$); dunque certamente esiste $\lim_{n \rightarrow \infty} P(A_n)$; analogamente esiste $\lim_{n \rightarrow \infty} P(B_n)$.

Inoltre, essendo \mathcal{F} una σ -algebra, risulta $A \in \mathcal{F}$ e quindi esiste $P(A)$; analogamente esiste $P(B)$.

Poniamo ora

$$C_1 = A_1, \quad C_2 = A_2 \setminus A_1 = A_2 \cap \bar{A}_1, \quad C_3 = A_3 \setminus A_2 = A_3 \cap \bar{A}_2, \dots$$

È facile verificare che $\forall n \in \mathbb{N}$, C_1, \dots, C_n sono a due a due disgiunti e che $A_n = \bigcup_{i=1}^n C_i$. Ne segue quindi, $\forall n \in \mathbb{N}$,

$$P(A_n) = \sum_{i=1}^n P(C_i).$$

Inoltre possiamo scrivere $A \equiv \bigcup_{n=1}^{\infty} C_n$ da cui segue, in virtù del fatto che P è σ -additiva,

$$P(A) = \sum_{i=1}^{\infty} P(C_i);$$

cioè, per definizione di somma di una serie,

$$P(A) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(C_i) = \lim_{n \rightarrow \infty} P(A_n).$$

La relazione

$$P(B) = \lim_{n \rightarrow \infty} P(B_n).$$

Si verifica in modo analogo, oppure si può verificare a partire dalla relazione $P(A) = \lim_{n \rightarrow \infty} P(A_n)$, definendo $A_n = \bar{B}_n$ e utilizzando la formula di De Morgan (99).

Prima di concludere questo sottoparagrafo, è opportuno prestare attenzione alla seguente osservazione, che riveste un'importanza fondamentale per la comprensione di alcuni successivi punti.

Osservazione 4. In uno spazio di probabilità finito, si può assumere, senza effettiva perdita di generalità, che tutti gli eventi elementari abbiano probabilità positiva; infatti, se un evento elementare avesse probabilità nulla, potremmo tranquillamente escluderlo dalla lista degli eventi possibili; notiamo inoltre che, anche eliminando tutti gli eventi di probabilità nulla, la probabilità dell'unione di tutti gli eventi elementari rimasti si mantiene uguale ad 1.

Dunque, nel caso finito, è ragionevole considerare situazioni in cui tutti gli eventi, siano essi elementari o composti, hanno tutti probabilità strettamente positiva, con la sola esclusione dell'evento impossibile.

Nel caso di uno spazio di probabilità (Ω, \mathcal{F}, P) con \mathcal{F} di potenza maggiore del numerabile ciò invece non è più necessariamente vero: in generale vi saranno degli eventi di probabilità nulla, ma che risultano possibili; infatti il generico evento elementare avrà probabilità zero, tranne al più per un insieme finito o numerabile di eventi elementari.

Notiamo che potrebbe anche accadere che tutti gli eventi elementari abbiano singolarmente probabilità zero e che gli unici eventi di probabilità strettamente positiva siano eventi composti da un'unione non numerabile di eventi elementari (ciò è connesso con il fatto che la proprietà di additività numerabile, che assicura che la probabilità di un'unione numerabile di eventi a due a due disgiunti sia uguale alla serie delle loro probabilità, non è estendibile al caso di unioni più che numerabili). In tal caso si ha che, eliminando tutti gli eventi di probabilità nulla, resta soltanto l'insieme vuoto.

Questo è un punto un pò delicato che potrebbe all'inizio apparire poco chiaro.

È opportuno in proposito pensare ad una analogia (che in effetti è molto di più che un'analogia) con la geometria elementare: sulla retta reale l'intervallo $[0, 1]$, che ha lunghezza finita uguale a 1, è composto da un'unione (più che numerabile) di punti, tutti di lunghezza 0; addirittura la stessa retta reale, che ha lunghezza infinita, è composta da un'unione (più che numerabile) di punti (tutti di lunghezza 0).

A tale stesso proposito è utile guardare con attenzione la successiva **Osservazione 7**.

Inseriamo qui, solo per lettori particolarmente interessati, la dimostrazione della seguente proposizione, cui si era già accennato in precedenza.

Proposizione 2 Se P è σ -additiva allora essa è anche, necessariamente, finitamente additiva.

Dimostrazione. Siano E_1, \dots, E_n sottoinsiemi appartenenti a \mathcal{F} e a due a due disgiunti; ponendo

$$E_{n+1} = \emptyset, \quad E_{n+2} = \emptyset, \dots, \quad E_{n+k} = \emptyset, \dots$$

otteniamo che anche la successione $\{E_i\}_{i=1,2,\dots}$ è composta di insiemi a due a due disgiunti.

D'altra parte si ha che

$$\bigcup_{i=1}^n E_i = \bigcup_{i=1}^{\infty} E_i$$

e dunque

$$P\left(\bigcup_{i=1}^n E_i\right) = P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

La dimostrazione si conclude basandosi sull'identità $P(\emptyset) = 0$, da cui possiamo far seguire

$$\sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^n P(E_i).$$

Ora osserviamo però che l'identità $P(\emptyset) = 0$ era stata ottenuta, come avevamo visto nella Lez. 2, dall'assioma $P(\Omega) = 1$ e da quello dell'additività finita della P . Non possiamo quindi darla per scontata a priori in questa dimostrazione, in cui si tratta proprio di verificare la validità della additività finita.

Possiamo però notare che $P(\emptyset) = 0$ segue immediatamente anche dall'assunzione che P sia σ -additiva.

Possiamo infatti scrivere, ponendo $P(\emptyset) = c$, $O_i = \emptyset, i = 1, 2, \dots$,

$$\emptyset = \bigcup_{i=1}^{\infty} O_i$$

da cui segue

$$c = \sum_{i=1}^{\infty} c$$

in quanto $O_i, i = 1, 2, \dots$, sono a due a due disgiunti e possiamo dunque concludere che deve essere $c = 0$, in quanto una serie a termini costanti può risultare convergente solo se i suoi termini sono tutti nulli.

13.1 Definizione generale di variabile aleatoria

Continuando secondo lo schema introdotto nella Lezione n. 7, consideriamo una variabile aleatoria come un'applicazione a valori reali e definita su uno spazio campione, cioè su un insieme di eventi elementari, Ω .

Consideriamo una generica applicazione $X : \Omega \rightarrow \mathbb{R}$.

Come abbiamo già visto, nel caso in cui Ω è un'insieme di cardinalità finita, allora necessariamente anche il codominio $X(\Omega)$ è un insieme finito, che finora abbiamo indicato, ad esempio, con il simbolo $\{x_1, \dots, x_n\}$.

Nel caso in cui Ω sia un insieme di cardinalità arbitraria, $X(\Omega)$ non è più necessariamente un insieme finito: $X(\Omega)$ potrebbe anche essere costituito da una successione $\{x_1, x_2, \dots\}$ o addirittura da un intervallo (limitato o illimitato) di numeri reali.

Ciò non è da vedere come un problema, piuttosto si tratta di una possibilità in accordo con le esigenze della teoria della probabilità. Difatti, come già detto, per motivi di tipo sia teorico che applicativo a noi serve poter considerare variabili aleatorie a valori in insiemi numerabili o con la potenza del continuo.

Esempio 13.5. Sia $\{E_1, E_2, \dots\}$ una successione di eventi e siano X_1, X_2, \dots le corrispondenti variabili indicatrici.

Consideriamo la variabile aleatoria

$$T \equiv \inf\{n \geq 1 : X_n = 1\}$$

Dunque l'evento $\{T = n\}$ è equivalente all'evento

$$\{X_1 = 0, X_2 = 0, \dots, X_{n-1} = 0, X_n = 1\}$$

e possiamo interpretare T come il numero (aleatorio) di prove necessarie fino ad ottenere il primo successo, nella successione di prove $\{E_1, E_2, \dots\}$.

L'insieme dei valori possibili per T coincide con l'insieme dei numeri naturali⁷³ $\{1, 2, \dots\}$.

Esempio 13.6. In un test di affidabilità un'apparecchiatura elettrica appena prodotta viene testata lasciandola funzionare ininterrottamente fino a quando non smetta di funzionare a causa di qualche guasto. Indicando con T la lunghezza complessiva del tempo di durata realizzato dall'apparecchiatura, otteniamo una variabile aleatoria i cui valori possibili sono in linea di principio i valori reali positivi.

Consideriamo ora specificamente il caso di una variabile aleatoria X definita su Ω e a valori in un insieme numerabile $X(\Omega) \equiv \{x_1, x_2, \dots\}$ e sia \mathcal{F} la σ -algebra costituita da quei sottoinsiemi di Ω che consideriamo come eventi, cioè su cui è definita la misura di probabilità P .

Fissiamo ora un generico elemento $x_j \in X(\Omega)$ e guardiamo al sottoinsieme

$$X^{-1}(\{x_j\}) \equiv \{\omega \in \Omega : X(\omega) = x_j\}$$

Naturalmente vogliamo che $X^{-1}(\{x_j\})$ sia effettivamente un evento, cioè vogliamo poter parlare della "probabilità" che X assuma il valore x_j e possiamo far questo se risulta

$$X^{-1}(\{x_j\}) \in \mathcal{F}. \quad (101)$$

Per tale motivo diremo che l'applicazione $X : \Omega \rightarrow X(\Omega) \equiv \{x_1, x_2, \dots\}$ costituisce effettivamente una variabile aleatoria solo se la condizione (101) risulta verificata, $\forall x_j \in X(\Omega)$. Notiamo ora che, se X è una tale variabile aleatoria, anche un sottoinsieme del tipo $\{\omega \in \Omega : X(\omega) \leq z\}$ (o, scrivendo più brevemente, $\{X \leq z\}$) risulta appartenere a \mathcal{F} , $\forall z \in \mathbb{R}$, in quanto

$$\{X \leq z\} = \bigcup_{j: x_j \leq z} \{X = x_j\}.$$

Esercizio proposto 13.1. Dimostrare che se \mathcal{F} è una σ -algebra allora $X : \Omega \rightarrow X(\Omega) \equiv \{x_1, x_2, \dots\}$ è una variabile aleatoria se e solo⁷⁴ se risulta $\{\omega \in \Omega : X(\omega) \leq z\} \in \mathcal{F}$, $\forall z \in \mathbb{R}$.

Consideriamo ora il caso in cui $X(\Omega)$ è un intervallo (a, b) della retta reale.

Anche in questo caso richiederemo, affinché X sia una variabile aleatoria ben definita, che risulti $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.

Osserviamo però che nel presente caso tale condizione non implica $\{\omega \in \Omega : X(\omega) \leq z\} \in \mathcal{F}$, $\forall z \in \mathbb{R}$; infatti un tale insieme non è più ottenibile, in generale, attraverso un'operazione di unione numerabile a partire da insiemi del tipo $\{\omega \in \Omega : X(\omega) = x\}$.

Giungiamo quindi alla seguente

Definizione 13.4 (variabile aleatoria). Un'applicazione $X : \Omega \rightarrow X(\Omega) \subseteq \mathbb{R}$ è una variabile aleatoria se un qualunque sottoinsieme del tipo $\{\omega \in \Omega : X(\omega) \leq z\}$ risulta appartenere a \mathcal{F} , per ogni $z \in \mathbb{R}$.

⁷³Il lettore più attento avrà notato che c'è qualche problema nel caso in cui

$$\{X_1 = 0, X_2 = 0, \dots, X_{n-1} = 0, X_n = 0, X_{n+1} = 0, \dots\} = \bigcap_{n=1}^{\infty} \{X_n = 0\}.$$

Infatti in tale caso T va definita in modo appropriato. Un modo potrebbe essere quello di includere il valore ∞ tra i valori che può assumere T , e allora bisognerebbe considerare anche le variabili aleatorie che assumono valori in $\mathbb{R} \cup \{\infty\}$. Un altro modo per eliminare il problema è invece di notare che l'evento $\bigcap_{n=1}^{\infty} \{X_n = 0\}$ ha probabilità nulla, e quindi la definizione di T su tale evento non cambia in alcun modo la probabilità degli eventi $\{T = k\}$ per $k \in \{1, 2, \dots\}$.

⁷⁴In riferimento alla nota precedente: se $X(\Omega)$ include anche il valore ∞ , allora la condizione che X sia una variabile aleatoria diviene $\{\omega \in \Omega : X(\omega) \leq z\} \in \mathcal{F}$, $\forall z \in \mathbb{R} \cup \{+\infty\}$.

13.2 Distribuzioni di probabilità, funzioni di distribuzione

È opportuno iniziare la trattazione di questo argomento con l'analisi del caso di variabili aleatorie (che diremo **discrete**), per metterne in evidenza alcune analogie con il caso di variabili aleatorie a valori in un insieme finito.

Sia dunque (Ω, \mathcal{F}, P) uno spazio di probabilità e sia X una variabile aleatoria

$$X : \Omega \rightarrow X(\Omega)$$

con $X(\Omega)$ insieme discreto di numeri reali:

$$X(\Omega) \equiv \{x_1, x_2, \dots\}.$$

Possiamo considerare a questo punto (analogamente a quanto avevamo già fatto nella Lezione n. 7) un nuovo spazio di probabilità $(X(\Omega), \mathcal{G}, P_X)$ dove \mathcal{G} coincide con $\mathcal{P}(X(\Omega))$ e P_X è la misura di probabilità su \mathcal{G} , individuata univocamente dalla seguenti posizioni

$$P_X(\{x_j\}) = P(\{X = x_j\}) = P(X^{-1}(\{x_j\})) = P(\{\omega \in \Omega : X(\omega) = x_j\})$$

$$P_X(E) = \sum_{i: x_i \in E} P_X(\{x_i\}), \quad \forall E \in \mathcal{G}$$

(si vedano a tale proposito anche le considerazioni svolte nella precedente **Osservazione 2**).

Tale misura di probabilità P_X è la misura indotta da X ed è anche detta **distribuzione di probabilità** di X .

Dobbiamo ricordare che siamo giustificati a considerare la probabilità dell'evento $P(X^{-1}(\{x_j\}))$ in quanto, proprio per definizione della nozione di variabile aleatoria, deve risultare valida la condizione (101).

Dal momento che la famiglia degli eventi della forma $X^{-1}(\{x_j\}) \equiv \{\omega \in \Omega : X(\omega) = x_j\}$, per $j = 1, 2, \dots$ costituisce una partizione numerabile di insiemi appartenenti ad \mathcal{F} , in virtù dell'assioma di additività numerabile di P , deve risultare

$$\sum_{j=1}^{\infty} P(X = x_j) = 1. \quad (102)$$

Possiamo riassumere dicendo che, analogamente al caso finito, la distribuzione di probabilità di una variabile aleatoria discreta X può essere espressa assegnando l'insieme dei valori possibili $X(\Omega) = \{x_1, x_2, \dots\}$ e le loro rispettive probabilità $P(X = x_1), P(X = x_2), \dots$

Le probabilità $P(X = x_1), P(X = x_2), \dots$ devono essere delle quantità non negative e deve essere rispettata la condizione di normalizzazione (102). $P(X = x)$, vista come funzione della variabile x , $x \in X(\Omega)$, viene chiamata "funzione di densità discreta" (o più brevemente "densità discreta" della variabile aleatoria X).

Esempio 13.7. Consideriamo di nuovo una successione di eventi $\{E_1, E_2, \dots\}$ ed assumiamo in particolare che, $\forall n \in \mathbb{N}$, E_1, E_2, \dots, E_n , costituiscano delle prove bernoulliane, di probabilità θ ($0 < \theta < 1$). Siano X_1, X_2, \dots le corrispondenti variabili indicatrici e sia T la variabile aleatoria definita da

$$T \equiv \inf\{n \geq 1 : X_n = 1\}$$

Come già osservato nell'Esempio 13.5, l'evento $\{T = n\}$ è equivalente all'evento

$$\{X_1 = 0, X_2 = 0, \dots, X_{n-1} = 0, X_n = 1\};$$

Da ciò segue ora

$$P(T = n) = (1 - \theta)^{n-1} \theta, \quad n = 1, 2, \dots$$

Si dice che T segue **una distribuzione geometrica di parametro** θ . Osserviamo che risulta

$$\sum_{r=0}^{\infty} P(T = r) = \theta \sum_{r=0}^{\infty} (1 - \theta)^r = \theta \frac{1}{1 - (1 - \theta)} = 1. \quad (103)$$

Nel caso in cui si tratti con una variabile aleatoria X tale che $X(\Omega)$ è un intervallo limitato o illimitato di numeri reali (quindi $X(\Omega)$ non ha più la potenza del discreto), potremo dire che **la distribuzione di probabilità di** X può essere specificata assegnando la probabilità di tutti gli eventi del tipo $\{X \in I\}$ dove I è un generico intervallo della retta⁷⁵

$$P_X(I) = P(\{X \in I\}).$$

Notiamo infatti che nel presente caso non ha più senso “elencare” tutti i valori possibili per X (cioè tutti i singoli elementi appartenenti a $X(\Omega)$) e specificare le loro rispettive probabilità. Fra l'altro, per $x \in X(\Omega)$, la probabilità $P(X = x)$ sarà uguale a 0 tranne al più che per un insieme numerabile di valori di x ; ciò verrà verificato fra breve (**Osservazione 7**) e può comunque essere intuito ricordando quanto accennato nella precedente **Osservazione 4**.

Già sappiamo che, essendo X una variabile aleatoria (si veda la precedente Definizione 13.4), tutti i sottoinsiemi del tipo

$$\{\omega \in \Omega : X(\omega) \leq x\}, \quad x \in \mathbb{R}$$

sono degli eventi ed ha senso quindi considerare $P(X \leq x)$.

Definizione 13.5 (funzione di distribuzione). *La funzione $F_X : \mathbb{R} \rightarrow [0, 1]$ definita dalla relazione*

$$F_X(x) \equiv P(X \leq x)$$

*viene detta **funzione di distribuzione** della variabile aleatoria X .*

In altre parole $F_X(x)$ esprime **la probabilità che X cada nell'intervallo** $(-\infty, x]$. Se conosciamo dunque la distribuzione di probabilità di X , possiamo in particolare ricavare $F_X(x)$, $\forall x \in \mathbb{R}$.

Viceversa assegnare la funzione di distribuzione di una variabile aleatoria X equivale a conoscere interamente la distribuzione di probabilità di X , come si può verificare facilmente:

- Possiamo infatti dire di conoscere la distribuzione di probabilità di X quando sappiamo assegnare la probabilità di tutti gli eventi del tipo $\{X \in I\}$, dove I è un generico intervallo della retta.

- In effetti è possibile calcolare $P_X(I) = P(\{X \in I\})$ in termini della funzione $F_X(x)$, e in particolare per ogni coppia di valori $a < b \in \mathbb{R}$, risulta

$$P(a < X \leq b) = F_X(b) - F_X(a). \quad (104)$$

Possiamo scrivere infatti

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

⁷⁵Gli intervalli I possono essere limitati o illimitati, aperti o chiusi, sia a destra che a sinistra:

$$(a, b), \quad (a, b), \quad [a, b), \quad [a, b], \\ (-\infty, b], \quad (-\infty, b), \quad (a, \infty), \quad [a, -\infty).$$

dove i due eventi $\{X \leq a\}$ e $\{a < X \leq b\}$ sono, ovviamente, fra loro incompatibili; dunque

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

da cui la (104) segue immediatamente ricordando la definizione di F_X .

Analogamente si può verificare che risulta⁷⁶

$$P(a < X < b) = F_X(b) - F_X(a) - P(X = b) \quad (105)$$

$$P(a \leq X < b) = F_X(b) - F_X(a) - P(X = b) + P(X = a) \quad (106)$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X = a). \quad (107)$$

D'ora in poi, quindi, la distribuzione di probabilità di una variabile aleatoria X verrà specificata assegnando la funzione $F_X(x)$.

Osservazione 5. Nel caso di una variabile aleatoria X con $X(\Omega)$ coincidente con un intervallo, la funzione di distribuzione $F_X(x)$ è lo strumento naturale per individuare la sua distribuzione di probabilità P_X . Ma, indipendentemente dal fatto che $X(\Omega)$ sia un insieme finito, un insieme numerabile, oppure un intervallo, è sempre possibile definire la $F_X(x)$.

Supponiamo ora in particolare che risulti $X(\Omega) \equiv \{x_1, \dots, x_n\}$, con $x_1 < x_2 < \dots < x_n$ e $P(X = x_1) = p_1, \dots, P(X = x_n) = p_n$. Si ha allora

$$F_X(x) = \begin{cases} 0 & \text{per } x < x_1 \\ p_1 & \text{per } x_1 \leq x < x_2 \\ p_1 + p_2 & \text{per } x_2 \leq x < x_3 \\ \dots & \dots \\ \dots & \dots \\ \sum_{j=1}^{n-1} p_j & \text{per } x_{n-1} \leq x < x_n \\ 1 & \text{per } x \geq x_n \end{cases}$$

Vediamo dunque che, in questo caso, $F_X(x) : \mathbb{R} \rightarrow [0, 1]$ è una funzione costante a tratti, con salti positivi, rispettivamente di ampiezza p_1, p_2, \dots, p_n , nei punti x_1, x_2, \dots, x_n . Ne segue che $F_X(x)$ è una funzione continua da destra e monotona non decrescente.

⁷⁶È possibile anche dare delle formule per ottenere le probabilità degli eventi del tipo precedente, che dipendono solo da F . Infatti è possibile dimostrare (si veda più avanti la formula (109)) che, indicando come consueto con $F_X(x^-)$ il limite da sinistra di F , ovvero

$$F_X(x^-) := \lim_{y \rightarrow x^-} F_X(y)$$

si ha

$$P(X < x) = F_X(x^-).$$

Da ciò è immediato ottenere che $P(X = x) = F_X(x) - F_X(x^-)$ (si veda più avanti l'**Osservazione 7**) e che

$$P(a < X < b) = F_X(b^-) - F_X(a)$$

$$P(a \leq X < b) = F_X(b^-) - F_X(a^-)$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a^-).$$

Esempio 13.8. Sia Z è una variabile aleatoria discreta finita, con

$$Z(\Omega) \equiv \{z_1 = -1, z_2 = 0, z_3 = 1\}$$

e con

$$P(Z = -1) = P(Z = 0) = P(Z = 1) = \frac{1}{3}.$$

Si ha allora

$$F_Z(x) = \begin{cases} 0 & \text{per } x < -1 \\ \frac{1}{3} & \text{per } -1 \leq x < 0 \\ \frac{2}{3} & \text{per } 0 \leq x < 1 \\ 1 & \text{per } x \geq 1 \end{cases}$$

il cui grafico è riportato in Figura 1.

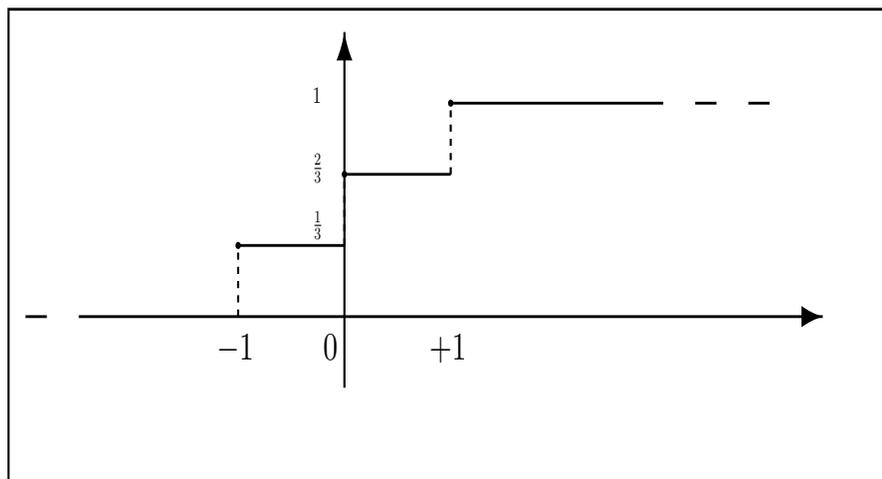


Figura 1: Grafico di $F_Z(x)$

Consideriamo ora una variabile aleatoria X arbitraria, cioè con $X(\Omega)$ non necessariamente finito. Oltre alla ovvia proprietà $0 \leq F_X(x) \leq 1$, vale in generale il seguente risultato

Proposizione 3 La funzione di distribuzione F_X deve necessariamente soddisfare le seguenti proprietà:

- (i) F_X è non decrescente
- (ii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$, $\lim_{x \rightarrow \infty} F_X(x) = 1$
- (iii) F_X è continua da destra, cioè, $\forall \tilde{x} \in \mathbb{R}$, risulta

$$\lim_{x \rightarrow \tilde{x}^+} F_X(x) = F_X(\tilde{x}).$$

Dimostrazione

- (i) Questa proprietà segue immediatamente dalla relazione (104). Infatti, per $a < b \in \mathbb{R}$, risulta

$$F_X(b) - F_X(a) = P(a < X \leq b) \geq 0.$$

(ii) Osserviamo che possiamo scrivere

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow \infty} P(X \leq -n);$$

ponendo

$$B_n \equiv \{X \leq -n\}$$

otteniamo una successione tale che, per $n = 1, 2, \dots$, $B_n \supseteq B_{n+1}$ e $\bigcap_{n=1}^{\infty} B_n = \emptyset$.

Dunque la condizione $\lim_{x \rightarrow -\infty} F_X(x) = 0$ si ottiene facilmente ricordando la **Proposizione 1**. Analogamente possiamo verificare la necessità della condizione $\lim_{x \rightarrow +\infty} F_X(x) = 1$, osservando che possiamo scrivere

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow \infty} P(\{X \leq n\});$$

ponendo

$$A_n \equiv \{X \leq n\}$$

otteniamo una successione tale che $A_1 \subseteq A_2 \subseteq \dots$ e $\bigcup_{n=1}^{\infty} A_n = \Omega$.

(iii) Notiamo innanzitutto che, essendo F_X una funzione monotona (non decrescente), i suoi eventuali punti di discontinuità possono essere soltanto punti con discontinuità di 1^a specie; si ha cioè che, $\forall \tilde{x} \in \mathbb{R}$, risultano esistere i due limiti

$$F_X(\tilde{x}^+) \equiv \lim_{x \rightarrow \tilde{x}^+} F_X(x), \quad F_X(\tilde{x}^-) \equiv \lim_{x \rightarrow \tilde{x}^-} F_X(x).$$

Si tratta ora di verificare che risulta

$$F_X(\tilde{x}^+) = F_X(\tilde{x}).$$

Essendo F_X monotona non decrescente basterà mostrare che risulta

$$\lim_{n \rightarrow \infty} F_X(\tilde{x} + \frac{1}{n}) = F_X(\tilde{x}).$$

Gli eventi $B_n = \{X \leq \tilde{x} + \frac{1}{n}\}$ costituiscono una successione non crescente e posto

$$B = \bigcap_{n=1}^{\infty} \{X \leq \tilde{x} + \frac{1}{n}\}$$

abbiamo, per la proprietà di continuità di P ,

$$\lim_{n \rightarrow \infty} F_X(\tilde{x} + \frac{1}{n}) = \lim_{n \rightarrow \infty} P(X \leq \tilde{x} + \frac{1}{n}) = P(B).$$

Bisogna ora dimostrare che risulta

$$B = \{X \leq \tilde{x}\}.$$

Notiamo allora che si ha ovviamente $\{X \leq \tilde{x}\} \subseteq B$, in quanto⁷⁷ $\{X \leq \tilde{x}\} \subseteq \{X \leq \tilde{x} + \frac{1}{n}\} = B_n$, $\forall n = 1, 2, \dots$. D'altra parte se $\omega \in B$, sarà anche⁷⁸ $X(\omega) \leq \tilde{x} + \frac{1}{n}$ per $\forall n = 1, 2, \dots$ e dunque $X(\omega) \leq \tilde{x} + \lim_{n \rightarrow \infty} \frac{1}{n} = \tilde{x}$, cioè $\omega \in \{X \leq \tilde{x}\}$.

⁷⁷Se $\omega \in \{X \leq \tilde{x}\}$, cioè se $X(\omega) \leq \tilde{x}$ allora ovviamente $X(\omega) \leq \tilde{x} + \frac{1}{n}$, cioè $\omega \in \{X \leq \tilde{x} + \frac{1}{n}\} = B_n$.

⁷⁸Si ha $\omega \in B$ se e solo se $\omega \in B_n$ per ogni $n = 1, 2, \dots$, ovvero se e solo se $X(\omega) \leq \tilde{x} + \frac{1}{n}$ per ogni $n = 1, 2, \dots$. Ma allora $X(\omega) \leq \lim_{n \rightarrow \infty} \tilde{x} + \frac{1}{n} = \tilde{x}$, ovvero $\omega \in \{X(\omega) \leq \tilde{x}\}$.

Pur senza darne qui dimostrazione, è opportuno citare che, viceversa, per qualunque funzione $F : \mathbb{R} \rightarrow [0, 1]$, che soddisfi le proprietà (i), (ii), (iii) di cui nella precedente **Proposizione 3**, è possibile costruire un'opportuno spazio di probabilità (Ω, \mathcal{F}, P) ed una variabile aleatoria X su Ω tali che risulti

$$F_X(x) = F(x), \quad \forall x \in \mathbb{R}.$$

Esempio 13.9 (variabili aleatorie di Cauchy). Sia $F : \mathbb{R} \rightarrow [0, 1]$ definita da

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}. \quad (108)$$

Si vede facilmente che questa funzione soddisfa le proprietà (i), (ii), (iii) della **Proposizione 3**: la prima e la terza sono banali, e la seconda discende dal fatto che

$$\lim_{x \rightarrow -\infty} \arctan(x) = -\frac{\pi}{2} \quad e \quad \lim_{x \rightarrow +\infty} \arctan(x) = \frac{\pi}{2}.$$

Sia X una variabile aleatoria, con funzione di distribuzione $F_X(x) = F(x)$, con F definita come in (108). Allora si dice che X **ha distribuzione di Cauchy** e inoltre, ad esempio,

$$\begin{aligned} P(X \in (-1, 1]) &= F(1) - F(-1) = \frac{1}{\pi} \arctan(1) + \frac{1}{2} - \left(\frac{1}{\pi} \arctan(-1) + \frac{1}{2} \right) \\ &= \frac{1}{\pi} \frac{\pi}{4} - \frac{1}{\pi} \left(-\frac{\pi}{4} \right) = \frac{1}{2}. \end{aligned}$$

Osservazione 6 Si verifica immediatamente, ancora in base alla (104), che F_X si mantiene costante in un intervallo (a, b) se e solo se risulta

$$P(a < X < b) = 0.$$

Osservazione 7 (Significato dei punti di discontinuità di F_X). Abbiamo già notato nel corso della dimostrazione della Proposizione 3 che, essendo F_X una funzione monotona risultano esistere, $\forall \tilde{x} \in \mathbb{R}$, i due limiti

$$F_X(\tilde{x}^+), \quad F_X(\tilde{x}^-).$$

Tenendo conto della continuità da destra di F_X , sappiamo che risulta

$$F_X(\tilde{x}^+) = F_X(\tilde{x}) = P(X \leq \tilde{x}).$$

Per quanto riguarda $F_X(\tilde{x}^-)$, notiamo che possiamo anche scrivere

$$F_X(\tilde{x}^-) = \lim_{x \rightarrow \tilde{x}^-} F_X(x) = \lim_{n \rightarrow \infty} F_X\left(\tilde{x} - \frac{1}{n}\right) = \lim_{n \rightarrow \infty} P\left(X \leq \tilde{x} - \frac{1}{n}\right)$$

e dunque⁷⁹ che, per la proprietà di continuità di P , possiamo concludere

$$F_X(\tilde{x}^-) = P(X < \tilde{x}). \quad (109)$$

⁷⁹Infatti la successione di eventi $A_n = \{X \leq \tilde{x} - \frac{1}{n}\}$ è monotona non decrescente e

$$\{X < \tilde{x}\} = \bigcup_{n=1}^{\infty} \{X \leq \tilde{x} - \frac{1}{n}\}$$

e quindi, per la proprietà di continuità delle probabilità

$$P(X < \tilde{x}) = \lim_{n \rightarrow \infty} P(X \leq \tilde{x} - \frac{1}{n}).$$

Questa osservazione è alla base della interessante proprietà:

i valori reali che una variabile aleatoria X assume con probabilità positiva sono tutti e soli i punti di discontinuità per F_X , e inoltre, se \tilde{x} è un punto di discontinuità per F , allora

$$P(X = \tilde{x}) = F_X(\tilde{x}) - F_X(\tilde{x}^-).$$

Verifichiamo immediatamente questa proprietà, osservando che

I Se esiste qualche valore $\tilde{x} \in \mathbb{R}$, tale che

$$P(X = \tilde{x}) = \tilde{p} > 0,$$

deve risultare allora

$$P(X \leq \tilde{x}) = P(X < \tilde{x}) + \tilde{p} > P(X < \tilde{x}) \quad (110)$$

in quanto⁸⁰, ovviamente,

$$P(X = \tilde{x}) = P(X \leq \tilde{x}) - P(X < \tilde{x}).$$

Dalla (110) segue

$$F_X(\tilde{x}) > F_X(\tilde{x}^-),$$

cioè \tilde{x} risulta essere un punto di salto (e quindi punto di discontinuità) per la funzione F_X .

II Sia ora \tilde{x} un qualunque punto di discontinuità per F_X . In virtù del fatto che F_X è non decrescente (e quindi può solo avere punti di discontinuità di 1^a specie), risulterà che

$$F_X(\tilde{x}^+) = F_X(\tilde{x}) > F_X(\tilde{x}^-)$$

cioè, ricordando ancora una volta la definizione di F_X e la (109)

$$P(X \leq \tilde{x}) = P(X < \tilde{x}) + \rho_{\tilde{x}},$$

e dunque

$$P(X = \tilde{x}) = \rho_{\tilde{x}},$$

essendo $\rho_{\tilde{x}} = F_X(\tilde{x}) - F_X(\tilde{x}^-)$ una quantità strettamente positiva.

Un risultato dell'Analisi matematica mostra che una funzione monotona (quale risulta essere la F_X) ammette, al più, un insieme numerabile di punti di discontinuità; quindi:

⁸⁰Si osservi che la relazione

$$P(X = \tilde{x}) = P(X \leq \tilde{x}) - P(X < \tilde{x}),$$

alla luce della definizione di funzione di distribuzione, e alla luce della (109), si può riscrivere come

$$P(X = \tilde{x}) = F_X(\tilde{x}) - F_X(\tilde{x}^-).$$

tutti i punti sulla retta reale hanno probabilità 0 per la X , tranne al più un insieme finito o una successione numerabile⁸¹.

Prima di passare oltre, vediamo un esempio naturale di distribuzione di probabilità, per una variabile aleatoria X , con funzione di distribuzione F_X ovunque continua (e quindi tale che $P(X = x) = 0, \forall x \in \mathbb{R}$)

Esempio 13.10 (v.a. uniformi). *Sia (α, β) un fissato intervallo della retta reale ed indichiamo con $A_{(\alpha, \beta)}(x_1, x_2)$, per $x_1 \leq x_2 \in \mathbb{R}$, la lunghezza dell'intervallo $(\alpha, \beta) \cap [x_1, x_2]$:*

$$A_{(\alpha, \beta)}(x_1, x_2) \equiv \begin{cases} \min(\beta, x_2) - \max(\alpha, x_1) & \text{se } \min(\beta, x_2) \geq \max(\alpha, x_1) \\ 0 & \text{altrimenti} \end{cases}$$

Sia ora X una variabile aleatoria tale che

$$P(x_1 \leq X \leq x_2) = \frac{A_{(\alpha, \beta)}(x_1, x_2)}{\beta - \alpha}.$$

Osserviamo innanzitutto che, $\forall x \in \mathbb{R}$, risulta

$$P(X = x) = 0$$

(e quindi F_X risulterà essere una funzione ovunque continua); inoltre si ha

$$P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{\beta - \alpha} \quad \text{nel caso in cui } \alpha \leq x_1 \leq x_2 \leq \beta,$$

ed inoltre

$$P(x_1 \leq X \leq x_2) = 0 \quad \text{nel caso in cui gli intervalli } (\alpha, \beta) \text{ e } (x_1, x_2) \text{ abbiano intersezione vuota.}$$

*In un tale caso diremo che X segue **una distribuzione uniforme nell'intervallo** (α, β) e scriveremo in simboli*

$$X \sim R(\alpha, \beta).$$

Determiniamo ora la funzione di distribuzione F_X .

Si ha

$$\begin{aligned} F_X(x) &= P(X \leq x) = \lim_{x_1 \rightarrow -\infty} P(x_1 \leq X \leq x) = \\ &= \lim_{x_1 \rightarrow -\infty} \frac{A_{(\alpha, \beta)}(x_1, x)}{\beta - \alpha}. \end{aligned}$$

Si verifica facilmente che risulta

$$F_X(x) = \begin{cases} 0 & \text{per } x \leq \alpha \\ \frac{x - \alpha}{\beta - \alpha} & \text{per } \alpha \leq x \leq \beta \\ 1 & \text{per } x \geq \beta \end{cases} \quad (111)$$

⁸¹Si può dare la dimostrazione del fatto che la funzione F_X ammette, al massimo, un insieme numerabile di punti di discontinuità, utilizzando l'interpretazione probabilistica dei punti di discontinuità e notando che, per ogni intero $n \geq 1$,

$$\{\tilde{x} \in \mathbb{R} : P(X = \tilde{x}) \geq \frac{1}{n}\}$$

può contenere al massimo n punti. Per terminare la dimostrazione bisogna poi osservare che la famiglia dei punti di discontinuità di F_X è l'unione numerabile degli insiemi di punti in cui $P(X = \tilde{x}) = F_X(\tilde{x}) - F_X(\tilde{x}^-) \geq \frac{1}{n}$.

Nel caso particolare $\alpha = 0$, $\beta = 1$, si ha

$$F_X(x) = \begin{cases} 0 & \text{per } x \leq 0 \\ x & \text{per } 0 \leq x \leq 1 \\ 1 & \text{per } x \geq 1 \end{cases} \quad (112)$$

Il grafico di F_X , per X uniforme è riportato in Figura 2.

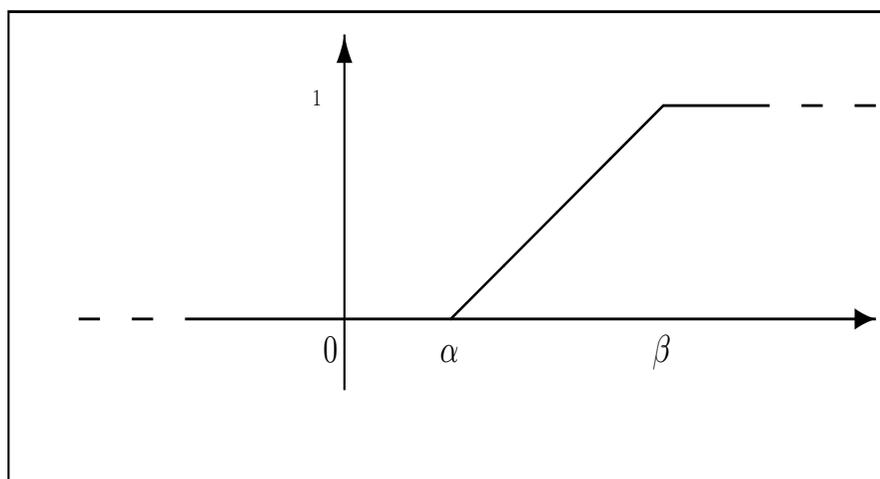


Figura 2: Grafico di $F_X(x)$, per X uniforme in (α, β)

13.3 Funzioni di distribuzione continue, funzioni di densità di probabilità

Come abbiamo già detto, **nel caso in cui F_X risulti essere una funzione continua**, avremo

$$P(X = x) = 0, \quad \forall x \in \mathbb{R}. \quad (113)$$

Questo fatto da una parte comporta una conseguenza piacevole:

$$P(a < X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b), \quad (114)$$

(come si vede subito, tenendo conto delle relazioni (104), (105), (106) e (107)), ma d'altra parte pone un importante problema:

Come facciamo allora ad esprimere sinteticamente quali sono le zone dei valori più o meno probabili per la X ?

Prima di dare la soluzione che in alcuni casi si può dare a questo problema (si veda la definizione 13.6), proviamo a vedere come si potrebbe ragionare. Può venire spontaneo di procedere come segue: fissiamo un numero δ finito, positivo, magari abbastanza piccolo e analizziamo, al variare di $x \in \mathbb{R}$, l'andamento della probabilità

$$P(x \leq X \leq x + \delta).$$

Nel caso di X distribuita in modo uniforme sull'intervallo $(0, 1)$, ad esempio, otterremo

$$P(x \leq X \leq x + \delta) = \begin{cases} 0 & \text{per } x \leq -\delta \\ \delta - |x| & \text{per } -\delta \leq x \leq 0 \\ \delta & \text{per } 0 \leq x \leq 1 - \delta \\ 1 - x & \text{per } 1 - \delta \leq x \leq 1 \\ 0 & \text{per } x \geq 1 \end{cases}$$

Può venire anche spontaneo di vedere che cosa succede prendendo la quantità δ sempre più piccola; ma ovviamente, sempre in virtù della continuità di F_X , avremo, $\forall x \in \mathbb{R}$,

$$\lim_{\delta \rightarrow 0} P(x \leq X \leq x + \delta) = \lim_{\delta \rightarrow 0} F_X(x + \delta) - F_X(x) = 0.$$

Al variare di $x \in \mathbb{R}$, studieremo allora piuttosto

$$\lim_{\delta \rightarrow 0} \frac{P(x \leq X \leq x + \delta)}{\delta}. \quad (115)$$

Vediamo ora cosa possiamo ottenere in un caso abbastanza generale.

Proposizione 4 Sia F_X una funzione continua e derivabile in ogni x . Allora il limite in (115) esiste e risulta

$$\lim_{\delta \rightarrow 0} \frac{P(x \leq X \leq x + \delta)}{\delta} = F'_X(x),$$

essendo $F'_X(x)$ il valore in x della derivata prima di F_X .

Sia inoltre F'_X , la derivata prima di F_X , una funzione continua. Allora, per ogni $a < b$, si ha

$$P(a < X < b) = P(a \leq X \leq b) = \int_a^b F'_X(x) dx. \quad (116)$$

Dimostrazione. Essendo F_X continua, risulta

$$P(x \leq X \leq x + \delta) = P(x < X \leq x + \delta)$$

e dunque, essendo F_X anche derivabile in x ,

$$\lim_{\delta \rightarrow 0} \frac{P(x \leq X \leq x + \delta)}{\delta} = \lim_{\delta \rightarrow 0} \frac{F_X(x + \delta) - F_X(x)}{\delta} = F'_X(x).$$

Per dimostrare la (116) basta ricordare che, come sappiamo dall'Analisi, una funzione G continua e derivabile, con derivata prima continua, risulta essere l'integrale della sua derivata G' , cioè risulta, per ogni intervallo $[a, b]$:

$$\int_a^b G'(x) dx = G(b) - G(a).$$

Infatti si può applicare la precedente relazione ad F_X , tenendo conto di (114), e del fatto che

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

La precedente **Proposizione 4** si applica immediatamente al caso di una variabile aleatoria X con distribuzione di Cauchy (confrontare l'Esempio 13.9) per il quale si ha

$$F'_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

Tuttavia la **Proposizione 4**, non si applica al caso della variabile aleatoria uniforme, la cui funzione di distribuzione è derivabile solo per $x \neq \alpha, \beta$. La derivata vale $\frac{1}{\beta-\alpha}$ per $x \in (\alpha, \beta)$, mentre vale 0 sia in $(-\infty, \alpha)$ che in (β, ∞) . Tuttavia, è possibile mostrare che in questo caso la relazione (116) è valida (confrontare la **Proposizione 4bis**). Queste osservazioni suggeriscono la seguente

Definizione 13.6 (funzione di densità di probabilità). *Sia X una variabile aleatoria, e sia f una funzione con $f(x) \geq 0 \forall x \in \mathbb{R}$, e per la quale, qualunque sia l'intervallo $[a, b]$, risulta*

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Diremo allora che la distribuzione di probabilità di X **ammette densità**. La funzione f viene detta **funzione di densità di probabilità** (o semplicemente **funzione di densità**) e viene usualmente indicata con il simbolo f_X .

Ovviamente se X ammette densità allora la sua funzione di distribuzione è continua, infatti

$$P(X = a) = P(X \in [a, a]) = P(a \leq X \leq a) = \int_a^a f(x) dx = 0.$$

In generale, la funzione di densità non è definita univocamente in tutti i punti, come mostra la seguente **Proposizione 4bis**. A questo proposito si veda anche la successiva **Osservazione 9**.

Proposizione 4bis Condizione sufficiente affinché X ammetta densità e che

- (i) $F_X(x)$ sia continua,
- (ii) $F_X(x)$ sia derivabile, tranne al più un numero finito m di punti $x_i, i = 1, \dots, m$,
- (iii) la derivata prima $F'_X(x)$ sia continua in ciascuno degli $m+1$ intervalli in cui i punti $x_i, i = 1, \dots, m$ dividono la retta, ovvero, supponendo $x_1 < x_2 < \dots < x_m$, in ciascuno degli intervalli

$$(-\infty, x_1), \quad (x_1, x_2), \quad \dots, (x_k, x_{k+1}), \dots, (x_{m-1}, x_m), \quad (x_m, \infty),$$

- (iv) la derivata prima $F'_X(x)$ sia prolungabile con continuità su ciascuno degli intervalli

$$(-\infty, x_1], \quad [x_1, x_2], \quad \dots, [x_k, x_{k+1}], \dots, [x_{m-1}, x_m], \quad [x_m, \infty),$$

preso separatamente.

In tale caso X ammette come densità di probabilità qualsiasi funzione f che coincida con la derivata prima di F_X , naturalmente dove quest'ultima esiste, ossia

$$f(x) = F'_X(x), \quad \forall x \neq x_i.$$

Invece nei punti x_i la funzione f non deve soddisfare nessuna condizione.

Dimostrazione: omessa.

Come applicazione abbiamo il caso delle variabili uniformi, ma si veda anche l'Esempio 13.14.

Esempio 13.11 (densità di una v.a. uniforme). *Sia X una variabile aleatoria con distribuzione uniforme nell'intervallo (α, β) , secondo quanto definito nel precedente Esempio 13.10. Derivando rispetto a x la funzione di distribuzione F_X , otteniamo la funzione di densità di probabilità:*

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & x \notin [\alpha, \beta] \end{cases}$$

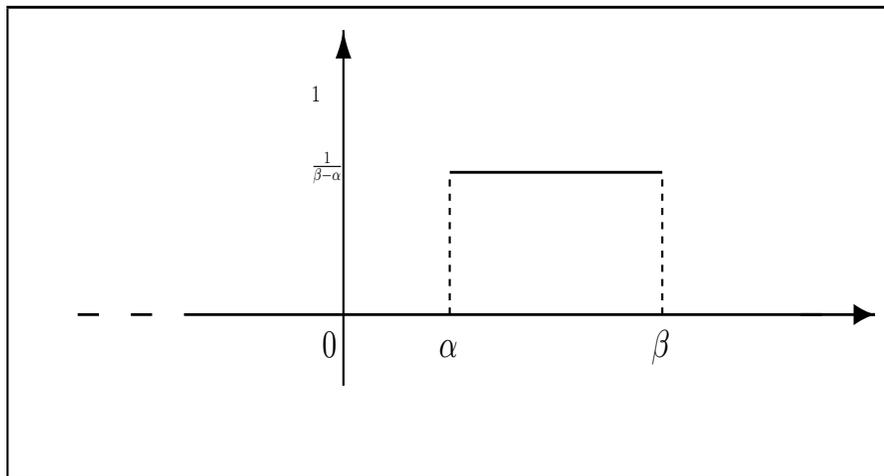


Figura 3: Grafico di $f_X(x)$, per X uniforme in (α, β)

Osservazione 8. Se la distribuzione di probabilità di X ammette densità $f_X(x)$ allora, per ogni valore $x \in \mathbb{R}$, risulta

$$P(X \leq x) = P(-\infty < X \leq x) = F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (117)$$

come si ottiene subito facendo il limite per a che tende a $-\infty$ in

$$P(a \leq X \leq x) = \int_a^x f_X(\xi) d\xi.$$

Inoltre, mandando x a $+\infty$ nella (117), si ottiene

$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(\xi) d\xi.$$

Infine va sottolineato che, per stabilire che X ammette densità $f_X = f$ è sufficiente verificare che

$$F_X(x) = \int_{-\infty}^x f(\xi) d\xi$$

dal momento che

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \int_a^b f(\xi) d\xi$$

Osservazione 9. Sia F_X una funzione di distribuzione con densità f_X e sia g una funzione tale che

$$g(x) = f_X(x)$$

per tutti i punti x della retta (tranne, al più, per quelli in un insieme numerabile). Allora risulta anche per ogni intervallo $[a, b]$

$$P(a \leq X \leq b) = \int_a^b g(x) dx.$$

In tal caso anche $g(x)$ viene detta densità per F_X ; possiamo dire dunque che non esiste un'unica funzione di densità, bensì un'intera famiglia di funzioni di densità. Tale famiglia costituisce una classe di equivalenza: essa contiene la classe di tutte le funzioni non negative che si differenziano dalla f_X soltanto su un insieme di punti finito o numerabile.

Da quanto sopra deriva anche la seguente

Osservazione 10. Una qualunque funzione di densità f gode delle seguenti proprietà

(i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$

Spesso quindi, invece di individuare la distribuzione di una variabile aleatoria attraverso la funzione di distribuzione si preferisce definire la distribuzione di probabilità attraverso una funzione che goda delle proprietà (i) e (ii), come nel caso degli Esempi 13.12 e 13.13, delle v.a. esponenziali e gaussiane, rispettivamente.

Altre volte invece si definisce la densità a meno di una costante di proporzionalità, come spiegato qui di seguito, e come nell'Esempio 13.15, alla fine di questa Lezione.

Sia $g(x)$ una funzione assegnata a valori non negativi e sia k una costante positiva; la funzione

$$f(x) = k \cdot g(x)$$

è una funzione di densità se e solo se esiste finito l'integrale $\int_{-\infty}^{+\infty} g(x) dx$ e risulta

$$k = \frac{1}{\int_{-\infty}^{+\infty} g(x) dx}$$

cioè

$$f(x) = \frac{g(x)}{\int_{-\infty}^{+\infty} g(x) dx}.$$

Esempio 13.12 (v.a. esponenziali). Sia λ una costante positiva assegnata e consideriamo la funzione di densità

$$f_X(x) = \begin{cases} \lambda \exp\{-\lambda x\} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

La corrispondente funzione di distribuzione è data da

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi = \begin{cases} \int_0^x \lambda \exp\{-\lambda x\} dx & x \geq 0 \\ 0 & x < 0 \end{cases}$$

e dunque

$$F_X(x) = \begin{cases} 1 - \exp\{-\lambda x\} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Di una variabile aleatoria X con tale funzione distribuzione si dice che segue una **distribuzione esponenziale di parametro λ** .

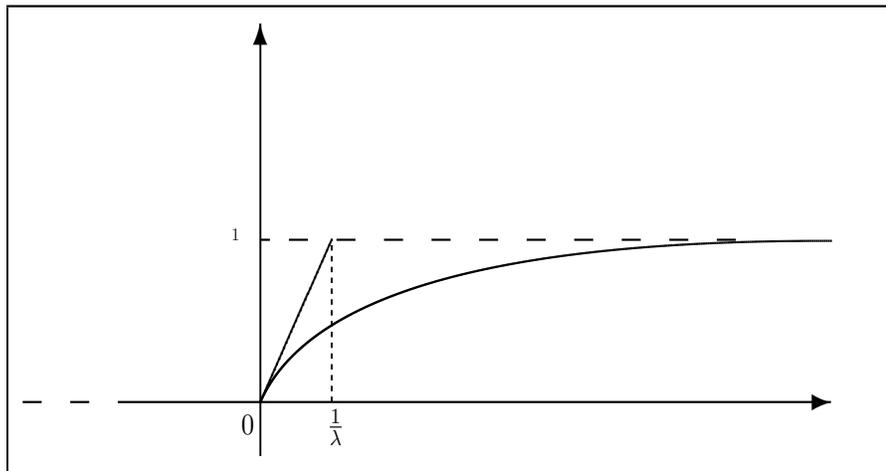


Figura 4: Grafico di $F_X(x)$, per X esponenziale di parametro λ (caso $\lambda > 1$)

Esempio 13.13 (v.a. gaussiane standard). Una variabile aleatoria segue una distribuzione **gaussiana standard**, e si scrive $X \sim \mathcal{N}(0, 1)$, se la sua funzione di densità è data da

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}, x \in \mathbb{R}.$$

Tale funzione è la ben nota funzione degli errori di Gauss ed è noto che non è possibile esprimere la corrispondente funzione distribuzione

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\xi^2}{2}\right\} d\xi$$

in modo esplicito, in termini di altre funzioni elementari.

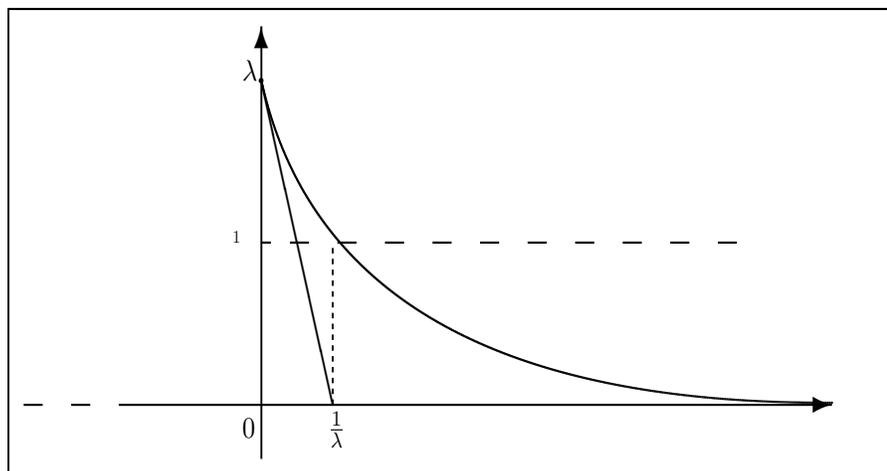


Figura 5: Grafico di $f_X(x)$, per X esponenziale di parametro λ (caso $\lambda > 1$)

Attraverso il calcolo di un opportuno integrale doppio è possibile verificare che il fattore $\frac{1}{\sqrt{2\pi}}$ ha il ruolo di costante di normalizzazione, cioè che vale l'identità

$$\int_{-\infty}^{+\infty} \exp\left\{-\frac{\xi^2}{2}\right\} d\xi = \sqrt{2\pi}.$$

13.4 Valori attesi per variabili aleatorie generali

Come abbiamo visto nella Lezione 9 sui valori attesi, nel caso degli spazi di probabilità finiti, ci sono diversi modi di calcolare il valore atteso di una variabile aleatoria X o di una sua funzione $h(X)$. In particolare (si veda la **Proposizione 10** della Lez. 9) si ha che se $X(\Omega) = \{x_1, \dots, x_n\}$

$$E(h(X)) = \sum_{i=1}^n h(x_i) p_X(x_i) = \sum_{i=1}^n h(x_i) P(X = x_i).$$

Questa espressione suggerisce un modo naturale di definire i valori attesi nel caso di variabili aleatorie discrete, ossia nel caso in cui $X(\Omega) = \{x_1, \dots, x_n, x_{n+1}, \dots\}$, come

$$E(h(X)) = \sum_{i=1}^{\infty} h(x_i) P(X = x_i).$$

Tuttavia è necessario assicurarsi che la serie precedente sia assolutamente convergente, per assicurarsi che la somma della serie non dipenda dall'ordine in cui la somma viene effettuata, cioè si deve richiedere che

$$E(|h(X)|) = \sum_{i=1}^{\infty} |h(x_i)| P(X = x_i) < \infty.$$

Appare anche naturale dare la definizione del valore atteso nel caso di variabili aleatorie con densità f_X come segue

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x) dx,$$

purché

$$E(|h(X)|) = \int_{-\infty}^{\infty} |h(x)| f_X(x) dx, < \infty.$$

È importante sottolineare che le proprietà di linearità e monotonia, viste nella Lezione 9, continuano a valere per il valore atteso definito in questo modo. Tuttavia non diamo qui la dimostrazione di questa affermazione

Si possono anche dare le definizioni della varianza come

$$\text{Var}(X) \stackrel{\text{def}}{=} E((X - E(X))^2),$$

ma nel caso di v.a. generali è necessario supporre che il valore atteso $E(X^2)$ sia finito, a differenza del caso finito, dove non era necessario fare alcuna ipotesi.

Grazie alla proprietà di linearità si dimostra, come nel caso finito, che

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

Esercizio proposto 13.2 (v.a. geometriche). Sia T una v.a. geometrica di parametro θ . Si verifichi che

$$E(T) = \frac{1}{\theta}, \quad e \quad \text{Var}(T) = \frac{1 - \theta}{\theta^2}.$$

Esercizio proposto 13.3 (v.a. di Poisson). Si dimostri che, per $\lambda > 0$

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1.$$

Di conseguenza ha senso considerare una v.a. X a valori in $\{0, 1, 2, \dots\}$ per la quale

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

In questo caso si dice che X ha **distribuzione di Poisson di parametro λ** .

Si dimostra anche che, se S_n sono v.a. binomiali $b(n, \theta_n)$ di parametro $\theta = \theta_n = \frac{\lambda}{n}$ allora,

$$\lim_{n \rightarrow \infty} P(S_n = k) = P(X = k) \quad \forall k = 0, 1, 2, \dots,$$

ossia vale il seguente risultato:

Teorema (di approssimazione) di Poisson Sia S_n una variabile aleatoria binomiale $b(n, \lambda/n)$, dove $\lambda > 0$ è un numero fissato. Allora si ha

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \forall k = 0, 1, 2, \dots$$

Dimostrazione. Basta osservare che

$$\begin{aligned} P(S_n = k) &= \binom{n}{k} \theta^k (1 - \theta)^{n-k} = \frac{n!}{k! (n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \frac{1}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

in quanto valgono le seguenti tre relazioni

$$(i) \quad \frac{n!}{(n-k)!} \frac{1}{n^k} = \frac{n(n-1)\cdots(n-(k-1))}{n^k} = \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-(k-1)}{n} \xrightarrow{n \rightarrow \infty} 1^k = 1,$$

$$(ii) \quad \left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\lambda},$$

ed infine

$$(iii) \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \xrightarrow{n \rightarrow \infty} 1.$$

Esercizio proposto 13.4 (v.a. Poisson, continuazione). *Sia X una v.a. di Poisson di parametro λ . Si verifichi che*

$$E(X) = \lambda, \quad e \quad Var(X) = \lambda.$$

*Soluzione*⁸²

Si osservi che, se con le stesse notazioni dell'Esercizio proposto 13.3, $S_n \equiv b(n, \frac{\lambda}{n})$

$$E(S_n) = \lambda, \quad e \quad Var(S_n) = \lambda \left(1 - \frac{\lambda}{n}\right) \rightarrow \lambda.$$

Esercizio proposto 13.5 (v.a. esponenziali). *Sia X una v.a. esponenziale di parametro λ . Si verifichi che*

$$E(X) = \frac{1}{\lambda}, \quad e \quad Var(X) = \frac{1}{\lambda^2}$$

Esercizio proposto 13.6 (v.a. gaussiane). *Sia X una v.a. gaussiana standard. Si verifichi che*

$$E(X) = 0, \quad e \quad Var(X) = 1$$

*Soluzione*⁸³

⁸²Sia $X \sim Poiss(\lambda)$, allora $\mathbb{E}(X) = \lambda$ e $\mathbb{E}(X^2) = \lambda^2 + \lambda$, da cui $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Infatti

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} k P(X=k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \sum_{h=0}^{\infty} \frac{\lambda^{h+1}}{h!} e^{-\lambda} = \lambda \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda. \end{aligned}$$

Analogamente si ha che $\mathbb{E}(X^2) = \lambda^2 + \lambda$: considerando che $X^2 = X + X(X-1)$, si ha

$$\begin{aligned} \mathbb{E}(X^2) &= \mathbb{E}(X + X(X-1)) = \mathbb{E}(X) + \mathbb{E}(X(X-1)) \\ &= \lambda + \sum_{k=0}^{\infty} k(k-1) P(X=k) = \lambda + \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda + \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda + \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} = \lambda + \sum_{h=0}^{\infty} \frac{\lambda^{h+2}}{h!} e^{-\lambda} = \lambda + \lambda^2 \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} e^{-\lambda} = \lambda + \lambda^2 e^{\lambda} e^{-\lambda} = \lambda + \lambda^2. \end{aligned}$$

⁸³Per mostrare che $E(X) = 0$, basta osservare che la funzione

$$g(x) := x f_X(x) \text{ è dispari, cioè } g(-x) = g(x)$$

Esercizio proposto 13.7 (v.a. di Cauchy). *Sia X una v.a. di Cauchy. Si dimostri che la condizione*

$$E(|X|) = \int_{-\infty}^{\infty} |x| f_X(x) dx, < \infty$$

non è soddisfatta, e quindi il valore atteso non esiste.

*Soluzione*⁸⁴

13.5 Esempi svolti

Esempio 13.14. *X è una variabile aleatoria con funzione di distribuzione data da*

$$F(x) = \begin{cases} 0 & \text{per } x < -1 \\ \frac{1}{2}(x+1) & \text{per } -1 \leq x < 0 \\ \frac{1}{2} + \frac{x^2}{2} & \text{per } 0 \leq x < 1 \\ 1 & \text{per } x \geq 1 \end{cases}$$

a) Calcolare $P(-\frac{1}{2} \leq X \leq \frac{1}{2})$.

soluzione di a) $P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = \frac{3}{8}$

Infatti, per $a \leq b$

$$P(a \leq X \leq b) = P(X = a) + P(a < X \leq b) = F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a),$$

*dove l'ultima uguaglianza dipende dal fatto che nel nostro caso la funzione $x \rightarrow F_X(x) = F(x)$ è continua.*⁸⁵

e che, essendo $d(\frac{1}{2}x^2) = x dx$, si ha

$$E(|X|) = \int_{-\infty}^{\infty} |x| f_X(x) dx = 2 \int_0^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}x^2} d(\frac{1}{2}x^2)$$

con il cambio di variabile $y = \frac{1}{2}x^2$

$$= 2 \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-y} dy = 2 \frac{1}{\sqrt{2\pi}} (-e^{-y})_0^{\infty} = 2 \frac{1}{\sqrt{2\pi}} (-0 + 1) < \infty$$

e quindi, il valore atteso vale 0. Di conseguenza la varianza coincide con $E(X^2)$ e sia ha

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2} d(\frac{1}{2}x^2)$$

essendo $e^{-\frac{1}{2}x^2} d(\frac{1}{2}x^2) = -d(e^{-\frac{1}{2}x^2})$, ed integrando per parti

$$= \frac{1}{\sqrt{2\pi}} \left\{ [-x e^{-\frac{1}{2}x^2}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \right\} = \frac{1}{\sqrt{2\pi}} \left\{ [-0 + 0] + \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \right\} = \int_{-\infty}^{\infty} f_X(x) dx = 1$$

⁸⁴Nel caso della distribuzione di Cauchy si ha $f_X(-x) = f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ e quindi

$$\begin{aligned} \int_{-\infty}^{\infty} |x| f_X(x) dx &= 2 \int_0^{\infty} x \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_0^{\infty} \frac{1}{1+x^2} d(x^2) \\ &= \frac{1}{\pi} \int_0^{\infty} \frac{1}{1+y} dy = \frac{1}{\pi} (\log(1+y))_0^{\infty} = \infty \end{aligned}$$

⁸⁵Per verificare la continuità basta considerare che

- $F(-1^-) = 0$, che coincide con $F(-1) = \frac{1}{2}(-1+1) = 0$;
- $F(0^-) = \frac{1}{2}$, che coincide con $F(0) = \frac{1}{2} + \frac{0^2}{2} = \frac{1}{2}$;
- $F(1^-) = \frac{1}{2} + \frac{1^2}{2} = 1$, che coincide con $F(1) = 1$.

Quindi

$$P(-\frac{1}{2} \leq X \leq \frac{1}{2}) = F(\frac{1}{2}) - F(-\frac{1}{2}) = \frac{1}{2} + \frac{\frac{1}{4}}{2} - \frac{1}{2} \left(-\frac{1}{2} + 1 \right) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}$$

b) Calcolare la funzione di densità di probabilità di X .

soluzione di b) Applicando la **Proposizione 4bis** si ha

$$f(x) = F'(x) = \begin{cases} 0 & \text{per } x < -1 \\ \frac{1}{2} & \text{per } -1 < x < 0 \\ x & \text{per } 0 < x < 1 \\ 0 & \text{per } x > 1 \end{cases}$$

Come verifica, si noti che $f(x)$ è effettivamente una densità di probabilità, infatti $f(x) \geq 0$ per ogni $x \in \mathbb{R}$, ed inoltre

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-1}^0 f(x) dx + \int_0^1 f(x) dx = \int_{-1}^0 \frac{1}{2} dx + \int_0^1 x dx = \frac{1}{2} + \frac{x^2}{2} \Big|_{x=0}^{x=1} = \frac{1}{2} + \frac{1}{2} = 1.$$

Anche senza utilizzare la **Proposizione 4bis**, che non è stata dimostrata, si può verificare che $f(x)$ è la densità di X : infatti (si ricordi l'**Osservazione 8**) **basterebbe verificare** che $F(x) = \int_{-\infty}^x f(y) dy$, ovvero che

- per $x < -1$

$$0 = F(x) \stackrel{?}{=} \int_{-\infty}^x f(y) dy = 0$$

il che è ovvio;

- per $-1 \leq x < 0$

$$\frac{1}{2}(x+1) = F(x) \stackrel{?}{=} \int_{-\infty}^x f(y) dy = \int_{-1}^x f(y) dy = \int_{-1}^x \frac{1}{2} dy$$

il che è ovvio;

- per $0 \leq x < 1$

$$\frac{1}{2} + \frac{x^2}{2} = F(x) \stackrel{?}{=} \int_{-\infty}^x f(y) dy = \int_{-1}^0 f(y) dy + \int_0^x f(y) dy = \frac{1}{2} + \int_0^x y dy = \frac{1}{2} + \frac{y^2}{2} \Big|_{y=0}^{y=x}$$

il che è ovvio;

- per $x \geq 1$

$$1 = F(x) \stackrel{?}{=} \int_{-\infty}^x f(y) dy = \int_{-1}^0 f(y) dy + \int_0^1 f(y) dy = 1,$$

il che è ovvio;

c) Calcolare $\mathbb{E}(X)$.

soluzione di c) $\mathbb{E}(X) = \frac{1}{12}$.

Infatti

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-1}^0 x f(x) dx + \int_0^1 x f(x) dx \\ &= \int_{-1}^0 x \frac{1}{2} dx + \int_0^1 x x dx = \frac{1}{2} \frac{x^2}{2} \Big|_{x=-1}^{x=0} + \frac{x^3}{3} \Big|_{x=0}^{x=1} = -\frac{1}{4} + \frac{1}{3} = \frac{-3+4}{12} = \frac{1}{12}. \end{aligned}$$

Esempio 13.15. Sia X una variabile aleatoria con funzione di densità di probabilità

$$f_X(x) = \begin{cases} kx & \text{per } 0 \leq x \leq 1 \\ k(2-x) & \text{per } 1 \leq x \leq 2 \\ 0 & \text{altrimenti} \end{cases}$$

a) Trovare il valore della costante k .

b) Trovare la funzione di distribuzione di X .

Soluzione di a) La costante $k = 1$.

Infatti si tratta di trovare k in modo che $\int_{-\infty}^{\infty} f_X(x) dx = 1$. Per cui,

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x) dx &= k \left(\int_0^1 x dx + \int_1^2 (2-x) dx \right) = k \left(\frac{x^2}{2} \Big|_0^1 + 2 - \frac{x^2}{2} \Big|_1^2 \right) \\ &= k \cdot 1 = 1 \quad \text{se e solo se } k = 1 \end{aligned}$$

Soluzione di b) La funzione di distribuzione F_X è data da

$$F_X(x) = \begin{cases} 0 & \text{per } -\infty < x < 0 \\ \frac{x^2}{2} & \text{per } 0 \leq x < 1 \\ 2x - \frac{x^2}{2} - 1 & \text{per } 1 \leq x < 2 \\ 1 & \text{per } 2 \leq x \end{cases}$$

Infatti chiaramente per ogni $x \leq 0$

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = 0,$$

per $0 \leq x \leq 1$

$$F_X(x) = F_X(0) + \int_0^x y dy = 0 + \frac{x^2}{2},$$

per $1 \leq x < 2$

$$\begin{aligned} F_X(x) &= F_X(1) + \int_1^x (2-y) dy = \frac{1}{2} + \int_1^x (2-y) dy = \frac{1}{2} + 2y \Big|_1^x - \frac{y^2}{2} \Big|_1^x \\ &= \frac{1}{2} + 2x - 2 - \frac{x^2}{2} + \frac{1}{2} = 2x - \frac{x^2}{2} - 1, \end{aligned}$$

ed infine per $x \geq 2$

$$F_X(x) = F_X(2) + \int_2^x f_X(y) dy = 1 + 0 = 1.$$

c) Trovare il valore atteso di X (attenzione: è possibile trovarlo senza fare calcoli?)

soluzione di c) Il valore atteso di X vale 1

Infatti si tratta di calcolare

$$\begin{aligned} \int_{-\infty}^{\infty} x f_X(x) dx &= \int_0^1 x^2 dx + \int_1^2 x(2-x) dx = \left. \frac{x^3}{3} \right|_0^1 + \left. x^2 \right|_1^2 - \left. \frac{x^3}{3} \right|_1^2 \\ &= \frac{1}{3} + 4 - 1 - \frac{8}{3} + \frac{1}{3} = 1 \end{aligned}$$

Tuttavia per motivi di simmetria si ha che tale integrale deve venire 1 infatti il grafico della densità è simmetrico rispetto all'asse $x = 1$ e quindi, tenendo presente l'analogia con il baricentro, si ha immediatamente che il valore atteso deve essere 1.

Se X ammette densità f_X , e se F_X è nota, allora è possibile trovare $f_X(x)$ derivando la funzione $F_X(x)$ (almeno nei punti in cui F_X è derivabile. Tuttavia questo procedimento, se usato senza le dovute cautele, può portare a degli errori, come mostra il seguente controesempio, in cui si vede come è possibile che una funzione di distribuzione di una v.a. possa ammettere derivata tranne in un numero finito di punti, senza che la v.a. ammetta densità.

Esempio 13.16. Sia Z la variabile aleatoria discreta, definita come nell'Esempio 13.8. Allora la derivata prima di F_Z esiste in ogni $x \neq -1, 0, 1$: infatti $F'_Z(x) = 0$ per ogni $x \neq -1, 0, 1$. Tuttavia, ovviamente, la funzione $f(x) \equiv 0$ (cioè identicamente uguale a 0) non può essere la densità di Z , in quanto palesemente

$$F_Z(x) \neq \int_{-\infty}^x f(y) dy \equiv 0.$$

13.6 Trasformazioni di variabili aleatorie e il caso delle trasformazioni affini

Sia X una variabile aleatoria e sia $h : \mathbb{R} \rightarrow \mathbb{R}$ una funzione reale. L'applicazione

$$\omega \mapsto Y(\omega) := h(X(\omega))$$

è una trasformazione della variabile aleatoria X .

Nel caso degli spazi finiti Y è sempre una variabile aleatoria e la sua distribuzione è individuata da $Y(\Omega) = h(X(\Omega)) = \{y_1, y_2, \dots, y_m\}$ e da

$$P(Y = y_k) = \sum_{i: h(x_i)=y_k} P(X = x_i), \quad k = 1, 2, \dots, m$$

(si veda la **Proposizione 12** della Lezione 9, in cui si considera la trasformazione $W = h(X)$).

Nel caso degli spazi generali ci sono due differenze:

- (i) non è sempre vero che Y sia una variabile aleatoria;
- (ii) può accadere che la distribuzione di Y non si possa calcolare nel modo precedente.

Il problema si pone in particolare se la variabile aleatoria X non è discreta, mentre se la variabile aleatoria X è discreta numerabile, allora si generalizza immediatamente la precedente relazione, in quanto necessariamente Y è discreta (finita o numerabile): infatti ancora⁸⁶ si ha $Y(\Omega) = h(X(\Omega)) = \{y_k, k \geq 1\}$ e, similmente al caso finito

$$\{Y = y_k\} = \cup_{i: h(x_i)=y_k} \{X = x_i\}, \quad k \geq 1.$$

Questo fatto implica che l'insieme $\{Y = y_k\}$ è un evento (cioè appartiene ad \mathcal{F}) in quanto unione finita o numerabile di eventi, e quindi

$$P(Y = y_k) = \sum_{i: h(x_i)=y_k} P(X = x_i), \quad k \geq 1.$$

Si noti che in questa formula, a differenza della formula precedente del caso finito, la somma può essere estesa ad un insieme numerabile.

Il problema si può risolvere anche nel caso generale sotto condizioni che riguardano la funzione h , ma qui vedremo solo alcuni casi particolari. In tutti questi esempi l'idea è quella di riscrivere l'insieme

$$\{Y \leq y\}$$

in termini di un evento del tipo

$$\{X \in I\}$$

dove I è un intervallo⁸⁷. Ad esempio se $h(x) = x^2$, allora

$$\{Y \leq y\} = \emptyset$$

per $y < 0$, mentre, per $y \geq 0$

$$\{Y \leq y\} = \{X^2 \leq y\} = \{-\sqrt{y} \leq X \leq \sqrt{y}\} = \{X \in I\}, \quad \text{dove } I = [-\sqrt{y}, \sqrt{y}].$$

⁸⁶L'unica differenza con il caso finito è che $Y(\Omega)$ può essere un insieme infinito.

⁸⁷Più in generale basterebbe scrivere

$$\{Y \leq y\} = \cup_{i=1}^k \{X \in I_i\}$$

oppure

$$\{Y \leq y\} = \cup_{i=1}^{\infty} \{X \in I_i\}$$

cioè scrivere $\{Y \leq y\}$ come unione finita o numerabile di eventi, per essere sicuri che sia un evento.

Da ciò si deduce immediatamente che l'insieme $\{Y \leq y\}$ è un evento in quanto $\{X \in I\}$ lo è ed inoltre si ottiene la funzione di distribuzione di Y in quanto

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0, \\ P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) & \text{per } y \geq 0. \end{cases}$$

Il caso in cui $h(x) = x^3$ si ottiene in modo simile, ma è più semplice: qualunque sia y

$$\{Y \leq y\} = \{X^3 \leq y\} = \{X \leq y^{\frac{1}{3}}\}.$$

Di conseguenza, qualunque sia y ,

$$F_Y(y) = P(Y \leq y) = P(X \leq y^{\frac{1}{3}}) = F_X(y^{\frac{1}{3}}).$$

Osservazione Supponiamo ora che X ammetta densità. In questo caso, ci si potrebbe anche chiedere se $Y = h(X)$ ammette densità. Il candidato naturale è ovviamente la derivata rispetto ad y della funzione di distribuzione $F_Y(y)$, che nell'esempio precedente di $h(x) = x^3$ è la funzione composta $F_X(g(y))$, con $g(y) = y^{\frac{1}{3}}$, ossia

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g(y)) = F'_X(g(y)) \frac{d}{dy} g(y) = F'_X(g(y)) \frac{1}{3} y^{-\frac{2}{3}}.$$

Tuttavia sorge qualche problema a percorrere questa strada anche se F_X fosse continua con derivata continua, infatti la funzione $g(y) = y^{\frac{1}{3}}$ non ha derivata in zero. Questo tipo di problema si potrebbe risolvere, ma lo tralasciamo in questo corso elementare.

Il prossimo paragrafo è dedicato invece al caso delle trasformazioni affini, cioè al caso in cui $h(x) = \alpha + \beta x$, ed in questo caso affronteremo anche in problema della determinazione della densità di $Y = h(X)$.

13.6.1 Il caso delle trasformazioni affini

Sia X una variabile aleatoria e siano α e β due numeri reali. Si indichi con Y la seguente trasformazione affine di X

$$Y = \alpha + \beta X.$$

Il **primo problema** che ci poniamo in questo paragrafo è il seguente: **data F_X , la funzione di distribuzione di X , calcolare F_Y , la funzione di distribuzione di Y .**

Successivamente ci occuperemo del **secondo problema: se X è una variabile aleatoria che ammette densità di probabilità f_X , la variabile aleatoria Y ammette densità di probabilità f_Y ? e se (come effettivamente è) la risposta è sì, come si calcola f_Y ?**

Considereremo solo il caso in cui $\beta \neq 0$, in quanto il caso $\beta = 0$ corrisponde al caso banale in cui $Y = \alpha$, cioè Y è una variabile aleatoria degenera⁸⁸.

⁸⁸In questo caso la $F_Y(y) = 0$ per $y < \alpha$, ed $F_Y(y) = 1$ per $y \geq \alpha$. In questo caso, ovviamente, la variabile aleatoria Y non ammette densità, qualunque sia la distribuzione di X .

Cominciamo dando la soluzione dei precedenti problemi:

Soluzione del primo problema.

La soluzione dipende dal segno di β e si ha che la funzione di distribuzione di $Y = \alpha + \beta X$ è

$$F_Y(y) = F_X\left(\frac{y-\alpha}{\beta}\right), \quad \beta > 0, \quad (118)$$

$$F_Y(y) = 1 - F_X\left(\frac{y-\alpha}{\beta}\right) + P\left(X = \frac{y-\alpha}{\beta}\right), \quad \beta < 0. \quad (119)$$

Soluzione del secondo problema.

Se X ammette densità f_X allora anche $Y = \alpha + \beta X$ ammette densità e si ha

$$f_Y(y) = f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{|\beta|}, \quad \beta \neq 0. \quad (120)$$

Proseguiamo con le verifiche delle precedenti soluzioni

Verifica della soluzione per il primo problema: funzione di distribuzione

Cominciamo con l'osservare che

$$F_Y(y) = P(Y \leq y) = P(\alpha + \beta X \leq y).$$

A questo punto dobbiamo distinguere tra i due casi $\beta > 0$ o $\beta < 0$

caso $\beta > 0$ In questo caso⁸⁹

$$\{\alpha + \beta X \leq y\} = \left\{X \leq \frac{y-\alpha}{\beta}\right\}$$

e quindi

$$F_Y(y) = P(Y \leq y) = P(\alpha + \beta X \leq y) = P\left(\left\{X \leq \frac{y-\alpha}{\beta}\right\}\right) = F_X\left(\frac{y-\alpha}{\beta}\right).$$

caso $\beta < 0$ In questo caso⁹⁰

$$\{\alpha + \beta X \leq y\} = \left\{X \geq \frac{y-\alpha}{\beta}\right\}$$

e quindi

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\alpha + \beta X \leq y) = P\left(\left\{X \geq \frac{y-\alpha}{\beta}\right\}\right) \\ &= 1 - P\left(\left\{X < \frac{y-\alpha}{\beta}\right\}\right) = 1 - F_X\left(\left(\frac{y-\alpha}{\beta}\right)^-\right) \\ &= 1 - F_X\left(\frac{y-\alpha}{\beta}\right) + P\left(X = \frac{y-\alpha}{\beta}\right). \end{aligned}$$

⁸⁹Infatti

$$Y(\omega) \leq y \Leftrightarrow \alpha + \beta X(\omega) \leq y \Leftrightarrow X(\omega) \leq \frac{y-\alpha}{\beta}$$

e quindi

$$\{\omega : Y(\omega) \leq y\} = \left\{\omega \in \Omega : X(\omega) \leq \frac{y-\alpha}{\beta}\right\}$$

⁹⁰Infatti

$$Y(\omega) \leq y \Leftrightarrow \alpha + \beta X(\omega) \leq y \Leftrightarrow X(\omega) \geq \frac{y-\alpha}{\beta}$$

e quindi

$$\{\omega : Y(\omega) \leq y\} = \left\{\omega \in \Omega : X(\omega) \geq \frac{y-\alpha}{\beta}\right\}$$

Verifica della soluzione per il secondo problema: funzione di densità

Cominciamo con il caso in cui $F_X(x)$ è continua, derivabile con derivata continua in tutto \mathbb{R} , ossia

$$f_X(x) = \frac{d}{dx} F_X(x)$$

caso $\beta > 0$ In questo caso, come visto in precedenza,

$$F_Y(y) = F_X\left(\frac{y-\alpha}{\beta}\right),$$

da cui $F_Y(y)$ è continua, derivabile, con derivata continua in ogni y , e la derivata è la densità di probabilità. Inoltre si ha, utilizzando la regola della derivazione della funzione composta⁹¹

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y-\alpha}{\beta}\right) = \frac{d}{dx} F_X(x) \Big|_{x=\frac{y-\alpha}{\beta}} \frac{d}{dy} \frac{y-\alpha}{\beta} = f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{\beta}$$

caso $\beta < 0$ In questo caso, essendo F_X una funzione continua,

$$F_Y(y) = 1 - F_X\left(\left(\frac{y-\alpha}{\beta}\right)^-\right) = 1 - F_X\left(\frac{y-\alpha}{\beta}\right),$$

da cui procedendo in modo simile al caso precedente

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} \left(1 - F_X\left(\frac{y-\alpha}{\beta}\right)\right) = - \frac{d}{dx} F_X(x) \Big|_{x=\frac{y-\alpha}{\beta}} \frac{d}{dy} \frac{y-\alpha}{\beta} = -f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{\beta}$$

Si osservi che, in questo caso ($\beta < 0$) si ottiene il segno negativo, come del resto doveva essere: infatti così si ottiene che⁹² $\frac{d}{dy} F_Y(y) \geq 0$, come deve essere, in quanto F_Y è una funzione crescente in senso lato (o non decrescente).

Si osservi ancora che possiamo esprimere il risultato ottenuto in entrambi i casi nel seguente modo:

$$\frac{d}{dy} F_Y(y) = f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{|\beta|}$$

Unificando così i due casi.

Se invece la funzione F_X non avesse derivata continua, i conti effettuati sarebbero validi solo nei punti in cui la derivata di F_X esiste. Tali conti mostrano come, se Y ammette densità, allora l'unica funzione candidata ad essere la funzione di densità di probabilità $f_Y(y)$ sia appunto

$$f_Y(y) = f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{|\beta|}.$$

Non diamo qui la dimostrazione generale che questo è effettivamente il caso e rimandiamo ad esempio al testo di Baldi per la dimostrazione. Ricordiamo solo che, in generale, se Z è una variabile aleatoria, con funzione di distribuzione $F_Z(x)$, allora può accadere che F_Z ammetta derivata in ogni x , esclusi un numero finito di punti, ma che la variabile aleatoria non ammetta densità, o in altre parole, che la derivata non possa essere la funzione di densità di Z , come accade nell'Esempio 13.16.

⁹¹Si ricordi che, se $h(\cdot)$ è derivabile in t , e $\varphi(\cdot)$ derivabile in $h(t)$, allora

$$\frac{d}{dt} \varphi(h(t)) = \frac{d}{dx} \varphi(x) \Big|_{x=h(t)} \frac{d}{dt} h(t).$$

⁹²Si consideri che $f_X\left(\frac{y-\alpha}{\beta}\right) \geq 0$ e che $-\frac{1}{\beta} \geq 0$.

Esempio 13.17. *Una trasformazione affine di X con distribuzione Uniforme in $(0, 1)$ è ancora uniforme.*

Innanzitutto ricordiamo che $Z \sim R(a, b)$, ovvero che Z ha distribuzione uniforme nell'intervallo (a, b) significa che la sua funzione di distribuzione $F_Z(z)$ vale

$$F_Z(z) = \begin{cases} 0 & \text{per } z < a \\ \frac{z-a}{b-a} & \text{per } a \leq z < b \\ 1 & \text{per } z \geq b \end{cases}$$

o equivalentemente che la sua densità di probabilità $f_Z(z)$ vale

$$f_Z(z) = \begin{cases} 0 & \text{per } z < a \\ \frac{1}{b-a} & \text{per } a < z < b \\ 0 & \text{per } z > b \end{cases}$$

In particolare quindi, per $X \sim R(0, 1)$, si ha

$$F_X(x) = \begin{cases} 0 & \text{per } x < 0 \\ x & \text{per } 0 \leq x < 1 \\ 1 & \text{per } x \geq 1 \end{cases} \quad e \quad f_X(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 & \text{per } 0 < x < 1 \\ 0 & \text{per } x > 1 \end{cases}$$

Sia ora

$$Y = \alpha + \beta X.$$

Innanzitutto notiamo che Y assume valori tra α e $\alpha + \beta$, se $\beta > 0$, mentre assume i valori tra $\alpha + \beta$ e α , se $\beta < 0$. Quindi possiamo immediatamente capire che $Y(\Omega) = (\min(\alpha, \alpha + \beta), \max(\alpha, \alpha + \beta))$. Ciò ci fa subito "sospettare" che Y sia appunto uniforme in tale intervallo. Ciò si può dimostrare immediatamente notando che la sua densità di probabilità $f_Y(y)$ vale

$$f_Y(y) = f_X\left(\frac{y-\alpha}{\beta}\right) \frac{1}{|\beta|} = \begin{cases} 0 & \text{per } \frac{y-\alpha}{\beta} < 0 \\ 1 \cdot \frac{1}{|\beta|} & \text{per } 0 < \frac{y-\alpha}{\beta} < 1 \\ 0 & \text{per } \frac{y-\alpha}{\beta} > 1 \end{cases}$$

ovvero, distinguendo a seconda del segno di β , e riscrivendo le condizioni su y in modo più esplicito

se $\beta > 0$	se $\beta < 0$
$f_Y(y) = \begin{cases} 0 & \text{per } y < \alpha \\ \frac{1}{ \beta } & \text{per } \alpha < y < \alpha + \beta \\ 0 & \text{per } y > \alpha + \beta \end{cases}$	$f_Y(y) = \begin{cases} 0 & \text{per } y > \alpha \\ \frac{1}{ \beta } & \text{per } \alpha + \beta < y < \alpha \\ 0 & \text{per } y < \alpha + \beta \end{cases}$

Analogamente si può procedere con la funzione di distribuzione. Consideriamo prima il caso $\beta > 0$,

$$F_Y(y) = F_X\left(\frac{y-\alpha}{\beta}\right) = \begin{cases} 0 & \text{per } \frac{y-\alpha}{\beta} < 0 \\ \frac{y-\alpha}{\beta} & \text{per } 0 \leq \frac{y-\alpha}{\beta} < 1 \\ 1 & \text{per } \frac{y-\alpha}{\beta} \geq 1 \end{cases}$$

e poi il caso $\beta < 0$,

$$F_Y(y) = 1 - F_X\left(\left(\frac{y-\alpha}{\beta}\right)^-\right) = \begin{cases} 1 & \text{per } \frac{y-\alpha}{\beta} < 0 \\ 1 - \frac{y-\alpha}{\beta} = \frac{\alpha+\beta-y}{\beta} = \frac{y-(\alpha+\beta)}{|\beta|} & \text{per } 0 \leq \frac{y-\alpha}{\beta} < 1 \\ 0 & \text{per } \frac{y-\alpha}{\beta} \geq 1 \end{cases}$$

Ovvero, riscrivendo le condizioni su y in modo più esplicito,

$se \beta > 0$	$se \beta < 0$
$F_Y(y) = \begin{cases} 0 & \text{per } y < \alpha \\ \frac{y-\alpha}{\beta} & \text{per } \alpha \leq y < \alpha + \beta \\ 1 & \text{per } y \geq \alpha + \beta \end{cases}$	$F_Y(y) = \begin{cases} 1 & \text{per } y > \alpha \\ \frac{y-(\alpha+\beta)}{ \beta } & \text{per } \alpha + \beta < y \leq \alpha \\ 0 & \text{per } y \leq \alpha + \beta \end{cases}$

Esempio 13.18. Una trasformazione lineare (cioè con $\alpha = 0$) di X con distribuzione esponenziale di parametro λ è ancora esponenziale, di parametro $\frac{\lambda}{\beta}$, purché $\beta > 0$.

Innanzitutto notiamo che se $Y = \beta X$, con $X \sim \text{Exp}(\lambda)$ e se β è positivo, allora $Y(\Omega) = \mathbb{R}^+$. Allora

$$F_Y(y) = F_X\left(\frac{y}{\beta}\right) = \begin{cases} 0 & \text{per } \frac{y}{\beta} < 0 \\ 1 - e^{-\lambda \cdot \frac{y}{\beta}} & \text{per } \frac{y}{\beta} \geq 0 \end{cases}$$

o meglio,

$$F_Y(y) = \begin{cases} 0 & \text{per } y < 0 \\ 1 - e^{-\frac{\lambda}{\beta} y} & \text{per } y \geq 0 \end{cases}$$

che dimostra l'asserto.

Esercizio proposto 13.8. Si ritrovi il risultato del precedente Esempio 13.18 attraverso il calcolo della densità.

Esempio 13.19. Una trasformazione affine di X con distribuzione gaussiana standard $N(0, 1)$ è ancora gaussiana: se $X \simeq N(0, 1)$ ed

$$Y = \mu + \sigma X$$

allora

$$Y \sim N(\mu, \sigma^2),$$

ovvero

$$f_Y(y) = \frac{1}{\sqrt{2\pi} \sigma^2} \exp\left\{-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right\}, \quad y \in \mathbb{R}.$$

Infatti, basta ricordare che

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} x^2\right\}, \quad x \in \mathbb{R},$$

in modo che, dalla formula generale, con $\alpha = \mu$ e $\beta = \sigma$ (da cui $|\beta| = \sqrt{\sigma^2}$) si ha

$$f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right) \frac{1}{\sqrt{\sigma^2}}.$$

È importante sottolineare il significato dei parametri: ricordando che $\mathbb{E}(X) = 0$ e che $\text{Var}(X) = \mathbb{E}(X^2) = 1$ si ha che il parametro μ rappresenta il valore atteso di Y

$$\mathbb{E}(Y) = \mathbb{E}(\mu + \sigma X) = \mu + \sigma \mathbb{E}(X) = \mu + \sigma \cdot 0 = \mu,$$

mentre il parametro σ^2 rappresenta la varianza di Y

$$\text{Var}(Y) = \mathbb{E}[(Y - \mu)^2] = \mathbb{E}[(\sigma X + \mu - \mu)^2] = \mathbb{E}(\sigma^2 X^2) = \sigma^2 \text{Var}(X) = \sigma^2.$$

Infine può essere utile notare che, per $\sigma > 0$, si ha

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right),$$

e quindi anche la funzione di distribuzione di una gaussiana di parametri μ e σ^2 , può essere calcolata attraverso l'uso delle tavole.

Invece per $\sigma < 0$, si ha

$$F_Y(y) = 1 - \Phi\left(\frac{y-\mu}{\sigma}\right) = \Phi\left(-\frac{y-\mu}{\sigma}\right), \quad (121)$$

dove l'uguaglianza deriva dal fatto che, qualunque sia $x \in \mathbb{R}$, vale la relazione

$$\Phi(-x) = 1 - \Phi(x). \quad (122)$$

Tale relazione deriva immediatamente dal fatto che, se $X \sim N(0, 1)$ allora

$$\Phi(-x) = P(X \leq -x) = P(-X \geq x) \stackrel{*}{=} P(X \geq x) = 1 - \Phi(x),$$

dove l'uguaglianza sovrastata dall'asterisco vale in quanto la variabile aleatoria $-X$ è ancora una variabile aleatoria gaussiana standard: $-X = \mu + \sigma X$, con $\mu = 0$ e $\sigma = -1$ e quindi $-X \sim N(0, (-1)^2) = N(0, 1)$.

Alternativamente (122) si può ottenere dal fatto che

$$\Phi(-x) = \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

cambiando variabile $z = -y$

$$= \int_{+\infty}^{+x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-z)^2} (-dz) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

e d'altra parte

$$\begin{aligned} 1 - \Phi(x) &= 1 - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\ &= \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \end{aligned}$$

Grafico della funzione di distribuzione e della densità di una gaussiana standard

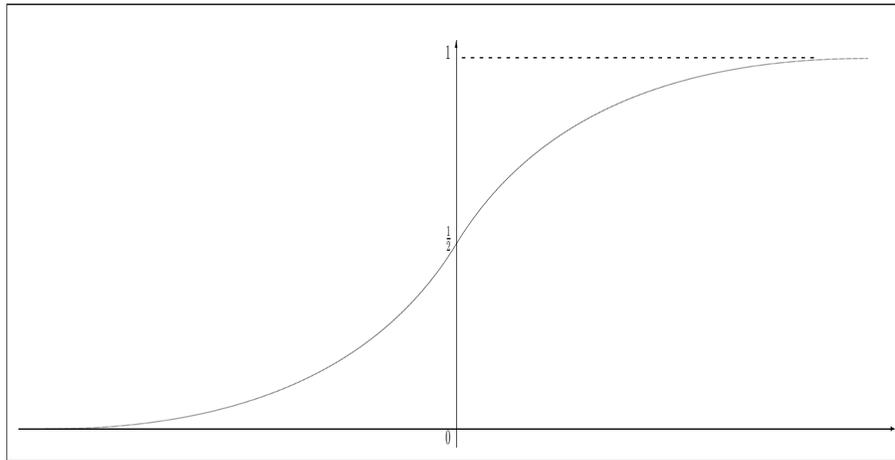


Figura 6: Grafico di $\Phi(x) = F_X(x)$, per X gaussiana standard

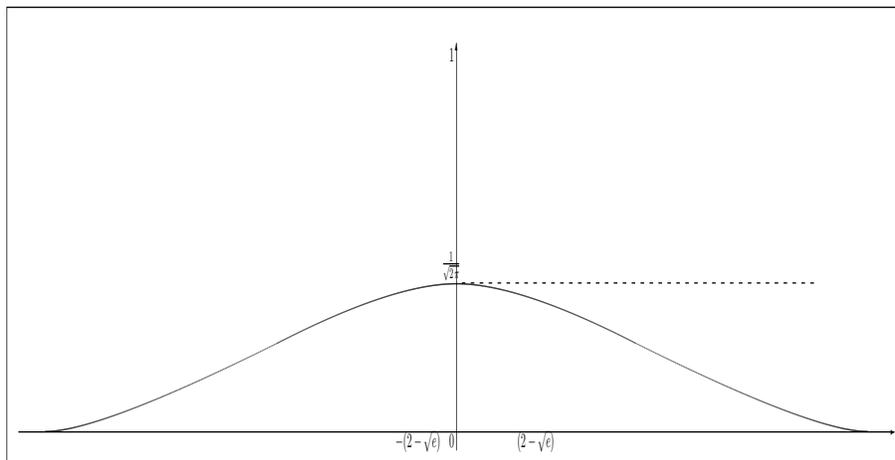


Figura 7: Grafico di $\varphi(x) = f_X(x)$, per X gaussiana standard

Studio della funzione $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$:

- $\varphi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \left(-\frac{2x}{2}\right) = -x \varphi(x)$
- $\varphi''(x) = -\varphi(x) + (-x)\varphi'(x) = -\varphi(x) + (-x)(-x)\varphi(x) = \varphi(x)(x^2 - 1)$ e di conseguenza
 - $\varphi(x)$ è convessa per $x < -1$
 - $\varphi(x)$ è concava per $-1 < x < 1$
 - $\varphi(x)$ è convessa per $x > 1$

Il massimo della funzione si ha per $x = 0$ e vale $\varphi(0) = \frac{1}{\sqrt{2\pi}} (\simeq 0,399)$.

Spiegazione dell'uso della tavola della gaussiana standard:

Per iniziare si noti che gli indici di riga sono i 35 numeri $\{0.0, 0.1, \dots, 3.3, 3.4\}$ che vanno da 0 a 3.4 e che differiscono tra loro di un decimo, mentre gli indici di colonna sono i 10 numeri $\{0.00, 0.01, \dots, 0.09\}$, che vanno da 0 a 0.09 e differiscono tra loro di un centesimo. Sommando un numero di riga, con uno di colonna si può ottenere uno tra i 350 valori di x che vanno da 0 a 3.49, e che differiscono tra loro di un centesimo. Viceversa ognuno di tali valori x , ad esempio $x = 1.43$, si può considerare come la somma della parte fino ai decimi più la parte dei centesimi, nell'esempio $x = 1.43 = 1.4 + 0.03$, individuando così un indice di riga, nell'esempio 1.4, ed uno di colonna, nell'esempio 0.03. Nella tavola, al posto di riga 1.4 e di colonna 0.03 si trova il valore di $\Phi(1.43) = 0.9236$, ovvero della funzione di distribuzione di una gaussiana standard, calcolata in $1.4+0.03$, e approssimato alla quarta cifra decimale.

I valori di $\Phi(x)$ per $x \geq 3.50$ si possono⁹³ approssimare con 1. Per quanto riguarda i valori di $\Phi(x)$ per valori negativi si usa la relazione (122), ossia

$$\Phi(-x) = 1 - \Phi(x),$$

ed in questo modo si può ottenere la funzione di distribuzione in⁹⁴ 699 valori tra $-3,49$ e $3,49$, equispaziati di un centesimo, ossia in

$$x = \frac{k}{100}, \quad \text{per } -349 \leq k \leq 349.$$

Dalla relazione precedente si ottiene ad esempio che $\Phi(-1.43) = 1 - \Phi(1.43) = 1 - 0.9236 = 0.0764$

Infine la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria Y con distribuzione $N(\mu, \sigma^2)$, previo una trasformazione affine di Φ : questo infatti ci assicura la relazione (121), ossia

$$P(Y \leq y) = \Phi\left(\frac{y-\mu}{\sigma}\right).$$

Ad esempio se $Z \sim N(1, 4)$ e si vuole calcolare $P(Z \leq 3.86)$, dalla (121) si ottiene che, essendo $Z \sim N(\mu, \sigma^2)$, con $\mu = 1$ e $\sigma^2 = 4$,

$$P(Z \leq 3.86) = \Phi\left(\frac{3.86-1}{2}\right) = \Phi(1.43) = 0.9236$$

Si noti infine che anche in questo la tavola ci permette di calcolare la funzione di distribuzione di una variabile aleatoria con distribuzione $N(\mu, \sigma^2)$ in 699 valori y , ossia i valori per i quali

$$-3,49 \leq \frac{y-\mu}{\sigma} \leq 3,49 \quad \Leftrightarrow \quad \mu - 3,49\sigma \leq y \leq \mu + 3,49\sigma$$

o piu' precisamente per $\frac{y-\mu}{\sigma} = \frac{k}{100}$, per $-349 \leq k \leq 349$, cioè

$$y = \mu + \frac{k}{100}\sigma, \quad \text{per } -349 \leq k \leq 349.$$

Problema inverso: trovare x_α tale che $\Phi(x_\alpha) = \alpha$

$\Phi(x_\alpha) = \alpha$.90	.95	.975	.99	.995	.999	.9995	.99995	.999995
x_α	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.891	4.417

La tabella è autoesplicativa, forse vale la pena solo di sottolineare che se $x \geq x_\alpha$ allora $\Phi(x) \geq \alpha$.

⁹³Ovviamente approssimare con 1 numeri maggiori o uguali a 0.9998 ha senso solo in problemi in cui la precisione non è fondamentale.

⁹⁴Si noti che $\Phi(0) = \Phi(-0) = 1/2$ e quindi non si tratta di 700 valori, ma solo di 699

14 Variabili aleatorie in casi più generali: indipendenza, Legge dei Grandi Numeri e Teorema Centrale del Limite.

In questa Lezione riprenderemo il discorso iniziato nella Lezione 10 a proposito della media aritmetica Y_n di n variabili aleatorie non correlate, iniziato con la **Proposizione 9** della Lez. 10, la cui dimostrazione è basata sulla disuguaglianza di Chebyshev.

Riprendiamo questo problema nella sezione 14.2 sulla Legge (debole) dei Grandi Numeri, sempre basandoci sulla disuguaglianza di Chebyshev, ma considerando invece una successione di variabili aleatorie.

Nella Lezione 10 ci siamo posti anche una importante domanda su quanto grande si dovesse prendere n in modo che la probabilità dell'evento

$$\{ \text{media aritmetica } Y_n \text{ e valore atteso differiscono di } \mathbf{poco} \}$$

sia **vicina ad uno**. La risposta era anch'essa basata sulla disuguaglianza di Chebyshev. Sempre nella sezione 14.2 si trova qualche approfondimento su questo problema.

Tuttavia, come viene osservato all'inizio della sezione 14.3, risultati più precisi alla domanda posta nella Lezione 10 si potrebbero ottenere se fosse nota la funzione di distribuzione della somma S_n delle variabili aleatorie, dato che $Y_n = \frac{1}{n} S_n$. In alcuni casi la distribuzione di S_n si può calcolare esplicitamente. Vedremo un paio di esempi nella sezione 14.3, ma è necessario di generalizzare la nozione di indipendenza, data in precedenza solo per variabili aleatorie negli spazi di probabilità finiti, e data solo per due variabili aleatorie, con una nozione di **indipendenza completa** per più variabili aleatorie (Definizioni 14.3 e 14.4). Per questo motivo tecnico la lezione inizia con diverse definizioni di indipendenza. Tuttavia anche nel caso di variabili (completamente) indipendenti, spesso non è facile, o addirittura non è possibile, ottenere esplicitamente la distribuzione della somma. Per la soluzione approssimata di questo problema ci aiuta il Teorema Centrale del Limite (**Proposizione 4**), come è illustrato nella **Proposizione 3**.

Terminiamo questo discorso introduttivo ricordando che, come preannunciato (sempre nella Lezione 10, e precisamente nella **Proposizione 10**) il Teorema Centrale del Limite, che riguarda successioni di variabili aleatorie **completamente indipendenti**, è connesso con una proprietà che riguarda la somma standardizzata di n variabili aleatorie. La **Proposizione 10** della Lez. 10, riguarda solo il caso di uno schema di n prove bernoulliane, in cui la media aritmetica diviene la frequenza relativa dei successi, ma si generalizza immediatamente al caso di variabili aleatorie più generali.

14.1 Famiglie di variabili aleatorie indipendenti

Molte delle definizioni e delle proprietà delle variabili aleatorie in spazi finiti valgono anche per le variabili aleatorie generali. Ad esempio si ha ancora che il valore atteso della somma di variabili aleatorie è la somma dei valori attesi e la regola per il calcolo della varianza della somma rimane identica.

In questo paragrafo ci chiediamo come si deve definire l'indipendenza per due variabili aleatorie X ed Y , nel caso generale, e daremo anche un'ulteriore definizione di indipendenza completa (o globale) per più di due variabili aleatorie.

Tra le varie caratterizzazioni di indipendenza, sicuramente non possiamo generalizzare⁹⁵ quella per cui $P(X = x, Y = y) = P(X = x)P(Y = y)$, in quanto, ad esempio, per le variabili aleatorie con funzione di distribuzione continua, la precedente relazione sarebbe solo una banalità: infatti si ridurrebbe alla relazione⁹⁶ $0 = 0$. Possiamo invece generalizzare quella data in **Proposizione 1** della Lezione 8, nel seguente modo.

Definizione 14.1 (indipendenza di due variabili aleatorie). *Due variabili aleatorie X ed Y si dicono **indipendenti** se e solo se comunque scelti due intervalli I e J , limitati o illimitati,*

$$P(X \in I, Y \in J) = P(X \in I) \cdot P(Y \in J).$$

Come nel caso discreto, anche nel caso generale vale il risultato che l'indipendenza di due variabili aleatorie implica⁹⁷ la non correlazione, mentre non è vero il viceversa.

Strettamente collegata alla precedente definizione, c'è la seguente

Definizione 14.2 (indipendenza a due a due di n variabili aleatorie). *Siano X_1, X_2, \dots, X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) . Esse si dicono **indipendenti a due a due** se comunque scelti $i \neq j$, con $i, j \in \{1, 2, \dots, n\}$, le due variabili aleatorie X_i ed X_j sono indipendenti, ovvero comunque scelti $i \neq j$, e comunque scelti I e J , intervalli (limitati o illimitati) di \mathbb{R} , si ha:*

$$P(X_i \in I, X_j \in J) = P(X_i \in I) \cdot P(X_j \in J).$$

Una condizione più forte dell'indipendenza a due a due è l'indipendenza globale

Un caso particolarmente interessante è quello in cui le variabili aleatorie X_i , per $i = 1, \dots, n$, sono completamente (o globalmente) indipendenti tra loro, ovvero

Definizione 14.3 (indipendenza di n variabili aleatorie). *Siano X_1, X_2, \dots, X_n n variabili aleatorie definite tutte sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) . Esse si dicono⁹⁸ **completamente (o globalmente) indipendenti tra loro** se comunque scelti J_1, J_2, \dots, J_n , intervalli (limitati o illimitati) di \mathbb{R} , si ha:*

$$P(X_1 \in J_1, X_2 \in J_2, \dots, X_n \in J_n) = P(X_1 \in J_1) \cdot P(X_2 \in J_2) \cdot \dots \cdot P(X_n \in J_n)$$

La precedente definizione implica l'indipendenza a due a due.

Proposizione 1 Se le n variabili aleatorie X_1, X_2, \dots, X_n sono completamente (o globalmente) indipendenti fra loro, allora lo sono anche a due a due.

Dimostrazione Per semplicità di notazione mostriamo solamente che X_1 ed X_2 sono indipendenti, ma la dimostrazione è essenzialmente la stessa nel caso generale di X_i ed X_j .

Il punto essenziale da osservare è che \mathbb{R} è un intervallo, e che gli eventi del tipo $\{X_k \in \mathbb{R}\}$ coincidono con l'evento certo, di conseguenza

$$\{X_1 \in J_1, X_2 \in J_2\} = \{X_1 \in J_1, X_2 \in J_2, X_3 \in \mathbb{R}, \dots, X_n \in \mathbb{R}\}$$

e quindi

$$\begin{aligned} P(X_1 \in J_1, X_2 \in J_2) &= P(X_1 \in J_1, X_2 \in J_2, X_3 \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\ &= P(X_1 \in J_1) \cdot P(X_2 \in J_2) \cdot P(X_3 \in \mathbb{R}) \cdot \dots \cdot P(X_n \in \mathbb{R}) \\ &= P(X_1 \in J_1) \cdot P(X_2 \in J_2). \end{aligned}$$

⁹⁵Tuttavia nel caso delle variabili aleatorie discrete questa caratterizzazione rimane valida, infatti le dimostrazioni della equivalenza delle caratterizzazioni rimangono sostanzialmente invariate, pur di sostituire a somme finite somme infinite, per cui ad esempio due variabili aleatorie X ed Y con $X(\Omega) = \{x_1, x_2, \dots\}$ ed $Y(\Omega) = \{y_1, y_2, \dots\}$ sono indipendenti se e solo se $P(X = x_h, Y = y_k) = P(X = x_k)P(Y = y_h)$ per ogni h e k .

⁹⁶Se $P(X = x) = 0$ per ogni $x \in \mathbb{R}$, allora anche $P(X = x, Y = y) = 0$, in quanto $\{X = x, Y = y\} \subseteq \{X = x\}$.

⁹⁷Ovviamente è necessario che le variabili aleatorie ammettano valore atteso finito.

⁹⁸A volte il termine completamente può essere trascurato, e si può parlare semplicemente di variabili aleatorie indipendenti tra loro.

Osservazione 1 Come già detto, quanto visto per le variabili aleatorie discrete vale anche per le variabili aleatorie in generale: in particolare se le variabili aleatorie sono indipendenti a due a due, allora la varianza della somma è la somma delle varianze. Alla luce della precedente **Proposizione 1**, lo stesso vale nel caso in cui le variabili aleatorie sono completamente (o globalmente) indipendenti tra loro.

Definizione 14.4 (indipendenza di una successione di variabili aleatorie). Sia $\{X_n; n = 1, 2, \dots\}$ una successione di variabili aleatorie, tutte definite sullo stesso spazio di probabilità (Ω, \mathcal{F}, P) . Si dice che sono **una successione di variabili aleatorie indipendenti** se comunque scelto un numero finito di esse, queste risultano completamente indipendenti tra loro.

14.2 Legge dei Grandi Numeri

Il risultato più importante di questa Lezione è noto come la **Legge debole dei grandi numeri**. Tale risultato è enunciato alla fine di questo paragrafo (**Proposizione 2**) e riguarda le successioni di variabili aleatorie indipendenti a due a due.

Prima di arrivare ad enunciare e dimostrare la legge debole dei grandi numeri, riprendiamo quanto visto utilizzando la disuguaglianza di Chebyshev nella **Proposizione 9** della Lezione 10, ma allargando un poco la prospettiva.

Prima di tutto va detto che la disuguaglianza di Chebyshev continua a valere anche nel caso di variabili aleatorie generali, con l'unica accortezza che **nel caso generale bisogna ipotizzare che esistano finiti valore atteso e varianza⁹⁹ della variabile aleatoria X** . Per cui, indicando come al solito $\mu = \mathbb{E}(X)$ e $\sigma^2 = \text{Var}(X)$ si ha

$$P(\{|X - \mu| > \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Anche la **Proposizione 9** continua a valere, **pur di assumere che esistano finiti valore atteso $\mathbb{E}(X_i)$ e varianza $\text{Var}(X_i)$** , che come al solito poniamo uguali rispettivamente a μ e σ^2 .

Proposizione 9(versione generale) Se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, e se esistano finiti valore atteso $\mathbb{E}(X_i) = \mu$ e varianza $\text{Var}(X_i) = \sigma^2$ allora

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2},$$

dove Y_n è la media aritmetica $Y_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$.

14.2.1 Approfondimenti sull'utilizzo della disuguaglianza di Chebyshev

Se X_1, X_2, \dots, X_n sono variabili aleatorie indipendenti a due a due, e con la stessa distribuzione, nell'**Osservazione 6** della Lezione 10 abbiamo visto come trovare il numero n di prove per cui la probabilità dell'evento "il valore atteso μ e la media aritmetica $Y_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ differiscono meno di una quantità prefissata ε " sia almeno $1 - \delta$, (nell'Esempio 10.6 ciò è stato applicato al caso di variabili binarie).

⁹⁹Sappiamo che possono esistere variabili aleatorie per le quali valore atteso e/o varianza non esistono, o valgono infinito. Questo problema non si pone nel caso finito in quanto in quel caso il calcolo del valore atteso e della varianza si riduce ad una somma finita e non presenta quindi nessun tipo di problema.

Infatti sappiamo che se

$$n \geq \frac{\sigma^2}{\delta \varepsilon^2} \quad (123)$$

allora

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \leq \delta \quad \Leftrightarrow \quad P(\{-\varepsilon \leq Y_n - \mu \leq \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} \geq 1 - \delta.$$

Nel caso particolare in cui le variabili X_i siano variabili binarie, con $P(X_i = 1) = \theta$ e $P(X_i = 0) = 1 - \theta$, allora $\mu = \theta$, $\sigma^2 = \theta(1 - \theta)$ e basta prendere

$$n \geq \frac{\theta(1 - \theta)}{\delta \varepsilon^2}$$

per ottenere che la probabilità dell'evento “la frequenza relativa dei successi¹⁰⁰ Y_n differisce dalla probabilità di successo θ meno di ε ” sia maggiore di $1 - \delta$.

Ovvero

$$n \geq \frac{\theta(1 - \theta)}{\delta \varepsilon^2} \quad \Rightarrow \quad P(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1 - \delta. \quad (124)$$

Nelle applicazioni si usa la frequenza relativa per “stimare” la probabilità θ : ovvero possiamo considerare il caso in cui possiamo osservare gli esiti di diversi esperimenti di uno stesso fenomeno, gli esperimenti sono condotti nelle stesse condizioni, per cui la probabilità di successo dell'esperimento è la stessa in tutte le prove, e infine **si assume che le prove siano stocasticamente indipendenti tra loro**, tuttavia **non si assume che sia noto esattamente il valore della probabilità di successo θ** .

In questo contesto la misura di probabilità dipende dal parametro θ ed è quindi più opportuno indicarla con P_θ , invece che con P .

Riprendendo quanto detto nell'**Osservazione 6** della Lezione 10 in questo contesto possiamo riscrivere

$$P_\theta(\{\theta - \varepsilon \leq Y_n \leq \theta + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\theta(1 - \theta)}{\varepsilon^2},$$

ma anche

$$P_\theta(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\theta(1 - \theta)}{\varepsilon^2}.$$

Questo secondo modo di scrivere è più interessante, in quanto, in questo contesto, mentre possiamo osservare Y_n , invece non conosciamo affatto θ . L'idea è che vorremmo poter “valutare” la probabilità θ con Y_n , con un errore al più di ε . Ovviamente in nessun caso, facendo degli esperimenti, avremo la garanzia che la frequenza relativa Y_n e la probabilità di successo θ differiscano meno di ε , tuttavia la disuguaglianza di Chebyshev ci permette di affermare che ciò accade con probabilità elevata, e permette anche di trovare delle limitazioni inferiori a tale probabilità.

A prima vista però sorge una difficoltà: sembra che per adoperare la disuguaglianza di Chebyshev sia necessario conoscere θ , mentre abbiamo assunto che θ non sia noto. Ma a questo problema si può ovviare osservando che la funzione $h(x) = x(1 - x)$ vale al massimo¹⁰¹ $\frac{1}{4}$ e quindi si ha

$$P_\theta(\{|Y_n - \theta| > \varepsilon\}) \leq \frac{1}{n} \frac{\theta(1 - \theta)}{\varepsilon^2} \leq \frac{1}{4\varepsilon^2 n},$$

¹⁰⁰Successo all' i -esima prova significa $X_i = 1$.

¹⁰¹La funzione $h(x) = x(1 - x)$ ha il suo punto di massimo in $x = \frac{1}{2}$ come si vede subito, e quindi $h(x) \leq h(\frac{1}{2}) = \frac{1}{4}$.

$$\Downarrow$$

$$P_{\theta}(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\varepsilon^2 n}.$$

Ciò permette di affermare che, **qualunque sia la probabilità di successo** θ , la probabilità che θ e la frequenza relativa Y_n differiscano meno di ε vale almeno $1 - \frac{1}{4\varepsilon^2 n}$.

Più interessante ancora, dal punto di vista operativo, è tuttavia il fatto che siamo in grado di rispondere **alla domanda**:

Quante prove si devono effettuare, ovvero quanto si deve prendere grande n , affinché, con probabilità almeno $1 - \delta$, la frequenza relativa differisca dalla probabilità di successo meno di ε ?

La risposta alla precedente domanda è molto semplice: **è sufficiente prendere**¹⁰²

$$n \geq \frac{1}{4\delta\varepsilon^2}, \quad (126)$$

in altre parole

$$n \geq \frac{1}{4\delta\varepsilon^2} \Rightarrow P_{\theta}(\{-\varepsilon \leq Y_n - \theta \leq \varepsilon\}) \geq 1 - \delta \quad \forall \theta \in (0, 1). \quad (127)$$

Infatti in tale caso (126) è equivalente a $\delta \geq \frac{1}{4\varepsilon^2 n}$ e quindi, **qualunque sia** θ

$$P_{\theta}(|Y_n - \theta| > \varepsilon) \leq \frac{1}{n} \frac{\theta(1-\theta)}{\varepsilon^2} \leq \frac{1}{4\varepsilon^2 n} \leq \delta,$$

\Downarrow

$$P_{\theta}(\{Y_n - \varepsilon \leq \theta \leq Y_n + \varepsilon\}) \geq 1 - \frac{1}{n} \frac{\theta(1-\theta)}{\varepsilon^2} \geq 1 - \frac{1}{4\varepsilon^2 n} \geq 1 - \delta.$$

Esempio 14.1. Sia Y_n la frequenza relativa dei successi in uno schema di Bernoulli di parametro θ . Si determini un n in modo che, **qualunque sia il valore di** θ , l'errore assoluto tra Y_n e θ sia minore di 0.1, con probabilità almeno 0.99.

Soluzione. Siamo nel caso precedente con $\varepsilon = 0.1 = \frac{1}{10}$ e con $1 - \delta = 0.99$, ovvero $\delta = \frac{1}{100}$. Quindi se

$$n \geq \frac{1}{4 \cdot \left(\frac{1}{10}\right)^2 \cdot \frac{1}{100}} = \frac{10000}{4} = 2500,$$

¹⁰²Si deve prendere

$$n \geq n(\varepsilon, \delta) \quad (125)$$

dove

$$n(\varepsilon, \delta) := \left\lceil \frac{1}{4\varepsilon^2\delta} \right\rceil,$$

cioè la parte intera superiore di $\frac{1}{4\varepsilon^2\delta}$. Si ricordi che la parte intera superiore di un numero reale a è l'intero k tale che $k - 1 < a \leq k$, ed è indicata appunto con $\lceil a \rceil$.

allora

$$P(-\varepsilon \leq Y_n - \theta \leq \varepsilon) \geq 0.99$$

E quindi, qualunque sia il valore di θ , è sufficiente prendere $n = 2500$.

Esempio 14.2. Calcolare il minimo valore di n per il quale, in uno schema di Bernoulli con probabilità θ , in base alla disuguaglianza di Chebyshev, si possa scrivere

$$P\left(\left\{\left|\frac{S_n}{n} - \theta\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{10},$$

qualunque sia il valore di θ .

Soluzione Si può procedere considerando che

$$\begin{aligned} P\left(\left\{\left|\frac{S_n}{n} - \theta\right| > \frac{1}{30}\right\}\right) &\leq \frac{\theta(1-\theta)}{n\left(\frac{1}{30}\right)^2} \leq \frac{1}{4n\left(\frac{1}{30}\right)^2} = \frac{900}{4n} \leq \frac{1}{10}, \\ &\Downarrow \\ \frac{900}{4\frac{1}{10}} &= \frac{9000}{4} = 2250 \leq n, \end{aligned}$$

oppure direttamente utilizzando la formula (127)

$$n \geq \frac{1}{4\varepsilon^2\delta} = \frac{1}{4\left(\frac{1}{30}\right)^2\frac{1}{10}} = \frac{900}{4\frac{1}{10}} = \frac{9000}{4} = 2250.$$

Osservazione 2. Si suggerisce di confrontare il risultato con quello dell'Esempio 10.6, in cui invece il valore di θ era dato, e quindi si era ottenuto, sempre utilizzando la disuguaglianza di Chebyshev¹⁰³, che bastava prendere $n = 2223$.

Osservazione 3 Si faccia attenzione al fatto che queste limitazioni inferiori sono date in base alla disuguaglianza di Chebyshev. I valori ottenuti per n sono sicuramente validi, ma sono eccessivamente grandi ed in genere più elevati del necessario. In realtà bastano valori di n più piccoli (daremo un'idea del motivo per cui i valori trovati sono eccessivi nella Lezione sul Teorema centrale del limite).

Osservazione 4 (Errore relativo) Va anche sottolineato che finora abbiamo valutato solo l'errore assoluto, tra Y_n e θ , mentre avrebbe più interesse l'errore relativo, ovvero $\left|\frac{Y_n - \theta}{\theta}\right|$: infatti se θ fosse dell'ordine di un centesimo, stimare θ con un errore assoluto dell'ordine di un decimo non sarebbe molto ragionevole¹⁰⁴. In questo caso la maggiorazione della disuguaglianza di Chebyshev permette di affermare che, per ogni θ

$$P_\theta\left(\left|\frac{Y_n - \theta}{\theta}\right| > \varepsilon\right) = P_\theta(|Y_n - \theta| > \varepsilon\theta) \leq \frac{1}{n} \frac{\theta(1-\theta)}{(\varepsilon\theta)^2} = \frac{1-\theta}{\theta} \frac{1}{n\varepsilon^2} \leq \delta,$$

per cui

$$n \geq \frac{1-\theta}{\theta} \frac{1}{\delta\varepsilon^2} \Rightarrow P_\theta\left(\left|\frac{Y_n - \theta}{\theta}\right| > \varepsilon\right) \leq \delta$$

¹⁰³In realtà nell'Esempio citato si è utilizzata la (124).

¹⁰⁴Se nel misurare la distanza fra due città si commette un errore dell'ordine di un metro, ci possiamo dichiarare completamente soddisfatti, mentre certamente non lo saremmo se l'errore dell'ordine di un metro riguardasse la misura di un tavolo da mettere in cucina!!!

Purtroppo, se θ non è noto, questa limitazione inferiore non è molto utile in quanto la funzione $h_1(x) = \frac{1-x}{x} = \frac{1}{x} - 1$ converge ad infinito per $x \rightarrow 0^+$, ed è quindi impossibile¹⁰⁵ trovare un valore di n che sia valido qualunque sia θ .

14.2.2 Formulazione della Legge dei Grandi Numeri

Nel formulare la domanda con la richiesta di scegliere n , c'è un punto che abbiamo volutamente trascurato fin qui. La possibilità di scegliere n presuppone di avere a disposizione un numero di eventi, (o di variabili aleatorie) completamente indipendenti potenzialmente grande a piacere¹⁰⁶.

Dal punto di vista matematico è più comodo poter affermare direttamente di avere a disposizione una successione di eventi completamente indipendenti e tutti con la stessa probabilità θ , o una successione di variabili aleatorie completamente indipendenti. Ciò presuppone uno spazio di probabilità Ω infinito, e quindi solo dopo aver introdotto gli spazi di probabilità generali e la nozione di successioni di variabili aleatorie, riformuliamo la **Proposizione 9** della Lezione 10 per successioni di variabili aleatorie. Tale formulazione è nota con il nome di Legge Debole dei Grandi Numeri.

Proposizione 2 (Legge Debole dei Grandi Numeri) Sia $\{X_i, i \geq 1\}$ una successione di v.a. indipendenti a due a due ed identicamente distribuite¹⁰⁷, per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu$ e $Var(X_i) = \sigma^2$, $S_n = \sum_{i=1}^n X_i$ e $Y_n = \frac{S_n}{n}$, si ha, qualunque sia $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) = \lim_{n \rightarrow \infty} P (|Y_n - \mu| > \varepsilon) = 0$$

Dimostrazione. Basta osservare che

$$0 \leq P (|Y_n - \mu| > \varepsilon) \leq \frac{1}{n} \frac{\sigma^2}{\varepsilon^2},$$

mandare n all'infinito ed usare il Teorema del confronto per le successioni numeriche:

$$0 \leq \lim_{n \rightarrow \infty} P (|Y_n - \mu| > \varepsilon) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \frac{\sigma^2}{\varepsilon^2} = 0.$$

Osservazione 5 Dalla **Proposizione 1** appare immediato che se $\{X_i, i \geq 1\}$ è una successione di variabili aleatorie completamente indipendenti, allora la Legge Debole dei Grandi Numeri continua a valere. Sotto questa ulteriore ipotesi vale anche il così detto Teorema centrale del limite che è oggetto del prossimo paragrafo. Nel prossimo paragrafo vedremo anche alcune relazioni tra questi due importantissimi risultati.

¹⁰⁵Diverso è il caso in cui, pur non conoscendo esattamente θ si sappia che $\theta \geq \theta_0$ con $\theta_0 > 0$: allora basterà prendere $n \geq \frac{1-\theta_0}{\theta_0} \frac{1}{\delta \varepsilon^2}$.

¹⁰⁶Si potrebbe ovviare al problema supponendo di avere una successione di spazi di probabilità $(\Omega_n, \mathcal{P}(\Omega_n), P^{(n)})$ e su ciascuno spazio n eventi $E_1^{(n)}, E_2^{(n)}, \dots, E_n^{(n)}$ che formano uno schema di Bernoulli con probabilità $\theta^{(n)} = \theta$ per ogni n .

¹⁰⁷Poiché le variabili aleatorie X_n hanno tutte la stessa distribuzione, si ha che se esistono finiti valore atteso e varianza di X_1 , allora esistono finiti valore atteso e varianza di X_i e coincidono con quelli di X_1 .

14.3 Somma di variabili aleatorie indipendenti e Teorema Centrale del Limite

Come abbiamo detto la disuguaglianza di Chebyshev permette di trovare delle limitazioni inferiori alle probabilità del tipo

$$P\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right)$$

che a loro volta permettono di dedurre la legge dei grandi numeri. Tuttavia se si conoscesse la funzione di distribuzione $F_{S_n}(x)$ della variabile aleatoria S_n , tale probabilità si potrebbe calcolare esattamente come

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| \leq \varepsilon\right) &= P(n(\mu - \varepsilon) \leq S_n \leq n(\mu + \varepsilon)) = F_{S_n}(n(\mu + \varepsilon)) - F_{S_n}(n(\mu - \varepsilon)) \\ &= F_{S_n}(n(\mu + \varepsilon)) - F_{S_n}(n(\mu - \varepsilon)) + P(\{S_n = n(\mu - \varepsilon)\}) \end{aligned}$$

Appare quindi chiaro che calcolare la distribuzione della somma di variabili aleatorie $S_n = X_1 + X_2 + \dots + X_n$ sia un problema interessante è, oltre che di per sé, anche per le connessioni con la legge dei grandi numeri e delle relazioni tra media aritmetica e valore atteso.

14.3.1 Esempi di calcolo della somma di variabili aleatorie indipendenti

Sappiamo calcolare esattamente la distribuzione della somma di variabili aleatorie in alcuni casi specifici. Ad esempio quando le X_i sono le indicatrici di eventi E_i che formano uno schema di Bernoulli di parametro θ , sappiamo che la distribuzione della somma è la distribuzione binomiale $b(n; \theta)$.

Esempio 14.3. Ancora sappiamo che se due variabili aleatorie X_1 ed X_2 sono indipendenti e hanno distribuzione binomiale di parametri n_i e θ (attenzione n_1 può essere diverso da n_2 , ma θ è lo stesso per $i = 1, 2$), allora la somma $X_1 + X_2$ ha distribuzione binomiale di parametri $n_1 + n_2$ e θ (confrontare lo svolgimento dell'Esercizio 8.4). Questo risultato si estende anche al caso di n variabili aleatorie completamente (o globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $\text{bin}(m; \theta)$ allora S_n ha distribuzione $\text{bin}(n \cdot m; \theta)$.

Esempio 14.4. Siano X_1 ed X_2 variabili aleatorie di Poisson di parametro $\lambda_1 (> 0)$ e $\lambda_2 (> 0)$ rispettivamente, ovvero per $i = 1, 2$

$$P(X_i = k) = \frac{\lambda_i^k}{k!} e^{-\lambda_i}, \quad k = 0, 1, 2, \dots$$

Si assuma che le variabili siano indipendenti, ovvero che

$$P(X_1 = h, X_2 = k) = P(X_1 = h)P(X_2 = k), \quad \text{per ogni } h, k \in \{0, 1, \dots\}$$

Si vede facilmente che la variabile aleatoria $X_1 + X_2$ ha distribuzione di Poisson di parametro $\lambda_1 + \lambda_2$, ovvero che

$$P(X_1 + X_2 = m) = \frac{(\lambda_1 + \lambda_2)^m}{m!} e^{-(\lambda_1 + \lambda_2)}, \quad m = 0, 1, 2, \dots \quad (128)$$

Infatti, per $m = 0, 1, \dots$ l'evento

$$\{X_1 + X_2 = m\} = \bigcup_{k=0}^{\infty} \{X_1 = k, X_2 = m - k\} = \bigcup_{k=0}^m \{X_1 = k, X_2 = m - k\},$$

in quanto $\{X_2 = m - k\} = \emptyset$ per $k = m + 1, m + 2, \dots$. Per cui

$$\begin{aligned} P(X_1 + X_2 = m) &= \sum_{k=0}^m P(X_1 = k, X_2 = m - k) = \sum_{k=0}^m P(X_1 = k)P(X_2 = m - k) \\ &= \sum_{k=0}^m \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{(m-k)}}{(m-k)!} e^{-\lambda_2} \\ &= \frac{1}{m!} \sum_{k=0}^m m! \frac{1}{k!} \frac{1}{(m-k)!} \lambda_1^k \lambda_2^{(m-k)} e^{-(\lambda_1 + \lambda_2)} \end{aligned}$$

dalla formula della potenza del binomio si ottiene la tesi, in quanto

$$\frac{1}{m!} \sum_{k=0}^m \binom{m}{k} \lambda_1^k \lambda_2^{(m-k)} = \frac{(\lambda_1 + \lambda_2)^m}{m!}, \quad m = 0, 1, 2, \dots$$

da cui si ottiene immediatamente la (128).

Anche questo risultato si estende anche al caso di n variabili aleatorie completamente (o globalmente) indipendenti tra loro: in particolare se le variabili aleatorie X_i hanno tutte la stessa distribuzione $Poiss(\lambda)$ allora S_n ha distribuzione $Poiss(n \cdot \lambda)$. Tuttavia va osservato che per calcolare $P(S_n \leq x)$, per $x \geq 0$, pur avendo a disposizione una formula esatta, ovvero

$$P(S_n \leq x) = \sum_{0 \leq k \leq [x]} P(S_n = k) = \sum_{k=0}^{[x]} \frac{(n \cdot \lambda)^k}{k!} e^{-n \cdot \lambda},$$

se n è “grande”, gli elementi della precedente sommatoria sono composti da fattori molto grandi $((n \cdot \lambda)^k)$ e molto piccoli $(e^{-n \cdot \lambda})$, e che quindi possono essere “scomodi” da calcolare. Nel prossimo paragrafo vedremo come ottenere un valore approssimato per $P(S_n \leq x)$ anche in questo esempio.

14.3.2 Approssimazione normale e Teorema Centrale del Limite

Più complesso risulta il calcolo della funzione di distribuzione della somma per altre variabili aleatorie¹⁰⁸, tuttavia si può innanzi tutto osservare come calcolare la funzione di distribuzione di S_n sia equivalente a calcolare la distribuzione di una sua trasformata affine¹⁰⁹ ovvero:

se a_n e b_n sono numeri reali, con $b_n > 0$, allora¹¹⁰

$$\{S_n \leq x\} = \left\{ \frac{S_n - a_n}{b_n} \leq \frac{x - a_n}{b_n} \right\}$$

¹⁰⁸Questo argomento viene svolto nel caso generale nei successivi corsi di Calcolo delle Probabilità, e richiede nozioni di Analisi, come ad esempio gli integrali di funzioni di più variabili.

¹⁰⁹Più in generale, data la funzione di distribuzione di una variabile aleatoria X , è sempre possibile ottenere la distribuzione della variabile aleatoria $Y = \alpha + \beta X$, il caso successivo è un caso particolare di questo, con $X = S_n$, $\alpha = -\frac{a_n}{b_n}$ e $\beta = \frac{1}{b_n}$.

A questo proposito si veda la sezione sulle Trasformazioni affini di variabili aleatorie.

¹¹⁰Infatti

$$\omega \in \{S_n \leq x\} \iff S_n(\omega) \leq x \iff \frac{S_n(\omega) - a_n}{b_n} \leq \frac{x - a_n}{b_n} \iff \omega \in \left\{ \frac{S_n - a_n}{b_n} \leq \frac{x - a_n}{b_n} \right\}$$

Una scelta naturale per a_n e per b_n è quella che trasforma S_n in una variabile aleatoria standard, ovvero quella di prendere $a_n = \mathbb{E}(S_n)$ e $b_n = \sqrt{\text{Var}(S_n)}$.

In questo modo infatti, per la disuguaglianza di Chebyshev, sappiamo che, qualunque siano n ed $\alpha > 0$

$$P(-\alpha \leq \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \leq \alpha) \geq 1 - \frac{1}{\alpha^2}.$$

Alla luce della seguente **Proposizione 4**, nota come Teorema Centrale del Limite (o anche Teorema del Limite Centrale), si può dimostrare il seguente risultato.

Proposizione 3 (approssimazione normale) Se le variabili aleatorie X_i , per $i = 1, 2, \dots, n$ sono (globalmente o completamente) indipendenti, hanno la stessa distribuzione, ammettono valore atteso finito $\mu = \mu_X$, varianza finita $\sigma^2 = \sigma_X^2$ e non nulla, allora $\mathbb{E}(S_n) = n\mu$, $\text{Var}(S_n) = n\sigma^2 > 0$, e

$$F_{S_n}(x) = P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq \frac{x - n\mu}{\sqrt{n\sigma^2}}\right) \simeq \Phi\left(\frac{x - n\mu}{\sqrt{n\sigma^2}}\right), \quad (129)$$

dove $\Phi(x)$ è la funzione di distribuzione di una variabile aleatoria gaussiana standard $N(0, 1)$.

A titolo di esempio riprendiamo il caso in cui le variabili aleatorie X_i hanno distribuzione di Poisson di parametro λ . Si vede facilmente che $E(X_i) = \lambda$ e che $\text{Var}(X_i) = \lambda$. In particolare, per $\lambda = 1$, $n = 100$ ed $x = 100$, ovvero se vogliamo calcolare approssimativamente $P(S_{100} \leq 100)$ attraverso $\Phi\left(\frac{x - n\mu}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{100 - 100}{\sqrt{100}}\right) = \Phi(0) = \frac{1}{2}$.

La dimostrazione della precedente **Proposizione 3** si basa sul seguente risultato basilare e che svolge un ruolo "centrale" nel Calcolo delle Probabilità.

Proposizione 4 (Teorema Centrale del Limite) Sia $\{X_i, i \geq 1\}$ una successione di v.a. indipendenti ed identicamente distribuite, per le quali esistano finiti valore atteso e varianza. Posto $\mathbb{E}(X_i) = \mu$ e $\text{Var}(X_i) = \sigma^2$, si assuma che $\sigma^2 > 0$. Allora indicando con S_n^* variabile aleatoria standardizzata di S_n , si ha

$$S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}, \quad (130)$$

e, indicando con $F_{S_n^*}(x)$ la funzione di distribuzione di S_n^* , si ha

$$\lim_{n \rightarrow \infty} F_{S_n^*}(x) = \lim_{n \rightarrow \infty} P(S_n^* \leq x) = \Phi(x), \quad (131)$$

dove Φ è la funzione di distribuzione di una variabile aleatoria Gaussiana standard: in altre parole

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy. \quad (132)$$

Inoltre il limite è uniforme per $x \in \mathbb{R}$, ovvero

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \left| P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) - \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \right| = 0. \quad (133)$$

Non diamo la dimostrazione di questo risultato, ma notiamo solo che la (130) si dimostra tenendo conto che $\mathbb{E}(S_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu$ e che per la completa indipendenza dalle variabili aleatorie X_i , si ha¹¹¹

$$\text{Var}(S_n) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2.$$

La precedente relazione sarebbe valida anche nel caso in cui le variabili aleatorie fossero solo indipendenti a due a due (o addirittura solo non correlate), ma sottolineiamo il fatto che, mentre la Legge Debole dei Grandi Numeri, vale sotto l'ipotesi di indipendenza a due a due, e non è necessario supporre $\sigma^2 > 0$, invece **per il Teorema Centrale del Limite, serve la condizione di completa indipendenza** e ovviamente è **necessario supporre** $\sigma^2 > 0$, altrimenti non si potrebbe nemmeno formulare la tesi.

Dimostrazione della Proposizione 3 (Approssimazione normale) Fondamentale per dimostrare l'approssimazione (129) della funzione di distribuzione della somma S_n è il fatto che la convergenza in (132) sia uniforme¹¹²: infatti, posto

$$E_n(x) = F_{S_n^*}(x) - \Phi(x), \text{ e } x_n = \frac{x - n\mu}{\sqrt{n}\sigma^2},$$

si ha

$$F_{S_n}(x) = P(S_n \leq x) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq \frac{x - n\mu}{\sqrt{n}\sigma^2}\right) = F_{S_n^*}(x_n) = \Phi(x_n) + E_n(x_n),$$

per cui

$$|F_{S_n}(x) - \Phi(x_n)| = |E_n(x_n)| \leq \sup_{x \in \mathbb{R}} |E_n(x)|.$$

Basta solo osservare che (133) garantisce che $\sup_{x \in \mathbb{R}} |E_n(x)|$ converge a zero¹¹³ per n che tende all'infinito.

¹¹¹Come già osservato nell'**Osservazione 1**

¹¹²Si osservi che in generale le condizioni che

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \quad \lim_{n \rightarrow \infty} x_n = x$$

non implicano che

$$\lim_{n \rightarrow \infty} f_n(x_n) = f(x).$$

Basta pensare al seguente **controesempio**:

$$\begin{cases} f_n(x) = 0 & x < \frac{1}{n}, \\ f_n(x) = 1 & x \geq \frac{1}{n} \end{cases} \quad \begin{cases} f(x) = 0 & x \leq 0, \\ f(x) = 1 & x > 0 \end{cases}$$

Chiaramente se $x \leq 0$ allora $f_n(x) = 0$ e quindi $\lim_{n \rightarrow \infty} f_n(x) = f(x) = 0$, analogamente, se $x > 0$, allora per $n > \frac{1}{x}$ si ha $f_n(x) = 1$, e quindi $\lim_{n \rightarrow \infty} f_n(x) = f(x) = 1$. Inoltre, posto $x_n = \frac{1}{n}$, si ha $\lim_{n \rightarrow \infty} x_n = 0$, tuttavia ovviamente $f_n(x_n) = f_n(\frac{1}{n}) = 1$ che non converge ad $f(x) = f(0) = 0$.

¹¹³Pur essendo assolutamente al di fuori dell'ambito di un corso elementare di probabilità, vale la pena di ricordare che esistono delle maggiorazioni per $\sup_{x \in \mathbb{R}} |E_n(x)|$, nel caso in cui si supponga che il valore atteso $\mathbb{E}(|X|^3)$ esista e sia finito. In particolare è stato dimostrato che

$$\sup_{x \in \mathbb{R}} |E_n(x)| \leq \frac{C}{\sqrt{n}} \frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

con C costante. Il valore di C non è noto esattamente ma è noto che $0.4097 \leq C \leq 0.7975$, in particolare quindi vale

$$\sup_{x \in \mathbb{R}} |E_n(x)| \leq \frac{1}{\sqrt{n}} \frac{\mathbb{E}(|X|^3)}{\sigma^3},$$

I primi a fornire maggiorazioni in questa direzione sono stati Berry ed Eessen all'inizio degli anni 40 dello scorso XX secolo.

14.3.3 Altre conseguenze del Teorema Centrale del Limite e relazioni con la legge dei grandi numeri

Si osservi che il Teorema Centrale del Limite implica che

$$\lim_{n \rightarrow \infty} P\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq b\right) = \Phi(b) - \Phi(a),$$

come si vede subito applicando la proprietà che per ogni variabile aleatoria X , con funzione di distribuzione $F(x)$, si ha $P(a < X \leq b) = F(b) - F(a)$.

Il Teorema Centrale del Limite implica anche che

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq b\right) = \Phi(b) - \Phi(a),$$

infatti, come si vede facilmente¹¹⁴,

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma^2} = a\right) = 0.$$

Dopo questa osservazione possiamo tornare indietro alle relazioni tra Legge dei Grandi Numeri e Teorema Centrale del Limite.

Indicando, come al solito, con Y_n la media aritmetica $\frac{S_n}{n}$, si ha

$$Y_n - \mu = \frac{S_n - n\mu}{n},$$

e quindi la standardizzata della media aritmetica Y_n coincide con la standardizzata della somma S_n , cioè

$$\sqrt{\frac{n}{\sigma^2}}(Y_n - \mu) = \sqrt{\frac{n}{\sigma^2}} \frac{S_n - n\mu}{n} = \frac{S_n - n\mu}{\sqrt{n}\sigma^2}$$

e inoltre

$$\{|Y_n - \mu| \leq \varepsilon\} = \left\{-\varepsilon \leq \frac{S_n - n\mu}{n} \leq \varepsilon\right\} = \left\{-\sqrt{\frac{n}{\sigma^2}} \varepsilon \leq \frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq \sqrt{\frac{n}{\sigma^2}} \varepsilon\right\}.$$

Di conseguenza

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - 1, \quad (134)$$

infatti

$$\begin{aligned} P(\{|Y_n - \mu| \leq \varepsilon\}) &= P\left(-\varepsilon \leq \frac{S_n - n\mu}{n} \leq \varepsilon\right) = P\left(-\sqrt{\frac{n}{\sigma^2}} \varepsilon \leq \frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq \sqrt{\frac{n}{\sigma^2}} \varepsilon\right) \\ &= P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma^2} = -\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) \\ &\quad + \Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - \Phi\left(-\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) + E_n\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - E_n\left(-\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) \\ &\simeq \Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - \Phi\left(-\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) = 2\Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - 1. \end{aligned}$$

Si ottiene di nuovo la stessa tesi della Legge debole dei Grandi Numeri, (**Proposizione 2**), ma sotto l'ipotesi più restrittiva che le variabili aleatorie siano completamente indipendenti. Infatti mandando n all'infinito nella precedente relazione (134) si ottiene

$$\begin{aligned} &\lim_{n \rightarrow \infty} P(\{|Y_n - \mu| \leq \varepsilon\}) \\ &= \lim_{n \rightarrow \infty} 2\Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - 1 + E_n\left(\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right)\right) - E_n\left(-\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) = 2 - 1 + 0 - 0 = 1. \end{aligned}$$

¹¹⁴Si osservi che

$$P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma^2} = a\right) \leq P\left(a - \frac{1}{n} \leq \frac{S_n - n\mu}{\sqrt{n}\sigma^2} \leq a\right) = \Phi(a) - \Phi\left(a - \frac{1}{n}\right) \xrightarrow[n \rightarrow \infty]{} 0$$

Esempio 14.5. Sia X_1 una variabile aleatoria che può assumere i valori $0, \frac{1}{2}, 1, \frac{3}{2}$ e con

$$p_{X_1}(0) = P(X_1 = 0) = \frac{1}{10}, \quad p_{X_1}(\frac{1}{2}) = P(X_1 = \frac{1}{2}) = \frac{1}{10},$$

$$p_{X_1}(1) = P(X_1 = 1) = \frac{4}{10}, \quad p_{X_1}(\frac{3}{2}) = P(X_1 = \frac{3}{2}) = \frac{4}{10}.$$

Si ponga il valore atteso di X_1 uguale a μ e la sua varianza uguale a σ^2 .

Siano $X_1, X_2, X_3, \dots, X_{100}$ delle variabili aleatorie con la stessa distribuzione di X_1 e completamente (o globalmente) indipendenti tra loro e si ponga $Y_{100} \equiv \frac{\sum_{j=1}^{100} X_j}{100}$. Utilizzando il Teorema Centrale del Limite, approssimare la probabilità

$$P\left(\left\{\mu - \frac{1}{10} \leq Y_{100} \leq \mu + \frac{1}{10}\right\}\right).$$

Soluzione Innanzi tutto come si trova facilmente si ha $\mu = \frac{21}{20}$ e $\sigma^2 = \frac{89}{400}$. Quindi la probabilità cercata è approssimata con

$$\begin{aligned} P\left(\left\{\mu - \frac{1}{10} \leq Y_{100} \leq \mu + \frac{1}{10}\right\}\right) &\simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\varepsilon\right) - 1 = 2\Phi\left(\sqrt{\frac{100}{\frac{89}{400}}}\frac{1}{10}\right) - 1 \\ &= 2\Phi\left(\sqrt{\frac{400}{89}}\right) - 1 = 2\Phi\left(\frac{20}{\sqrt{89}}\right) - 1 \\ &\simeq 2\Phi(2,1199) - 1 \simeq 2 \cdot 0.9826 - 1 = 1,9652 - 1 = 0.9652 \end{aligned}$$

Finora ci siamo posti il problema del tipo: **fissati n (grande) ed $\varepsilon > 0$, quanto vale approssimativamente la probabilità che media aritmetica e valore atteso differiscano di meno di ε ?**

Supponiamo ora di voler rispondere **in modo approssimato** alla domanda: **siano n (grande) e $\delta \in (0, 1)$ fissati, per quale valore di $\varepsilon = \varepsilon(n, \delta)$ posso affermare che**

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 1 - \delta?$$

Il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ . Tuttavia permette di dare una buona valutazione del tipo di comportamento di ε : (trascurando E_n) andiamo a mostrare che $\varepsilon = \varepsilon(n, \delta)$ è un infinitesimo dell'ordine di $\frac{1}{\sqrt{n}}$.

Prima di tutto invece di valutare esattamente

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 1 - \delta$$

consideriamo la (134), ossia che, per n sufficientemente grande,

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\varepsilon\right) - 1 \quad \Leftrightarrow \quad P(\{|Y_n - \mu| \leq \sqrt{\frac{\sigma^2}{n}}x\}) \simeq 2\Phi(x) - 1$$

con

$$x = \sqrt{\frac{n}{\sigma^2}}\varepsilon,$$

e quindi cerchiamo invece ε , o equivalentemente x , in modo che

$$2\Phi\left(\sqrt{\frac{n}{\sigma^2}}\varepsilon\right) - 1 = 2\Phi(x) - 1 = 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\frac{n}{\sigma^2}}\varepsilon\right) = 2\Phi(x) = \frac{2 - \delta}{2} = 1 - \frac{\delta}{2}.$$

Sicuramente esiste¹¹⁵ un valore $x = x_{1-\delta/2}$ per cui

$$\Phi(x_{1-\delta/2}) = 1 - \frac{\delta}{2};$$

Inoltre possiamo trovare un valore approssimato di $x_{1-\delta/2}$ utilizzando le tavole della gaussiana. Ad esempio per $\delta = 0.1$ si ottiene $1 - \delta/2 = 1 - 0.05 = 0.95$ ed $x_{1-\delta/2} = x_{0.95} = 1.645$.

A questo punto basta porre

$$\sqrt{\frac{n}{\sigma^2}} \varepsilon = x_{1-\delta/2} \quad \Leftrightarrow \quad \varepsilon = \varepsilon(n, \delta) = x_{1-\delta/2} \cdot \sqrt{\frac{\sigma^2}{n}} = \frac{x_{1-\delta/2} \cdot \sigma}{\sqrt{n}},$$

per ottenere il risultato desiderato.

Osservazione Possiamo riassumere quanto appena provato con l'affermazione che per n (grande) e $x_{1-\delta/2}$ tale che $\Phi(x_{1-\delta/2}) = 1 - \frac{\delta}{2}$, si ha

$$P(\{|Y_n - \mu| \leq x_{1-\delta/2} \frac{\sigma}{\sqrt{n}}\}) \simeq 1 - \delta.$$

Terminiamo questa sezione tornando invece al problema¹¹⁶ **in cui sia ε che δ sono fissati**, e supponiamo di voler rispondere **in modo approssimato** alla domanda: **per quali n posso affermare che**

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta?$$

Anche il seguente procedimento non è del tutto rigoroso, perché trascura l'errore di approssimazione E_n tra $F_{S_n^*}$ e Φ . Tuttavia permette di dare una buona valutazione del tipo di richiesta vada fatta su n per ottenere la limitazione inferiore richiesta.

Anche in questo problema, invece di cercare una limitazione inferiore esatta

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \geq 1 - \delta$$

sempre considerando che, per n sufficientemente grande,

$$P(\{|Y_n - \mu| \leq \varepsilon\}) \simeq 2\Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - 1$$

cerchiamo invece una limitazione inferiore

$$2\Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) - 1 \geq 1 - \delta \quad \Leftrightarrow \quad \Phi\left(\sqrt{\frac{n}{\sigma^2}} \varepsilon\right) \geq \frac{2 - \delta}{2} = 1 - \frac{\delta}{2}$$

Come nel caso precedente possiamo trovare un valore $x_{1-\delta/2}$ per cui

$$\Phi(x_{1-\delta/2}) = 1 - \frac{\delta}{2}.$$

Si osservi che, essendo Φ una funzione non decrescente¹¹⁷,

$$\Phi(x) \geq \Phi(x_{1-\delta/2}) = 1 - \frac{\delta}{2}, \quad \text{per ogni } x \geq x_{1-\delta/2},$$

¹¹⁵La funzione di distribuzione Φ è una funzione strettamente crescente e continua, e assume quindi tutti i valori $(0, 1)$.

¹¹⁶Di questo tipo di problema ci siamo occupati nella sezione degli approfondimenti sull'uso della disuguaglianza di Chebyshev, e la risposta è stata: basta prendere $n \geq n(\varepsilon, \delta)$, dove $n(\varepsilon, \delta)$ è definito in (123).

¹¹⁷In realtà basta trovare sulla tavola della gaussiana standard un valore $\bar{x}_{1-\delta/2}$ tale che

$$\Phi(\bar{x}_{1-\delta/2}) \geq 1 - \frac{\delta}{2}.$$

Il ragionamento fatto con $x_{1-\delta/2}$ si può ripetere mettendo $\bar{x}_{1-\delta/2}$ al posto di $x_{1-\delta/2}$.

A questo punto basta imporre che

$$\sqrt{\frac{n}{\sigma^2}} \varepsilon \geq x_{1-\delta/2} \Leftrightarrow \sqrt{n} \geq \frac{x_{1-\delta/2} \sqrt{\sigma^2}}{\varepsilon} \Leftrightarrow n \geq n_{TCL}(\varepsilon, \delta) := \frac{x_{1-\delta/2}^2 \sigma^2}{\varepsilon^2} \quad (135)$$

per ottenere il risultato desiderato.

Osservazione 6 Si confrontino tra loro (123) e (135): come si vede (123) e (135) sono molto simili, la seconda si ottiene sostituendo al posto di $\frac{1}{\delta}$, il valore $x_{1-\delta/2}^2$.

Quindi a parità di valori di ε e σ^2 si ottiene che la limitazione inferiore con la disuguaglianza di Chebyshev, pur essendo esatta, chiede

$$n \geq n_{Ch}(\varepsilon, \delta) = \frac{x_{1-\delta/2}^2}{\delta} n_{TCL}(\varepsilon, \delta)$$

Per capire quindi la differenza si osservi che se $\delta = 0,01$, allora $\frac{1}{\delta} = 100$, mentre, essendo $x_{1-\delta/2} = x_{1-0,005} = x_{0,995} = 2,576$ (come si può trovare dalle tavole) si ha che $x_{1-\delta/2}^2 = 6,635776$.

In questo caso

$$n_{Ch}(\varepsilon, \delta) = \frac{1}{\delta x_{1-\delta/2}^2} n_{TCL}(\varepsilon, \delta) = \frac{16}{0,01 \cdot 6,63577} n_{TCL}(\varepsilon, \delta) \simeq 15,0698 n_{TCL}(\varepsilon, \delta).$$

Se invece $\delta = 0,001$, allora $\frac{1}{\delta} = 1000$, mentre, essendo $x_{1-\delta/2} = x_{1-0,0005} = x_{0,9995} = 3,291$ (come si può trovare dalle tavole) si ha che $x_{1-\delta/2}^2 = 10,830681$.

In questo caso

$$n_{Ch}(\varepsilon, \delta) = \frac{x_{1-\delta/2}^2}{\delta} n_{TCL}(\varepsilon, \delta) = \frac{1}{0,001 \cdot 10,830681} n_{TCL}(\varepsilon, \delta) \simeq 92,3302 n_{TCL}(\varepsilon, \delta).$$

e quindi il valore di $n_{Ch}(\varepsilon, \delta)$ è circa 92 volte più grande di $n_{TCL}(\varepsilon, \delta)$, che è calcolato con il Teorema Centrale del Limite.

ALFABETO GRECO

α	A	alfa
β	B	beta
γ	Γ	gamma
δ	Δ	delta
ϵ o anche ε	E	epsilon
ζ	Z	zeta
η	H	eta
θ o anche ϑ	Θ	theta
ι	I	iota
κ	K	kappa
λ	Λ	lambda
μ	M	mu o anche mi
ν	N	nu o anche ni
ξ	Ξ	xi (csi)
o	O	omicron
π o anche ϖ	Π	pi greco
ρ o anche ϱ	R	rho
σ o, in fine parola, ς	Σ	sigma
τ	T	tau
υ	Υ o anche Y	üpsilon
ϕ o anche φ	Φ	phi (fi)
χ	X	chi
ψ	Ψ	psi
ω	Ω	omega