Università degli Studi di Roma "La Sapienza"
Anno Accademico 2003-2004

Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Triennale in Matematica

INTRODUZIONE AL CALCOLO DELLE PROBABILITÀ

(Versione incompleta e provvisoria, maggio 2004)

Fabio Spizzichino

A.A. 2003/04

versione del 25 maggio 2004

Indice

Introduzione

L'introduzione sara' scritta in seguito.

NOTAZIONI Per il momento sono state fatte alcune correzioni, rispetto alla versione di gennaio 2004, attualmente presente in rete, fino alla Sezione 8 inclusa, nella versione Marzo 2004. Tali correzioni sono evidenziate in blu solo sulla nuova in copia in rete[1], tranne per gli errori di stampa. Questa nuova versione (Maggio 2004) contiene, oltre alle correzioni delle sezioni 1−−8, anche le correzioni e gli ampliamenti delle Sezioni 9, 10, 11 e 12, ma contiene anche alcune ulteriori correzioni o aggiunte alle sezioni da 1 ad 8, rispetto alla versione di Marzo 2004. Le correzioni rispetto alla versione di gennaio 2004 sono ancora in blu, le correzioni o le aggiunte che riguardano le sezioni da 1 ad 8 sono invece in verde (sempre solo sulla copia in rete).

Inoltre, per comodita' del lettore e' stato inserito alla fine del libro l'alfabeto greco.

Sono in preparazione anche delle Lezioni relative agli spazi di probabilita' piu' generali, alle variabili aleatorie continue e alla Legge dei Grandi Numeri e al Teorema Centrale del Limite, si spera di finirle al piu' presto: nel frattempo si consultino i testi consigliati:

Dall'Aglio, "Calcolo delle Probabilita'"

P.Baldi, "Calcolo delle Probabilita' e Statistica"

Non sono evidenziati neanche alcuni cambiamenti di scrittura relativi all'uso delle parentesi: si e' 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 e' 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 probabilita' condizionate.

Rispetto alla precedente versione del gennaio 2004, il complementare di un evento E, pensato come sottoinsieme di Ω, viene indicato sempre (si spera) con $\overline{E}$.

Un altro cambiamento riguarda la notazione $A \subset B$ che e' stata sostituita da $A \subseteq B$.

Un cambiamento di notazioni riguarda infine la Sezione 11 in cui invece di $P_S(s|R = r)$ o $(P_R(r|S = s))$ si e' usato $p_S(s|R = r)$ (o $p_R(r|S = s)$)

Va infine detto che le correzioni sono a cura di Giovanna Nappo e 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 **molto gradite**, grazie.

[1]indirizzo web: http://www.mat.uniroma1.it/people/nappo/nappo.html#CPE2003-04

1 Fenomeni aleatori; spazio dei risultati elementari di un esperimento

Iniziamo con una discussione euristica mirante a giustificare la successiva definizione della nozione di *spazio finito di probabilita'*, che verra' 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: $A \equiv \{X_1 \leq X_2\}, B \equiv \{X_1 + X_2 \text{ pari}\}$, $C \equiv \{X_1 > 3\}, \ldots$.

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

E' anche facile rendersi conto che, all'interno della famiglia $\mathcal{E}$, e' 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 $\bigvee, \bigwedge, \widetilde{\cdot}$; siano E_1, E_2, E eventi appartenenti ad $\mathcal{E}$, allora la somma logica $E_1 \bigvee E_2$ coincide con l'evento:

$$\{\textit{si e' verificato almeno uno dei due eventi } E_1 \ \textit{ed} E_2\}$$

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

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

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

$$\{\textit{non si e' verificato l'evento } E\}.$$

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*

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

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* $E = \{X_1 > 3\}$ *e' un evento composto. In tale esperimento gli eventi semplici sono dati da*

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

e l'evento $E = \{X_1 > 3\} = \{X_1 = 4\} \cup \{X_1 = 5\} \cup \{X_1 = 6\}$.

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

$$\{X_1 = h,\ X_2 = k\} \qquad h = 1, 2, ..., 6;\ k = 1, 2, ..., 6$$

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. E' 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 $\omega_1, ..., \omega_N$ gli eventi elementari in un esperimento.

Definizione 1.2. $\Omega \equiv \{\omega_1, ..., \omega_N\}$ *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 e, per $E \in \mathcal{P}(\Omega)$, indichiamo con $|E|$ la ***cardinalita'*** 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 $\Omega = \{\text{permutazioni}^2 \text{ di } \{1, 2, 3, 4\}\}$, $|\Omega| = 24$,

Noi vogliamo analizzare quei casi in cui vi sia una *situazione di incertezza* (cioe' di mancanza di completa informazione) circa il modo di risultare dell'esperimento stesso. Cio' significa che non sappiamo a priori quale effettivamente si realizzera' 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 e' 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* $\Omega = \{\omega_1, ..., \omega_6\}$*. 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* $\{X > 5\}, \{X \text{ dispari}\}$, $\{X \leq 3\}$, $\{X \text{ numero primo}\}, ...$.

[2] Per la definizione di permutazione vedere piu' avanti la Sezione 3.2

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

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

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

(c) $\widetilde{E}_1$ si verifica se e solo se e' 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 piu' 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 comodita' 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***:

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

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

$$\mathcal{H}\left(E_1 \bigvee E_2\right) = \mathcal{H}(E_1) \cup \mathcal{H}(E_2),$$

$$\mathcal{H}\left(E_1 \bigwedge E_2\right) = \mathcal{H}(E_1) \cap \mathcal{H}(E_2),$$

$$\mathcal{H}\left(\widetilde{E}\right) = \complement(\mathcal{H}(E)).$$

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

Dovremo pero' 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, risultera' naturale attribuire un'interpretazione di tipo "logico" a varie semplici nozioni di tipo insiemistico; a tale proposito vediamo intanto lo specchietto presentato qui di seguito.

Interpretazione "logica" di nozioni di tipo insiemistico:

$\cdot$ $A \subseteq B$ 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***"

$\cdot\cdot$ Ω e' un evento vero qualunque evento elementare si verifichi, in quanto esso contiene tutti gli eventi elementari e dunque interpretiamo Ω come l'***evento certo***

$\cdots$ $\emptyset$, non contenendo alcuno degli eventi elementari possibili, e' un evento che non e' mai verificato; dunque interpretiamo $\emptyset$ come l'***evento impossibile***

$\cdots$ $A \cup B = \Omega$ 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*** *(e' certo che se ne verifichi* **almeno** *uno dei due)*

$\cdots$ $A \cap B = \emptyset$ 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*** *(e' certo che se ne verifichi* **al piu'** *uno dei due).*

1.1 Esercizi di verifica

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

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

e vedere i risultati *manque, passe, noir, rouge, pair, unpair,* come altrettanti eventi composti[3]. Supponiamo che nell'esperimento si verifichi l'evento elementare $\{16\}$. Quale degli eventi composti sopra elencati e' verificato e quale no?

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

$$\{\textit{si verifica esattamente un solo evento fra } A \text{ e } B\}.$$

Esercizio 1.3. Siano A, B e C eventi. Scrivete le espressioni degli eventi:
a) Almeno due tra questi si verificano;
b) Esattamente due tra questi si verificano;
c) Al piu' due tra questi si verificano;
d) 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:

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

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

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

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

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

[3] Si ricorda che la Roulette e' 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. Si puo' vedere la composizione colorazione di una roulette, sul sito web della professoressa Nappo: http://www.mat.uniroma1.it/people/nappo/.

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

$$\{\textit{almeno due elementi di tipo } B\} \cup \{\textit{l'elemento estratto alla seconda estrazione e' di tipo } B\}?$$

Esercizio 1.6. Consideriamo di nuovo l'urna di cui nell'esercizio precedente. Se ne effettuano sette successive estrazioni senza reinserimento (cioe' 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 pero' le tre estrazioni sono effettuate con reinserimento.
a) Elencate anche in questo caso gli eventi elementari.
b) 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 lezione[4]).

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

[4] Questo tipo di problemi sara' esaminato in generale nella Sezione 6

2 Spazi finiti di probabilita'

Introduciamo ora il concetto di **probabilita'**. 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. *Una* **misura di probabilita'** *o, piu' semplicemente, una* **probabilita'** *su* $(\Omega, \mathcal{P}(\Omega))$ *e' una funzione di insieme* $P : \mathcal{P}(\Omega) \to [0,1]$ *che soddisfa i seguenti assiomi*

i) $P(E) \geq 0$ *(***proprieta' di non-negativita'***)*[5]

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

iii) $P(E_1 \cup E_2) = P(E_1) + P(E_2)$, *per* $E_1 \cap E_2 = \emptyset$ *(***proprieta' di additivita'***).*

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

Osservazione 1. Prima di proseguire e' opportuno citare il fatto che esistono diverse possibili *interpretazioni* del termine probabilita': ad esempio probabilita' *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, e' giustificato imporre che la probabilita' 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*

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

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:

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

Elenchiamo ora alcune proprieta' della probabilita', che risultano conseguenze immediate degli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1.

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

[5] La proprieta' di non-negativita' e' evidentemente superflua se $P : \mathcal{P}(\Omega) \to [0,1]$.

Prima di tutto notiamo che la proprieta' *iii)* di additivita' si generalizza al caso di n eventi disgiunti a due a due

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

$$E_i \cap E_j = \emptyset, \qquad per\ i, j \in \{1, 2, \cdots, n\},\ con\ i \neq j; \tag{1}$$

allora si ha la condizione

$$P\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{i=1}^{n} P(E_i). \tag{2}$$

(la dimostrazione si ottiene facilmente per induzione su n)

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

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

$$p(\omega_i) = P(\{\omega_i\}), \qquad i = 1, ..., N,$$

risulta[6]

$$P(E) = \sum_{\omega_i \in E} p(\omega_i). \tag{3}$$

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

$$P\left(\overline{E}\right) = 1 - P(E). \tag{4}$$

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

$$P(\emptyset) = 0 \tag{5}$$

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

$$P(A) \leq P(B). \tag{6}$$

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

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{7}$$

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

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

allora si ha la condizione

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

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

[6] La somma $\sum_{\omega_i \in E} p(\omega_i)$ va intesa come la somma sugli indici $i \in \{1, \ldots, N\}$ tali che $\omega_i \in E$.

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

Ulteriori conseguenze degli assiomi della probabilita' verranno viste in seguito.

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

1) Prima definiamo P come una funzione di insieme, cioe'

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

che soddisfi gli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1, e poi definiamo la funzione di punto $p: \Omega \to [0,1]; \ \omega_i \to p(\omega_i) = P(\{\omega_i\})$. Questa funzione dovra' soddisfare le condizioni

$$p(\omega_i) \geq 0, \qquad \sum_{\omega_i \in \Omega} p(\omega_i) = 1. \tag{11}$$

2) Prima definiamo una funzione di punto

$$p: \Omega \to [0,1]; \ \omega_i \to p(\omega_i),$$

che soddisfi le condizioni (11) e poi definiamo una funzione di insieme $P : \mathcal{P}(\Omega) \to [0,1]$, attraverso la precedente formula (3).

E' facile verificare che tale funzione di insieme soddisfa gli assiomi *i)*, *ii)*, *iii)* della Definizione 2.1.

Osservazione 3 (**Probabilita' definite a meno di un fattore di proporzionalita'; normalizzazione**). Una misura di probabilita' sullo spazio $\Omega = \{\omega_1, ..., \omega_N\}$ e' individuata quando vengano assegnati i numeri $p(\omega_i) = P(\{\omega_i\})$, $(i = 1, 2, ..., N)$ soddisfacenti le condizioni (11). Supponiamo ora che $p(\omega_i)$, $(i = 1, 2, ..., N)$ siano assegnati a meno di una costante di proporzionalita'; supponiamo cioe' che siano assegnati dei numeri g_i $(i = 1, 2, ..., N)$, tali che

$$p(\omega_i) = K \cdot g_i \tag{12}$$

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

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

Notiamo che si usa esprimere brevemente la condizione (12) usando il seguente simbolismo:

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

Esempio 2.1 (dado non equilibrato). *Un dado ha sei facce numerate da 1 a 6; esso e' pesato in modo tale che ciascuna faccia abbia una probabilita' di presentarsi (in un singolo lancio) proporzionale al suo valore. Sia*

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

Trovare $P(A)$.

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

$$p(\omega_i) = K \cdot i, \qquad i = 1, ..., 6,$$

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

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

e

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

2.1 Esercizi di verifica

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

Esercizio 2.2. Siano A e B due eventi tali che

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

a) Quanto vale $P(A \cap B)$?
b) Qual e' la probabilita' che, fra A e B, se ne verifichi almeno uno?
c) Qual e' la probabilita' che se ne verifichi esattamente uno?

Esercizio 2.3. Una moneta viene lanciata due volte e poniamo

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

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

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

$$P(A \cup B \cup C) =$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

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

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

$$P\left(A \cap B \cap C\right) = P\left(A \cap \overline{B} \cap C\right) = 0.1,$$

$$P\left(\overline{A} \cap B \cap C\right) = P\left(A \cap B \cap \overline{C}\right) = 0.15,$$

$$P\left(A \cap \overline{B} \cap \overline{C}\right) = P\left(\overline{A} \cap B \cap \overline{C}\right) = P\left(\overline{A} \cap \overline{B} \cap C\right) = 0.05.$$

Calcolare
a) $P\left(A\right), P\left(B\right), P\left(C\right)$
b) $P\left(A \cup B\right), P\left(A \cup C\right)$
c) $P\left(A \cup B \cup C\right).$

3 Probabilita' "classiche" e calcolo combinatorio

3.1 Probabilita' "classiche"

Qui ci soffermiamo a trattare dei casi particolari, ma molto rilevanti, di spazi di probabilita' finiti. Sia dunque

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

e supponiamo che si voglia porre

$$p(\omega_i) = K, \qquad \forall \omega_i \in \Omega \tag{13}$$

per un'opportuna costante positiva K. Si vuole cioe' imporre che tutti i risultati elementari siano, fra di loro, equiprobabili. Ci riferiremo a tale caso dicendo che si ha una ***distribuzione di probabilita' uniforme sugli eventi elementari***.

Confrontando la posizione (13) con la condizione di normalizzazione (10), otteniamo immediatamente

$$p(\omega_i) = \frac{1}{N}, \qquad i = 1, ..., N$$

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

$$P(E) = \frac{|E|}{N}, \qquad \forall E \in \mathcal{P}(\Omega). \tag{14}$$

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

Soluzione. Si ha che Ω e' costituito dalle permutazioni[7] di n elementi; dunque $|\Omega| = n!$ L'evento

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

e' 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 (14) esprime il fatto che, nel caso in cui tutti gli eventi elementari di uno spazio finito sono equiprobabili, la probabilita' di un generico evento composto si calcola quale ***rapporto fra casi favorevoli e casi possibili***.

Si faccia attenzione al fatto che la (14) *non costituisce una definizione del concetto di probabilita', ma solo un caso particolare:* nel precedente paragrafo abbiamo gia' introdotto tale concetto in modo assiomatico e la suddetta formula e' 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 (13), il calcolo della probabilita' di un evento composto E si riduce al problema, ***combinatorio***, di individuare $N = |\Omega|$ e $|E|$.

[7] Per la definizione di permutazione vedere piu' avanti la Sezione 3.2

3.2 Calcolo combinatorio

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 probabilita'.

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 cardinalita' se e solo se fra essi e' 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:

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

Disposizioni con ripetizione di classe k di n elementi.

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

Tali disposizioni costituiscono dunque l'insieme $A^k = \overbrace{A \times A \times \cdots \times A}^{k\,volte}$, 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 cardinalita'

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

dove si e' usata la notazione *n fattoriale*, ovvero $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$.

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 e' $n!$.

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 [8].

[8] Alternativamente una combinazione di classe k di n elementi si puo' definire come ***un sottoinsieme di cardinalita' k di un insieme di cardinalita' n***.

Se C^n_k indica il numero delle combinazioni di classe k di n elementi, D^n_k indica il numero delle disposizioni senza ripetizione di classe k di n elementi, e P_k indica il numero delle permutazioni di k elementi, e' facile convincersi che

$$D^n_k = C^n_k \cdot P_k.$$

Il numero complessivo di tali combinazioni e' dunque dato da

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

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*.*

Soluzioni. 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 e' assegnata con una votazione indipendente dalle altre, si possono avere anche delle ripetizioni (cioe' e' 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 e' utile introdurre la seguente convenzione: per ogni numero naturale n, si pone come e' ovvio, tenendo presente la convenzione $0!=1$, [9]

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

E' ben noto (e comunque si verifica immediatamente) 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

$$(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^k\cdot b^{n-k}. \tag{15}$$

Infatti, si osservi che le disposizioni di n elementi di classe k si possono raggruppare considerando quelle 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 ciascun gruppo e' composto dallo stesso numero 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 ha come conseguenza appunto la relazione $D^n_k = C^n_k\cdot P_k$, da cui, tenendo conto che $P_k=k!$ e $D^n_k=\frac{n!}{(n-k)!}$, si ricava immediatamente

$$\frac{n!}{(n-k)!}=C^n_k\cdot k!,\qquad \text{ovvero}\qquad C^n_k=\frac{n!}{k!\,(n-k)!}$$

[9] Preferiamo invece **non utilizzare** la convenzione che

$$\binom{n}{k}=0$$

qualora $k<0$, oppure $k>n$.

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 cardinalita' k contenuti in un insieme composto da n elementi, otteniamo che 2^n e' uguale alla cardinalita' della famiglia delle parti di un insieme di n elementi[10]

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.

Ponendo invece nella (15) $a = x, b = 1$, otteniamo l'identita'

$$(x+1)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{16}$$

3.3 Alcuni classici esempi

Consideriamo ora qualche semplice e classico esempio di probabilita' combinatorie.

Esempio 3.3 (Problema del compleanno). *Qual e' la probabilita' 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 probabilita' dell'evento complementare

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

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

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

e la probabilita' cercata e' fornita da $P(E) = 1 - P(\overline{E})$. Indichiamo ora tale probabilita' con $P_M(E)$ per mettere in evidenza la sua dipendenza dal valore di M. Ovviamente $P_M(E)$ e' una funzione crescente di M ed e' 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"). *E' piu' 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?*

---

[10] La relazione $\sum_{k=0}^{n} \binom{n}{k} = 2^n$ si puo' dedurre anche pensando che l'insieme delle parti $\mathcal{P}(\{a_1, \cdots, a_n\}$ di $\{a_1, \cdots, a_n\}$ e' l'unione delle famiglie dei sottoinsiemi di cardinalita' k al variare di k da 0 ad n, per cui $\sum_{k=0}^{n} C_k^n = |\mathcal{P}(\{a_1, \cdots, a_n\}|$, la cardinalita' dell'insieme delle parti. Inoltre $\mathcal{P}(\{a_1, \cdots, a_n\}$ ha cardinalita' 2^n, in quanto e' in corrispondenza biunivoca con l'insieme delle funzioni $f : \{a_1, \cdots, a_n\} \mapsto \{0, 1\}$. La corrispondenza e' data da $A \longleftrightarrow f = \mathbf{1}_A$, dove

$$\mathbf{1}_A(a_i) = 1, \qquad \textit{se} \quad a_i \in A$$
$$\mathbf{1}_A(a_i) = 0, \qquad \textit{se} \quad a_i \notin A.$$

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

$$P(\{\textit{almeno un asso in 4 lanci}\}) = 1 - P(\{\textit{nessun asso in 4 lanci}\})$$

$$P(\{\textit{almeno un doppio asso in 24 lanci}\}) =$$
$$= 1 - P(\{\textit{nessun doppio asso in 24 lanci}\}).$$

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 $x_i \in \{1, ..., 6\}$. Dunque $|\Omega| = 6^4$. Gli eventi elementari che costituiscono l'evento composto $\{\textit{nessun asso in 4 lanci}\}$ corrispondono, invece, alle disposizioni con ripetizione di classe 4 dei 5 elementi $\{2, ..., 6\}$. Si ha quindi

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

Analogamente si ottiene

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

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 e' la probabilita' che tutte le ragazze si trovino nella stessa squadra e tutti i ragazzi nella squadra avversaria?

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

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

Nel caso a) un solo evento elementare e' favorevole e dunque la probabilita' cercata e' $\frac{1}{\binom{4N}{2N}}$.

Nel caso b) gli eventi elementari favorevoli 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 probabilita' cercata e' 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 e' la probabilita' di nessuna testa sui primi h lanci?*

Soluzione. Possiamo schematizzare gli eventi elementari in questo esperimento come gli elementi dell'insieme $\{0, 1\}^n$ cioe' come n-uple con elementi uguali a 0 (croce) o uguali a 1 (testa), (ad esempio l'evento elementare $\omega \equiv (0, 0, 1, 1, 0, ..., 0)$ 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 $\{\textit{nessuna testa sui primi h lanci}\}$ e' allora l'evento composto

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

Traduciamo la condizione che la moneta sia perfetta con la posizione

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

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 e' la probabilita' di k risultati testa negli n lanci di una moneta?*

Soluzione. Si ha lo stesso spazio di probabilita' 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 probabilita' ***binomiali***.

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

Soluzione. Si e' sottinteso che gli n elettori siano stati selezionati senza reinserimento. L'esperimento consiste dunque nel selezionare un sottoinsieme di cardinalita' n (*il campione*) dall'insieme degli M elettori (*la popolazione*) e quindi $|\Omega| = \binom{M}{n}$. Si sottointende che il sondaggio sia condotto in modo casuale, cioe' che ogni "campione" abbia uguale probabilita' $\frac{1}{\binom{M}{n}}$ 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 $(n-k)$ elettori fra i votanti per B da' luogo ad una n-upla di elettori (un campione) che contiene k elettori per A.

Dunque la probabilita' cercata e' 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 probabilita' ***ipergeometriche***. Osserviamo che i valori possibili per k devono rispettare la condizione

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

che, dopo semplici passaggi, diviene

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

3.4 Alcune proprieta' dei coefficienti binomiali.

Nello studio del Calcolo delle Probabilita' e' opportuno tenere presente alcune identita' fondamentali riguardanti i coefficienti binomiali. Ne presentiamo intanto alcune qui di seguito.

Una prima semplice identita' e' la seguente

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}, \tag{17}$$

purche' $k - 1 \ge 0$ e $k \le n-1$, ovvero per $1 \le k \le n-1$. Si noti che tuttavia per $k = 0$ e per $k = n$ ovviamente si ha $\binom{n}{0} = \binom{n}{n} = 1$. La dimostrazione di tale formula e' immediata; bastera' infatti sviluppare i coefficienti binomiali[11] (provare come esercizio).

[11] Prendendo come interpretazione di $\binom{n}{k}$ il numero dei sottoinsiemi di cardinalita' k di un insieme $\{a_1, a_2, \cdots a_{n-1}, a_n\}$ di cardinalita' n si puo' ragionare anche nel seguente modo. I sottoinsiemi di cardinalita' k si possono dividere in due classi: i sottoinsiemi C che contengono a_n e i sottoinsiemi D che non lo contengono, e quindi il numero dei sottoinsiemi di cardinalita' k si puo' esprimere come la somma del numero dei sottoinsiemi del primo tipo e del numero dei sottoinsiemi del secondo tipo. D'altra parte quelli del primo tipo si possono esprimere come $C = C' \cup \{a_n\}$, con C' sottoinsieme di $\{a_1, a_2, \cdots a_{n-1}\}$ di cardinalita' $k-1$, e quindi sono tanti quanti i sottoinsiemi di cardinalita' $k-1$ di un insieme di cardinalita' $n-1$, ovvero $\binom{n-1}{k-1}$, mentre quelli del secondo tipo, D sono sottoinsiemi di $\{a_1, a_2, \cdots a_{n-1}\}$ di cardinalita' k e quindi sono esattamente $\binom{n-1}{k}$.

Qui vogliamo comunque anche darne una semplice dimostrazione probabilistica, ricordando quanto visto nel precedente Esempio 3.7, relativo ad n lanci di una moneta.

Poniamo $E \equiv \{$*si ottengono k risultati testa in n lanci di una moneta perfetta*$\}$. Sappiamo che la probabilita' di ottenere tale risultato e' uguale a $\frac{\binom{n}{k}}{2^n}$. D'altra parte, ponendo

$$E_1 \equiv \{\textit{(k-1) teste sui primi (n-1) lanci}\} \cap \{\textit{testa all'n-esimo lancio}\},$$

$$E_2 \equiv \{\textit{k teste sui primi (n-1) lanci}\} \cap \{\textit{croce all'n-esimo lancio}\},$$

possiamo anche scrivere

$$E \equiv E_1 \cup E_2$$

e, essendo chiaramente $E_1 \cap E_2 \emptyset$,

$$P(E) = P(E_1) + P(E_2). \tag{18}$$

Ora possiamo notare che gli eventi composti E_1 ed E_2 hanno rispettivamente cardinalita' uguale a $\binom{n-1}{k-1}$ e $\binom{n-1}{k}$; e dunque la (18) diventa

$$\frac{\binom{n}{k}}{2^n} = \frac{\binom{n-1}{k-1}}{2^n} + \frac{\binom{n-1}{k}}{2^n}.$$

L'identita' (17) e' in particolare alla base della costruzione del ben noto ***Triangolo di Tartaglia***.

Utilizzando (17) e' anche facile, per $0 \leq k \leq n$, ottenere la seguente

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

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

$$\sum_{k=0 \bigvee (n-s)}^{n \bigwedge r} \binom{r}{k}\binom{s}{n-k} = \binom{r+s}{n}, \tag{19}$$

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

Quest'ultima puo' essere verificata ad esempio osservando intanto quanto segue: $\forall x$

$$(1+x)^{r+s} = (1+x)^r \cdot (1+x)^s = \left(\sum_{k=0}^{r} \binom{r}{k} x^k\right) \cdot \left(\sum_{h=0}^{s} \binom{s}{h} x^h\right)$$

$$= \sum_{k=0}^{r}\sum_{h=0}^{s} \binom{r}{k}\binom{s}{h} x^{k+h} = \sum_{n=0}^{r+s} \sum_{\substack{0\leq k\leq r,\, 0\leq h\leq s \\ k+h=n}} \binom{r}{k}\binom{s}{h} x^{k+h}$$

$$= \sum_{n=0}^{r+s} \sum_{0\leq k\leq r,\, 0\leq n-k\leq s} \binom{r}{k}\binom{s}{n-k} x^n = \sum_{n=0}^{r+s} x^n \left(\sum_{k=0 \bigvee (n-s)}^{r \bigwedge n} \binom{r}{k}\binom{s}{n-k}\right).$$

D'altra parte, ricordando la (16),

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

e la (19) si ottiene confrontando termine a termine tali due sviluppi. Ponendo in particolare, ad esempio, $r \le s$ 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

$$\sum_{k=0}^{r \wedge s} \binom{r}{k}\binom{s}{k} = \binom{r+s}{s} = \binom{r+s}{r}, \tag{20}$$

che per $r = s = n$ diviene

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}. \tag{21}$$

3.5 Esercizi di verifica

Esercizio 3.1. Le lettere AAMMM vengono ordinate a caso. Qual e' la probabilita' di ottenere la parola MAMMA?

Esercizio 3.2. Si fanno n lanci di una moneta perfetta. Per $1 \le h \le n$, qual e' la probabilita' 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 e' la probabilita' 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 e' la probabilita' di vincere almeno una di queste puntate?

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

Esercizio 3.6. Qual e' la probabilita' 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, ..., 90}).

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

Esercizio 3.8. Qual e' la probabilita' 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 probabilita' degli eventi elencati qui di seguito:
a) $\{$*tutti i dadi danno punteggi diversi fra loro*$\}$
b) $\{$*due dadi danno punteggi uguali fra loro e gli altri tre danno punteggi tutti diversi*$\}$
("coppia")
c) $\{$*tre dadi danno punteggi uguali fra loro e gli altri due danno due punteggi diversi*$\}$
("tris")
d) $\{$*quattro dadi danno punteggi uguali fra loro e uno da un punteggio diverso*$\}$
("poker")
e) $\{$*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 puo' saltare se non si sono trovate eccessive difficolta' 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:

$$\{\textit{Nessuna tazzina è su un piattino dello stesso colore}\}$$

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

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

Calcolare la probabilità dell'evento

$$\{\textit{Nessuna tazzina su un piattino dello stesso colore}\}$$

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

Esercizio 3.11. Siano M, m_1, n e k numeri assegnati e tali che $m_1 < M, n < M$, $\max(0, m_1 + n - M) \leq k \leq \min(n, r)$. Verificate l'identita'

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

e datene un'interpretazione probabilistica.

Esercizio 3.12. Una moneta perfetta viene lanciata r volte da Renato e s volte da Stefano $(r > 3, s > 3)$. Si ponga

$$X = \text{ numero dei lanci in cui Renato ottiene il risultato testa}$$

$$Y = \text{ numero dei lanci in cui Stefano ottiene il risultato testa}$$

Calcolare la probabilita' dei seguenti eventi.
(a) $\{X = 3\} \cap \{Y = 3\}$
(b) $\{X = 3\} \cup \{Y = 3\}$
(c) $\{\max(X, Y) = 3\}$
(d) $\{X = Y\}$.

4 Probabilita' condizionate

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

Abbiamo gia' visto che, in uno spazio di probabilita' finito, la teoria della probabilita' e' in effetti gia' tutta contenuta nella formula (3), che mostra come si ottenga la probabilita' di un evento composto, una volta assegnate le probabilita' a ciascun evento elementare. Si vedra' comunque che le formule che verranno ottenute nel seguito (e la nozione di indipendenza stocastica che verra' illustrata a partire dalla prossima lezione) costituiscono spesso una guida al ragionamento probabilistico, che puo' rivelarsi complementare all'uso della formula (3). Tali nozioni infatti permettono, alcune volte, di assegnare probabilita' ad eventi composti (oppure di calcolarle sulla base di probabilita' assegnate ad altri eventi) in modo piu' 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 e' opportuno analizzare il significato "logico" della nozione di ***partizione di un insieme***.

Consideriamo una collezione di sottoinsiemi $H_1, ..., H_m$ dello spazio Ω ($H_l \in \mathcal{P}(\Omega), l = 1, ..., m$), che costituisca una partizione di Ω, cioe' 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 $H_1, ..., H_m$ come eventi, abbiamo che essi sono a ***due a due incompatibili*** (cioe' e' *certo che non se ne possono verificare due contemporaneamente*) e, d'altra parte, essi sono ***esaustivi*** (*e' certo che se ne verifichi almeno uno*); dunque: e' certo che si verifichi uno ed uno soltanto degli eventi $H_1, ..., H_m$ (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 probabilita' si ha che, se $H_1, ..., H_m$ 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

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

e dunque

$$P(E) = P(E \cap H_1) + ... + P(E \cap H_m). \tag{22}$$

4.1 Definizione di probabilita' condizionata

Cominciamo con un esempio

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

Soluzione. Tutti gli eventi elementari

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

sono inizialmente ritenuti equiprobabili.

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

$$\{X = 2\}, ..., \{X = 6\} \quad \text{(non sappiamo pero' “quale”)}.$$

E' naturale, a questo punto, **assumere** quanto segue:
l'informazione che si e' verificato l'evento B non modifica la situazione di equiprobabilita' fra gli eventi $\{X = 2\}, ..., \{X = 6\}$.

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

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

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

Cio' suggerisce la seguente

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

$$P(E|H) = \frac{P(E \cap H)}{P(H)}. \tag{23}$$

Osservazione 1 (**di carattere euristico**). All'interno di ciascuna delle interpretazioni della probabilita' (classica, frequentista, soggettivista, ...) cui si e' accennato in precedenza, il numero $P(E|H)$ definito nella (23) coincide effettivamente con la probabilita' che, coerentemente con tale interpretazione, dovremmo attribuire al verificarsi di E, se sapessimo che si e' verificato H.

Cio' costituisce la motivazione per definire “assiomaticamente” la nozione di probabilita' condizionata attraverso la (23).

Esercizio proposto 4.1 (continuazione dell'Esempio 2.1, dado non equilibrato). *Un dado ha sei facce numerate da 1 a 6; esso e' pesato in modo tale che ciascuna faccia abbia una probabilita' di presentarsi (in un singolo lancio) proporzionale al suo valore. Siano*

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

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

Vediamo ora le semplici, ma importanti, conseguenze della definizione di probabilita' condizionata, gia' menzionate in precedenza.

4.2 Conseguenze immediate della definizione di probabilita' condizionata

4.2.1 Formula delle probabilita' composte

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

$$P(E_1 \cap E_2) = P(E_2|E_1)P(E_1). \tag{24}$$

Questa formula si puo' generalizzare.

***Proposizione 1*. (Formula delle probabilita' composte)** Consideriamo n eventi E_1, E_2,..., E_n, per i quali si abbia $P(E_1) > 0$, $P(E_1 \cap E_2) > 0$, ..., $P(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$. Si ha

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

Dimostrazione Segue immediatamente dalla definizione di probabilita' condizionata: possiamo scrivere

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

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

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

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

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

La precedente dimostrazione si puo' formalizzare utilizzando il principio di induzione.[12]

Si osservi inoltre che le condizioni richieste sono sovrabbondanti[13] : infatti basta che

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

[12] Il caso $n = 2$ corrisponde alla formula (24). Supposta vera l'affermazione della Proposizione per $m - 1$ eventi, ovvero

$$P(E_1 \cap E_2 \cap \cdots \cap E_{m-1}) =$$
$$= P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \cdots \cdot P(E_{m-1}|E_1 \cap E_2 \cap \cdots \cap E_{m-2}).$$

mostriamo ora che vale per m eventi, e infatti

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

[13] Essendo $E_1 \supseteq E_1 \cap E_2 \supseteq \cdots \supseteq E_1 \cap E_2 \cap \cdots \cap E_{n-1}$, per la proprieta' di monotonia della probabilita' si ha $P(E_1) \geq P(E_1 \cap E_2) \geq \cdots \geq P(E_1 \cap E_2 \cap \cdots \cap E_{n-1}) > 0$.
Inoltre, sotto questa ipotesi, e' chiaro che si potrebbe scegliere un qualunque ordine per i primi $n - 1$ eventi, ad esempio, per $n = 4$, se $P(E_1 \cap E_2 \cap E_3 > 0$, allora

$$P(E_1 \cap E_2 \cap E_3 \cap E_4) = P(E_3)P(E_2|E_3)P(E_1|E_3 \cap E_2)P(E_{i_n}|E_1 \cap E_2 \cap E_3)$$

Infine, se si sapesse che $P(E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap E_n) > 0$, allora la formula delle probabilita' composte potrebbe essere scritta per un ordine qualunque, ovvero: per ogni permutazione $(i_1, i_2, \cdots, i_n)$ di $1, 2, \cdots n$

$$\begin{aligned}
P(E_1 \cap E_2 \cap \cdots \cap E_n) &= P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n})\\
&= P(E_{i_1}) \cdot P(E_{i_2}|E_1) \cdot P(E_{i_3}|E_{i_1} \cap E_{i_2}) \cdot \cdots \cdot P(E_{i_n}|E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_n-1}).
\end{aligned}$$

La formula delle probabilita' composte di solito viene usata per trovare la probabilita' dell'intersezione di un numero finito di eventi, specialmente quando e' piu' facile valutare le probabilita' condizionate rispetto alle probabilita' 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 gia' provate, finche' non trova la chiave giusta. Vogliamo trovare la probabilita' dell'evento*

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

Soluzione. Scriviamo E come intersezione di diversi eventi come segue:
$E \equiv \{\textit{chiave errata al primo tentativo}\} \cap \{\textit{ chiave errata al secondo tentativo}\} \cap ... \cap \{\textit{chiave giusta al } k\textit{-esimo tentativo}\}$.

Utilizzando la formula delle probabilita' composte possiamo scrivere dunque

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

$$P(\{\textit{chiave errata al secondo tentativo }\} \,|\, \{\textit{ chiave errata al primo tentativo}\}) \cdot ... \cdot$$

$$\cdot ... \cdot P(\{\textit{chiave giusta al } k\textit{-esimo tentativo}\} \,|\, \{\textit{chiave errata al primo tentativo}\} \cap ... \cap$$

$$... \cap \{\textit{chiave errata al } (k-1)\textit{-esimo tentativo}\}) =$$

$$= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot ... \cdot \frac{n-k+1}{n-k+2} \cdot \frac{1}{n-k+1} = \frac{1}{n}.$$

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

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

e quindi

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

4.2.2 Formula delle probabilita' totali

Proposizione 2. Sia $\{H_1, ..., H_m\}$ una partizione di Ω, con $P(H_l) > 0$, $l = 1, ..., m$. Per un qualunque evento $E \in \mathcal{P}(\Omega)$, risulta

$$P(E) = P\left(E|H_1\right) P\left(H_1\right) + ... + P\left(E|H_m\right) P\left(H_m\right),$$

o piu' brevemente

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

Dimostrazione.
Basta ricordare la precedente formula (22), ovvero $P(E) = \sum_{k=1}^{m} P(E \cap H_k)$, e tener conto che, essendo $P(H_l) > 0$, si ha, per $l = 1, ..., m$, $P\left(E \cap H_l\right) = P\left(E|H_l\right) P\left(H_l\right)$.

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

Esempio 4.3 (Estrazione di un numero al lotto). *Qual e' la probabilita' 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 $\{H, \overline{H}\}$, dove

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

e applichiamo la formula delle probabilita' totali; si ha cosi'

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

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 e' la probabilita' di ottenere un numero pari di risultati testa?*

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

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

Si capisce subito che, se la moneta e' perfetta, si avra' per motivi di simmetria,

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

E' pero' utile anche ragionare come segue, adoperando la formula delle probabilita' totali:

$$P(E_n) = P(E_{n-1}) \cdot P(E_n|E_{n-1}) + P(\overline{E}_{n-1}) \cdot P(E_n|\overline{E}_{n-1}), \tag{25}$$

e risulta

$$P(E_n|E_{n-1}) = P(\{\textit{risultato croce all'n-esimo lancio}\}|E_{n-1})$$
$$P(E_n|\overline{E}_{n-1}) = P(\{\textit{risultato testa all'n-esimo lancio}\}|\overline{E}_{n-1})$$

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

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

Otteniamo dunque dalla (25)

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

4.2.3 Formula di Bayes

Applicando la definizione di probabilita' condizionata e poi la formula delle probabilita' 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 $\{H_1, ..., H_m\}$ una partizione dell'evento certo questa formula si generalizza nel seguente modo.

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

$$P(H_l|E) = \frac{P(E|H_l)\,P(H_l)}{\sum_{r=1}^{m} P(E|H_r)\,P(H_r)}, \quad l = 1, ..., m. \tag{26}$$

Dimostrazione.

Per la definizione di probabilita' condizionata, si ha

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

Applicando la formula delle probabilita' composte al numeratore del membro a destra otteniamo $P(H_l \cap E) = P(E|H_l)\,P(H_l)$; applicando la formula delle probabilita' 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 $\Omega, \mathcal{H} \equiv \{H_1, ..., H_m\}$: sappiamo che e' verificato uno ed uno soltanto degli eventi $H_1, ..., H_m$, ma non sappiamo quale.

Attribuiamo, rispettivamente, probabilita' $P(H_1), ..., P(H_m)$ a ciascuno di tali eventi (possiamo pensare che tali probabilita' esprimono il nostro stato di informazione "iniziale" su tale partizione).

Supponiamo poi di avere l'informazione che e' verificato l'evento E e ci chiediamo come, in conseguenza di cio', si debbano modificare le probabilita' da attribuire agli eventi $H_1, ..., H_m$ (cioe' come cio' modifichi il nostro stato di informazione "iniziale" su $\mathcal{H}$).

Tali "nuove" probabilita' coincideranno con le probabilita' condizionate $P(H_l|E)$ $(l = 1, ..., m)$, che vanno calcolate attraverso la formula (26). Dunque la formula di Bayes puo' essere vista come la regola secondo cui lo stato di informazione su $\mathcal{H}$ si modifica sulla base dell'osservazione dell'evento E.

Piu' in generale, si puo' considerare la nuova probabilita'

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

E' facile verificare che P_E cosi' definita e' una misura di probabilita' si $(\Omega, \mathcal{P}(\Omega))$. Si lascia al lettore la verifica, di queste proprieta'.

Esercizio proposto 4.2 (P_E e' una probabilita'). *Controllare che P_E verifica gli assiomi i), ii) e iii) della Definizione 2.1, ovvero che*
i) $P_E(A) \geq 0$,
ii) $P_E(\Omega) = 1$,
iii) se A_1 e A_2 sono eventi incompatibili, cioe' $A_1 \cap A_2 = \emptyset$, 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 puo' essere applicata in molti altri problemi, sostanzialmente analoghi, suggeriti in diversi campi di applicazione della teoria delle probabilita'.

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

Soluzione. Poniamo

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

$$H_1 \equiv \{\textit{il pezzo scelto e' di tipo } A\}$$

$$H_2 \equiv \{\textit{il pezzo scelto e' di tipo } B\}$$

$$H_3 \equiv \{\textit{il pezzo scelto e' di tipo } C\}.$$

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

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

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 probabilita' 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 (26) puo' essere piu' brevemente scritta nella forma

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

Notiamo che, essendo $\mathcal{H}$ una partizione di Ω, sappiamo gia' 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 quantita' $K = \frac{1}{P(E)} = \frac{1}{\sum_{r=1}^{m} P(E|H_r)P(H_r)}$ ha dunque il ruolo di ***costante di normalizzazione***.

4.3 Esercizi di verifica

Esercizio 4.1. A e B sono due eventi tali che

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

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

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

$$A_i \equiv \{X_i \textit{ pari}\}, \; i = 1, 2.$$

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

Esercizio 4.3. Nel lancio di due dadi, qual e' la probabilita' condizionata che nessuno dei due punteggi sia superiore a 4 sapendo che la somma dei due punteggi e' 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 probabilita' dell'evento

$$E \equiv \{\textit{la pallina estratta e' gialla}\}$$

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

Esercizio 4.5. Nel gioco della roulette, qual e' la probabilita' condizionata del risultato *pair*, dato che si e' ottenuto il risultato *passe*? (Ricordiamo che il risultato $\{0\}$, non e' *passe*, 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 e' la probabilita' di una pallina rossa alla seconda estrazione?
b) Sapendo che le palline estratte nelle due successive estrazioni sono dello stesso colore, qual e' la probabilita' 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. Do' le seguenti valutazioni di probabilita'

$$P(\{\textit{Emilio leggera' la sua posta elettronica}\}) = 0.7$$

$$P(\{\textit{Emilio ricevera' il messaggio sul suo cellulare}\}) = 0.8$$

$$P(\{\textit{Emilio leggera' la posta elettronica e ricevera' il messaggio sul cellulare}\}) = 0.56.$$

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

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

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

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

5 Correlazione e indipendenza fra eventi

Riprendiamo l'Esempio 4.5 della precedente Lezione 4, sull'estrazione da un lotto di pezzi apparentemente identici. Come ci si poteva gia' 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 e' equivalente a

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

e che, dunque, tale relazione e' 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 e' 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 probabilita' $\frac{1}{36}$. Consideriamo gli eventi composti $E_1 \equiv \{X_1 \text{ pari}\}$, $E_2 \equiv \{X_1 + X_2 \text{ pari}\}$, $E_3 \equiv \{X_1 + X_2 \leq 4\}$, $E_4 \equiv \{X_1 \leq 2\}$, $E_5 \equiv \{\max(X_1, X_2) > 3\}$.*

E' 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.

E' importante a questo punto tener presente quanto segue.

Osservazione 1. Consideriamo ancora, a mo' di esempio, l'esperimento del lancio di due dadi. Si verifica immediatamente che assegnare uguali probabilita' $\frac{1}{36}$ a tutti gli eventi elementari implica che gli eventi (composti) del tipo $\{X_1 = i\}$, $\{X_2 = j\}$ sono tutti equiprobabili con probabilita' 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.

E' d'altra parte immediato verificare anche il viceversa, cioe' che la condizione $P(X_1 = i)$=$P(X_2 = j)$=$\frac{1}{6}$ insieme all'indipendenza stocastica per le coppie $\{X_1 = i\}$, $\{X_2 = j\}$ implica che tutti gli eventi elementari $\{X_1 = i\} \cap \{X_2 = j\}$ hanno probabilita' uguale ad $\frac{1}{36}$.

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

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

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

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

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

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

Calcolare la probabilita' 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 probabilita' imponendo i valori delle probabilita' 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), cioe' 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 e' piu' sintetico, e cio' puo' costituire una caratteristica importante nei casi in cui $|\Omega|$ e' 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 probabilita' 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 $E \equiv \{$*Tizio vince almeno un premio*$\}$ ed $E_i \equiv \{$ *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 probabilita' composte, si ottiene[14]

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

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

5.1 Indipendenza fra partizioni e fra algebre di eventi

In uno spazio di probabilita', consideriamo due eventi A e B e le loro rispettive negazioni $\overline{A}$ e $\overline{B}$.

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

$$A \perp\!\!\!\perp B, \quad \overline{A} \perp\!\!\!\perp B, \quad A \perp\!\!\!\perp \overline{B}, \quad \overline{A} \perp\!\!\!\perp \overline{B},$$

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

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

e dunque $\overline{A} \perp\!\!\!\perp B$.

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

Cio' suggerisce la seguente definizione.

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

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

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

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

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

[14] Posto A_1^i l'evento all'estrazione i-sima non viene estratto nessuno dei due biglietti si ha

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

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.

E' interessante notare anche che a questo risultato si puo' arrivare anche pensando che invece si tratti 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 puo' 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 $\overline{E}_1$ diviene l'evento $\{$*nelle due estrazioni si estraggono solo palline arancioni*$\}$: la probabilita' di non vincere diviene allora immediatamente $\frac{697}{700} \cdot \frac{696}{699}$.

Osservazione 2. La nozione di indipendenza fra partizioni ha, nella teoria della probabilita', 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 servira' per definire in modo semplice e concettualmente efficiente la nozione di indipendenza fra due *variabili aleatorie* (Lez. 7 e 8). In tale prospettiva e' utile presentare qui le seguenti nozioni.

Definizione 5.5 (Algebra). *Una famiglia $\mathcal{G}$ di eventi di Ω (dunque $\mathcal{G} \subseteq \mathcal{P}(\Omega)$) e' un'***algebra***, se sono verificate le seguenti proprieta'*
i) $\Omega \in \mathcal{G}$
ii) $E \in \mathcal{G} \Rightarrow \overline{E} \in \mathcal{G}$
iii) $E_1, E_2 \in \mathcal{G} \Rightarrow E_1 \cup E_2 \in \mathcal{G}$

E' ovvio che se $\mathcal{G} \subseteq \mathcal{P}(\Omega)$ e' 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 $\mathcal{A} \equiv \{A_1, A_2, ..., A_n\}$ 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 proprieta':*
* $\mathcal{G}(\mathcal{A})$ *e' 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 cioe' che $\mathcal{G}(\mathcal{A})$ e' la piu' piccola famiglia di sottoinsiemi di Ω che abbia contemporaneamente le due proprieta' di essere un'algebra e di contenere al suo interno tutti i sottoinsiemi della famiglia $\mathcal{A}$.

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

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

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

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

Vale in proposito la seguente
Proposizione 1. Siano $\mathcal{A}$ e $\mathcal{B}$ sono 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 e' utile svolgere il seguente esercizio[15].

[15] Inoltre va ricordato che

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

Esercizio proposto 5.3. *Siano $\mathcal{A} \equiv \{A_1, A_2, ..., A_n\}$ e $\mathcal{B} \equiv \{B_1, B_2, ..., B_m\}$ 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*

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

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

5.2 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 e' verificato B, o anche il sapere con certezza che non si e' 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 (e' 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 e' gia' detto, tale nozione e' fondamentale nella costruzione di modelli probabilistici. Infatti, in pratica, nell'assegnare una misura di probabilita' su uno spazio campione, si parte sempre dall'individuazione di famiglie di eventi a ciascuno dei quali si impone uguale probabilita' e a coppie di eventi entro cui si impone l'indipendenza stocastica. In base a tali posizioni si deduce quale debba essere la probabilita' dei vari eventi composti, interessanti nel problema stesso (o almeno si deducono delle condizioni cui tali probabilita' debbono soddisfare).

Per rimanere sul piano dell'esemplificazione piu' 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 e' palesemente contraddetta; per ora accenniamo soltanto che cio' accade nei casi in cui una situazione di mancanza di informazione fa si' che ciascun evento osservato contiene un forte valore informativo, che si riflette sulle aspettative relative ad altri eventi connessi. Tale punto verra' 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.

Cio' e' 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 equiprobabilita' fra gli eventi elementari abbiamo le relazioni: $A \perp\!\!\!\perp B$, $A \perp\!\!\!\perp C$, $B \perp\!\!\!\perp C$. Notiamo pero' 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 piu' di due eventi.

Si da' allora le seguente definizione. Sia $\{E_1, E_2, ..., E_n\}$ una famiglia di eventi in uno stesso spazio di probabilita' e consideriamo le partizioni $\mathcal{P}_1 \equiv \{E_1, \overline{E}_1\}, ..., \mathcal{P}_n \equiv \{E_n, \overline{E}_n\}$.

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

$$P(A_1 \cap A_2 \cap ... \cap A_m) = P(A_1) \cdot P(A_2) \cdot ... \cdot P(A_m). \tag{27}$$

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

$$P(E_{j_1} \cap E_{j_2} \cap ... \cap E_{j_m}) = P(E_{j_1}) \cdot P(E_{j_2}) \cdot ... \cdot P(E_{j_m}) \qquad \text{per ogni } \{j_1, ..., j_m\} \subseteq \{1, 2, ..., n\}, \text{ con } 2 \leq m \leq n. \tag{28}$$

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

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1) \cdot P(A_2) \cdot ... \cdot P(A_n), \qquad \text{con } A_i = E_i \text{ oppure } A_i = \overline{E}_i \tag{29}$$

Osservazione 3 È importante notare che in alcuni testi la definizione di famiglia di eventi completamente indipendenti e' data attraverso la relazione (28), mentre in altri e' data attraverso la (29). Cio' e' dovuto al fatto che le relazioni (28) e (29) sono equivalenti, ed entrambe implicano la (27). Cio' e' ovvio per $n = 2$, come detto all'inizio del paragrafo. Non diamo qui la dimostrazione di questa proprieta' di equivalenza tra (28) e (29), ma proponiamo al lettore il seguente esercizio, la cui soluzione e' basata sull'osservazione che per tre eventi A, B, C si ha $A \cap B = (A \cap B \cap C) \cup (A \cap B \cap \overline{C})$, e che quindi $P(A \cap B) = P(A \cap B \cap C) + P(A \cap B \cap \overline{C})$, da cui anche $P(A \cap B \cap \overline{C}) = P(A \cap B) - P(A \cap B \cap C)$.

Esercizio proposto 5.4. *Dimostrare l'equivalenza di (28) e (29) per* $n = 3$.

Un caso assai particolare ma di notevole interesse e' 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* $E_1, E_2, ..., E_n$ *costituiscono delle* **prove bernoulliane** *se sono completamente indipendenti ed hanno tutti una stessa probabilita'* θ, *con* $0 < \theta < 1$.

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

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

5.3 Esercizi di verifica

Esercizio 5.1. A e B sono due eventi tali che

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

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 probabilita' 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 *passe* 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

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

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

Esercizio 5.7. a) Qual e' la probabilita' che il numero 16 venga estratto su una data ruota del lotto in una fissata giornata?
b) Qual e' la probabilita' che il numero 16 non venga mai estratto su una data ruota del lotto per n giornate consecutive?
c) Qual e' la probabilita' condizionata che il numero 16 venga estratto l'$(n+1)$-esima giornata, dato che non e' 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 inviera' allora un messaggio di posta elettronica, Bruno gli inviera' un messaggio sul telefono cellulare, e interverra' anche Carla, cercando di avvertire di persona la sorella di Emilio. Essi avranno successo rispettivamente con probabilita' $P(A) = 0.8$, $P(B) = 0.7$ e $P(C) = 0.6$; i tre eventi, inoltre, sono completamente indipendenti.
a) Trovare la probabilita' che Emilio venga informato dell'appuntamento.
b) Dato che Emilio si presenta effettivamente all'appuntamento, come devo valutare la probabilita' 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

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

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

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

Trovate una misura di probabilita' su $(\Omega, \mathcal{P}(\Omega))$, diversa da quella uniforme considerata nell'***Esempio 4*** (cioe' tale che gli eventi elementari non siamo tutti fra di loro siano equiprobabili), per la quale le partizioni $\mathcal{A} \equiv \{A_1, A_2, ..., A_6\}$ e $\mathcal{B} \equiv \{B_1, B_2, ..., B_6\}$ siano comunque indipendenti.

Esercizio 5.12. Dimostrate la precedente ***Proposizione 1***.

6 Probabilita' binomiali e ipergeometriche; estrazioni casuali da urne

In questo paragrafo vogliamo discutere in modo piu' sistematico due particolari modelli probabilistici, che sono gia' comparsi in precedenti esempi.

6.1 Probabilita' binomiali

Per quanto riguarda il primo caso consideriamo, su uno spazio finito di probabilita', n prove bernoulliane $E_1, ..., E_n$; cioe', ricordando la Definizione 5.8, assumiamo che $E_1, ..., E_n$ sono completamente indipendenti ed equiprobabili: ponendo $P(E_i) = \theta$, per $i = 1, 2..., n$ e ponendo

$$X_i = \begin{cases} 1 & \text{se si verifica } E_i \\ 0 & \text{se si verifica } \overline{E}_i \end{cases} \qquad i = 1, 2, \cdots, n, \tag{30}$$

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

$$P(\{X_1 = x_1, ..., X_n = x_n\}) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, \tag{31}$$

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

Poniamo ora

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

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

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

Due eventi del tipo

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

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

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

essi risultano equiprobabili, entrambi di probabilita' $\theta^k(1-\theta)^{n-k}$, in virtu' dell'equazione (31).

E dunque, dal momento che la cardinalita' dell'insieme

$$\{\mathbf{x} \in \{0, 1\}^n : \sum_{i=1}^n x_i = k\} \tag{32}$$

e' uguale a $\binom{n}{k}$, potremo scrivere, per $k = 0, 1, ..., n$,

$$\begin{aligned} P(\{S_n = k\}) &= \sum_{\mathbf{x}\in\{0,1\}^n : \sum_{i=1}^n x_i = k} P(\{X_1 = x_1, ..., X_n = x_n\}) \\ &= \binom{n}{k} \theta^k (1-\theta)^{n-k}. \end{aligned} \tag{33}$$

Probabilita' del tipo in (33) prendono il nome di ***probabilita' 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 probabilita' $\frac{1}{6}$; 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 e' effettivamente un fumatore (o fumatrice) con probabilita' uguale al* 70%. *Se lo scompartimento contiene* 5 *posti (oltre a quello da me prenotato), qual e' la probabilita' che vi 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 avra':

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

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 probabilita'* $p = 0.6$ *ed indipendentemente da quello che accade per gli altri errori. Inoltre i due correttori lavorano indipendentemente uno dall'altro.*

Trovare la probabilita' 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 probabilita'

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

(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 bernuolliane, in ognuna delle quali vi e' una probabilita' di successo uguale a θ. Indicando dunque con S il numero complessivo degli errori individuati, si avra'

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

6.2 Estrazioni casuali da urne

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

Pensiamo ad una popolazione composta di M oggetti, diciamo $c_1, ..., c_M$, di cui m_1 di tipo A e i rimanenti $m_2 = M - m_1$ di un diverso tipo B. Poniamo, per $r = 1, ..., M = m_1 + m_2$,

$$\tau_r = \begin{cases} 1 & \text{se } c_r \text{ e' di tipo } A \\ 0 & \text{se } c_r \text{ e' di tipo } B \end{cases},$$

e supponiamo ad esempio di aver numerato $c_1, ..., c_M$ in modo tale che

$$\tau_r = \begin{cases} 1 & \text{per } 1 \leq r \leq m_1 \\ 0 & \text{per } m_1 + 1 \leq r \leq M = m_1 + m_2 \end{cases} . \tag{34}$$

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

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

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

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

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

In virtu' della posizione (34) potremo equivalentemente scrivere anche

$$\begin{aligned} E_i &\equiv \{(j_1, ..., j_n) \in \Omega : \ 1 \leq j_i \leq m_1\} \\ &= \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{i-1\, volte} \times \{1, 2, \cdots, m_1\} \times \overbrace{\{1, ..., M\} \times \cdots \times \{1, ..., M\}}^{n-i\, volte}, \end{aligned} \tag{35}$$

ed in modo analogo

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

Proposizione 1. Gli eventi $E_1, ..., E_n$ definiti in (35) 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 probabilita' $\frac{1}{M^n}$ a ciascuno degli eventi elementari $(j_1, ..., j_n) \in \Omega$. Dunque, considerando per $1 \leq i \leq n$, la quantita' X_i come definita nella (30), si ha

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

e, analogamente

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

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

$$A_i \in \mathcal{P}_i = \{E_i, \overline{E}_i\} = \Big\{\{X_i = 1\}, \ \{X_i = 0\}\Big\}$$

si ha

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

che corrisponde alla relazione (29), che come osservato in Osservazione 3 della Lez. 5, e' equivalente[16], alla relazione (27).

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 (33) ed in virtu' della ***Proposizione 1***, possiamo concludere scrivendo

$$P(\{S_n = k\}) = \binom{n}{k} \left(\frac{m_1}{M}\right)^k \left(\frac{m_2}{M}\right)^{n-k}, \qquad \text{per } k = 0, 1, \cdots, n. \tag{36}$$

6.3 Probabilita' 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

$$E_i \equiv \{\textit{oggetto di tipo A alla i-esima estrazione}\}, \qquad i = 1, 2, ..., n$$

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

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

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

$$\Omega \equiv \{(j_1, ..., j_n) : \ 1 \le j_i \le M, \ j_1 \ne ... \ne j_n\} = \{(j_1, ..., j_n) : \ 1 \le j_i \le M, \ j_1, \cdots, j_n \text{ tutti distinti}\}.$$

Dunque

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

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

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

[16] Tuttavia va osservato che, in modo assolutamente analogo, si potrebbe dimostrare direttamente la (27) per ogni $2 \le m \le n$, notando che

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

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

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

e dunque

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

da cui

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

riscrivendo in forma piu' compatta tale ultima frazione, attraverso la notazione dei coefficienti binomiali, possiamo concludere:

$$P(\{S_n = k\}) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}} \quad \text{per } \max(0, n + m_1 - M) \leq k \leq \min(n, m_1). \tag{37}$$

In effetti gia' conosciamo questo risultato dall'Esempio 3.8.

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

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

cioe'

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

In effetti quest'ultima identita' coincide con la (19).

Osservazione 1. Si puo' pervenire al risultato (37) anche in un modo alternativo, applicando direttamente la formula delle probabilita' composte agli eventi del tipo $\{X_1 = x_1, ..., X_n = x_n\} = \{X_1 = x_1\} \cap \cdots \cap \{X_n = x_n\}$:

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

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

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

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

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

cioe' 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 probabilita' 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.*

Cio' costituisce un esempio di quanto era stato gia' accennato piu' in generale in merito alla nozione di probabilita' condizionata: *si giunge cioe' a calcolare delle probabilita' di eventi composti non tramite un calcolo combinatorio, bensi' assegnando delle probabilita' condizionate e delle condizioni di simmetria fra eventi; tali considerazioni sono analoghe a quelle gia' svolte nell'Esempio 6.2.*

Osservazione 2. Il problema qui affrontato riguarda il calcolo della probabilita' 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. E' interessante confrontare fra lore le due formule (36) e (37). Entrambe risolvono il problema detto; la prima riguarda pero' 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 puo' aspettare che non vi sia grande differenza fra estrazioni con o senza reinserimento. Cio' puo' essere formalizzato come segue: e' possibile dimostrare che, per fissati valori di n, k (con $0 \leq k \leq n$) e θ (con $0 < \theta < 1$) ed indicando con $\lfloor x \rfloor$ la *parte intera*[17] di un numero reale x, in effetti risulta

$$\lim_{M \to \infty} \frac{\binom{\lfloor \theta M \rfloor}{k}\binom{\lfloor (1-\theta)M \rfloor}{n-k}}{\binom{M}{n}} = \binom{n}{k}\theta^k(1-\theta)^{n-k}.$$

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 rechera' effettivamente a votare per lui con una probabilita' del 40%, indipendentemente dal comportamento degli altri (e si sottointende che egli certamente non ottiene voti da coloro cui non ha inviato la lettera).

a) Qual e' la probabilita' che egli riceva esattamente 51 voti se invia la lettera a 200 persone?

b) Qual e' la probabilita' di essere eletto se invia la lettera a 100 persone?

c) Qual e' il numero minimo di persone cui deve inviare copia della lettera affinche' la probabilita' di essere eletto sia superiore all'80%?

[17] Ricordiamo che la parte intera $\lfloor x \rfloor$ di x e' quel numero intero k tale che $k \leq x < k+1$

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.
a) Supponendo che la scatola contiene $m_1 = 20$ viti nuove e $m_2 = 6$ viti usurate, calcolare la probabilita' che almeno quattro delle cinque viti scelte siano nuove.
b) Si supponga ora di non conoscere inizialmente il numero M_1 delle viti nuove nella scatola e si ponga

$$P(\{M_1 = h\}) = \binom{26}{h} \left(\frac{4}{5}\right)^h \left(\frac{1}{5}\right)^{26-h}, \qquad h = 0, 1, 2, ..., 26$$

Dopo aver verificato che tutte le cinque viti scelte sono nuove, come va calcolata la probabilita' 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 probabilita' dei seguenti eventi:
a) nessuna lampadina difettosa fra le tre estratte
b) esattamente una lampadina difettosa fra le tre estratte
c) almeno una lampadina difettosa fra le tre estratte.
Si considerino separatamente i due diversi casi in cui
i) si estraggano le tre lampadine contemporaneamente
ii) 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 pero' 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 probabilita' che un fissato TC abbia entrambe le batterie cariche.

Esercizio 6.5. Riottenere la formula (37) seguendo le indicazioni contenute nella precedente ***Osservazione 1***.

7 Variabili aleatorie e distribuzioni di probabilita'

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 e' 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 sara' lo specifico valore che X effettivamente assume: il valore che essa assume sara' connesso (in qualche preciso modo) al risultato elementare di un qualche esperimento aleatorio; in base alla misura di probabilita' assegnata sullo spazio campione in tale esperimento, potremo valutare la probabilita' 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 $X: \Omega \to X(\Omega) \subseteq \mathbb{R}$ viene detta* **variabile aleatoria** *(definita su $(\Omega, \mathcal{P}(\Omega))$).*

Osservazione 1. Essendo Ω un insieme finito, l'immagine di X, ovvero $X(\Omega)$, sara' un insieme del tipo $\{x_1, ..., x_n\}$, con $x_j \in \mathbb{R}$, per $j = 1, ..., n$.

Consideriamo ora gli eventi

$$X^{-1}(\{x_j\}) \equiv \{\omega \in \Omega : X(\omega) = x_j\}, \quad j = 1, 2, ..., n.$$

Tali eventi vengono anche indicati brevemente con i simboli

$$\{X = x_1\}, ..., \{X = x_n\}.$$

E' immediato verificare che la famiglia degli eventi

$$\big\{\{X = x_1\}, ..., \{X = x_n\}\big\}$$

costituisce una partizione di Ω; per tale motivo si ha ovviamente che, ponendo[18]

$$p_j \equiv P(X = x_j), \quad j = 1, ..., n,$$

risulta

$$\sum_{j=1}^{n} p_j = 1, \qquad p_j \geq 0, \qquad j = 1, ..., n.$$

[18] Per mettere in evidenza la dipendenza dalla variabile aleatoria X si usa anche la notazione

$$p_X(x_i) \equiv P(X = x_i), \quad i = 1, ..., n.$$

La funzione $p_X : X(\Omega) \to \mathbb{R};\ x \mapsto p_X(x) \equiv P(X = x)$, cosi' definita viene anche detta ***densita' discreta*** di X. Ovviamente vale

$$\sum_{i=1}^{n} p_X(x_i) = 1, \qquad p_X(x_i) \geq 0, \qquad i = 1, ..., n.$$

Possiamo dunque considerare il nuovo spazio di probabilita'

$$(X(\Omega), \mathcal{P}(X(\Omega)), P_X)$$

dove, per $A \in \mathcal{P}(X(\Omega))$

$$P_X(A) \equiv \sum_{j: x_j \in A} p_j = P\left(X^{-1}(A)\right). \tag{38}$$

Definizione 7.2. *Interpreteremo*[19] $P_X(A)$ *come* $P(\{X \in A\})$*; la misura di probabilita'*$P_X(\cdot)$ *su* $\mathcal{P}(X(\Omega))$ *prende il nome di* **distribuzione di probabilita'** *della variabile aleatoria* X*.*

Osservazione 2. Ovviamente a qualunque variabile aleatoria definita su uno spazio di probabilita' possiamo associare la sua distribuzione di probabilita'. Due diverse variabili aleatorie, definite o meno su uno stesso spazio di probabilita', possono dar luogo ad una stessa distribuzione di probabilita' (vedi i due successivi Esempi 7.1 e 7.2).

E' opportuno innanzitutto richiamare l'attenzione sui due particolari tipi di variabili aleatorie: le variabili aleatorie ***degeneri*** e le variabili aleatorie ***binarie***.

Definizione 7.3. *Diciamo che* X*, variabile aleatoria definita su* Ω*, e' una variabile aleatoria* **degenere** *se esiste* $\overline{x} \in \mathbb{R}$*, tale che* $X(\Omega) = \{\overline{x}\}$*, cioe' se* X *e' costante su* Ω*.*

Definizione 7.4. *Diciamo che* X*, variabile aleatoria definita su* Ω*, e' una variabile aleatoria* **binaria** *se* $X(\Omega) = \{0, 1\}$*.*

Definizione 7.5. *Sia* $E \in \mathcal{P}(\Omega)$ *un evento e sia* $\mathcal{X}_E$ *la* **funzione caratteristica***, (o anche* **funzione indicatrice***) di* E *:*

$$\mathcal{X}_E(\omega) = \begin{cases} 1 & per\ \omega \in E \\ 0 & per\ \omega \notin E \end{cases}$$

$\mathcal{X}_E$ *e' dunque una variabile aleatoria binaria*[20]*, che viene indicata con il termine* **indicatore** *di* E*.*

Osservazione 3. Per qualunque v.a. binaria X esiste $E \subseteq \Omega = \{\omega_1, \cdots, \omega_N\}$ tale che

$$X(\omega_i) = \mathcal{X}_E(\omega_i), \qquad i = 1, ..., N.$$

Poniamo infatti

$$E = \{\omega_i \in \Omega : \ X(\omega_i) = 1\}$$

Si avra' allora

$$\overline{E} = \{\omega_i \in \Omega : \ X(\omega_i) = 0\}$$

[19] Infatti l'evento $\{X \in A\}$ si puo' scrivere come

$$\{X \in A\} = \bigcup_{i: x_i \in A} \{X = x_i\},$$

e di conseguenza

$$P(\{X \in A\}) = \sum_{i: x_i \in A} P(\{X = x_i\}) = \sum_{i: x_i \in A} p_X(x_i) = \sum_{i: x_i \in A} p_i = P_X(A).$$

[20] Per la funzione indicatrice $\mathcal{X}_E$ si usa anche il simbolo $\mathbf{1}_E$.

e dunque possiamo scrivere $X(\omega_i) = \mathcal{X}_E(\omega_i), \qquad i = 1, ..., N.$

Proposizione 1. Una qualunque variabile aleatoria X si scrive come combinazione lineare di variabili aleatorie binarie.

Dimostrazione.

Sia $X(\Omega) = \{x_1, ..., x_n\}$ l'insieme dei valori assumibili da X; consideriamo gli eventi

$$E_j \equiv \{X = x_j\}, \quad j = 1, ..., n$$

e le variabili aleatorie binarie $X_1, ..., X_n$ definite come indicatori di tali eventi, ovvero $X_j = \mathcal{X}_{E_j}$. E' ovvio[21] allora che possiamo scrivere

$$X(\omega_i) = \sum_{j=1}^{n} x_j \cdot X_j(\omega_i), \qquad i = 1, ..., N.$$

Osservazione 4. Spesso si mira a determinare direttamente la distribuzione di probabilita' 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 probabilita' Ω su cui la variabile puo' essere definita, ne' come tale variabile vi possa essere definita, ne' quale sia la misura di probabilita' su $(\Omega, \mathcal{P}(\Omega))$.

Vediamo ora qualche esempio di distribuzione di probabilita'.

Esempio 7.1. *Consideriamo ancora una volta l'esperimento legato al lancio di due dadi, in cui*

$$\Omega \equiv \{(h,k) : \ 1 \le h \le 6, \ 1 \le k \le 6\}.$$

$$P(\{(h,k)\}) = \frac{1}{36}, \qquad 1 \le h \le 6, \quad 1 \le k \le 6$$

Su questo spazio possiamo definire diverse variabili aleatorie, ad esempio:

$X_1 : (h,k) \to h$ *("punteggio del primo dado"),* $X_2 : (h,k) \to k$ *("punteggio del secondo dado"),* $X : (h,k) \to h+k$ *("somma dei due punteggi"),* $W : (h,k) \to \frac{h}{k}$*, etc...*

X_1 *e* X_2 *hanno la stessa distribuzione di probabilita', data da :*

$$P(\{X_1 = x\}) = P(\{X_2 = x\}) = \frac{1}{6}, \qquad x = 1, 2, ...6;$$

questa e' la **distribuzione uniforme** *su* $\{1,2,...,6\}$.

La distribuzione di probabilita' di $X = X_1 + X_2$ *e' invece data da*

$$P(\{X = x\}) = \frac{x-1}{36}, \quad x = 2, 3, ..., 7, \qquad P(\{X = x\}) = \frac{13-x}{36}, \quad x = 8, 9, ..., 12.$$

[21] Per la dimostrazione basta mostrare che la funzione $Y(\omega)$, definita da

$$Y(\omega) \equiv \sum_{j=1}^{n} x_j \cdot X_j(\omega_i)$$

coincide con $X(\omega)$ per ogni ω. A questo scopo basta osservare che la famiglia di eventi E_j forma una partizione dell'evento certo Ω e di conseguenza, qualunque sia $\omega \in \Omega$ esiste ed e' unico un E_j tale che $\omega \in E_j$. Allora per tale ω ovviamente $X(\omega) = x_j$, e anche $Y(\omega) = x_j$, come si vede subito, tenendo conto che $\mathcal{X}_{E_i}(\omega) = 0$ se $i \neq j$ e che ovviamente $\mathcal{X}_{E_j}(\omega) = 1$:

$$Y(\omega) = x_j \cdot \mathcal{X}_{E_j}(\omega) + \sum_{i \neq j} x_i \cdot \mathcal{X}_{E_i}(\omega) = x_j \cdot 1 + 0 = x_j.$$

Esempio 7.2. *Riprendiamo l'Esempio 4.2, considerando il caso $n = 6$. La variabile aleatoria*

$$T \equiv \textit{numero dei tentativi fino a trovare la chiave giusta}$$

puo' prendere i valori $1, 2, ..., 6$ *e risulta, grazie a quanto avevamo visto,*

$$P(T = x) = \frac{1}{6}, \quad x = 1, 2, ...6.$$

Confrontando tale risultato con quanto visto prima, troviamo l'esempio di due variabili aleatorie (cioe' X_1 e T), definite su spazi diversi, che hanno la stessa distribuzione di probabilita'.

Prima di considerare il successivo esempio e' utile fare mente locale sulla seguente semplice
Osservazione 4. Siano $E_1, ..., E_n$ degli eventi in uno spazio $(\Omega, \mathcal{P}(\Omega))$ e indichiamo rispettivamente con $X_1, ..., X_n$ i loro indicatori. Ovviamente $X_1, ..., X_n$ 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 cioe'

$$S_n(\omega_i) \equiv \sum_{h=1}^n X_h(\omega_i) = \sum_{h=1}^n 1_{E_h}(\omega_i).$$

Ovviamente S_n ha il significato di ***numero di successi fra gli eventi*** $E_1, ..., E_n$.

Esempio 7.3. *Consideriamo n prove bernoulliane, cioe' n eventi completamente indipendenti, ciascuno di probabilita' 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 $0, 1, ..., n$ e, come abbiamo visto nella lezione precedente, si ha,*

$$P(\{S_n = k\}) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \qquad k = 0, 1, ..., n.$$

Si dice che S_n segue una **distribuzione binomiale di parametri $\boldsymbol{n}$** *e* $\boldsymbol{\theta}$*; cio' 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*

$$P(\{S_n = k\}) = \frac{\binom{m_1}{k}\binom{m_2}{n-k}}{\binom{M}{n}}, \qquad \max(0, n + m_1 - M) \leq k \leq \min(n, m_1).$$

Si dice che S_n segue una **distribuzione ipergeometrica di parametri** *M, m_1, n e cio' si indica con il simbolo $S_n \sim Hyp(M_1, m_1, n)$.*

Finora abbiamo quasi esclusivamente considerato variabili aleatorie ***a valori interi*** (cioe' 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, e' interessante considerare anche la variabile aleatoria Y_n ("frequenza dei successi") definita dalla relazione

$$Y_n = \frac{S_n}{n}.$$

A questo proposito e' interessante piu' in generale, date n variabili aleatorie $X_1, ..., X_n$, 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 verra' 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, cioe' $X(\Omega) \subseteq \mathbb{Z}$, puo' essere spesso conveniente calcolare la distribuzione di probabilita' tenendo conto della relazione[22]

$$P(X = k) = P(X \leq k) - P(X \leq k-1), \qquad \text{per } k \in \mathbb{Z}.$$

Altre volte puo' essere conveniente tenere conto invece che

$$\begin{aligned} P(X = k) &= P(X \geq k) - P(X \geq k+1) \\ &= P(X > k-1) - P(X > k), \qquad \text{per } k \in \mathbb{Z}. \end{aligned}$$

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 puo' assumere Z e con quali probabilita'.*

Soluzione. I valori possibili per Z sono ovviamente $\{1, 2, ..., 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} \equiv \big\{\{X_1 = i\},\ i = 1, \cdots, 6\big\}$ e $\mathcal{B} \equiv \big\{\{X_2 = j\},\ j = 1, \cdots, 6\big\}$ sono indipendenti[23], (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti), risulta

$$\begin{aligned} P(Z = x) &= P(Z \leq x) - P(Z \leq x-1) \\ &= P(\{X_1 \leq x\} \cap \{X_2 \leq x\}) - P(\{X_1 \leq x-1\} \cap \{X_2 \leq x-1\}) \\ &= P(X_1 \leq x) \cdot P(X_2 \leq x) - P(X_1 \leq x-1) \cdot P(X_2 \leq x-1) \\ &= \left(\frac{x}{6}\right)^2 - \left(\frac{x-1}{6}\right)^2 = \frac{2x-1}{36}, \qquad x = 1, 2, ..., 6. \end{aligned}$$

[22] La dimostrazione delle relazioni seguenti e' basato sul fatto che

$$\{X \leq k\} = \{X = k\} \cup \{X \leq k-1\}$$

e

$$\{X \geq k\} = \{X = k\} \cup \{X \geq k+1\},$$

da cui

$$P(X \leq k) = P(X = k) + P(X \leq k-1)$$

e

$$P(X \geq k) = P(X = k) + P(X \geq k+1)$$

e infine che $\{X \geq h\} = \{X > h-1\}$.

[23] Questa proprieta' sara' alla base della definizione di indipendenza stocastica per le v.a. X_1 e X_2

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 puo' assumere W e con quali probabilita'.*

Soluzione. I valori possibili per W sono ovviamente $\{1, 2, ..., 6\}$; tenendo conto che le famiglie di eventi $\mathcal{A} \equiv \big\{\{X_1 = i\},\ i = 1, \cdots, 6\big\}$ e $\mathcal{B} \equiv \big\{\{X_2 = j\},\ j = 1, \cdots, 6\big\}$ sono indipendenti (e quindi anche gli eventi del tipo $\{X_1 \in I\}$ e $\{X_2 \in J\}$ sono indipendenti), risulta

$$\begin{aligned}
P(W = x) &= P(W \geq x) - P(W \geq x + 1) \\
&= P(\{X_1 \geq x\} \cap \{X_2 \geq x\}) - P(\{X_1 \geq x + 1\} \cap \{X_2 \geq x + 1\}) \\
&= P(X_1 \geq x) \cdot P(X_2 \geq x) - P(X_1 \geq x + 1) \cdot P(X_2 \geq x + 1) \\
&= P(X_1 > x - 1) \cdot P(X_2 > x - 1) - P(X_1 > x) \cdot P(X_2 > x) \\
&= (1 - P(X_1 \leq x - 1)) \cdot (1 - P(X_2 \leq x - 1)) - (1 - P(X_1 \leq x)) \cdot (1 - P(X_2 \leq x)) \\
&= \left(1 - \frac{x-1}{6}\right)^2 - \left(1 - \frac{x}{6}\right)^2 = \frac{49 - 14x + x^2 - (36 - 12x + x^2)}{36} \\
&= \frac{13 - 2x}{36}, \qquad x = 1, 2, ..., 6.
\end{aligned}$$

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 $i = 1, \cdots, 6$, ed $l = 1, 2$, e si assuma l'indipendenza delle partizioni $\mathcal{A}$ e $\mathcal{B}$.*

Osservazione 7. Il concetto di variabile aleatoria cosi' come introdotto in questa lezione (vedi Definizione 7.1), puo' talvolta apparire un po' forzato o artificiale a chi sia all'inizio dello studio della teoria assiomatica della probabilita'. Di fatto, invece, esso si rivela di importanza fondamentale sia nella formalizzazione rigorosa che nella comprensione di numerose questioni specifiche del calcolo delle probabilita'.

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 e' che la funzione definita dalla relazione

$$X(\omega) = X_1(\omega) + X_2(\omega), \text{ per } \omega \in \Omega.$$

Avevamo gia' avvertito comunque che la Definizione 7.1 di variabile aleatoria, cosi' come e' stata formulata, e' provvisoria. Come si vedra', essa va infatti adeguatamente modificata e completata quando si passi a trattare il caso in cui Ω non e' un insieme finito.

7.1 Esercizi di verifica

Esercizio 7.1. Indichiamo con $X_1, ..., X_5$ i 5 numeri estratti su una ruota del lotto (si estrae senza reinserimento da un'urna contenente i numeri {1,2,..., 90}) e sia inoltre X il valore più alto fra $X_1, ..., X_5$.
Calcolate $P(X \leq k)$ e $P(X = k)$ per $k = 1, 2...., 90$.

Esercizio 7.2. Supponiamo che X sia una variabile aleatoria a valori nell'insieme $\{0, 1, ..., n\}$ e che, per una coppia di costanti positive A e ρ, risulti

$$P(X = k) = A \cdot \frac{\rho^k}{k!\,(n-k)!}, \qquad k = 0, 1, ..., n,$$

Dimostrate che X segue una distribuzione binomiale ed individuatene i parametri.

Esercizio 7.3. Individuate una distribuzione di probabilità (non degenere) per una variabile aleatoria X in modo tale che risulti degenere 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 probabilita' binomiale $b(6, \frac{1}{3})$. Trovare qual e' il valore piu' probabile per X.

Esercizio 7.6. Consideriamo una variabile aleatoria con distribuzione di probabilità ipergeometrica

$$X \ \propto Hyp(6, 3; 3).$$

Qual e' il piu' 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.

8 Distribuzioni congiunte di piu' variabili aleatorie

In questa lezione esaminiamo alcune definizioni relative al caso in cui si considerino contemporaneamente, su uno stesso spazio di probabilita', due variabili aleatorie.

I concetti che verranno introdotti si possono estendere senza difficolta' al caso di un numero di variabili maggiore di due.

Sia dunque $(\Omega, \mathcal{P}(\Omega), P)$ uno spazio di probabilita' e X, Y una coppia di variabili aleatorie definite su di esso; si avra', diciamo, $X : \Omega \to X(\Omega) \equiv \{x_1, ..., x_n\}, Y : \Omega \to Y(\Omega) \equiv \{y_1, ..., y_m\}$ e possiamo considerare, su Ω, la partizione costituita dagli eventi del tipo

$$\{X = x_i\} \cap \{Y = y_j\} \equiv \{\omega \in \Omega : \ X(\omega) = x_i, Y(\omega) = y_j\}, \qquad i = 1, ..., n; \quad j = 1, ..., m.$$

Spesso, d'ora in poi, la scrittura $\{X = x_i\} \cap \{Y = y_j\}$ verra' piu' semplicemente sostituita da $\{X = x_i, Y = y_j\}$, come del resto gia' fatto nella Lez. precedente; porremo[24]

$$p_{ij} \equiv P(\{X = x_i, Y = y_j\}), \qquad i = 1, ..., n; \quad j = 1, ..., m$$

e ovviamente risultera'

$$p_{ij} \geq 0, \quad i = 1, ..., n; \ j = 1, ..., m; \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1.$$

Possiamo dunque considerare, analogamente a quanto visto sopra per il caso di una singola variabile aleatoria, il nuovo spazio di probabilita', 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[25]

$$\begin{aligned} P_{X,Y}(E) &\equiv P(\{(X, Y) \in E\}) = P(\{\omega \in \Omega : \ (X(\omega), Y(\omega)) \in E\}) \\ &= \sum_{(i,j):(x_i, y_j) \in E} p_{ij}; \end{aligned}$$

la misura di probabilita' $P_{X,Y}$ su $(X(\Omega) \times Y(\Omega), \mathcal{P}(X(\Omega) \times Y(\Omega)))$ prende il nome di ***distribuzione di probabilita' congiunta di*** X, Y.

---

[24] Anche qui utilizzeremo anche la notazione

$$p_{X,Y}(x_i, y_j) \equiv P(\{X = x_i, Y = y_j\}), \qquad i = 1, ..., n; \quad j = 1, ..., m$$

per mettere in evidenza le variabili aleatorie X ed Y coinvolte, ed i valori $\{x_1, \cdots, x_n\}$ e $\{x_1, \cdots, y_m\}$ che possono assumere. La funzione $p_{X,Y}$ assume il nome di ***densita' discreta congiunta*** di X e Y. Ovviamente, essendo $\{\{X = x_i, Y = y_j\}, \ i = 1, ..., n; \ j = 1, ..., m; \}$ una partizione, risultera'

$$p_{X,Y}(x_i, y_j) \geq 0, \quad i = 1, ..., n; \ j = 1, ..., m; \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1.$$

[25] Analogamente al caso di una variabile aleatoria unidimensionale si ha

$$\{(X, Y) \in E\} = \bigcup_{i,j:(x_i, y_j) \in E} \{(X, Y) = (x_i, y_j)\} = \bigcup_{i,j:(x_i, y_j) \in E} \{X = x_i, Y = y_j\},$$

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 $\{x_1, ..., x_n\}$, $\{y_1, ..., y_m\}$ rispettivamente e sia la loro distribuzione di probabilita' congiunta individuata dalle probabilita'

$$p_{ij} \equiv P(\{X = x_i, Y = y_j\}), \qquad i = 1, ..., n; \quad j = 1, ..., m.$$

La ***distribuzione marginale*** per la v.a. X e' la distribuzione di probabilita' concentrata sui valori $\{x_1, ..., x_n\}$ e che loro attribuisce le probabilita'

$$p'_i = \sum_{j=1}^{m} P(\{X = x_i, Y = y_j\}) = \sum_{j=1}^{m} p_{ij}, \quad i = 1, ..., n, \tag{39}$$

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 probabilita' per una variabile aleatoria, in quanto risulta

$$p'_i \geq 0, \quad i = 1, ..., n; \qquad \sum_{i=1}^{n} p'_i = 1.$$

Analogamente la ***distribuzione marginale*** per la v.a. Y e' la distribuzione di probabilita' concentrata sui valori $\{y_1, ..., y_m\}$ e che loro attribuisce le probabilita'

$$p''_j = \sum_{i=1}^{n} P(\{X = x_i, Y = y_j\}) = \sum_{i=1}^{n} p_{ij}, \qquad j = 1, ..., m. \tag{40}$$

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 probabilita' congiunte indicate nella seguente tabella*[26]

$Y \diagdown X$	**−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

Applichiamo la formula (39) per trovare la distribuzione di probabilita' 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 cioe' ottenuto le probabilita' 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 :

$$P\left(Y = \frac{1}{4}\right) = 0.3, \quad P\left(Y = \frac{1}{2}\right) = 0.17, \quad P\left(Y = \frac{3}{4}\right) = 0.19, \quad P\left(Y = 1\right) = 0.34.$$

[26] Si noti che tutti i calcoli che seguono non dipendono dallo spazio di probabilita' $(\Omega, \mathcal{P}(\Omega), P)$ su cui le due variabili aleatorie X ed Y sono definite, ma possono essere effettuati utilizzando solamente la distribuzione congiunta.

Riportiamo allora tali distribuzioni marginali, inserendole in una riga ed in una colonna aggiunte rispettivamente nei margini in basso e a destra della tabella.

Cioe' completiamo la tabella riportandovi anche le somme di riga e le somme di colonna; otteniamo dunque

$Y \diagdown X$	-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 probabilita' congiunta di attraverso una *tabella a doppia entrata* (cioe', insomma, una matrice), come segue:

$Y \diagdown X$	x_1	$\cdots$	x_i	$\cdots$	x_n	P_Y
y_1	$p_{1,1}$	$\cdots$	$p_{i,1}$	$\cdots$	$p_{n,1}$	p''_1
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
y_j	$p_{1,j}$	$\cdots$	$p_{i,j} = P(X = x_i, Y = y_j)$	$\cdots$	$p_{n,j}$	$p''_j = P(Y = y_j)$
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
y_m	$p_{1,m}$	$\cdots$	$p_{i,m}$	$\cdots$	$p_{n,m}$	p''_m
P_X	p'_1	$\cdots$	$p'_i = P(X = x_i)$	$\cdots$	p'_n	1

In questa tabella:

• la prima riga e la prima colonna rispettivamente indicano i valori possibili $\{x_i, i = 1, ..., n\}$, $\{y_j, j = 1, ..., m\}$ per le due variabili aleatorie;

• gli elementi nelle righe e colonne interne indicano le probabilita' 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, cioe' le probabilita' 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 $1 \le j \le m$ e consideriamo l'evento $\{Y = y_j\}$, che assumiamo avere probabilita' strettamente positiva.

Supponiamo ora che sia stato osservato questo evento, mentre non e' stato osservato il valore assunto dalla variabile X.

Ci possiamo allora domandare quale sia, data questa informazione, la distribuzione di probabilita' che esprime lo stato di informazione parziale circa X.

A questo quesito, risulta naturale rispondere con la seguente:

Definizione 8.1. *La* **distribuzione di probabilita' condizionata della variabile** X, **dato l'evento** $\{Y = y_j\}$ *e' la distribuzione di probabilita' che concentra, sui valori* x_i $(1 \le i \le n)$*, le* *probabilita' condizionate*[27] *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 $P(X = x_i|Y = y_j)$ *e' un modo tipograficamente rapido per scrivere* $P(\{X = x_i\}|\{Y = y_j\})$.

E' ovvio che le $p'_{i|j}$ si ricavano dalle probabilita' congiunte p_{ij} tramite la formula[28]

$$p'_{i|j} = \frac{p_{ij}}{p''_j} = \frac{p_{ij}}{\sum_{i=1}^{n} p_{ij}}, \tag{41}$$

e che, qualunque sia $j = 1, \cdots, m$

$$p'_{i|j} \geq 0, \quad i = 1, \cdots n, \qquad \sum_{i=1}^{n} p'_{i|j} = 1.$$

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 probabilita' che concentra, sui valori y_j $(1 \leq j \leq m)$, le probabilita' condizionate date da

$$p''_{j|i} \equiv P(Y = y_j|X = x_i) = \frac{p_{ij}}{p'_i} = \frac{p_{ij}}{\sum_{j=1}^{m} p_{ij}}. \tag{42}$$

Anche in questo caso, qualunque sia $i = 1, \cdots, n$

$$p''_{j|i} \geq 0, \quad j = 1, \cdots m, \qquad \sum_{j=1}^{m} p''_{j|i} = 1.$$

Osservazione 1. Consideriamo una coppia di variabili aleatorie X ed Y, per le quali gli insiemi di valori possibili siano $X(\Omega) \equiv \{x_1, ..., x_n\}$ e $Y(\Omega) \equiv \{y_1, ..., y_m\}$. La distribuzione di probabilita' congiunta e' allora individuata dall'insieme delle probabilita' congiunte

$$\{p_{ij};\ i = 1, ..., n;\ j = 1, ..., m\}.$$

In base a $\{p_{ij};\ i = 1, ..., n;\ j = 1, ..., m\}$, attraverso le formule (39), (40), (41), (42), si determinano univocamente

• le probabilita' marginali

$$\{p'_i; i = 1, ..., n\}, \qquad \{p''_j; \quad j = 1, ..., m\}$$

• e le probabilita' condizionate

$$\{p'_{i|j};\ j = 1, ..., m,\ i = 1, ..., n\}, \qquad \{p''_{j|i};\ i = 1, ..., n,\ j = 1, ..., m\}.$$

[27] Useremo anche le seguenti notazioni

$$p_X(x_i|Y = y_j) \equiv P(X = x_i|Y = y_j) \qquad \text{oppure} \qquad p_{X|Y}(x_i|y_j) \equiv P(X = x_i|Y = y_j)$$

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.

[28] Le formule (41) e (42) si possono scrivere rispettivamente anche nel seguente modo:

$$p_{X|Y}(x_i|y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{i=1}^{n} p_{X,Y}(x_i, y_j)},$$

$$p_{Y|X}(y_j|x_i) \equiv P(Y = y_j|X = x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} = \frac{p_{X,Y}(x_i, y_j)}{\sum_{j=1}^{m} p_{X,Y}(x_i, y_j)}.$$

Supponiamo ora di assegnare la coppia $\{p'_i;\ i = 1, ..., n\}$, $\{p''_j;\ j = 1, ..., m\}$, (con $p'_i \geq 0$ e $p''_j \geq 0$); si ricordi che vanno rispettati i vincoli dati dalle condizioni di normalizzazione

$$\sum_{i=1}^{n} p'_i = 1; \qquad \sum_{j=1}^{m} p''_j = 1.$$

Vi sono ovviamente diverse distribuzioni congiunte che ammettono $\{p'_i;\ i = 1, ..., n\}$, $\{p''_j;\ j = 1, ..., m\}$ come probabilita' marginali; guardiamo infatti il sistema lineare

$$\begin{cases} \sum_{j=1}^{m} p_{ij} = p'_i, & i = 1, ..., n \\ \sum_{i=1}^{n} p_{ij} = p''_j, & j = 1, ..., m \\ \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1 \end{cases}$$

nelle variabili p_{ij}: si tratta di un sistema di $(n + m + 1)$ equazioni in $n \cdot m$ variabili che risulta indeterminato[29], nei casi $n \geq 2, m \geq 2$.

Invece $\{p_{ij};\ i = 1, ..., n; j = 1, ..., m\}$ risultano univocamente determinate quando si impongano ad esempio sia le probabilita' marginali $\{p'_i;\ i = 1, ..., n\}$ che le probabilita' condizionate $\{p''_{j|i};\ i = 1, ..., n,\ j = 1, ..., m\}$; infatti si deve avere, proprio per la definizione di distribuzione condizionata,

$$p_{ij} = p'_i \cdot p''_{j|i}, \qquad i = 1, ..., n, \quad j = 1, ..., m.$$

Cio' mostra anche che la conoscenza di $\{p'_i;\ i = 1, ..., n\}$ e $\{p''_{j|i};\ i = 1, ..., n,\ j = 1, ..., m\}$ determina la coppia $\{p''_j;\ j = 1, ..., m\}$ $\{p'_{i|j};\ i = 1, ..., n,\ j = 1, ..., m\}$.

Notiamo anche che la relazione che lega fra loro tali probabilita' non e' nient'altro che la Formula di Bayes, che, riadattata al simbolismo qui introdotto, puo' essere riscritta nella forma:

$$p'_{i|j} = \frac{p'_i \cdot p''_{j|i}}{p''_j},$$

[29] In realta', date $\{p'_i;\ i = 1, ..., n\}$ e $\{p''_j;\ j = 1, ..., m\}$, con le condizioni di normalizzazione

$$\sum_{i=1}^{n} p'_i = 1; \qquad \sum_{j=1}^{m} p''_j = 1,$$

si tratta del sistema nelle $n \cdot m$ incognite p_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$,

$$\begin{cases} \sum_{j=1}^{m} p_{ij} = p'_i, & i = 1, ..., n \\ \sum_{i=1}^{n} p_{ij} = p''_j, & j = 1, ..., m \\ p_{ij} \geq 0, & i = 1, ..., n \quad j = 1, ..., m. \end{cases}$$

La condizione $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$ e' infatti automaticamente soddisfatta, grazie alle condizioni di normalizzazione per $\{p'_i;\ i = 1, ..., n\}$ e $\{p''_j;\ j = 1, ..., m\}$. I gradi di liberta' 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 e' analogo, ma, ad esempio, si possono fissare i valori di $p_{i,j}$ per $i = 1, ..., n-1$ e $j = 1, ..., m-1$, in modo che

$$p_{i,j} \geq 0,\ i = 1, ..., n-1,\ j = 1, ..., m-1 \qquad \sum_{j=1}^{m-1} p_{ij} \leq p'_i, \quad i = 1, ..., n-1 \quad e \quad \sum_{i=1}^{n-1} p_{ij} \leq p''_j, \quad j = 1, ..., m-1.$$

I rimanenti $n - 1 + m - 1 + 1 = n + m - 1$ valori di $p_{n,j}$ per $j = 1, ..., m-1$, $p_{i,m}$, per $i = 1, ..., n-1$, e $p_{n,m}$ sono automaticamente ricavati dalle equazioni del sistema.

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 gia' state calcolate (nell'Esempio 7.1 e nell'Esempio 7.5 della precedente lezione) le distribuzioni marginali. Costruite ora la tabella delle probabilita' congiunte e calcolate la distribuzione condizionata di Y, dato l'evento $\{X = 9\}$.

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 aleatori 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

Esercizio proposto 8.2. *Siano X ed Y due variabili aleatorie. Verificare*[30] *che, le seguenti condizioni sono fra di loro equivalenti:*

(i) le probabilita' 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*

$$\mathcal{A} \equiv \big\{\{X = x_1\}, ..., \{X = x_n\}\big\} \quad e \quad \mathcal{B} \equiv \big\{\{Y = y_1\}, ..., \{Y = y_m\}\big\}.$$

[30] Le 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 cio', sommando sugli indici $i = 1, \ldots, n$ 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).

In altre parole le variabili aleatorie X ed Y sono stocasticamente indipendenti se e solo se, vale[31]

$$p_{ij} = p'_i \cdot p''_j, \qquad \forall\, 1 \le i \le n, \quad 1 \le j \le m,$$

ovvero

$$P\big(\{X = x_i\} \cap \{Y = y_j\}\big) = P\big(\{X = x_i\}\big) \cdot P\big(\{Y = y_j\}\big), \qquad \forall\, 1 \le i \le n, \quad 1 \le j \le m$$

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

$$P(X \in I,\, Y \in J) = P(X \in I)P(Y \in J), \qquad \text{qualunque siano } I \subseteq X(\Omega) \text{ e } J \subseteq Y(\Omega) \tag{43}$$

o anche, posto $\mathcal{G}(X) = \mathcal{G}(\mathcal{A})$, l'algebra generata dalla partizione $\mathcal{A} \equiv \big\{\{X = x_1\}, ..., \{X = x_n\}\big\}$ e $\mathcal{G}(Y) = \mathcal{G}(\mathcal{B})$, l'algebra generata dalla partizione $\mathcal{B} \equiv \big\{\{Y = y_1\}, ..., \{Y = y_m\}\big\}$, si ha

$$P(A \cap B) = P(A)P(B), \qquad \text{qualunque siano } A \in \mathcal{G}(X) \text{ e } B \in \mathcal{G}(Y).$$

Inoltre vale anche il viceversa, ovvero se vale (43), allora X ed Y sono indipendenti secondo la precedente Definizione 8.2.

Dimostrazione. In realta', come gia' 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

$$\{X \in I\} = \bigcup_{i:\, x_i \in I} \{X = x_i\}, \qquad \{Y \in J\} = \bigcup_{j:\, y_j \in J} \{Y = y_j\},$$

e quindi

$$\{X \in I\} \cap \{Y \in J\} = \left(\bigcup_{i:\, x_i \in I} \{X = x_i\} \right) \bigcap \left(\bigcup_{j:\, y_j \in J} \{Y = y_j\} \right) = \bigcup_{i:\, x_i \in I}\, \bigcup_{j:\, y_j \in J} \{X = x_i\} \cap \{Y = y_j\}.$$

Di conseguenza, se X ed Y sono indipendenti secondo la precedente Definizione 8.2, allora[32]

$$\begin{aligned}
P(\{X \in I\} \cap \{Y \in J\}) &= \sum_{i:\, x_i \in I} \sum_{j:\, y_j \in J} P(\{X = x_i\} \cap \{Y = y_j\}) \\
&= \sum_{i:\, x_i \in I} \sum_{j:\, y_j \in J} P(\{X = x_i\}) \cdot P(\{Y = y_j\}) \\
&= \sum_{i:\, x_i \in I} P(\{X = x_i\}) \cdot \sum_{j:\, y_j \in J} P(\{Y = y_j\}) \\
&= \sum_{i:\, x_i \in I} P(\{X = x_i\}) \cdot P(\{Y \in J\}) \\
&= P(X \in I)P(Y \in J).
\end{aligned}$$

Il viceversa e' ovvio, basta prendere $I = \{x_i\}$ e $J = \{y_j\}$.

---

[31] Si noti che si tratta della condizione (iii) dell'Esercizio proposto 8.2.
[32] Si 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

$Y \diagdown X$	$\mathbf{-1}$	$\mathbf{0}$	$\mathbf{1}$
$\mathbf{-1}$	0.1	0.1	0
$\mathbf{0}$	0	0.1	0.3
$\mathbf{1}$	0.1	0.15	θ

essendo θ un opportuno valore $0 < \theta < 1$.
a) Determinare il valore di θ.
b) Determinare la distribuzione di probabilita' congiunta della coppia (X, Z), dove $Z \equiv X \cdot Y$
c) Determinare la distribuzione di probabilita' marginale della variabile Z
d) Determinare la distribuzione di probabilita' condizionata di X, dato $\{Z = 1\}$.

Esercizio 8.2. Verificare se sono stocasticamente indipendenti le variabili aleatorie X ed Y, la cui distribuzione di probabilita' congiunta e' 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 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 $X(\Omega) \equiv \{x_1, ..., x_n\}$ e $Y(\Omega) \equiv \{y_1, ..., y_m\}$ e siano $\{p'_i;\ i = 1, ..., n\}$ e $\{p''_j;\ j = 1, ..., m\}$ le relative probabilita' marginali. Trovare la distribuzione di probabilita' 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.
i) Determinare la distribuzione di probabilita' di $Z \equiv X + Y$.
ii) Determinare la distribuzione di probabilita' 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. Iniziamo notando immediatamente che $Z(\Omega) = \{z \in \mathbb{R} : \exists x_k \in X(\Omega),\, y_h \in \Omega :\ x_k + y_h = z\}$, e

$$\begin{aligned}\{Z = z\} = \{X + Y = z\} &= \bigcup_{x_k,\, y_h \colon x_k + y_h = z} \{X = x_k,\, Y = y_h\} \\ &= \bigcup_k \{X = x_k,\, Y = z - x_k\} = \bigcup_h \{X = z - y_h,\, Y = y_h\}\end{aligned}$$

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

$$\begin{aligned}P(\{Z = z\}) = P(\{X + Y = z\}) &= \sum_{x_k,\, y_h \colon x_k + y_h = z} P(\{X = x_k,\, Y = y_h\}) \\ &= \sum_k P(\{X = x_k,\, Y = z - x_k\}) = \sum_h P(\{X = z - y_h,\, Y = y_h\}).\end{aligned}$$

Se inoltre si ha 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. 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

$$Z(\Omega) = 0, 1, 2, \dots r+s,$$

$$\begin{aligned}P(Z=\ell) &= \sum_{\substack{k+h=\ell\\ 0\le k\le r,\ 0\le h\le s}} \binom{r}{k}\theta^k\,(1-\theta)^{r-k}\binom{s}{h}\theta^h\,(1-\theta)^{s-h}\\ &= \sum_{\substack{k+h=\ell\\ 0\le k\le r,\ 0\le h\le s}} \theta^\ell\,(1-\theta)^{r+s-\ell}\binom{r}{k}\binom{s}{h}\\ &= \theta^\ell\,(1-\theta)^{r+s-\ell} \sum_{0\le k\le r,\ 0\le \ell-k\le s}\binom{r}{k}\binom{s}{\ell-k},\end{aligned}$$

da cui, ricordando la (19),

$$P(Z=\ell) = \theta^\ell\,(1-\theta)^{r+s-\ell}\binom{r+s}{\ell} \qquad \ell = 0, 1, \dots, r+s.$$

Soluzione dell'Esercizio 8.5. Punto ii) Utilizzando le definizioni di probabilita' 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 $0 \le i \le r$ e $0 \le k-i \le s$

$$\begin{aligned}P(X=i|Z=k) &= \frac{P(X=i,\, Y=k-i)}{P(Z=k)} = \frac{P(X=i)P(Y=k-i)}{P(Z=k)}\\ &= \frac{\binom{r}{i}\theta^i\,(1-\theta)^{r-i}\binom{s}{k-i}\theta^{k-i}\,(1-\theta)^{s-(k-i)}}{\binom{r+s}{k}\theta^k\,(1-\theta)^{r+s-k}}\\ &= \frac{\binom{r}{i}\binom{s}{k-i}}{\binom{r+s}{k}} \qquad 0\le i\le r \text{ e } 0\le k-i\le s,\end{aligned}$$

cioe' la distribuzione condizionata di X data $Z=k$ e' 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 probabilita' θ, cioe' 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 proprieta':

a) X' ha la stessa distribuzione di X, cioe' $b(r, \theta)$,
b) Y' ha la stessa distribuzione di Y, cioe' $b(s, \theta)$,
c) X' ed Y' sono indipendenti[33],

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 cio' 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

$$(E_1^* \cap E_2^* \cap \ldots \cap E_r^*) \cap (E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*)$$

nelle quali per esattamente k tra gli E_ℓ^* si ha $E_\ell^* = E_\ell$, mentre per i restanti $r+s-k$ si ha $E_\ell^* = \overline{E}_\ell$.

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 $K \subseteq \{1, 2, \ldots, r + s\}$ di cardinalita' k. Ciascuno degli $\binom{r+s}{k}$ eventi di questo tipo ha la stessa probabilita' (ed esattamente $\theta^k \cdot (1 - \theta)^{r+s-k}$).
L'evento $X' = i$, per $i = 0, 1, \ldots r$ corrisponde al caso in cui la cardinalita' di $K_A := K \cap \{1, 2, \ldots, r\}$ e' uguale ad i (e quindi la cardinalita' di $K_B := K \cap \{r + 1, r + 2, \ldots, r + s\}$ e' uguale ad $k - i$).

Tenendo conto di cio' ci si puo' convincere facilmente che la distribuzione condizionata di X' data $Z' = k$ e'

$$P(X' = i | Z' = k) = \frac{\binom{r}{i}\binom{s}{k-i}}{\binom{r+s}{k}} \qquad \text{per } 0 \le i \le r, \quad 0 \le k - i \le s,$$

cioe' e' una $Hyp(r + s, r; k)$.

---

[33]Per dimostrare la proprieta' c) bisogna utilizzare la ***Proposizione 1*** della Lezione 5, considerando che
i) la partizione $\mathcal{A}$ generata dagli eventi $\{E_1, E_2, \ldots, E_r\}$ e la partizione $\mathcal{A}$ generata dagli eventi $\{E_{r+1}, E_{r+2}, \ldots, E_{r+s}\}$ sono indipendenti:
infatti un evento $A \in \mathcal{A}$ se e solo se $A = E_1^* \cap E_2^* \cap \ldots \cap E_r^*$ dove $E_i^* = E_i$ oppure $E_i^* = \overline{E}_i$, e analogamente un evento $B \in \mathcal{B}$ se e solo se $B = E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*$ dove $E_j^* = E_j$ oppure $E_j^* = \overline{E}_j$, e quindi

$$P\big((E_1^* \cap E_2^* \cap \ldots \cap E_r^*) \cap (E_{r+1}^* \cap E_{r+2}^* \cap \ldots \cap E_{r+s}^*)\big) = P(E_1^*) \cdot P(E_2^*) \cdot \cdots \cdot P(E_r^*) \cdot P(E_{r+1}^*) \cdot P(E_{r+2}^*) \cdot \cdots \cdot P(E_{r+s}^*)$$

$$\Downarrow$$

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

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 proprieta'

In questa lezione verra' introdotta la nozione di "valore atteso" di una variabile aleatoria e ne verranno messe in evidenza proprieta' ed aspetti fondamentali.

In molti problemi di tipo probabilistico, data una variabile aleatoria X, sorge la necessita' di individuare una quantita' 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 finanziaria[34], nasce spesso l'esigenza di valutare una cifra deterministica, che risulti equa quale importo da pagare per avere il diritto di godere di tale ricavo[35]. Similmente, in un gioco d'azzardo vi e' la necessita' di verificare se il gioco sia, o meno, equo, e cosi' via...

Come si comincera' 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* $X:\ \Omega \equiv \{\omega_1, ..., \omega_N\} \to X(\Omega) \subseteq \mathbb{R}$ *una variabile aleatoria definita su* $(\Omega, \mathcal{P}(\Omega))$. *Si definisce* **valore atteso**[36] *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 e' bene elencare alcune proprieta' immediate di tale definizione; in quanto segue X, Y, Z, ... sono variabili aleatorie definite sullo stesso spazio $\Omega = \{\omega_1, ..., \omega_N\}$.

La prima proprieta' e' immediata[37] (ricordando che deve essere, ovviamente, $\sum_{i=1}^{N} p(\omega_i) = 1$):

Proposizione 1. Se X e' una variabile aleatoria degenere, cioe' tale che, per un valore $\widehat{x} \in \mathbb{R}$, vale $X(\omega_i) = \widehat{x}$, $i = 1, ..., N$, allora si ha

$$\mathbb{E}(X) = \widehat{x}.$$

Anche le dimostrazioni delle seguenti due proprieta' seguono banalmente dalla definizione.

---

[34] Ad esempio nel caso di un'***opzione call*** l'operazione consiste nel comprare la possibilita', 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 possibilita' viene esercitata se il prezzo (aleatorio) al tempo T e' maggiore di K, altrimenti non viene esercitata: non e' conveniente pagare K quello che si puo' ottenere sul mercato ad un prezzo minore.

[35] Piu' in generale quando ci si assicura contro eventuali danni o furti, quando si scommette o si gioca in borsa, si deve pagare una quantita' di denaro certa (il premio di un'assicurazione, l'ammontare della scommessa) in cambio di una quantita' aleatoria (il denaro che si potrebbe ottenere in caso di danno o furto, l'importo della scommessa, in caso di vincita).

E' interessante considerare anche il caso del gioco in borsa, in cui la quantita' certa e', ad esempio, il prezzo di un'***opzione call***, ovvero il prezzo per ottenere la possibilita' di pagare una azione ad un prezzo prefissato K, nell'istante T, mentre la quantita' aleatoria e' il ricavo tra il prezzo dell'azione al tempo T e K, se questa differenza e' positiva.

[36] Oltre a valore atteso si usano anche i termini ***valore medio***, ***aspettazione***, ***speranza matematica***, e a volte anche ***media***, che pero' e' meglio evitare, perche' spesso quando si parla di media ci si riferisce alla media aritmetica.

[37] E del resto, pensando all'interpretazione di $\mathbb{E}(X)$ come quantita' certa che si e' disposti a scambiare con X, si capisce che se X e' a sua volta una quantita' certa $\widehat{x}$, ci si aspetta che $\mathbb{E}(X)$ sia uguale a $\widehat{x}$.

Proposizione 2. Sia $E \subseteq \Omega$ un evento e $X = \mathcal{X}_E$ la variabile aleatoria indicatore di E. Si ha

$$\mathbb{E}(X) = P(E).$$

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

$$a \leq \mathbb{E}(X) \leq b.$$

Piu' in generale vale la ***proprieta' di monotonia***: se X ed Y sono due variabili aleatorie con la proprieta' che $X(\omega) \leq Y(\omega)$ per ogni $\omega \in \Omega$, allora

$$\mathbb{E}(X) \leq \mathbb{E}(Y).$$

Una proprieta' assolutamente fondamentale del valore atteso e' la ***linearita'***, che verra' 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, cioe' poniamo

$$Z(\omega_i) \equiv a \cdot X(\omega_i) + b \cdot Y(\omega_i), \qquad \forall 1 \leq i \leq N.$$

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 virtu' della definizione, abbiamo

$$\mathbb{E}(Z) = \sum_{i=1}^{N} p(\omega_i) \left[a \cdot X(\omega_i) + b \cdot Y(\omega_i)\right] =$$

$$a \cdot \sum_{i=1}^{N} p(\omega_i) X(\omega_i) + b \cdot \sum_{i=1}^{N} p(\omega_i) Y(\omega_i) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y).$$

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 degenere con $Y(\omega_i) = 1$ per ogni $\omega_i \in \Omega$.)

Infine va ricordato che la proprieta' della ***Proposizione 4*** si estende immediatamente al caso di n variabili aleatorie:

$$\mathbb{E}\left(\sum_{k=1}^{n} a_i X_i\right) = \sum_{k=1}^{n} a_i \mathbb{E}(X_i)$$

Di fatto, per calcolare il valore atteso $\mathbb{E}(X)$ di una variabile aleatoria X, occorre conoscere la distribuzione di probabilita' di X ma non e' necessariamente richiesto di conoscere quale sia lo spazio di probabilita' Ω su cui X e' definita, ne' come X vi possa essere definita, ne' quale sia la misura di probabilita' su $(\Omega, \mathcal{P}(\Omega))$.

Il valore atteso $\mathbb{E}(X)$ dipende infatti soltanto dalla distribuzione di probabilita' 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 $X(\Omega) \equiv \{x_1, ..., x_n\}$. Risulta allora

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j \cdot P\big(\{X = x_j\}\big). \tag{44}$$

Dimostrazione.
I modo
Dal momento che $\{\{X = x_1\}, ..., \{X = x_n\}\}$ 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:

$$\sum_{i:X(\omega_i)=x_j} p(\omega_i) \cdot X(\omega_i) = \sum_{i:X(\omega_i)=x_j} p(\omega_i) \cdot x_j =$$

$$= x_j \left(\sum_{i:X(\omega_i)=x_j} p(\omega_i) \right) = x_j \cdot P\big(\{X = x_j\}\big).$$

La dimostrazione e' quindi completata.

II modo
La precedente dimostrazione e' autocontenuta. Tuttavia e' interessante notare che c'e' 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 linearita'

$$\mathbb{E}(X) = \sum_{j=1}^{n} x_j \mathbb{E}(X_j) = \sum_{j=1}^{n} x_j P(E_j) = \sum_{j=1}^{n} x_j P(\{X = x_j\}).$$

Questa dimostrazione si estende immediatamente anche al caso in cui si voglia calcolare $\mathbb{E}\big(h(X)\big)$. 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 probabilita', uniforme su $\{1, 2, ..., 6\}$. Il loro comune valore atteso e' dato da*

$$\mathbb{E}(X_1) = \sum_{j=1}^{6} \frac{j}{6} = 3.5\big(= \mathbb{E}(T)\big).$$

Questo esempio si generalizza immediatamente al caso in cui X e' una variabile aleatoria uniforme su $\{1, 2, \ldots, n\}$, ovvero con $P(X = j) = \frac{1}{n}$, per $j = 1, 2, \ldots, n$ per cui[38]

$$\mathbb{E}(X) = \sum_{j=1}^{n} j\frac{1}{n} = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Nel caso specifico di una variabile aleatoria X a valori in $\{0, 1, 2, ..., n\}$ o in $\{1, 2, ..., n\}$, il calcolo del valore atteso puo' anche essere convenientemente eseguito in termini della funzione

$$\overline{F}\,(j) \equiv P(\{X > j\});$$

vale infatti la seguente

Proposizione 8. Sia $X\,(\Omega) \subseteq \{0, 1, ..., n\}$. Allora si ha

$$\mathbb{E}(X) = \sum_{j=0}^{n-1} P(\{X > j\}) = \sum_{j=0}^{n-1} \overline{F}\,(j)$$

In particolare se $X\,(\Omega) = \{1, ..., n\}$, allora si ha

$$\mathbb{E}(X) = 1 + \sum_{j=1}^{n-1} P(\{X > j\})$$

Dimostrazione.

Come gia' notato nel corso della Lezione 7, si puo' 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

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{j=0}^{n} j \cdot P(\{X = j\}) = \sum_{j=1}^{n} j \cdot P(\{X = j\}) = \\
&= \sum_{j=1}^{n} j \cdot [P(\{X > j - 1\}) - P(\{X > j\})] \\
&= \sum_{j=1}^{n} j \cdot P(\{X > j - 1\}) - \sum_{j=1}^{n} j \cdot P(\{X > j\})
\end{aligned}$$

[38]Si ricordi che

$$\sum_{j=1}^{n} j = \frac{n(n+1)}{2}.$$

ponendo $h = j - 1$, e tenendo conto che $P(\{X > n\}) = 0$,

$$\begin{aligned}
&= \sum_{h=0}^{n-1} (h+1) \cdot P(\{X > h\}) - \sum_{j=1}^{n-1} j \cdot P(\{X > j\}) \\
&= 1 \cdot P(\{X > 0\}) + 2 \cdot P(\{X > 1\}) + 3 \cdot P(\{X > 2\}) + \cdots + nP(\{X > n-1\}) \\
&\qquad\qquad - 1 \cdot P(\{X > 1\}) - 2 \cdot P(\{X > 2\}) - \cdots - (n-1)P(\{X > n-1\}) \\
&= P(\{X > 0\}) + (2-1)P(\{X > 1\}) + \cdots + \big(n - (n-1)\big)P(\{X > n-1\}) \\
&= P(\{X > 0\}) + \sum_{h=1}^{n-1} \big((h+1) - h\big) \cdot P(\{X > h\}) \\
&= \sum_{h=0}^{n-1} P(\{X > h\}),
\end{aligned}$$

il che prova la prima affermazione. La seconda affermazione dipende solo dal fatto che se $X(\Omega) = \{1, ..., n\}$, 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 $Z = X \bigvee Y$ (cioe' $Z = \max(X, Y)$). Calcolare $\mathbb{E}(Z)$.*

Soluzione. Avevamo visto nell'Esempio 7.5 della precedente Lezione 7 che $P(\{Z \leq x\}) = \left(\frac{x}{6}\right)^2$, $x = 1, 2, ..., 6$. Dunque $P(\{Z > x\}) = 1 - P(\{Z \leq x\}) = 1 - \left(\frac{x}{6}\right)^2$. 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 $\{1, 2, \ldots, n\}$:

$$\begin{aligned}
\mathbb{E}(X \vee Y) &= \sum_{j=0}^{n-1} P(\{X \vee Y > j\}) = \sum_{j=0}^{n-1} \big(1 - P(\{X \vee Y \leq j\})\big) \\
&= \sum_{j=0}^{n-1} 1 - \sum_{j=0}^{n-1} P(\{X \leq j, Y \leq j\}) = n - \sum_{j=0}^{n-1} P(\{X \leq j, Y \leq j\}).
\end{aligned}$$

Si noti che finora si e' usato solo il fatto che le due variabili aleatorie sono a valori in $\{1, 2, \ldots, n\} \subseteq \{0, 1, 2, \ldots, n\}$.

A questo punto per l'indipendenza delle due variabili aleatorie X ed Y si ha[39]

$$P(\{X \leq j, Y \leq j\}) = P(\{X \leq j\})P(\{Y \leq j\}),$$

da cui

$$\mathbb{E}(X \vee Y) = n - \sum_{j=0}^{n-1} P(\{X \leq j\})P(\{Y \leq j\}). \tag{45}$$

[39] Si noti che la formula (45) vale in generale per variabili aleatorie indipendenti X e Y, a valori in $\{0, 1, 2, \ldots, n\}$.

Essendo sia X che Y uniformi in $\{1, 2, \dots, n\}$ si ha[40]

$$P(\{X \leq j\})P(\{Y \leq j\}) = \left(\frac{j}{n}\right)^2, \qquad \text{per } j = 0, 1, \dots n,$$

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'e' da notare comunque che, ***in molti casi, il valore atteso di una variabile aleatoria puo' essere ottenuto in modo semplice, senza neanche calcolare la distribuzione di probabilita'*** (o, nel caso di variabili a valori interi positivi, la funzione $\overline{F}(j)$).

Piuttosto si tratta di sfruttare adeguatamente la proprieta' di linearita'. Il ruolo di tale proprieta' verra' 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 (44), potremo scrivere:

$$\mathbb{E}(X) = \sum_{k=0}^{n} k \binom{n}{k} \theta^k (1-\theta)^{n-k},$$

e fare i conti relativi[41]. 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 $X_1, ..., X_n$, con $P(\{X_j = 1\}) = \theta$; per la proprieta' di linearita' del valore atteso si ha dunque

$$\mathbb{E}(X_j) = \theta, \qquad \mathbb{E}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathbb{E}(X_j) = n\theta.$$

[40] Si tenga conto del fatto che

$$\sum_{k=1}^{n} k^2 = \frac{n(n+\frac{1}{2})(n+1)}{3}.$$

[41] Si 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 $X_1, ..., X_n$ 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 probabilita', 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 probabilita' di successo e' uguale a* 0.80*, qual e' il valore atteso del numero di studenti che passano la prova?*

Naturalmente possiamo avere distribuzioni di probabilita' 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 e' una variabile aleatoria con distribuzione* $b(4, \frac{1}{6})$*; dunque il suo valore atteso e' dato da* $\frac{1}{6} \times 4 = \frac{2}{3}$*. Tale valore coincide anche con il valore atteso del numero di volte in qui 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 probabilita' di ottenere almeno un successo, come avevamo visto).

Osservazione 1. Dati n eventi, A_1; A_2, ...,A_n, consideriamo la variabile S_n che conta il numero dei successi su tali eventi, ovvero

$$S_n = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n} = X_1 + X_2 + \cdots + X_n,$$

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 probabilita' θ, il valore atteso $\mathbb{E}(S_n)$ e' uguale, in forza della proprieta' di linearita', 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 avra' influenza sulla distribuzione di probabilita' 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

$$= \sum_{h=0}^{n-1} n \frac{(n-1)!}{h!(n-1-h)!} \theta^{h+1} (1-\theta)^{n-1-h}$$

$$= n\theta \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} \theta^{h} (1-\theta)^{n-1-h} = n\theta \sum_{h=0}^{n-1} \binom{n-1}{h} \theta^{h} (1-\theta)^{n-1-h}$$

tenendo conto dello sviluppo del binomio

$$= n\theta \left(\theta + (1-\theta)\right)^{n-1} = n\theta \cdot 1.$$

La variabile aleatoria
$$Y_n \equiv \frac{S_n}{n}$$
puo' essere interpretata come la ***frequenza relativa dei successi*** sugli n eventi. Ancora per la proprieta' di linearita', avremo, in ogni caso,
$$\mathbb{E}\left(Y_n\right) = \theta.$$

Esempio 9.4 (valore atteso di una ipergeometrica). *Consideriamo in particolare* ***una distribuzione ipergeometrica di parametri*** M, m_1, n. *Come sappiamo questa e' la distribuzione di probabilita' di una variabile aleatoria S che conta il numero di successi su n eventi (non indipendenti) ciascuno di probabilita'* $\frac{m_1}{M}$*; e quindi si ha che* ***il suo valore atteso e' dato da*** $n \cdot \frac{m_1}{M}$.

Estendiamo ora la discussione su un punto, gia' introdotto nella precedente ***Osservazione 1***, che risulta di notevole importanza nella teoria della probabilita'.

Osservazione 2 **(valore atteso di una media aritmetica)**. Consideriamo n variabili aleatorie $X_1, ..., X_n$ definite su uno stesso spazio $(\Omega, \mathcal{P}(\Omega))$ e che abbiano tutte lo stesso valore atteso; cioe', usando il simbolo μ per indicare brevemente $\mathbb{E}(X_j)$,
$$\mathbb{E}(X_1) = ... = \mathbb{E}(X_n) = \mu.$$

Consideriamo ora, sullo stesso spazio $(\Omega, \mathcal{P}(\Omega))$, la variabile aleatoria definita come ***media aritmetica*** delle $X_1, ..., X_n$,
$$Y_n(\omega_i) = \frac{\sum_{j=1}^{n} X_j(\omega_i)}{n}, \qquad i = 1, ..., N;$$
si ha ovviamente, in virtu' della proprieta' di linearita' del valore atteso,
$$\mathbb{E}(Y_n) = \mu,$$
ottenendo quindi una generalizzazione del caso considerato nella precedente ***Osservazione 1***.

A parita' di valore atteso possono pero' sussistere situazioni assai diverse per quanto riguarda la distribuzione di probabilita' di Y_n, a seconda delle proprieta' di ***correlazione*** fra le variabili $X_1, ..., X_n$. Questo e' 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 e' 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 entita'* c_1,

r_2 *"secondi premi" di entita'* c_2 *e*

r_3 *"terzi premi" di entita'* c_3

(dunque $r = r_1 + r_2 + r_3$ *e* $c_1 > c_2 > c_3 > 0$*).*

Indichiamo con $X_1, ..., X_n$ *le vincite rispettivamente associate al biglietto 1, al biglietto 2, ..., al biglietto* n*.*

Ovviamente $X_1, ..., X_n$ *sono delle variabili aleatorie i cui valori possibili costituiscono l'insieme* $\{0, c_1, c_2, c_3\}$*.* $X_1, ..., X_n$ *hanno tutte, la stessa distribuzione di probabilita', data da*[42]

$$P(\{X_j = c_1\}) = \frac{r_1}{n}, \quad P(\{X_j = c_2\}) = \frac{r_2}{n}, \quad P(\{X_j = c_3\}) = \frac{r_3}{n},$$

mentre

$$P(\{X_j = 0\}) = \frac{n-r}{n}\big(= 1 - P(\{X_j = c_1\}) - P(\{X_j = c_2\}) - P(\{X_j = c_3\})\big);$$

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 puo' pervenire anche piu' semplicemente a quest'ultima conclusione, sfruttando di nuovo la proprieta' di linearita' del valore atteso e ragionando come segue: per motivi di simmetria, e' ovvio che $X_1, ..., X_n$ *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 degenere, di valore* $r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3$*, ovvero*

$$S_n(\omega) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3, \qquad \text{qualunque sia } \omega \in \Omega$$

quindi

$$\mathbb{E}\,(S_n) = r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3.$$

Per la proprieta' di linearita', 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

$$\mu = \frac{\mathbb{E}\,(S_n)}{n} = \frac{r_1 \cdot c_1 + r_2 \cdot c_2 + r_3 \cdot c_3}{n}. \tag{46}$$

Ovviamente anche il valore atteso della variabile aleatoria $Y_n \equiv \frac{S_n}{n}$ *(*Y_n *media aritmetica di* $X_1, ..., X_n$*) e' uguale alla quantita'* μ*, come sappiamo deve essere; ma in questo speciale caso* Y_n *e' una variabile aleatoria con distribuzione degenere, tutta concentrata sul valore* μ*.*

Dal confronto cioe' 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 degenere, come invece accade per* Y_n*.*

---

[42] Infatti si puo' pensare che vengano effettivamente messe in un'urna n palline con i numeri di tutti i possibili biglietti 1, 2 $\dots, n$ 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* $(r_1 + r2) - sima$ *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* $(r_1 + r2 + r_2) - sima$". Basta poi notare che la probabilita' 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 proprieta' di linearita' del valore atteso puo' essere illustrata dal seguente esempio. Supponiamo di possedere un biglietto di una lotteria A ed un biglietto della lotteria B.*

Supponiamo per semplicita' che A distribuisca

r_1' *premi di entita'* c_1' *e* r_2' *premi di entita'* c_2'*, su un totale di* n' *biglietti;*

mentre B distribuisce

r_1'' *premi di entita'* c_1' *e* r_2'' *premi di entita'* 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, cosicche' $X = X' + X''$.

X' e' una variabile aleatoria che puo' assumere i valori $\{0, c_1', c_2'\}$ e X'' e' una variabile aleatoria che puo' assumere i valori $\{0, c_1'', c_2''\}$ e risulta, procedendo come nel precedente Esempio 9.5,

$$\mathbb{E}(X') = \frac{c_1' r_1' + c_2' r_2'}{n'}, \qquad \mathbb{E}(X'') = \frac{c_1'' r_1'' + c_2'' r_2''}{n''}.$$

È importante sottolineare che, sempre grazie alla proprieta' di linearita' del valore atteso, ***per calcolare il valore atteso*** $\mathbb{E}(X)$ ***non c'e' bisogno di calcolare interamente la distribuzione di probabilita' di*** X; possiamo infatti ottenere immediatamente, in virtu' della proprieta' di linearita',

$$\mathbb{E}(X) = \mathbb{E}(X') + \mathbb{E}(X'') = \frac{c_1' r_1' + c_2' r_2'}{n'} + \frac{c_1'' r_1'' + c_2'' r_2''}{n''}.$$

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 e' $\mathbb{E}(X_j) = \mu$, dove μ e' ottenuto come in (46) 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) e' certo, nel senso che non dipende dal risultato aleatorio della lotteria (cioe' da quali saranno i biglietti estratti). Su tale base, si puo' 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 e' fondamentale nel contesto della finanza matematica, in relazione al concetto di "non arbitraggio"[43].

Il ruolo della nozione di valore atteso nella comprensione del concetto di gioco equo e' anche illustrata nel seguente

---

[43] Si dice che in una operazione finanziaria c'e' opportunita' di arbitraggio se l'operazione finanziaria di sicuro non comporta perdite, e comporta un guadagno strettamente positivo con probabilita' 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 puo' 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 cosi' di seguito finche' non vince la prima volta o finche' non esaurisce il suo capitale iniziale.*

Supponiamo che, in ciascuna puntata, il giocatore vince o perde con probabilita' 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 e' 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

$$P(\{X=0\}) = (1-\theta)^5\,, \qquad P(\{X=32\}) = 1-(1-\theta)^5\,,$$

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, cosi' 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 $\theta > \frac{1}{2}$, cioe' 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, ..., p_n, ponendole rispettivamente sui punti di ascissa $x_1, x_2, ..., x_n$. Osserviamo allora che la quantita' $\sum_{j=1}^{n} p_j \cdot x_j$ equivale al baricentro[44] di tale distribuzione di masse.

Cosi' come il baricentro costituisce un valore che, a certi fini, risulta "riassuntivo" di tutta la distribuzione di masse, cosi' il valore atteso di una variabile aleatoria costituisce un valore che, entro certi fini, riassume la conoscenza completa della distribuzione di probabilita'.

Una proprieta' di fondamentale importanza e' data dalla seguente

Proposizione 9. Siano X, Y due variabili aleatorie indipendenti definite su uno stesso spazio di probabilita'. Allora

$$\mathbb{E}\left(X\cdot Y\right) = \mathbb{E}(X)\cdot\mathbb{E}(Y).$$

Dimostrazione. Siano rispettivamente

$$X(\Omega) = \{x_1, ..., x_n\} \quad \text{e} \quad Y(\Omega) = \{y_1, ..., y_m\}.$$

Possiamo allora considerare la partizione di Ω costituita dagli eventi del tipo

$$E_{h,k} = \{X = x_h\} \cap \{Y = y_k\} \equiv \{\omega_i \in \Omega : \ X(\omega_i) = x_h, Y(\omega_i) = y_k\},$$

[44] Infatti in generale se le masse sono m_1, m_2, ..., m_n e se sono poste rispettivamente sui punti di ascissa $x_1, x_2, ..., x_n$, allora il baricentro e'

$$\hat{x} = \frac{x_1\cdot m_1 + x_2\cdot m_2 + \cdots + x_n\cdot m_n}{m_1+m_2+\cdots+m_n}.$$

Nel caso in cui $m_i = p_i$, con p_i densita' discreta, la precedente espressione diviene

$$\hat{x} = \frac{x_1\cdot p_1 + x_2\cdot p_2 + \cdots + x_n\cdot p_n}{p_1+p_2+\cdots+p_n} = x_1\cdot p_1 + x_2\cdot p_2 + \cdots + x_n\cdot p_n = \mathbb{E}(X).$$

in quanto il denominatore $p_1+p_2+\cdots+p_n$ vale 1.

per $h = 1, 2, \dots, n$, e $k = 1, 2, \dots, m$, e scrivere

$$\begin{aligned}\mathbb{E}\left(X \cdot Y\right) &= \sum_{i=1}^{N} p(\omega_i)\left[X\left(\omega_i\right) \cdot Y\left(\omega_i\right)\right] = \sum_{h=1}^{n}\sum_{k=1}^{m}\left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i)\left[X\left(\omega_i\right) \cdot Y\left(\omega_i\right)\right]\right] \\ &= \sum_{h=1}^{n}\sum_{k=1}^{m}\left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i) x_h \cdot y_k\right] = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \left[\sum_{i:\omega_i \in \{X=x_h\}\cap\{Y=y_k\}} p(\omega_i)\right] \\ &= \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k P\left(\{X = x_h\} \cap \{Y = y_k\}\right).\end{aligned}$$

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 proprieta' di linearita' del valore atteso si ha

$$\mathbb{E}(X \cdot Y) = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \mathbb{E}\left(\mathbf{1}_{E_{h,k}}\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k P\left(\{X = x_h\} \cap \{Y = y_k\}\right). \tag{47}$$

Si noti inoltre che fino a questo punto non si e' usata l'ipotesi di indipendenza, mentre ora, tenendo conto del fatto che, per l'indipendenza $P\left(\{X = x_h\} \cap \{Y = y_k\}\right) = P\left(\{X = x_h\}\right) \cdot P\left(\{Y = y_k\}\right)$, si ottiene[45]

$$\begin{aligned}\mathbb{E}\left(X \cdot Y\right) &= \sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \cdot P\left(X = x_h\right) \cdot P(Y = y_k) \\ &= \sum_{h=1}^{n} x_h \cdot P\left(X = x_h\right) \cdot \sum_{k=1}^{m} y_k \cdot P(Y = y_k) = \mathbb{E}(X) \cdot \mathbb{E}(Y),\end{aligned}$$

e la dimostrazione e' 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 definite su uno stesso spazio di probabilita' finito, e sia data la funzione $h: \mathbb{R} \to \mathbb{R};\ x \mapsto h(x)$. Allora

$$\mathbb{E}(h(X)) = \sum_{j=1}^{n} h(x_j) \cdot P\big(\{X = x_j\}\big). \tag{48}$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilita' finito, e sia data la

[45]Se i passaggi che seguono risultassero difficili, si noti che

$$\begin{aligned}&\sum_{h=1}^{n}\sum_{k=1}^{m} x_h \cdot y_k \cdot P\left(X = x_h\right) \cdot P(Y = y_k) = \sum_{h=1}^{n}\left(\sum_{k=1}^{m} x_h \cdot y_k \cdot P\left(X = x_h\right) \cdot P(Y = y_k)\right) \\ &= \sum_{h=1}^{n} x_h \cdot P\left(X = x_h\right)\left(\sum_{k=1}^{m} y_k \cdot P(Y = y_k)\right) = \sum_{h=1}^{n} x_h \cdot P\left(X = x_h\right)\mathbb{E}(Y) = \mathbb{E}(Y)\sum_{h=1}^{n} x_h \cdot P\left(X = x_h\right) = \mathbb{E}(Y)\mathbb{E}(X)\end{aligned}$$

funzione $g: \mathbb{R}^2 \to \mathbb{R};\ (x,y) \mapsto g(x,y)$. Allora

$$\mathbb{E}\left(g(X,Y)\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h,y_k) P\left(\{X=x_h\}\cap\{Y=y_k\}\right) \tag{49}$$

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 e' necessario calcolare la sua distribuzione di probabilita'**, ma basta applicare la (48) o la (49) rispettivamente.

Dimostrazione. La dimostrazione di (48) non e' altro che l'immediata estensione della dimostrazione della ***Proposizione 7***: infatti, con le stesse notazioni usate nel (*II modo*) di tale dimostrazione, basta notare[46] 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 linearita'

$$\mathbb{E}(h(X)) = \sum_{j=1}^{n} h(x_j)\mathbb{E}(X_j) = \sum_{j=1}^{n} h(x_j) P(E_j) = \sum_{j=1}^{n} h(x_j) P(\{X = x_j\}).$$

La dimostrazione di (49) e' 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 proprieta' di linearita' del valore atteso si ha

$$\mathbb{E}\big(g(X,Y)\big) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h,y_k)\mathbb{E}\left(\mathbf{1}_{E_{h,k}}\right) = \sum_{h=1}^{n}\sum_{k=1}^{m} g(x_h,y_k) P\left(\{X=x_h\}\cap\{Y=y_k\}\right). \tag{50}$$

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 probabilita', e siano $h_i: \mathbb{R} \to \mathbb{R};\ x \mapsto h_i(x)$, per $i = 1, 2$ due funzioni reali, allora

$$\mathbb{E}\left(h_1(X)\cdot h_2(Y)\right) = \mathbb{E}\left(h_1(X)\right)\cdot\mathbb{E}\left(h_2(Y)\right).$$

Dimostrazione. Basta prendere $g(x,y) = h_1(x)\cdot h_2(y)$ in (50) ottenendo cosi'

$$\mathbb{E}\big(h_1(X)\cdot h_2(Y)\big) = \sum_{h=1}^{n}\sum_{k=1}^{m} h_1(x_h)\cdot h_2(y_k)\cdot P\left(\{X=x_h\}\cap\{Y=y_k\}\right)$$

[46]Ovviamente se $\omega \in E_j = \{X = x_j\}$, ovvero e' tale che $X(\omega) = x_j$, allora $W(\omega) = h(X(\omega)) = h(x_j)$.

e sfruttando l'indipendenza tra X ed Y

$$= \sum_{h=1}^{n} \sum_{k=1}^{m} h_1(x_h) \cdot h_2(y_k) \cdot P\left(\{X = x_h\}\right) \cdot P\left(\{Y = y_k\}\right).$$

A questo punto la dimostrazione e' 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 gia' 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 e' necessario calcolare la sua distribuzione di probabilita'. Tuttavia questo problema e' un problema interessante, e abbiamo gia' 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 variabili aleatorie.

Proposizione 12 Sia X una variabile aleatoria definite su uno stesso spazio di probabilita' finito, e sia data la funzione $h:\ \mathbb{R} \to \mathbb{R};\ x \mapsto h(x)$. Allora, posto $W = h(X)$, e $W(\Omega) = \{w_1, w_2, \ldots, w_\ell\}$ si ha

$$\mathbb{P}(W = w_k) = \mathbb{P}(h(X) = w_k) = \sum_{j:\ h(x_j) = w_k} P\big(\{X = x_j\}\big), \qquad k = 1, 2, \ldots, \ell \tag{51}$$

Siano X, Y due variabili aleatorie definite su uno stesso spazio di probabilita' finito, e sia data la funzione $g:\ \mathbb{R}^2 \to \mathbb{R};\ (x,y) \mapsto g(x,y)$. Allora, posto $Z = g(X,Y)$ e $Z(\Omega) = \{z_1, z_2, \ldots, z_r\}$ si ha

$$\mathbb{P}\left(g(X,Y) = z_j\right) = \sum_{h,k:\ g(x_h,y_k) = z_j}^{1 \le h \le n,\ 1 \le j \le m} \sum P\left(\{X = x_h\} \cap \{Y = y_k\}\right), \qquad j = 1, 2, \ldots, r. \tag{52}$$

9.1 Esercizi di verifica

Esercizio 9.1. Sia X una variabile aleatoria tale che

$$P(X = -1) = q, \qquad P(X = 1) = p$$

con $0 < q = 1 - p < 1$. Calcolate $\mathbb{E}\left(X\right)$.

Esercizio 9.2. Sia X una variabile aleatoria a valori nell'insieme $\{1, 2, ..., n\}$ con distribuzione di probabilita' data dalla posizione

$$P(\{X = k\}) \propto k, \qquad k = 1, ..., n.$$

Calcolate[47] $P(\{X = k\})$, per $k = 1, ..., n$, e $\mathbb{E}(X)$.

[47] Si ricordi che

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \qquad \sum_{k=1}^{n} k^2 = \frac{1}{3} n (n + \frac{1}{2})(n+1)$$

Esercizio 9.3. Calcolate il valore atteso di una variabile aleatoria X a valori nell'insieme

$$\{0, 1, 2, ..., n\},$$

sapendo che risulta

$$P(\{X = k\}) \propto \frac{\rho^k}{k!(n-k)!}, \qquad k = 0, 1, ..., n.$$

con ρ costante positiva assegnata.

Esercizio 9.4. Siano $g_1, ..., g_n$ dei numeri positivi assegnati e $x_1 < ... < x_n$ assegnati numeri reali. Sia X una variabile aleatoria a valori nell'insieme $\{x_1, ..., x_n\}$ e tale che

$$P(\{X = x_k\}) \propto g_k, \qquad k = 1, ..., n.$$

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 probabilita' di vincere almeno un premio, cioe' $P(X > 0)$.
(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 probabilita' 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

$$\mathbb{E}\left(X_r^2 - X_r \cdot X_s\right).$$

Esercizio 9.8. Renato e Stefano lanciano una moneta perfetta n volte ciascuno ed il vincitore e' 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, cioe' 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 $n = 2, 3, ...,$

$$2^n - 1 = \sum_{k=0}^{n-1} 2^k$$

e riformulate piu' in generale il testo e la soluzione dell'Esempio 9.7 (sostituendo un valore n generico al posto di $n = 5$).

10 Varianza, Covarianza e comportamento delle medie aritmetiche di variabili aleatorie

Nella precedente lezione abbiamo visto che e' spesso interessante instaurare un confronto fra due variabili aleatorie che ammettano uguale valore atteso. Una nozione utile a tale riguardo e' quella di ***varianza***.

In questa lezione introdurremo dunque il concetto di varianza di una variabile aleatoria e ne vedremo alcune proprieta' fondamentali; verra' introdotto, a tale proposito, anche il concetto di covarianza di una coppia di variabili aleatorie.

Come si vedra', 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* $X : \Omega \equiv \{\omega_1, ..., \omega_N\} \to X(\Omega) \equiv \{x_1, ..., x_n\}$ *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* $Var(X)$*, cioe'*

$$Var\,(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \equiv \sum_{\omega_i \in \Omega} p(\omega_i)\,(X\,(\omega_i) - \mu)^2\,.$$

Le proprieta' fondamentali di tale definizione sono elencate qui di seguito.

Proposizione 1. La varianza di una variabile aleatoria X e' 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 e' una variabile aleatoria degenere[48].

La dimostrazione e' immediata e viene lasciata quale esercizio.

Proposizione 2. Per ogni variable aleatoria X

$$Var\,(X) = \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \tag{53}$$

Dimostrazione.

Applicando la proprieta' di linearita' del valore atteso, si ha immediatamente

$$Var\,(X) = \mathbb{E}\left(X^2 - 2\mu X + \mu^2\right) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 = \mathbb{E}(X^2) - \mu^2.$$

---

[48] Nel caso in cui Ω e' 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 $Var\,(X) = 0$ se e solo se esiste un $\hat{x} \in X(\Omega)$ tale che $P(X = \hat{x}) = 1$. In tal caso $\hat{x}$ coincide con il valore atteso.

Infatti, posto $\mathbb{E}(X) = \mu$, e

$$Var\,(X) = \sum_{i=1}^{N} p(\omega_i)\big(X(\omega_i) - \mu\big)^2 = 0$$

se e solo se

$$p(\omega_i)\big(X(\omega_i) - \mu\big)^2 = 0,$$

in quanto una somma di termini non negativi e' nulla se e solo se tutti gli addendi sono nulli; nel qual caso

$$\begin{cases} \text{o } X(\omega_i) - \mu = 0 & \text{e allora puo' essere sia } p(\omega_i) > 0, \text{ sia } p(\omega_i) = 0, \\ \text{o } X(\omega_i) - \mu \neq 0 & \text{e allora necessariamente } p(\omega_i) = 0. \end{cases}$$

Basta poi osservare che $P(\{X(\omega_i) \neq \mu\}) = \sum_{i:\, X(\omega_i) - \mu \neq 0} p(\omega_i) = 0$, e che cio' equivale a $P(\{X(\omega_i) = \mu\}) = 1$.

Un modo equivalente di esprimere la precedente proprieta' e' il seguente: $Var(X) = 0$ se e solo se X coincide con la variabile aleatoria degenere identicamente uguale a $\mu := \mathbb{E}(X)$ con probabilita' 1.

Anche per le seguenti due proprieta' le dimostrazioni sono immediate, tenendo conto ad esempio di (53), e vengono lasciate per esercizio.

Proposizione 3. Sia $Y = X + b$, essendo $b \in \mathbb{R}$. Allora

$$Var\,(Y) = Var\,(X)\,.$$

Proposizione 4. Sia $Y = a \cdot X$, essendo $a \in \mathbb{R}$. Allora

$$Var\,(Y) = a^2 \cdot Var\,(X)\,.$$

Proposizione 5. Siano X, Y due variabili aleatorie definite su $(\Omega, \mathcal{P}(\Omega))$. Allora

$$Var\,(X+Y) = Var\,(X) + Var\,(Y) + 2\,[\mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)]\,. \tag{54}$$

Dimostrazione.

Anche in questo caso la dimostrazione e' immediata; ricordando la (53) e applicando la proprieta' di linearita' del valore atteso, si ha infatti

$$\begin{aligned}
Var\,(X+Y) &= \mathbb{E}\left[(X+Y)^2\right] - [\mathbb{E}\,(X+Y)]^2 \\
&= \mathbb{E}\left(X^2 + 2XY + Y^2\right) - \left[\mathbb{E}(X)^2 + 2\mathbb{E}(X)\cdot\mathbb{E}(Y) + \mathbb{E}(Y)^2\right] \\
&= \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 + \mathbb{E}\left(Y^2\right) - \mathbb{E}(Y)^2 + 2\,[\mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)] \\
&= Var\,(X) + Var\,(Y) + 2\,[\mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)]\,.
\end{aligned}$$

Osservazione 1. Come il valore atteso (X), anche la varianza $Var\,(X)$ dipende soltanto dalla distribuzione di probabilita' 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 anche[49] *la **Proposizione 2** della precedente Lezione 9*

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X) - \mathbb{E}(X)^2 = \mathbb{E}(X)\,(1 - \mathbb{E}(X)) = p\,(1-p)\,.$$

Esempio 10.2. *Consideriamo una variabile aleatoria X tale*

$$P(X = -1) = q, \qquad P(X = 1) = p$$

con $0 < q = 1 - p < 1$. Allora si ha $\mathbb{E}\,(X) = 2p - 1$ e, osservando che X^2 e' una variabile aleatoria con distribuzione degenere su 1,

$$Var\,(X) = \mathbb{E}\left[X^2\right] - (2p-1)^2 = 4p\,(1-p)\,.$$

Consideriamo ora, per un dato $a > 0$, la variabile $W = a \cdot X$:

$$P(W = -a) = q, \qquad P(W = a) = p;$$

[49] Si 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

$$\mathbb{E}(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p.$$

allora W^2 e' una variabile aleatoria con distribuzione degenere su a^2 e si ha

$$\mathbb{E}(W) = a\mathbb{E}(X) = a(2p-1);$$

$$Var(W) = a^2 - a^2(2p-1)^2 = a^2 \cdot 4p(1-p) = a^2 \cdot Var(X),$$

come del resto era ovvio, tenendo conto della Proposizione 4.

Definizione 10.2 (Covarianza). *Siano date, su uno stesso spazio di probabilita', due variabile aleatorie X ed Y e, per comodita', indichiamo brevemente con μ_X e μ_Y i loro rispettivi valori attesi. Si definisce* **covarianza** *fra X, Y la quantita'*

$$Cov(X,Y) \equiv \mathbb{E}[(X-\mu_X)\cdot(Y-\mu_Y)] \Big(= \mathbb{E}\big[\big(X-\mathbb{E}(X)\big)\cdot\big(Y-\mathbb{E}(Y)\big)\big]\Big). \tag{55}$$

Osservazione 2. Si osservi che se $Y = X$, allora $Cov(X,Y = Cov(X,X) = Var(X)$. Inoltre, svolgendo il prodotto nel secondo membro della (55) ed applicando la proprieta' di linearita' del valore atteso, si ottiene immediatamente

$$Cov(X,Y) = \mathbb{E}(X \cdot Y) - \mu_X \cdot \mu_Y \Big(= \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y)\Big). \tag{56}$$

Dunque, ricordando (54), possiamo scrivere

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y). \tag{57}$$

Esercizio proposto 10.1. *Verificare*[50] *che, date tre variabili aleatorie X, Y, Z e due costanti α, β risulta*

$$Cov(X, \alpha Y + \beta Z) = \alpha Cov(X,Y) + \beta Cov(X,Z).$$

Come corollario della ***Proposizione 9*** della precedente Lezione 9, ed in virtu' della (57), otteniamo immediatamente

Proposizione 6. Siano date X ed Y due variabili aleatorie definite su uno stesso spazio di probabilita'. Se X ed Y sono stocasticamente indipendenti, allora

$$Var(X+Y) = Var(X) + Var(Y).$$

Dimostrazione. Basta ricordare la (56) e che, per la ***Proposizione 9*** della Lezione 9, se X ed Y sono indipendenti allora $\mathbb{E}(X\cdot Y) = \mathbb{E}(X)\cdot\mathbb{E}(Y)$.

[50] Per la simmetria della covarianza, la proprieta' che

$$Cov(X, \alpha Y_1 + \beta Y_2) = \alpha Cov(X, Y_1) + \beta Cov(X, Y_2).$$

ovviamente vale anche per

$$Cov(aX_1 + bX_2, Y) = aCov(X_1, Y) + bCov(X_2, Y).$$

e insieme sono note come la proprieta' di bilinearita' della covarianza:

$$Cov(aX_1 + bX_2, \alpha Y_1 + \beta Y_2) = a\cdot\alpha Cov(X_1,Y_1) + a\cdot\beta Cov(X_1,Y_2) + b\cdot\alpha Cov(X_2,Y_1) + b\cdot\beta Cov(X_2,Y_2).$$

Osservazione 3. Siano X ed Y due variabili aleatorie binarie definite su uno stesso spazio di probabilita'. Osserviamo che anche il loro prodotto[51] $X \cdot Y$ e' 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

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) \tag{58}$$

implica in tale caso che X ed Y sono stocasticamente indipendenti[52] 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, cioe' per coppie di variabili non entrambe binarie, ***la condizione (58) (ovvero $Cov(X, Y) = 0$) e' 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 $Y \equiv X^2$, cosicche' $X \cdot Y = X^3$ e ovviamente[53] X, Y non possono essere stocasticamente indipendenti; notiamo che, in questo caso, X^3 ha la stessa distribuzione di probabilita' 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 probabilita', sono* **non correlate** *(o anche* **scorrelate***) se risulta verificata la condizione (58).*

---

[51] A questo proposito e' 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 $\omega \in A \cap B$.

[52] Si osservi che, se come prima $X = \mathbf{1}_A$ ed $Y = \mathbf{1}_B$ allora la distribuzione congiunta di X e di Y e' 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 $B = \{Y = 1\}$, e' sufficiente per ottenere l'indipendenza delle variabili aleatorie X ed Y.

[53] Il fatto che X ed $Y = X^3$ non siano indipendenti si vede immediatamente: ad esempio $P(Y = 0) = P(X^3 = 0) = P(X = 0) = 1/3 > 0$, e quindi $P(Y = 0) \cdot P(X = 1) = (1/3) \cdot (1/3) > 0$, mentre $P(\{Y = 0\} \cap \{X = 1\}) = P(\emptyset) = 0$.

ATTENZIONE ORA, DA QUI IN POI, LA NUMERAZIONE DELLE DEFINIZIONI SARA' CAMBIATA!!!!

Prima di proseguire e' utile generalizzare la relazione (57) che permette di calcolare la varianza della somma di due variabili aleatorie al caso della somma di un numero finito $X_1, X_2, \ldots, X_n$ di variabili aleatorie:

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{h=1}^{n}\sum_{k=1}^{n} Cov(X_h, X_k) \tag{59}$$

$$= \sum_{k=1}^{n} Var(X_k) + \sum_{h\neq h}^{1\leq h,k\leq n}\sum Cov(X_h, X_k) \tag{60}$$

$$= \sum_{k=1}^{n} Var(X_k) + 2\sum_{h=1}^{n-1}\sum_{k=h+1}^{n} Cov(X_h, X_k). \tag{61}$$

Bastera' mostrare (59), in quanto le (60) e (61), sono solo forme differenti della prima espressione. Tuttavia rimandiamo la dimostrazione della (59), 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 $Var(X)$?*

Soluzione. Conviene ragionare lungo la linea gia' svolta nella soluzione dell'Esempio 9.3 della precedente lezione, cioe' *riguardiamo*[54] X come la somma di n variabili aleatorie binarie indipendenti $X_1, ..., X_n$, ciascuna di valore atteso θ e di varianza $\theta \cdot (1-\theta)$; dunque $Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) = n\theta \cdot (1-\theta)$, per la (60), in quanto $Cov(X_h, X_k) = 0$, 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 $X_1, ..., X_n$, ciascuna di valore atteso $\frac{m_1}{M}$ e dunque di varianza $\frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right)$.*

Nel presente caso pero' $X_1, ..., X_n$ non sono indipendenti; e' chiaro che esse, prese a due a due, esse hanno una stessa covarianza, cioe' $Cov(X_h, X_k) = Cov(X_1, X_2)$ e quindi possiamo scrivere

$$Var(X) = n \cdot \frac{m_1}{M} \cdot \left(1 - \frac{m_1}{M}\right) + n \cdot (n-1) \cdot Cov(X_1, X_2).$$

Ora, essendo X_1, X_2 variabili aleatorie binarie, si ha

$$\begin{aligned} Cov(X_1, X_2) &= P(\{X_1 = 1\} \cap \{X_2 = 1\}) - P(\{X_1 = 1\}) \cdot P(\{X_2 = 1\}) \\ &= P(\{X_1 = 1\}) \cdot P(\{X_2 = 1\} | \{X_1 = 1\}) - \left(\frac{m_1}{M}\right)^2 \\ &= \frac{m_1}{M} \cdot \frac{m_1 - 1}{M - 1} - \left(\frac{m_1}{M}\right)^2 = -\frac{m_1}{M} \cdot \frac{M - m_1}{M(M-1)} \end{aligned}$$

[54] Riguardare X come la somma di n variabili aleatorie binarie indipendenti (e quindi non correlate: $Cov(X_i, X_j) = 0$ per $i \neq j$) $X_1, ..., X_n$, 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

$$Var(X) = Var(S) = Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{h\neq k}^{1\leq h,k\leq n}\sum Cov(X_h, X_k) = \sum_{i=1}^{n} Var(X_i) = n\theta(1-\theta).$$

e possiamo concludere scrivendo

$$Var(X) = n \cdot \frac{m_1}{M} \cdot \frac{M - m_1}{M} \left(1 - \frac{n-1}{M-1}\right) = np(1-p)\left(1 - \frac{n-1}{M-1}\right),$$

dove si e' posto $p = \frac{m_1}{M}$, la percentuale di palline bianche presenti nell'urna[55].

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 $Cov\,(X_h, X_k)$, potremmo procedere semplicemente come segue. Innanzitutto, come nel precedente esempio, possiamo osservare che $Cov\,(X_h, X_k)$ non dipende dalla coppia di indici h, k (purche' $h \neq k$) e da cio' segue*

$$Var\,(S_n) = n \cdot Var\,(X_1) + n(n-1)Cov\,(X_1, X_2)\,;$$

si ha d'altra parte che la distribuzione di probabilita' di S_n e' degenere; ne segue

$$Var\,(S_n) = 0.$$

Possiamo dunque concludere

$$Cov\,(X_1, X_2) = -\frac{Var\,(X_1)}{n-1}.$$

Consideriamo ora n variabili aleatorie $X_1, ..., X_n$ (non necessariamente binarie) e, per semplicita', le assumiamo tali che

$$\mathbb{E}(X_1) = ... = \mathbb{E}(X_n) = \mu, \tag{62}$$

$$Var(X_1) = ... = Var(X_n) = \sigma^2 \tag{63}$$

$$Cov(X_h, X_k) = \varphi, \qquad 1 \leq h \neq k \leq 1. \tag{64}$$

Che cosa possiamo dire circa la loro ***media aritmetica***

$$Y_n \equiv \frac{1}{n}\sum_{h=1}^{n} X_h? \tag{65}$$

A tale proposito e' immediato verificare la seguente proposizione (bastera' tener presente sia la linearita' del valore atteso che le regole viste qui sopra circa il calcolo della varianza).

Proposizione 7. Siano date n variabili aleatorie $X_1, ..., X_n$, che verificano (62), (63) e (64). Si ha

$$\mathbb{E}(Y_n) = \mu$$

e

$$Var(Y_n) = \frac{1}{n}\left[\sigma^2 + (n-1)\varphi\right],$$

dove Y_n e' definito come sopra in (65).

Osservazione 5. La varianza $Var(X)$ di una variabile aleatoria X e' un indice del grado di dispersione della distribuzione di probabilita' rispetto al valore atteso; e' chiaro dalla definizione di

[55] Si noti il fatto che quindi se il numero totale M delle palline presenti nell'urna e' molto grande rispetto al numero n delle estrazioni, allora la varianza del numero di palline estratte (***senza reinserimento***) dall'urna e' molto vicina alla varianza del numero delle palline estratte nel caso in cui le estrazioni siano ***con reinserimento***.

varianza che, a parita' di valore atteso μ, scarti grandi (in modulo) di X rispetto a μ sono tanto piu' probabili quanto piu' risulta grande $Var(X)$.

A tale proposito e' utile osservare quanto segue: cosi' come $\mathbb{E}(X)$ corrisponde al concetto di baricentro quando le probabilita' $p_j = P(X = x_j)$ vengono interpretate come delle masse concentrate sui diversi punti $x_1, ..., x_n$, analogamente $Var(X)$, nella stessa interpretazione, corrisponde al concetto di ***momento di inerzia*** della distribuzione stessa.METTERE UNA FOOTNOTE??

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 probabilita' e se ne possono ricavare alcune utili disuguaglianze; in particolare si ha la seguente[56]

Proposizione 8. **(Diseguaglianza di Chebyshev)**. Sia X una variabile aleatoria, con valore atteso uguale a μ e varianza uguale a σ^2. Allora, $\forall \varepsilon > 0$,

$$P(\{|X - \mu| > \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Dimostrazione.

La dimostrazione segue immediatamente dalla definizione stessa di varianza, infatti possiamo scrivere

$$\begin{aligned}\sigma^2 = \sum_{\omega_i \in \Omega} p(\omega_i) \cdot (X(\omega_i) - \mu)^2 &\geq \sum_{\omega_i : |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot (X(\omega_i) - \mu)^2 \geq \\ &\geq \sum_{\omega_i : |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) \cdot \varepsilon^2 = \varepsilon^2 \cdot \sum_{\omega_i : |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) = \varepsilon^2 \cdot P(\{|X - \mu| > \varepsilon\}).\end{aligned}$$

La dimostrazione e' 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'e' il segno di minore uguale e non di minore stretto, in quanto e' possibile che $\sum_{\omega_i : |X(\omega_i) - \mu| > \varepsilon} p(\omega_i) = 0$, nel qual caso si ha che i membri della diseguaglianza sono entrambi uguali a zero.

Proposizione 9. Siano date n variabili aleatorie $X_1, ..., X_n$ ed indichiamo con Y_n la loro media aritmetica. Se $X_1, ..., X_n$ verificano (62), (63) e (64), ovvero sono tali che

$$\begin{aligned}\mathbb{E}(X_1) = ... = \mathbb{E}(X_n) = \mu,\\ Var(X_1) = ... = Var(X_n) = \sigma^2\\ Cov(X_h, X_k) = 0, \qquad 1 \leq h \neq k \leq 1,\end{aligned}$$

allora, qualunque sia $\varepsilon > 0$

$$P(\{|Y_n - \mu| > \varepsilon\}) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}.$$

[56] Si 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 e' interessante solo se $\frac{\sigma^2}{\varepsilon^2} < 1$: in caso contrario si ottiene solo una banalita' (ovvero che $P(\{|X - \mu| > \varepsilon\})$ e' minore o uguale di un numero strettamente maggiore di 1, il che e' ovvio in quanto $P(\{|X - \mu| > \varepsilon\})$ e' un numero minore o uguale ad 1).

Il lettore piu' attento puo' notare che la dimostrazione rimane invariata se si considera la probabilita' dell'evento $\{|X - \mu| \geq \varepsilon\}$, per cui vale anche

$$P(\{|X - \mu| \geq \varepsilon\}) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Dimostrazione.
La dimostrazione segue immediatamente ricordando le precedenti Proposizioni 7 ed 8.

Osservazione 6 La tesi della ***Proposizione 9*** si puo' 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 puo' dire che l'evento "*la media aritmetica Y_n di X_1, X_2,..., X_n, differisce dal valore atteso μ (comune a tutte le v.a. X_i) meno di ε*" ha probabilita' maggiore o uguale a $1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}$. Se n e' "*molto grande*", in modo che $1 - \frac{1}{n}\frac{\sigma^2}{\varepsilon^2}$ sia "*vicino*" ad 1, la tesi si puo' parafrasare anche dicendo che, se n e' "*molto grande*", media aritmetica e valore atteso differiscono tra loro meno di ε, con probabilita' "*vicina*" ad 1.

Piu' interessante ancora, dal punto di vista applicativo, e' il fatto che siamo in grado di rispondere ***alla domanda***:
Quante prove si devono effettuare, ovvero quanto si deve prendere grande n, affinche', con probabilita' almeno $1 - \delta$, la media aritmetica differisca dal valore atteso μ meno di ε?

La risposta alla precedente domanda e' molto semplice: ***e' 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 e' 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 $\sigma^2 = \theta(1-\theta) = \frac{5}{9}\frac{4}{9}$, e infine $\varepsilon = \frac{1}{30}$, di conseguenza

$$P\left(\left\{\left|\frac{S_n}{n} - \frac{5}{9}\right| > \frac{1}{30}\right\}\right) \leq \frac{1}{n}\frac{\sigma^2}{\varepsilon^2} = \frac{1}{n}\frac{\theta(1-\theta)}{\varepsilon^2} = \frac{1}{n}\frac{\frac{5}{9}\frac{4}{9}}{\left(\frac{1}{30}\right)^2} = \frac{1}{n}\frac{5\cdot 4\cdot 3^2\cdot 10^2}{9\cdot 9} = \frac{1}{n}\frac{2000}{9} \leq \frac{1}{10}$$

$$\Updownarrow$$

$$n \geq \frac{20000}{9} \simeq 2222,22$$

Definizione 10.4 (Variabili aleatorie standard). *Una variabile aleatoria Z si dice* **standard** *quando*

$$\mathbb{E}(Z) = 0, \qquad Var(Z) = 1.$$

Come applicazione della disuguaglianza di Chebyshev, si ha che, se Z e' una variabile aleatoria standard allora[57] , $\forall \varepsilon > 0$,

$$P(\{|Z| > \varepsilon\}) \leq \frac{1}{\varepsilon^2}. \tag{66}$$

Data una variabile aleatoria X, con valore atteso μ e varianza σ^2, possiamo costruire una variabile aleatoria standard Z, funzione[58] di X, ponendo

$$Z = \frac{X-\mu}{\sigma} \left(= \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}} \right)$$

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 puo' dire che una variabile aleatoria X, con valore atteso μ e varianza σ^2 si puo' sempre scrivere nella forma

$$X = \sigma Z + \mu,$$

essendo Z una variabile aleatoria standard; e inoltre la (66) diviene

$$P\left(\left\{\left|\frac{X-\mu}{\sigma}\right| > \varepsilon\right\}\right) \leq \frac{1}{\varepsilon^2}.$$

Consideriamo ora n prove bernoulliane $E_1, ..., E_n$ di probabilita' θ.

Indichiamo con $X_1, ..., X_n$ gli indicatori di $E_1, ..., E_n$ e con $S_n = \sum_{j=1}^n X_j$ il numero di successi sulle n prove, e quindi la loro media aritmetica

$$Y_n = \frac{S_n}{n},$$

e' 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

[57] Ovviamente la (66) ha interesse solo se $\frac{1}{\varepsilon^2} \leq 1$.

[58] Si 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 $\left(\frac{1}{\sigma}, -\frac{\mu}{\sigma}\right)$ e' l'unica coppia di valori (a,b) per cui $aX + b$ e' una variabile aleatoria standard.

Proposizione 10. Per ogni $\varepsilon > 0$

$$P\Big(\{|\sqrt{n}\frac{Y_n - \theta}{\sqrt{\theta\,(1-\theta)}}| > \varepsilon\}\Big) \leq \frac{1}{\varepsilon^2}.$$

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

$$\mathbb{E}(Y_n) = \theta, \qquad Var(Y_n) = \frac{\theta\,(1-\theta)}{n}$$

e quindi applicare la (66) 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 probabilita' di differire dal valore atteso (cioe' zero) piu' di ε non si puo' rendere piccola (nemmeno prendendo n grande), come invece accade per la variabile Y_n. Questa proprieta' e' connessa con il Teorema Centrale del Limite, che verra' discusso piu' avanti.

10.1 Diseguaglianza di Cauchy e coefficiente di correlazione

Siano date due variabili aleatorie X ed Y, vale allora la seguente diseguaglianza

$$|Cov(X,Y)| \leq \sqrt{Var(X)} \cdot \sqrt{Var(Y)}. \tag{67}$$

Tale diseguaglianza e' nota come ***diseguaglianza di Cauchy*** e generalizza la diseguaglianza di Cauchy-Schwartz per vettori, ovvero, se $\boldsymbol{u}, \boldsymbol{u}$ sono vettori di $\mathbb{R}^n$, allora

$$|<\boldsymbol{u},\boldsymbol{v}>| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|\Big(\sqrt{u_1^2+u_2^2+\cdots+u_n^2}\cdot\sqrt{v_1^2+v_2^2+\cdots+v_n^2}\Big),$$

dove con $<\boldsymbol{u},\boldsymbol{v}> = \sum_{i=1}^n u_i \cdot v_i$ si indica il prodotto scalare tra i vettori $\boldsymbol{u}$ e $\boldsymbol{v}$, e con $\|\boldsymbol{u}\| = \sqrt{u_1^2+u_2^2+\cdots+u_n^2}$ si indica il modulo del vettore $\boldsymbol{u}$.

Infatti si consideri il caso particolare in cui $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ ed $Y(\Omega) = \{y_1, y_2, \ldots, y_n\}$, e si abbia

$$P(X = x_i, Y = y_i) = \frac{1}{n}, \quad i = 1, 2, \ldots, n$$
$$P(X = x_i, Y = y_j) = 0, \qquad i \neq j, i, j = 1, 2, \ldots, n.$$

Allora, posto $\mu = \mathbb{E}(X)$ e $\nu = \mathbb{E}(Y)$, la diseguaglianza (67) 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 e' esattamente la diseguaglianza di Cauchy-Schwartz per i vettori $\boldsymbol{u}$ e $\boldsymbol{v}$ con $u_i = x_i - \mu$ e $v_i = y_i - \nu$, per $i = 1, 2, \ldots, n$, a parte per il fattore $\frac{1}{n}$.

La dimostrazione della diseguaglianza (67) di Cauchy e' basata sull'osservazione che la funzione

$$\varphi(x) := \mathbb{E}\big[\big((X-\mu)-x(Y-\nu)\big)^2\big]$$

gode di due proprieta':
$\varphi(x) \geq 0$ per ogni x, in quanto valore atteso di una variabile aleatoria non negativa,
$\varphi(x) = Var(X) - 2x\,Cov(X,Y) + x^2\,Var(Y)$, come si vede subito per la linearita' del valore atteso e considerando che

$$\big((X-\mu)-x(Y-\nu)\big)^2 = (X-\mu)^2 - 2x(X-\mu)(Y-\nu) + x^2(Y-\nu)^2.$$

Di conseguenza $\varphi(x) = a\,x^2 + b\,x + c$, con $a = Var(Y)$, $b = -2Cov(X,Y)$ e $c = Var(X)$, ed il discriminante $b^2 - 4ac = 4\big(Cov(X,Y)\big)^2 - 4Var(X)Var(Y)$ e' minore o uguale a zero, ovvero

$$\big(Cov(X,Y)\big)^2 \leq Var(X)Var(Y).$$

Per ottenere la (67) basta estrarre la radice quadrata.

Come conseguenza il rapporto

$$\rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

che e' detto ***coefficiente di correlazione*** tra X ed Y e' sempre minore o uguale ad 1 in valore assoluto.

Seguitando l'analogia con il caso vettoriale, in qualche senso il coefficiente di correlazione $\rho_{X,Y}$ generalizza il coseno tra due vettori, infatti e' noto che, se $\widehat{\boldsymbol{uv}}$ indica l'angolo formato fra due vettori,

$$\cos(\widehat{\boldsymbol{uv}}) = \frac{<\boldsymbol{u},\boldsymbol{v}>}{\|\boldsymbol{u}\|\cdot\|\boldsymbol{v}\|}.$$

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 $\cos(\widehat{\boldsymbol{uv}}) = \pm 1$, nel caso in cui $\boldsymbol{v} = \alpha\boldsymbol{u}$.

10.2 Appendice: Covarianza della somma di n variabili aleatorie

Se $X_1, X_2, \ldots, X_n$ sono n variabili aleatorie, definite sullo stesso spazio di probabilita', allora

$$\begin{aligned} Var\left(\sum_{k=1}^n X_k\right) &= \sum_{h=1}^n\sum_{k=1}^n Cov(X_h, X_k) \\ &= \sum_{k=1}^n Var(X_k) + \sum\sum_{h\neq k}^{1\leq h,k\leq n} Cov(X_h,X_k) \\ &= \sum_{k=1}^n Var(X_k) + 2\sum_{h=1}^{n-1}\sum_{k=h+1}^{n} Cov(X_h,X_k). \end{aligned}$$

Come gia' osservato basta mostrare la prima uguaglianza, in quanto le altre due sono solo forme differenti della prima espressione.

Per iniziare si ponga per semplicita' $\mathbb{E}(X_i) = \mu_i$, cosi' $\mathbb{E}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mu_i$. 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 da fatto che, come e' facile convincersi, in generale

$$
\begin{aligned}
(a_1 + a_2 + \cdots + a_n) \cdot (a_1 + a_2 + \cdots + a_n) = {} & a_1 \cdot a_1 + a_1 \cdot a_2 + \quad \cdots \quad + a_1 \cdot a_n \\
& + a_2 \cdot a_1 + a_2 \cdot a_2 + \quad \cdots \quad + a_2 \cdot a_n \\
& + \ \cdots \ + \ \cdots \ + \ \cdots \ + \ \cdots \\
& + a_n \cdot a_1 + a_n \cdot a_2 + \quad \cdots \quad + a_n \cdot a_n \\
& = \sum_{h=1}^{n} \sum_{k=1}^{n} a_h \cdot a_k .
\end{aligned}
$$

A questo punto basta passare al valore atteso e sfruttarne la proprieta' di linearita':

$$
\begin{aligned}
Var\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right)^2\right] \\
&= \mathbb{E}\left[\sum_{h=1}^{n} \sum_{k=1}^{n} (X_h - \mu_h)(X_k - \mu_k)\right] \\
&= \sum_{h=1}^{n} \sum_{k=1}^{n} \mathbb{E}\left[(X_h - \mu_h)(X_k - \mu_k)\right] \\
&= \sum_{h=1}^{n} \sum_{k=1}^{n} Cov(X_h, X_k).
\end{aligned}
$$

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 diseguaglianza di Chebyshev alla variabile aleatoria S_{100}?

Esercizio 10.5. Il Dipartimento di Matematica acquista 20 copie di un software; ciascuna copia ha probabilita' $\frac{1}{100}$ 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 *diseguaglianza 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

$$|Cov(X, Y)| \leq 1.$$

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 puo' risultare interessante per varie applicazioni e per le connessioni con problematiche di tipo statistico; essa permettera' 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 e' gia' 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 sara' dunque una variabile aleatoria a valori in $\{0, 1, ..., M\}$.

Lo stato di informazione su R viene descritto attraverso la distribuzione di probabilita':

$$p_R(r) \equiv P(\{R = r\}), \qquad r = 0, 1, ..., M. \tag{68}$$

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 e', in generale, una variabile aleatoria; la distribuzione di probabilita' di S, condizionata al valore assunto da R, e' data dalle probabilita' condizionate

$$p_S(s|\{R = r\}) \equiv P(\{S = s\}|\{R = r\}), \qquad s = 0, 1, ..., n. \tag{69}$$

E' chiaro che tale distribuzione condizionata sara' determinata in base alle modalita' 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 probabilita' congiunta della coppia (R, S), attraverso la formula[59]

$$p_{R,S}(r, s) \equiv P(\{R = r, S = s\}) = p_R(r) \cdot p_S(s|R = r), \qquad 0 \le s \le n, \quad 0 \le r \le M. \tag{70}$$

A partire da tale distribuzione congiunta possiamo ottenere la distribuzione marginale di S attraverso la formula

$$p_S(s) = \sum_{r=0}^{M} p_{r,s} = \sum_{r=0}^{M} p_R(r) \cdot p_S(s|R = r), \qquad s = 0, 1, ..., n. \tag{71}$$

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 $p_R(r|S = s) = P(\{R = r\}|\{S = s\})$,

$$p_R(r|S = s) = \frac{p_R(r) \cdot p_S(s|R = r)}{p_S(s)}, \qquad r = 0, 1, ..., M. \tag{72}$$

[59] Si tratta della solita formula delle probabilita' composta:

$$P(\{R = r, S = s\}) = P(\{R = r\})P(\{S = s\}|\{R = r\}).$$

***Osservazione 1* (di carattere euristico)**. Il problema risolto dalla formula (72), cioe' quello di calcolare la distribuzione di probabilita' 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 e' 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 bensi' scrutinandone soltanto una parte. Problemi di tale genere si presentano frequentemente in molti campi applicativi, ad esempio nel controllo di qualita' 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 e' quello piu' puramente probabilistico) si puo' riassumere come segue:

$*$ assegnando una distribuzione di probabilita' (marginale) ad R, si esprime ***lo stato di informazione***, circa il valore che puo' assumere tale variabile, di cui si dispone ***prima di fare il "campionamento"*** (cioe' 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 modalita' 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 probabilita' 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

$$\mathbb{E}(R|S = s) := \sum_{i=1}^{N} R(\omega_i) P(\{\omega_i\}|\{S = s\}),$$

cioe' il ***valore atteso condizionato*** di R data l'osservazione di un valore s per S.

Si puo' 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 quantita' non e' altro che il valore atteso calcolato rispetto alla distribuzione espressa nella (72) e cioe'

$$\mathbb{E}(R|S = s) = \frac{\sum_{r=0}^{M} r \cdot p_R(r) \cdot p_S\,(s|R = r)}{p_S(s)}. \tag{73}$$

Consideriamo ora in dettaglio due particolari modalita' 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)

$$p_S\,(s|R = r) = \binom{n}{s} \left(\frac{r}{M}\right)^s \left(\frac{M-r}{M}\right)^{n-s}, \qquad s = 0, 1, ..., n. \tag{74}$$

cioe' la distribuzione condizionata di S dato $(R = r)$ e' 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)

$$p_S\,(s|R = r) = \frac{\binom{r}{s}\binom{M-r}{n-s}}{\binom{M}{n}}, \qquad \max(0, r + n - M) \leq s \leq \min(r, n), \tag{75}$$

cioe' la distribuzione condizionata e' una distribuzione ipergeometrica $Hyp(M, r, n)$.

Imponendo rispettivamente tali due condizioni, le precedenti formule (71), (72) e (73) diventano dunque:

(i) estrazioni casuali con reinserimento

$$p_S(s) = \frac{\binom{n}{s}}{M^n}\sum_{r=0}^{M} p_R\,(r) \cdot r^s \cdot (M - r)^{n-s} \tag{76}$$

$$p_R\,(r|S = s) = \frac{p_R(r) \cdot \binom{n}{s}\left(\frac{r}{M}\right)^s\left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}, \qquad r = 0, 1, ..., M \tag{77}$$

e di conseguenza

$$\mathbb{E}(R|S = s) = \frac{\sum_{r=0}^{M} r \cdot p_R(r) \cdot \binom{n}{s}\left(\frac{r}{M}\right)^s\left(\frac{M-r}{M}\right)^{n-s}}{p_S(s)}. \tag{78}$$

(ii) estrazioni casuali senza reinserimento

$$p_S(s) = \sum_{r=s}^{M-n+s} p_R\,(r) \cdot \frac{\binom{r}{s}\binom{M-r}{n-s}}{\binom{M}{n}} \tag{79}$$

$$p_R\,(r|S = s) = \frac{p_R(r) \cdot \binom{r}{s}\binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}}, \qquad r = s, s + 1, ..., M - n + s, \tag{80}$$

e di conseguenza[60]

$$\mathbb{E}(R|S = s) = \frac{\sum_{r=s}^{M-n+s} r \cdot p_R(r) \cdot \binom{r}{s}\binom{M-r}{n-s}}{p_S(s) \cdot \binom{M}{n}} \qquad . \tag{81}$$

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.

E' interessante ora analizzare in particolare il caso in cui R segue una distribuzione binomiale.

[60]Si noti che la condizione che $r = s, s + 1, ..., M - n + s$ in (80), deriva immediatamente dalle condizioni su s in (75), 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 e' almeno s, e che il numero totale di elementi di tipo B e' almeno $n - s$, e quest'ultima condizione implica che gli elementi di tipo A non possono essere piu' 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 puo' prendere valori nell'insieme $\{0, 1, ..., M\}$ allora, se R segue una distribuzione binomiale, sara' $R \sim bin(M, \theta)$ per un qualche valore $0 < \theta < 1$:

$$p_R(r) = \binom{M}{r} \theta^r (1-\theta)^{M-r}, \qquad r = 0, 1, ..., M.$$

Sostituendo tale espressione nelle (79) e (80) 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 ***e' binomiale***, piu' precisamente $S \sim bin(n, \theta)$.

Per quanto riguarda la distribuzione condizionata di R data S, abbiamo

$$\begin{aligned}
p_R(r|S=s) &= \frac{\binom{M}{r} \ \theta^r \ (1-\theta)^{M-r} \cdot \binom{r}{s} \ \binom{M-r}{n-s}}{\binom{n}{s} \ \theta^s \ (1-\theta)^{n-s} \cdot \binom{M}{n}} \qquad r = s, s+1, ..., M-(n-s) \\
&= \theta^{r-s} \ (1-\theta)^{M-n-(r-s)} \ \frac{M! \ r! \ (M-r)! \ s! \ (n-s)! \ n! \ (M-n)!}{r! \ (M-r)! \ s! \ (r-s)! \ (n-s)! \ [(M-r)-(n-s)]! \ n! \ M!} \\
&= \theta^{r-s} \ (1-\theta)^{M-n-(r-s)} \ \frac{(M-n)!}{(r-s)! \ [(M-r)-(n-s)]!}
\end{aligned}$$

Possiamo concludere dunque scrivendo ***la distribuzione condizionata di*** R ***data*** S ***e' individuata da***

$$p_R(r|S=s) = \binom{M-n}{r-s} \cdot \theta^{r-s} (1-\theta)^{M-n-(r-s)}, \qquad r = s, s+1, ..., M-(n-s). \tag{82}$$

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 e' fatto per S possiamo dedurre che anche T deve avere una distribuzione binomiale, esattamente di parametri $M-n$ e θ. Dalla (82) possiamo poi dedurre[61] ***la distribuzione di probabilita' condizionata di T, dato $\{S = s\}$,***

$$\begin{aligned} p_T\,(t|S=s) &= P(\{R-S=t\}|\{S=s\}) = P(\{R=s+t\}|\{S=s\} \\ &= \binom{M-n}{t}\cdot \theta^t\,(1-\theta)^{M-n-t}\,, \qquad t=0,1,\ldots,M-n. \end{aligned}$$

Cioe' la distribuzione di probabilita' condizionata di T, dato $\{S = s\}$, qualunque sia il valore s, e' uguale alla distribuzione di probabilita' marginale di T e dunque[62] S e T sono variabili aleatorie stocasticamente indipendenti!

Osservazione 2* (di carattere euristico)**. E' utile soffermarsi ad illustrare lo specifico significato intuitivo che e' 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 $bin(M,\theta)$ alla variabile R traduce la seguente condizione: ogni elemento nella popolazione ha probabilita' θ di essere di tipo A e probabilita' $(1-\theta)$ 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 probabilita' 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 $bin(n,\theta)$.

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 $R \sim bin(M,\theta)$, possiamo vedere intuitivamente che anche T deve avere una distribuzione binomiale $bin(M-n,\theta)$ e che T, S debbono essere stocasticamente indipendenti; dunque possiamo scrivere

$$\begin{aligned} p_R(r|S=s) &= P\,(R=r|S=s) = P(R-S=r-s|S=s) \\ &= P(T=r-s|S=s) = P(T=r-s) \\ &= \binom{M-n}{r-s}\cdot\theta^{r-s}\,(1-\theta)^{M-n-(r-s)} \qquad r=s,s+1,\ldots,M-(n-s) \end{aligned}$$

e cioe' ritrovare la (82).

Consideriamo ora il valore atteso di R dato $\{S = s\}$.

[61] Il fatto che $P(\{R-S=t\}|\{S=s\}) = P(\{R=s+t\}|\{S=s\})$ e' intuitivo, ma puo' 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

$$P(\{R-S=t\}|\{S=s\}) = \frac{P(\{R-s=t\}\cap\{S=s\})}{P(\{S=s\})} = P(\{R=s+t\}|\{S=s\}$$

[62] Il fatto che la distribuzione di probabilita' 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 8.2, ed in particolare l'equivalenza tra le condizioni *(i)* e *(iii)*, ivi indicate.

Possiamo scrivere, in virtu' dell'indipendenza fra S e T,

$$\mathbb{E}(T|S = s) = \mathbb{E}(T) = \theta(M - n)$$

e possiamo facilmente concludere, sfruttando la proprieta' di linearita' 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 cio' non accade, se si attribuisce ad** R **un tipo di distribuzione di probabilita' 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 e' molto realistica nel problema del campionamento***, fin qui illustrato. In altre parole potremo dire che, se si assume che R sia binomiale, allora non e' 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 gia' scrutinati.

Cio' puo' risultare abbastanza evidente, ad esempio, nel problema del sondaggio elettorale: supponiamo di assumere che ogni elettore (in un gruppo di M elettori) abbia una probabilita' 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 $bin(M, p)$). E' chiaro intuitivamente che in un tal caso e' poco utile eseguire un sondaggio, in quanto la risposta di un elettore non fornisce indicazioni circa il comportamento degli altri.

La situazione piu' frequente, comunque, e' 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 e' ne' possibile ne' particolarmente opportuno chiarire completamente tali aspetti, essenzialmente connessi a problematiche di tipo statistico, sara' pero' utile avviare in proposito un discorso circa il ruolo della nozione di ***indipendenza condizionata***. Cio' sara' l'argomento del prossimo sottoparagrafo.

11.2 Esempi

Esaminiamo intanto un esempio per illustrare un caso in cui la distribuzione di R non e' 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 e' a noi noto, bensi' e' 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 e' la distribuzione di probabilita' 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 e' 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 ha[63]

$$P(\{R=3\}|\{S=2\}) = 0;$$

infine

$$\begin{aligned} P(\{R=2\}|\{S=2\}) &= \frac{\frac{1}{20}\cdot\frac{3}{10}}{\frac{63}{200}} = \frac{3}{63}, \\ P(\{R=4\}|\{S=2\}) &= \frac{\frac{5}{10}\cdot\frac{6}{10}}{\frac{63}{200}} = \frac{60}{63}. \end{aligned}$$

E' 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 e' la distribuzione di probabilita' di S?

Soluzione. Si ha

$$P(\{S=k\}) = \sum_{m=0}^{5} p_m \binom{3}{k}\left(\frac{m}{5}\right)^k \left(\frac{5-m}{5}\right)^{3-k}.$$

[63] Si osservi che $P(\{R=3\}|\{S=2\})=0$, in quanto per ipotesi $P(\{R=3\})=0$ e quindi

$$0 \le P(\{R=3\}\cap\{S=2\}) \le P(\{R=3\}) = 0.$$

Di conseguenza

$$P(\{R=3\}|\{S=2\}) = \frac{P(\{R=3\}\cap\{S=2\})}{P(\{S=2\})} = 0.$$

In particolare[64]

$$P(\{S=2\}) = \frac{1}{125}\left(0 + \frac{2}{10}\cdot 12 + \frac{1}{20}\cdot 36 + \frac{5}{10}\cdot 48\right)$$

Esempio 11.4. *Nelle stesse condizioni del precedente Esempio 11.4 qual e' 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 probabilita' condizionate sono nulle, e si ha[65]

$$P(\{R=1\}|\{S=2\}) = \frac{P(\{R=1\})P(\{S=2\}|\{R=1\})}{P(\{S=2\})} = \frac{4}{47}$$

analogamente

$$P(\{R=2\}|\{S=2\}) = \frac{3}{47}, \qquad P(\{R=4\}|\{S=2\}) = \frac{40}{47}$$

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.

Poniamo

$$E_j \equiv \{\textit{oggetto di tipo A alla j-esima estrazione}\}, \qquad j=1,2,...,n;$$

[64]Si consideri che per $k=2$, si ha $3-k=1$ e quindi, per $m=0$ si ha $\left(\frac{0}{5}\right)^2=0$, che, per $m=5$ si ha $\left(\frac{5-5}{5}\right)^1=0$, e che $p_3=0$. Di conseguenza

$$\begin{aligned}
P(\{S=k\}) &= p_1\binom{3}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^1 + p_2\binom{3}{2}\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^1 + p_4\binom{3}{2}\left(\frac{4}{5}\right)^2\left(\frac{1}{5}\right)^1 \\
&= \frac{3}{125}\left(\frac{2}{10}\cdot 4 + \frac{1}{20}\cdot 2^2\cdot 3 + \frac{5}{10}\cdot 4^2\right) = \frac{1}{125}\left(\frac{2}{10}\cdot 12 + \frac{1}{20}\cdot 36 + \frac{5}{10}\cdot 48\right) \\
&= \frac{3}{125}\left(\frac{1}{5}\cdot 4 + \frac{1}{5}\cdot 3 + \frac{5}{5}\cdot 8\right) = \frac{3}{625}\,(4+3+40) \\
&= \frac{3}{625}\cdot 47.
\end{aligned}$$

[65]Infatti

$$\begin{aligned}
P(\{R=1\}|\{S=2\}) &= \frac{P(\{R=1\})P(\{S=2\}|\{R=1\})}{P(\{S=2\})} \\
&= \frac{\frac{3}{625}\cdot 4}{\frac{3}{625}\cdot(4+3+40)} = \frac{4}{47}; \\
P(\{R=2\}|\{S=2\}) &= \frac{P(\{R=2\})P(\{S=2\}|\{R=1\})}{P(\{S=2\})} \\
&= \frac{\frac{3}{625}\cdot 3}{\frac{3}{625}\cdot(4+3+40)} = \frac{3}{47}; \\
P(\{R=4\}|\{S=2\}) &= \frac{P(\{R=4\})P(\{S=2\}|\{R=1\})}{P(\{S=2\})} \\
&= \frac{\frac{3}{625}\cdot 40)}{\frac{3}{625}\cdot(4+3+40)} = \frac{40}{47};
\end{aligned}$$

poniamo 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'e' ***dipendenza stocastica*** fra $E_1, ..., E_n$. Inoltre, S ha ***distribuzione ipergeometrica*** $Hyp(M, r, n)$.

Sotto la condizione che le estrazioni siano con reinserimento, $E_1, ..., E_n$ sono ***eventi stocasticamente indipendenti***, di probabilita' $\frac{r}{M}$, 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 $E_1, ..., E_n$ sussiste in virtu' 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 (cioe' 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 puo' essere in generale indipendenza stocastica fra*** $E_1, ..., E_n$ ***se viene a mancare la condizione (b)***.

Notiamo d'altra parte che, come abbiamo visto prima, la motivazione per effettuare un campionamento e' data proprio dall'esigenza di ricavare informazioni circa la composizione di una popolazione; e' 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 e' in generale indipendenza stocastica fra $E_1, ..., E_n$.

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 (cio' significa che stiamo facendo delle estrazioni da una popolazione in cui il numero delle palline verdi e' una variabile aleatoria R che puo' assumere, con uguale probabilita', il valore 3 oppure il valore 7). Poniamo*

$$E_j \equiv \{\textit{pallina verde alla j-esima estrazione}\}, \qquad j = 1, 2$$

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:

$$H_1 \equiv \{\textit{abbiamo eseguito le estrazioni da } U_1\}$$

$$H_2 \equiv \{\textit{abbiamo eseguito le estrazioni da } U_2\};$$

certamente una di queste due ipotesi e' vera, ma non sappiamo quale ed attribuiamo le probabilita' (visto che l'urna e' stata scelta "a caso")

$$P(H_1) = P(H_2) = \frac{1}{2}.$$

Abbiamo intanto, applicando la definizione di probabilita' condizionata e poi la formula delle probabilita' 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 gia' 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 probabilita' di ottenere una pallina verde in una singola estrazione e' uguale a $\frac{3}{10}$ e, visto che le estrazioni avvengono con reinserimento da un'urna contenente 3 palline verdi e 7 blu, la probabilita' di ottenere due volte pallina verde in due successive estrazione e' uguale a $\left(\frac{3}{10}\right)^2$; possiamo scrivere in formule

$$P(E_1|H_1) = P(E_2|H_1) = \frac{3}{10}, \qquad P(E_1 \cap E_2|H_1) = P(E_1|H_1)\,P(E_2|H_1) = \left(\frac{3}{10}\right)^2.$$

Analogamente, per quanto riguarda il condizionamento all'ipotesi H_2, possiamo scrivere

$$P(E_1|H_2) = P(E_2|H_2) = \frac{7}{10}, \qquad P(E_1 \cap E_2|H_2) = P(E_1|H_2)\,P(E_2|H_2) = \left(\frac{7}{10}\right)^2.$$

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

$$P(E_2)\Big(= P(H_1)P(E_2|H_1) + P(H_2)P(E_2|H_2)\Big) = P(E_1) = \frac{1}{2},$$

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**, *bensi'* **positivamente correlati**.

Esercizio proposto 11.1. *Calcolare la probabilita' 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 probabilita' 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 e' 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 probabilita' data da

$$p_0 = P(\{R = 0\}), \quad p_1 = P(\{R = 1\}), \; \ldots \; , \; p_M = P(\{R = M\})$$

con $\sum_{r=0}^{M} p_r = 1$. Eseguiamo delle estrazioni casuali con reinserimento dalla popolazione e poniamo

$$E_j \equiv \{\textit{oggetto di tipo A alla } j\textit{-esima estrazione}\}, \qquad j = 1, 2, \ldots$$

Vogliamo calcolare $P(E_2|E_1)$. Estendendo quanto svolto nel precedente esempio, abbiamo

$$P(E_1) = P(E_2) = \sum_{r=0}^{M} P(E_i|R = r) \cdot p_r = \sum_{r=0}^{M} \frac{r}{M} \cdot p_r = \frac{\mathbb{E}(R)}{M}$$

$$\begin{aligned} P(E_2|E_1) &= \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{r=0}^{M} P(E_1 \cap E_2|R = r) \cdot p_r}{\sum_{r=0}^{M} P(E_1|R = r) \cdot p_r} \\ &= \frac{\sum_{r=0}^{M} \left(\frac{r}{M}\right)^2 \cdot p_r}{\sum_{r=0}^{M} \frac{r}{M} \cdot p_r} = \frac{1}{M} \frac{\sum_{r=0}^{M} r^2 \cdot p_r}{\sum_{r=0}^{M} r \cdot p_r} = \frac{1}{M} \frac{\mathbb{E}(R^2)}{\mathbb{E}(R)}. \end{aligned}$$

E' facile verificare che E_1, E_2 sono stocasticamente indipendenti se e solo se R e' una variabile degenere[66] con $R(\omega_i) = \hat{r} \in \{0, 1, \ldots, M\}$, per ogni $\omega_i \in \Omega$, nel qual caso $Var(R) = 0$ 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 $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ una partizione dell'evento certo. Diremo che E_1 ed E_2 **sono condizionatamente indipendenti data la partizione** $\mathcal{H}$ *se risulta*

$$P(E_1 \cap E_2|H_k) = P(E_1|H_k) \cdot P(E_2|H_k), \qquad \textit{per ogni } k = 1, 2, \ldots, m.$$

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*

$$P(E_1 \cap E_2|\{Z = z\}) = P(E_1|\{Z = z\}) \cdot P(E_2|\{Z = z\}), \qquad \textit{per ogni } z \in Z(\Omega) = \{z_1.z_2, \ldots, z_\ell\}.$$

[66]Infatti $P(E_2) = P(E_2|E_1)$ se e solo se $\frac{\mathbb{E}(R)}{M} = \frac{1}{M}\frac{\mathbb{E}(R^2)}{\mathbb{E}(R)}$, cioe' se e solo se $(\mathbb{E}(R))^2 = \mathbb{E}(R^2)$.

Si osservi che la precedente definizione equivale all'indipendenza condizionata di E_1 ed E_2 data la partizione $\mathcal{H} = \Big\{\{Z = z_k\},\ k = 1, 2, \ldots, \ell\Big\}$.

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*

$$P\left(\{X = x\} \cap \{Y = y\}\right) |\{Z = z\}\} = P(X = x|Z = z) \cdot P(Y = y|Z = z),$$

per ogni $x \in X(\Omega) = \{x_i,\ i = 1, 2, \ldots, n\}$, per ogni $y \in Y(\Omega) = \{y_j,\ j = 1, 2, \ldots, m\}$, per ogni $z \in Z(\Omega) = \{z_k,\ k = 1, 2, \ldots, \ell\}$.

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. Cio' 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 puo' essere spedito da ciascuno dei tre canali. La probabilita' di trasmettere il messaggio correttamente tramite C_1 e' uguale a $p^{(1)} = 0.9$. Le analoghe probabilita' 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 e' stato riservato per quel messaggio) al fine di aumentare l'affidabilita' della trasmissione.*

a) Trovare la probabilita' che il messaggio sia trasmesso correttamente in almeno una delle due volte.

b) Si supponga di sapere in seguito che il messaggio e' stato trasmesso correttamente la seconda volta, ma non la prima volta. Condizionatamente a questa osservazione, come bisogna valutare le probabilita' che sia stato utilizzato il canale C_1, C_2 e C_3, rispettivamente?

Soluzione. Poniamo

$$\begin{aligned} E_i &\equiv \{\textit{messaggio trasmesso correttamente nell'i-esimo tentativo}\}, \qquad i = 1, 2 \\ H_j &\equiv \{\textit{e' stato assegnato il canale } C_j\}, \qquad j = 1, 2, 3. \end{aligned}$$

a) Dobbiamo calcolare $P(E_1 \cup E_2)$. Applicando la formula delle probabilita' 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

$$P(H_j) = \frac{1}{3}, \qquad j = 1, 2, 3.$$

Ora possiamo osservare che gli eventi E_1, E_2 ***non sono indipendenti*** bensi' ***sono condizionatamente indipendenti***, dati gli eventi della partizione $\{H_1, H_2, H_3\}$; quindi

$$P(E_1 \cup E_2|H_j) = P(E_1|H_j) + P(E_2|H_j) - P(E_1 \cap E_2|H_j) =$$

$$= P(E_1|H_j) + P(E_2|H_j) - P(E_1|H_j) \cdot P(E_2|H_j) = p^{(j)} \left(2 - p^{(j)}\right).$$

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 $P(H_j|\overline{E}_1 \cap E_2)$. Si ha, ancora in virtu' della condizione di indipendenza condizionata

$$P(\overline{E}_1 \cap E_2) = \sum_{j=1}^{3} P(\overline{E}_1 \cap E_2|H_j) \cdot P(H_j) = \sum_{j=1}^{3} P(\overline{E}_1|H_j) \cdot P(E_2|H_j)P(H_j)$$
$$= (0.1 \times 0.91 + 0.4 \times 0.6 + 0.3 \times 0.7)\,\frac{1}{3} = \frac{1}{3}\,0.54 \;\; = 0.18;$$

utilizzando la Formula di Bayes abbiamo dunque

$$P(H_j|\overline{E}_1 \cap E_2) = \frac{P(H_j)P(\overline{E}_1 \cap E_2|H_j)}{P(\overline{E}_1 \cap E_2)} = \frac{\frac{1}{3}P(\overline{E}_1|H_j) \cdot P(E_2|H_j)}{P(\overline{E}_1 \cap E_2)} = \frac{\left(1 - p^{(j)}\right) p^{(j)}}{0.54}.$$

e quindi

$$P(H_1|\overline{E}_1 \cap E_2) = \frac{9}{54} = \frac{3}{18}, \qquad P(H_2|\overline{E}_1 \cap E_2) = \frac{24}{54} = \frac{8}{18}, \qquad P(H_3|\overline{E}_1 \cap E_2) = \frac{21}{54} = \frac{7}{18}.$$

Una proprieta' fondamentale della nozione di indipendenza condizionata e' mostrata dalla seguente ***Proposizione 1***; la dimostrazione e' lasciata al lettore per esercizio[67].

Proposizione 1. Siano E_1, E_2 due eventi condizionatamente indipendenti data una variabile aleatoria Z con $Z\,(\Omega) = \{z_1, z_2, ..., z_n\}$. Allora

$$P(E_2|E_1 \cap \{Z = z_j\}) = P(E_2|\{Z = z_j\})$$

$$P(E_2|E_1) = \sum_{j=1}^{n} P(E_2|\{Z = z_j\})P(\{Z = z_j\}|E_1).$$

[67]Diamo 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

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} = \frac{\sum_{j=1}^{n} P(E_1|\{Z = z_j\})P(E_2|\{Z = z_j\})P(\{Z = z_j\})}{\sum_{i=1}^{n} P(E_1|\{Z = z_i\})P(\{Z = z_i\})}$$
$$= \sum_{j=1}^{n} P(E_2|\{Z = z_j\})\,\frac{P(E_1|\{Z = z_j\})P(\{Z = z_j\})}{\sum_{i=1}^{n} P(E_1|\{Z = z_i\})P(\{Z = z_i\})},$$

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 e' la probabilita' che sia stata scelta l'urna U_1?

Esercizio 11.2. Una pianta produce R semi, dove R e' una variabile aleatoria binomiale con parametri n e p. Supponiamo inoltre che ciascun seme, fra gli R prodotti, germogli con probabilita' 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 puo' contenere un errore con probabilita' 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 probabilita' 0.8. Gli errori inizialmente presenti e poi rilevati vengono corretti.
(a) Determinare la probabilita' che, dopo il controllo, non vi siano piu' esercizi contenenti errori
(b) Condizionatamente al fatto che non vi siano piu' esercizi contenenti errori dopo il controllo, qual e' la probabilita' 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 probabilita':

$$P(R = 0) = 0.5, \quad P(R = 10) = 0.4, \quad P(R = 20) = 0.1.$$

Qual e' la probabilita' 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 probabilita' $\frac{1}{2}$ ed indipendentemente da quello che accade alle altre.
(a) Calcolare la probabilita' 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 e' regolare oppure se e' truccato. Truccato qui significa che esso fornisce il risultato "sei" con probabilita' doppia rispetto a quella di tutti gli altri risultati. Attribuiamo probabilita' 0.9 all'ipotesi che esso sia regolare e probabilita' 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 e' la probabilita' che il dado sia truccato?
(b) Condizionatamente a questo risultato, qual e' la probabilita' 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 $bin\left(100, \frac{3}{4}\right)$. 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 probabilita' di S
(b) Trovare la distribuzione di probabilita' 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 diseguaglianza di Chebichev 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 $\{0, 1, 2, ..., 9\}$:

$$P(\{R = r\}) = \frac{1}{10}, \quad r = 0, 1, ..., 9.$$

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 probabilita' di S
(b) Trovare la distribuzione di probabilita' 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 $A_{n,r}$ $(n, r \in N)$ costituito dalle n-uple ordinate $\boldsymbol{x} \equiv (x_1, ..., x_n)$, con $x_j \geq 0$, interi, e tali che $\sum_{j=1}^{n} x_j = r$, in simboli

$$A_{n,r} = \left\{ \boldsymbol{x} \equiv (x_1, ..., x_n) : \sum_{j=1}^{n} x_j = r,\ x_i \in \mathbb{N} \cup \{0\},\ i = 1, 2, \ldots, n \right\}.$$

Ad esempio per $n = 3$ ed $r = 4$,

$$\begin{aligned} A_{3,4} = \Big\{ &(4,0,0),\ (0,4,0),\ (0,0,4), \\ &(3,1,0),\ (3,0,1),\ (1,3,0),\ (1,0,3),\ (0,3,1),\ (0,1,3), \\ &(2,1,1),\ (1,2,1),\ (1,1,2), \\ &(2,2,0),\ (2,0,2),\ (0,2,2) \Big\} \end{aligned}$$

Qui vogliamo illustrare il significato delle distribuzioni di probabilita' su $A_{n,r}$ ed esaminare alcuni casi notevoli.

Vediamo subito una naturale interpretazione di una distribuzione di probabilita' su $A_{n,r}$: siano dati r ***soggetti*** $O_1, ..., O_r$ ed n ***diversi siti*** $1, ..., n$.

Supponiamo che gli r soggetti si dispongano negli n siti in modo aleatorio[68] e consideriamo per $i = 1, 2, ..., r$ e $j = 1, 2, ..., n$ 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:

$$X_j \equiv \sum_{i=1}^{r} B_i^{(j)}, \qquad j = 1, ..., n.$$

La variabile aleatoria X_j indica dunque il ***numero complessivo dei soggetti nel sito*** j.

Il vettore aleatorio

$$\boldsymbol{X} \equiv (X_1, ..., X_n)$$

e' quindi un vettore aleatorio a valori in $A_{n,r}$ e ciascuna distribuzione di probabilita' su $A_{n,r}$ puo' essere vista come un modello probabilistico di scelta dei siti da parte dei soggetti.

Le variabili aleatorie $X_1, ..., X_n$ vengono detti ***numeri di occupazione*** e le distribuzioni di probabilita' su $A_{n,r}$ sono indicate con il termine ***modelli di occupazione***.

Prima di illustrare i casi piu' notevoli di modelli di occupazione, e' opportuno premettere le seguenti osservazioni.

Osservazione 1. ***L'insieme*** $A_{n,r}$ ***ha cardinalita' data da*** $\binom{n+r-1}{n-1}$.

Infatti il generico elemento di $A_{n,r}$ puo' essere ad esempio rappresentato attraverso il seguente tipo di "disegno"

$$\underbrace{**\ldots**}_{1^0\ sito}^{x_1} \Big| \underbrace{**\ldots**}_{2^0\ sito}^{x_2} \Big| \ldots\ldots \Big| \underbrace{**\ldots**}_{n^{simo}\ sito}^{x_n} \tag{83}$$

[68] La 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***.

dove, guardando da sinistra verso destra, x_1 e' il numero dei simboli $*$ a sinistra della prima barretta $|$, x_2 e' il numero dei simboli $*$ compresi fra la prima e la seconda barretta ... e cosi' via; x_n e' 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 $\boldsymbol{x} \in A_{n,r}$. 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 $* * \ | \ * * |$ o piu' esplicitamente,

$$\underset{1^0\,sito}{\overbrace{\ * * \ }^{2}} \ \big| \ \underset{2^0\,sito}{\overbrace{\ * * \ }^{2}} \ \big| \ \underset{3^0\,sito}{\overbrace{\qquad}^{0}}$$

corrisponde alla tripletta dei numeri di occupazione $(2, 2, 0)$.

Analogamente, la tripletta dei numeri di occupazione $(1, 0, 3)$ corrisponde al disegno $* \ | \ \ | * **$, o piu' esplicitamente a

$$\underset{1^0\,sito}{\overbrace{\quad * \quad}^{1}} \ \big| \ \underset{2^0\,sito}{\overbrace{\qquad}^{0}} \ \big| \ \underset{3^0\,sito}{\overbrace{\ * ** \ }^{3}}$$

La cardinalita' dell'insieme $A_{n,r}$ e' quindi la stessa dell'insieme di tutti i disegni del tipo (83)

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 e' anche la cardinalita' dell'insieme $A_{n,r}$.

Osservazione 2. Consideriamo qui il caso in cui gli r soggetti siano fra di loro distinguibili. Pensiamo cioe' che ciascuno dei simboli $*$ abbia un suo indice distintivo ed avremo dunque gli r simboli distinti $*_1, ..., *_r$.

In tal caso il generico risultato elementare dell'esperimento aleatorio in questione (consistente nel disporre i soggetti nei siti) si puo' rappresentare come una "configurazione" data da un disegno del tipo in (83) in cui vengono indicati pero' anche gli indici distintivi dei soggetti; avremo dunque ***una configurazione*** del tipo

$$\underset{1^0\,sito}{\overbrace{*_{i_1} \ *_{i_2} \cdots \ *_{i_{x_1-1}} \ *_{i_{x_1}}}^{x_1}} \ \Big| \ \underset{2^0\,sito}{\overbrace{*_{i_{x_1+1}} \ *_{i_{x_1+2}} \cdots *_{i_{x_1+x_2}}}^{x_2}} \ \Big| \ \ldots\ldots \ \Big| \ \underset{n^{simo}\,sito}{\overbrace{*_{i_{x_1+\ldots+x_{n-1}+1}} \ \cdots *_{i_r}}^{x_n}} \qquad (84)$$

In questo caso si considerano diverse due configurazioni anche se danno luogo alla stessa n-upla per i numeri di occupazione (cioe' allo stesso disegno), purche' vi sia almeno un soggetto che, nelle due configurazioni, cade in due siti diversi. Una tale ***configurazione*** si puo' anche vedere come una funzione che associa a ciascuno dei soggetti (distinguibili) il suo sito, cioe' come ***una applicazione dall'insieme*** $\{O_1, O_2, ..., O_r\}$ ***all'insieme*** $\{1, 2, ..., n\}$.

Il numero complessivo di tali configurazioni e' dunque dato da n^r***.***

E' chiaro che i disegni sono delle classi di equivalenza nell'insieme delle configurazioni[69]: vi sono in generale piu' configurazioni che danno luogo ad una stessa n-upla $\boldsymbol{x} \in A_{n,r}$ ed il loro numero e' dato dal ***coefficiente multinomiale***

$$\binom{r}{x_1\, x_2 \dots x_n} \equiv \frac{r!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!}. \tag{85}$$

Tale quantita' infatti esprime il numero dei possibili modi[70] in cui un insieme di r elementi puo' essere suddiviso in n sottoinsiemi, di rispettive cardinalita' x_1, x_2, ..., x_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 $\{1, 2, 3\}$

$$O_1 \mapsto 1$$
$$O_2 \mapsto 1$$
$$O_3 \mapsto 2$$
$$O_4 \mapsto 2$$

Questa funzione e' una delle possibili fra quelle che danno luogo al vettore dei numeri di occupazione

[69]Un modo di ragionare per arrivare a calcolare il coefficiente multinomiale e' il seguente: una configurazione potrebbe essere data da una permutazione di $\{O_1, O_2, ..., O_r\}$ che da' luogo ad un vettore di numeri di occupazione $\boldsymbol{x} \in A_{n,r}$ 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 cosi' via fino agli ultimi x_n elementi che vengono messi nel sito n (ovviamente se uno degli x_i e' nullo nessun elemento viene messo nel sito i). E' chiaro che permutazioni che hanno gli stessi elementi ai primi x_1 posti, gli stessi elementi ai successivi x_2 posti, e cosi' via fino agli ultimi x_n posti, sono associati allo stesso vettore $\boldsymbol{x}$ dei numeri di occupazione. Le classi di equivalenza sono dunque

$$\frac{r!}{x_1! \cdot x_2! \cdot \dots \cdot x_n!}.$$

Per rendere piu' 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 $\boldsymbol{x} = (2, 2, 0)$ 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).

[70]Per arrivare a tale coefficiente multinomiale si puo' anche ragionare come segue:
Prima di tutto va scelto un insieme B_1 di cardinalita' $\binom{r}{x_1}$. Per ciascuna scelta di B_1 si passa poi a scegliere un sottoinsieme B_2 di cardinalita' x_2, tra gli $r - x_1$ elementi rimasti. Cio' puo' essere fatto in $\binom{r-x_1}{x_2}$ modi diversi. Si prosegue scegliendo il sottonsieme B_3 di cardinalita' x_3 tra gli $r - x_1 - x_2 = r - (x_1 + x_2)$ elementi rimasti, e cosi' via fino a che non rimangono $r - (x_1 + x_2 + \ldots + x_{n-1}) = x_n$ elementi, che rappresentano il sottoinsieme B_n. Si ottengono quindi

$$\begin{aligned}
&\binom{r}{x_1} \cdot \binom{r-x_1}{x_2} \cdot \ldots \cdot \binom{r-(x_1+\ldots+x_{n-2})}{x_{n-1}} \cdot \binom{r-(x_1+\ldots+x_{n-1})}{x_n} \\
&= \frac{r!}{x_1!\,(r-x_1)!}\, \frac{(r-x_1)!}{x_2!\,(r-x_1-x_2)!} \cdots \frac{(r-x_1-\ldots-x_{n-2})!}{x_{n-1}!\,(r-x_1-\ldots-x_{n-1})!} \\
&= \frac{r!}{x_1!\,x_2! \cdots x_{n-1}!\,(r-x_1-\ldots-x_{n-1})!} = \frac{r!}{x_1!\,x_2! \cdots x_{n-1}!\,x_n!}
\end{aligned}$$

La configurazione ottenuta va interpretata nel senso che gli elementi di B_i vengono messi nel sito i, per $i = 1, \ldots, n$.

$(2,2,0)$. La famiglia completa di tali funzioni e' 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 $\{O_1, O_2, O_3, O_4\}$

$$\begin{gathered} B_1 = \{O_1, O_2\},\ B_2 = \{O_3, O_4\},\ B_3 = \emptyset, \\ B_1 = \{O_1, O_3\},\ B_2 = \{O_3, O_4\},\ B_3 = \emptyset, \\ B_1 = \{O_1, O_4\},\ B_2 = \{O_2, O_3\},\ B_3 = \emptyset, \\ B_1 = \{O_2, O_3\},\ B_2 = \{O_1, O_4\},\ B_3 = \emptyset, \\ B_1 = \{O_2, O_4\},\ B_2 = \{O_1, O_3\},\ B_3 = \emptyset, \\ B_1 = \{O_3, O_4\},\ B_2 = \{O_1, O_2\},\ B_3 = \emptyset. \end{gathered}$$

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 possibilita' di scegliere l'insieme B_2 in $\binom{4-2}{2}$ modi ed infine B_3 e' 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 probabilita' $\frac{1}{n^r}$ a ciascuna configurazione del tipo (84).

Per $\boldsymbol{x} \equiv (x_1, x_2, ..., x_n) \in A_{n,r}$ si ha dunque[71]

$$\begin{aligned} P(\{\boldsymbol{X} = \boldsymbol{x}\}) &= P(\{X_1 = x_1,\ X_2 = x_2,\ \dots,\ X_n = x_n\}) \\ &= \frac{\binom{r}{x_1 x_2 \dots x_n}}{n^r} = \frac{r!}{x_1!\,x_2!\,\cdots\,x_{n-1}!\,x_n!}\frac{1}{n^r}, \qquad \text{per ogni } \boldsymbol{x} \in A_{n,r}. \end{aligned}$$

Osservazione 3. Da quanto sopra, segue

$$\sum_{\boldsymbol{x} \in A_{n,r}} \binom{r}{x_1 x_2 \dots x_n} = n^r.$$

ovvero

$$\sum_{\boldsymbol{x} \in A_{n,r}} \frac{r!}{x_1!\,x_2!\,\cdots\,x_{n-1}!\,x_n!} = n^r.$$

[71] Si noti che sarebbe piu' evocativo mettere al posto del simbolo P un simbolo del tipo P_{MB} per mettere in evidenza che si tratta delle probabilita' relative al modello di Maxwell-Boltzmann. Tuttavia lasciamo come al solito la scrittura P per non appesantire le notazioni.

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 probabilita' a ciascun vettore di numeri di occupazione, cioe' a ciascuna n-upla $\boldsymbol{x} \in A_{n,r}$.

Si pone dunque[72]

$$P(\{\boldsymbol{X} = \boldsymbol{x}\}) = P(\{X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n\}) = \frac{1}{\binom{n+r-1}{n-1}}, \qquad \text{per ogni } \boldsymbol{x} \in A_{n,r}.$$

12.3 Modello di Fermi-Dirac

Supponiamo $r \leq n$ e poniamo

$$\widehat{A}_{n,r} \equiv \{\boldsymbol{x} \in A_{n,r} : \; x_i = 0, \text{ oppure } x_i = 1, \; i = 1, 2, ..., n\}$$

La cardinalita' di $\widehat{A}_{n,r}$ e' ovviamente uguale a $\binom{n}{r}$.

Ora si suppone che gli r soggetti siano distribuiti nei siti in modo da assegnare uguale probabilita' a ciascuno dei disegni in $\widehat{A}_{n,r}$ (cio' in particolare implica che in ciascun sito non puo' cadere piu' di un oggetto). Si ha allora[73]

$$\begin{aligned} P(\{\boldsymbol{X} = \boldsymbol{x}\}) = P(\{X_1 = x_1,\, X_2 = x_2,\, \ldots,\, X_n = x_n\}) &= \frac{1}{\binom{n}{r}} \qquad \text{per } \boldsymbol{x} \in \widehat{A}_{n,r} \\ &= \; 0 \qquad \text{per } \boldsymbol{x} \notin \widehat{A}_{n,r} \end{aligned}$$

12.4 Schemi di estrazioni da urne

Verra' illustrato qui di seguito che ai modelli di occupazione si puo' anche dare, equivalentemente, un'interpretazione in termini di schemi di estrazioni casuali da urne.

Supponiamo di avere inizialmente n oggetti di n diversi tipi[74] in un'urna e di eseguire in modo aleatorio r estrazioni dall'urna.

Tale esperimento ha come risultati elementari le n^r r-uple $(i_1, i_2, ..., i_r)$ con $i_j \in \{1, 2, ..., n\}$ e poniamo, per $i = 1, ..., n$,

$$\widetilde{X}_i \equiv \text{numero di volte in cui si presenta un oggetto di tipo } i \text{ nelle } r \text{ estrazioni}$$

Attenzione poi in realta' non si dice almeno in formule che alcune r-ple hanno probabilita' 0, ad esempio nel caso FD, dove sembrerebbe piu' ovvio prendere le disposizioni senza ripetizione.

Anche qui dunque $\widetilde{\boldsymbol{X}} \equiv \left(\widetilde{X}_1, ..., \widetilde{X}_n\right)$ e' 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 cio' 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 e' il

---

[72] Si noti che sarebbe piu' evocativo mettere al posto del simbolo P un simbolo del tipo P_{BE} per mettere in evidenza che si tratta delle probabilita' relative al modello di Bose-Einstein. Tuttavia lasciamo come al solito la scrittura P per non appesantire le notazioni.

[73] Si noti che sarebbe piu' evocativo mettere al posto del simbolo P un simbolo del tipo P_{FD} per mettere in evidenza che si tratta delle probabilita' relative al modello di Fermi-Dirac. Anche in questo caso lasciamo la scrittura P per non appesantire le notazioni.

[74] Gli n diversi tipi sono denotati come tipo 1, tipo 2, ..., tipo n.

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 momento di una estrazione, abbia la stessa probabilita' di presentarsi, indipendentemente dal suo tipo.

Le differenze fra i diversi modelli di estrazione risiedono nelle modalita' 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 probabilita' di presentarsi.

Ciascuno degli risultati elementari possibili ha la stessa probabilita' $\frac{1}{n^r}$.

Cio' corrisponde dunque, per il vettore $\widetilde{\boldsymbol{X}}$, 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 piu' 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 puo' presentarsi al piu' una sola volta nel complesso delle estrazioni.

Tutte le r-uple $(i_1, i_2, ..., i_r)$, con $i_1 \neq i_2 \neq ... \neq i_r$ hanno la stessa probabilita'

$$\frac{1}{n\,(n-1)\,...\,(n-r+1)}.$$

Tutte le n-uple $\boldsymbol{x} \in \widehat{A}_{n,r}$ sono equiprobabili per $\widetilde{\boldsymbol{X}}$. Cio' corrisponde dunque al ***modello di Fermi-Dirac***.

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). 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 cardinalita' k, corrisponde allo stesso vettore di numeri di occupazione, ovvero alla stessa n-pla $\boldsymbol{x} \in \widehat{A}_{n,r}$. Questo e' il motivo per cui si ottiene esattamente che

$$P(\widetilde{\boldsymbol{X}} = \boldsymbol{x}) = \frac{r!}{n\,(n-1)\,...\,(n-r+1)} = \frac{1}{\binom{n}{r}}, \qquad \text{per } \boldsymbol{x} \in \widehat{A}_{n,r}.$$

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***, cosicche' alla j-esima estrazione vi sono $(n + j - 1)$ individui nell'urna.

Vogliamo ora calcolare $P(\{\boldsymbol{X} = \boldsymbol{x}\})$ per $\boldsymbol{x} \equiv (x_1, x_2, ..., x_n) \in A_{n,r}$ (ricordiamo che risulta $\sum_{h=1}^{n} x_h = r$); cominciamo a considerare la probabilita' del risultato elementare $(i_1, i_2, ..., i_r)$ definito da

$$\begin{aligned}
&i_j = 1, && 1 \leq j \leq x_1\\
&i_j = 2, && x_1 + 1 \leq j \leq x_1 + x_2\\
&\vdots\\
&i_j = n, && \sum_{h=1}^{n-1} x_h + 1 \leq j \leq \sum_{h=1}^{n} x_h = r
\end{aligned}$$

(cioe': tutti i primi x_1 elementi estratti sono di tipo 1, poi segue l'estrazione di x_2 elementi tutti di tipo 2 e cosi' via). Tale probabilita' sara' data da

$$\begin{aligned}
&\frac{1}{n} \cdot \frac{2}{n+1} \cdot \frac{3}{n+2} \cdot ... \cdot \frac{x_1}{n + x_1 - 1}\\
&\cdot \frac{1}{n + x_1} \cdot \frac{2}{n + x_1 + 1} \cdot ... \cdot \frac{x_2}{n + x_1 + x_2 - 1}\\
&\cdot \frac{1}{n + x_1 + x_2} \cdot ... \cdot \frac{x_n - 1}{n + \sum_{h=1}^{n} x_h - 2} \cdot \frac{x_n}{n + \sum_{h=1}^{n} x_h - 1}\\
&= \frac{x_1! x_2! ... x_n!}{\frac{(n+r-1)!}{(n-1)!}}.
\end{aligned}$$

Possiamo ora osservare che qualunque altro risultato elementare $(i_1, i_2, ..., i_r)$, 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 probabilita' dell'r-upla precedentemente considerata.

Tale probabilita' e' data cioe' ancora da

$$\frac{x_1! x_2! ... x_n!}{\frac{(n+r-1)!}{(n-1)!}}.$$

Dunque

$$\begin{aligned}
P(\{\boldsymbol{X} = \boldsymbol{x}\}) &= \binom{r}{x_1 x_2 ... x_n} \cdot \frac{x_1! x_2! ... x_n!}{\frac{(n+r-1)!}{(n-1)!}}\\
&= \binom{r}{x_1 x_2 ... x_n} \cdot \frac{x_1! x_2! ... x_n!}{r! \binom{n+r-1}{n-1}} = \frac{1}{\binom{n+r-1}{n-1}}.
\end{aligned}$$

Cio' 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 pero' 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 probabilita' nel discreto.

In particolare il modello di Maxwell-Boltzmann si ripresenta in moltissime situazioni diverse, ed e' 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 verra' 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 puo' dar luogo, con uguali probabilita' $p = \frac{1}{3}$, a tre diversi risultati, che per semplicita' indichiamo con 1, 2, 3.*

Supponiamo che l'esperimento venga condotto per 10 successive volte, sempre con le stesse modalita' 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 probabilita' 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 puo' essere risolto guardando a X_1, X_2, X_3 come a dei numeri di occupazione, con $r = 10$ ed $n = 3$. (ogni prova e' vista come un soggetto ed ogni possibile risultato come un sito dentro cui si inserisce ciascun 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'equiprobabilita' nel comportamento delle diverse prove si ha che la distribuzione di probabilita' 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 ($i = 1, ..., 6$) 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 probabilita' cercata e' data da

$$\begin{aligned}
&P(\{X_1 = 3,\ X_2 = 3,\ X_3 = 3,\ X_4 = 3\}) \\
&= P(\{X_1 = 3,\ X_2 = 3,\ X_3 = 3,\ X_4 = 3,\ X_5 = 0,\ X_6 = 0\}) \\
&= \frac{12!}{(3!)^4\ (0!)^2\, 6^{12}} = \frac{12!}{6^{16}}.
\end{aligned}$$

Esempio 12.3. *Un testo contiene 20 caratteri. Supponiamo di sapere che esso contiene 5 errori (cioe' 5 caratteri errati) e di valutare che ciascun carattere abbia uguale probabilita' di essere errato rispetto a ciascun altro. Qual e' la probabilita' che gli errori si trovino nel* 3^0, 5^0, 10^0, 11^0, 12^0 *carattere?*

Questo e' un esempio di modello di Fermi-Dirac, con $n = 20$ *ed* $r = 5$, *in quanto in ogni posto ci puo' stare al piu' un errore. E dunque la probabilita' cercata e' 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 e' 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" e' equivoca: potrebbe trattarsi di modello di Maxwell-Boltzmann oppure di Bose-Einstein, a seconda che le persone vengano considerate o meno ***indipendenti***[75] fra di loro (o diciamo, ***distinguibili***).

Nei due diversi casi si avrebbe, rispettivamente,

$$P(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{5!}{(0!)^3\, 4!\, 4^5} = \frac{5!}{4! 4^5} = \frac{5}{4^5}$$

oppure

$$P(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{1}{\binom{8}{3}} = \frac{3!5!}{8!} = \frac{3 \cdot 2}{8 \cdot 7 \cdot 6} = \frac{1}{56}.$$

Riguardo al modello di Maxwell-Boltzmann in particolare, osserviamo che esso si rivela, per questo caso, piuttosto irrealistico: e' 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 equiprobabilita', vediamo che e' 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 equiprobabilita' fra i diversi siti. Tale estensione porta alla definizione di distribuzione multinomiale.

12.6 Distribuzione multinomiali

Consideriamo ancora un modello di occupazione con n oggetti ed r siti. Supponiamo ancora che vi sia indipendenza stocastica completa circa la scelta dei siti da parte dei diversi oggetti, ma non vi sia equiprobabilita' fra i diversi siti: supponiamo che ciascun oggetto scelga il sito j con fissata probabilita' p_j ($j = 1, ..., n$; $p_j \geq 0$, $\sum_{j=1}^{n} p_j = 1$).

Generalizzando quanto arguito per il modello di Maxwell-Boltzmann, possiamo concludere che, per i numeri di occupazione $X_1, ..., X_n$, risulta:

$$P(\{X_1 = x_1, ..., X_n = x_n\}) = \binom{r}{x_1\ x_2\ ...\ x_n} (p_1)^{x_1} \cdot (p_2)^{x_2} \cdot ... \cdot (p_n)^{x_n}\,. \tag{86}$$

$$= \frac{r!}{x_1!\ x_2!\ ...\ x_n!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_n^{x_n}, \qquad \text{per } (x_1, x_2, \dots x_n) \in A_{n,r}. \tag{87}$$

[75] La 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.

Diciamo in tal caso che la distribuzione congiunta di $X_1, ..., X_n$ e' una *distribuzione multinomiale* di parametri $(r, n; p_1, ..., p_n)$.

E' 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 $bin(r, p)$ e' 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 (86) in realta' e' piu' comodo pensare alla distribuzione multinomiale proprio come a una generalizzazione della binomiale[76].

Osserviamo infine che essendo ovviamente $P(\boldsymbol{X} \in A_{n,r}) = 1$, si ha che

$$\sum_{(x_1,x_2,...,x_n)\in A_{n,r}} \frac{r!}{x_1!\, x_2! \,...\, x_n!} p_1^{x_1} \cdot p_2^{x_2} \cdot \cdots \cdot p_n^{x_n} = 1,$$

che non e' altro che un caso particolare della formula della potenza del multinomio, ovvero

$$\sum_{(x_1,x_2,...,x_n)\in A_{n,r}} \frac{r!}{x_1!\, x_2! \,...\, x_n!} a_1^{x_1} \cdot a_2^{x_2} \cdot \cdots \cdot a_n^{x_n} = (a_1 + a_2 + \cdots + a_n)^r, \tag{88}$$

che a sua volta e' 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'e' un singolo vincitore: A, B e C hanno probabilita' 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 probabilita' dell'evento

$$\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}.$$

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

$$P(\{X_A = 1, X_B = 1, X_C = 1, X_D = 2\}) = \frac{5!}{(1!)^3\, 2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{5!}{2!} \left(\frac{1}{5}\right)^3 \left(\frac{2}{5}\right)^2 = \frac{48}{5^4} = \frac{48}{625}.$$

---

[76] L'idea e' che ci sono r prove ad n esiti possibili, ovvero

$$E_k^{(j)}, \qquad \text{per } j = 1, 2, \ldots, n,\ k = 1, 2, \ldots r,$$

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. Inoltre si suppone che, qualunque sia $k = 1, 2, \ldots, r$, la probabilita' $P(E_k^{(j)}) = p_j$. Infine si suppone l'indipendenza delle prove, cioe' che, qualunque sia $(j_1, j_2, \ldots, j_r) \in \{1, 2, \ldots, n\}^r$,

$$P(E_1^{(j_1)} \cap E_2^{(j_2)} \cap E_3^{(j_3)} \cap \cdots \cap E_r^{(j_r)}) = P(E_1^{(j_1)}) \cdot P(E_2^{(j_2)}) \cdot P(E_3^{(j_3)}) \cdot \cdots \cdot P(E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \cdots \cdot p_{j_n}.$$

Il caso della binomiale $bin(r, \theta)$ e dello schema di Bernoulli per r prove indipendenti, con probabilita' di successo θ rientra in questo modello, con $n = 2$, $E_k^{(1)} = E_k$, $E_k^{(2)} = \overline{E}_k$, $p_1 = \theta$, $p_2 = 1 - \theta$.

L'evento $\{X_1 = x_1, ..., X_n = x_n\}$, con $\boldsymbol{x} \in A_{n,r}$ si verifica se e solo se si verificano eventi del tipo $E_1^{(j_1)} \cap E_2^{(j_2)} \cap \cdots \cap E_r^{(j_r)}$, con x_1 indici di tipo 1, x_2 indici di tipo 2, ... , x_n indici di tipo n. Per calcolare la probabilita' dell'evento $\{X_1 = x_1, ..., X_n = x_n\}$, basta allora considerare che gli eventi di questo tipo hanno tutti la stessa probabilita':

$$P(E_1^{(j_1)} \cap E_2^{(j_2)} \cap \cdots \cap E_r^{(j_r)}) = p_{j_1} \cdot p_{j_2} \cdot p_{j_3} \cdot \cdots \cdot p_{j_r} = p_1^{x_1} \cdot p_2^{x_2} \cdot \cdots \cdot p_n^{x_n},$$

e che inoltre gli eventi di questo tipo sono esattamente $\frac{r!}{x_1!\, x_2! \,...\, x_n!}$.

Esempio 12.6. *In una giornata del campionato di calcio si attribuisce probabilita' 0.5 alla vittoria della squadra che gioca in casa (risultato 1), probabilita' 0.2 alla sconfitta della squadra che gioca in casa (risultato 2), probabilita' 0.3 al pareggio (risultato* x*), e il risultato di ciascuna partita e' giudicato essere indipendente dai risultati delle altre. Si consideri la colonna "vincente" della schedina del totocalcio*[77]*; 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* $(13, 3; \frac{5}{10}, \frac{2}{10}, \frac{3}{10})$.

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 probabilita' marginale della variabile aleatoria X_1.

Si avra', per definizione di distribuzione marginale,

$$P(\{X_1 = x\}) = \sum_{(z_2,...,z_n)\in A_{n-1,r-x}} P(\{X_1 = x, X_2 = z_2, ..., X_n = z_n\}) \tag{89}$$

Un'analoga formula vale per le distribuzioni marginali di $X_2, ..., X_n$.

Consideriamo ora la distribuzione di probabilita' 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 (89),

$$P(\{Y = y\}) = \sum_{(z_2,...,z_n)\in A_{n-1,y}} P(\{X_1 = r - y,\ X_2 = z_2,\ ...,\ X_n = z_n\}).$$

Supponiamo ora di voler calcolare la ***distribuzione condizionata di*** X_2 ***data*** X_1. Osserviamo allora che vale

$$P(\{X_1 = x_1, X_2 = x_2\}) = \sum_{(z_3,...,z_n)\in A_{n-2,r-x_1-x_2}} P(\{X_1 = x_1,\ X_2 = x_2,\ X_3 = z_3,\ ...,\ X_n = z_n\}),$$

da cui

$$P(\{X_2 = x_2\}|\{X_1 = x_1\}) = \frac{\sum_{(z_3,...,z_n)\in A_{n-2,r-x_1-x_2}} P(\{X_1 = x_1,\ X_2 = x_2,\ X_3 = z_3,\ ...,\ X_n = z_n\})}{\sum_{(\xi,\zeta_3,...,\zeta_n)\in A_{n-1,r-x_1}} P(\{X_1 = x_1,\ X_2 = \xi,\ X_3 = \zeta_3,\ ...,\ X_n = \zeta_n\})}.$$

[77] Una schedina del totocalcio e' 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.

Esempio 12.7. *Consideriamo il modello di Maxwell-Boltzmann con r soggetti ed n siti. Allora, per $x_1 = 0, 1, ..., r$, avremo (ricordando che $P(\{\boldsymbol{X} = \boldsymbol{x}\}) = \frac{\binom{r}{x_1\, x_2\, \dots\, x_n}}{n^r}$)*

$$
\begin{aligned}
P(\{X_1 = x\}) &= \sum_{(z_2,\dots,z_n)\in A_{n-1,r-x}} P(\{X_1 = x,\, X_2 = z_2,\, ...,\, X_n = z_n\} \\
&= \frac{1}{n^r} \sum_{(z_2,\dots,z_n)\in A_{n-1,r-x}} \frac{r!}{x!\,(r-x)!} \frac{(r-x)!}{z_2!...z_n!} \\
&= \frac{1}{n^r} \frac{r!}{x!\,(r-x)!} \sum_{(z_2,\dots,z_n)\in A_{n-1,r-x}} \frac{(r-x)!}{z_2!...z_n!},
\end{aligned}
$$

*da cui, ricordando l'***Osservazione 3***, o equivalentemente la formula della potenza del multinomio (88)*

$$
\begin{aligned}
P(\{X_1 = x\}) &= \binom{r}{x} \frac{1}{n^r} (n-1)^{r-x} = \\
&= \binom{r}{x} \left(\frac{1}{n}\right)^x \left(1 - \frac{1}{n}\right)^{r-x} \qquad \text{per } x = 0, 1, \dots, r.
\end{aligned}
$$

Cioè, come ci si doveva immaginare[78]*, la distribuzione marginale di X_1 e' binomiale di parametri r e $\frac{1}{n}$.*

Per quanto riguarda la distribuzione condizionata di X_2 data X_1 si ha innanzitutto che

$$
\begin{aligned}
P(\{X_1 = x_1, X_2 = x_2\}) &= \sum_{(z_3,\dots,z_n)\in A_{n-2,r-x_1-x_2}} \frac{\binom{r}{x_1\, x_2\, z_3\, \dots\, z_n}}{n^r} \\
&= \frac{1}{n^r} \frac{r!}{x_1!\, x_2!\, (r - x_1 - x_2)!} \sum_{(z_3,\dots,z_n)\in A_{n-2,r-x_1-x_2}} \frac{(r - x_1 - x_2)!}{z_3!\, \dots\, z_n!} \\
&= \binom{r}{x_1\; x_2\; (r - x_1 - x_2)} \left(\frac{1}{n}\right)^{x_1} \left(\frac{1}{n}\right)^{x_2} \left(1 - \frac{1}{n}\right)^{r-x_1-x_2},
\end{aligned}
$$

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 $0 \leq x_1 \leq r$) e' data da

$$
\begin{aligned}
P(\{X_2 = x_2\}|\{X_1 = x_1\}) &= \frac{P(\{X_1 = x_1, X_2 = x_2\})}{P(\{X_1 = x_1\})} \\
&= \binom{r - x_1}{x_2} \left(\frac{1}{n-1}\right)^{x_2} \left(1 - \frac{1}{n-1}\right)^{r-x_1-x_2};
\end{aligned}
$$

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 $\frac{1}{n-1}$.

[78]Il modello di Maxwell-Boltzmann e' 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 e' 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).

Analoghe formule si possono facilmente ottenere per quanto riguarda le distribuzioni marginali e condizionate di piu' di due fra le n variabili $X_1, ..., X_n$, per le distribuzioni di probabilita' delle loro somme parziali etc... ; e' 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 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ segua una distribuzione multinomiale di parametri $(r, n; p_1, p_2, \ldots, p_n)$.
È facile convincersi che la distibuzione marginale di X_i e' binomiale di parametri r e p_i, cosi' come la distribuzione di $X_i + X_j$ e' $bin(r, p_i + p_j)$, e cosi' via.

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 pero' 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

$$\begin{aligned}
P(\{X_1 = x_1, X_2 = x_2\}) &= \sum_{(z_3,\ldots,z_n) \in A_{n-2, r-x_1-x_2}} \binom{r}{x_1\ x_2\ z_3\ \ldots\ z_n} p_1^{x_1}\ p_2^{x_2}\ p_3^{z_3} \cdots\ p_n^{z_n} \\
&= \frac{r!}{x_1!\ x_2!\ (r - x_1 - x_2)!}\ p_1^{x_1}\ p_2^{x_2} \sum_{(z_3,\ldots,z_n) \in A_{n-2, r-x_1-x_2}} \frac{(r - x_1 - x_2)!}{z_3!\ \ldots\ z_n!}\ p_3^{z_3} \cdots\ p_n^{z_n} \\
&= \binom{r}{x_1\ x_2\ (r - x_1 - x_2)} p_1^{x_1}\ p_2^{x_2}\ \left(p_3^{z_3} + \cdots + p_n^{z_n}\right)^{r - x_1 - x_2} \\
&= \frac{r!}{x_1!\ x_2!\ (r - x_1 - x_2)!}\ p_1^{x_1}\ p_2^{x_2}\ \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2},
\end{aligned}$$

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 $0 \leq x_1 \leq r$) e' data da

$$\begin{aligned}
P(\{X_2 = x_2\} | \{X_1 = x_1\}) &= \frac{P(\{X_1 = x_1, X_2 = x_2\})}{P(\{X_1 = x_1\})} \\
&= \frac{\binom{r}{x_1\ x_2\ (r - x_1 - x_2)}\ p_1^{x_1}\ p_2^{x_2}\ \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2}}{\binom{r}{x_1}\ p_1^{x_1} (1 - p_1)^{r - x_1}} \\
&= \frac{\frac{r!}{x_1!\ x_2!\ (r - x_1 - x_2)!}\ p_1^{x_1}\ p_2^{x_2}\ \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2}}{\frac{r!}{x_1!\ (r - x_1)!}\ p_1^{x_1} (1 - p_1)^{r - x_1}} \\
&= \frac{\frac{1}{x_2!\ (r - x_1 - x_2)!}\ p_2^{x_2}\ \left(1 - (p_1 + p_2)\right)^{r - x_1 - x_2}}{\frac{1}{(r - x_1)!} (1 - p_1)^{r - x_1}} \\
&= \binom{r - x_1}{x_2} \left(\frac{p_2}{1 - p_1}\right)^{x_2} \left(1 - \frac{p_2}{1 - p_1}\right)^{r - x_1 - x_2}
\end{aligned}$$

per $x_2 \geq 0$, $x_1 + x_2 \leq r$, ovvero per $0 \leq x_2 \leq r - x_1$. 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 e' avuto esito di tipo 1. Infine la probabilita' di esito 2 va valutata condizionatamente a sapere che su ciascuna di tali prove non e' verificato un esito di tipo 1: in fondo questa e' l'interpretazione di $\frac{p_2}{1-p_1}$.

12.9 Esercizi di verifica

Esercizio 12.1. Sia $r = 5, n = 4$. Rappresentate la quaterna (1,2,2,1) sotto forma di un disegno del tipo in (83).

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 e' l'esame che e' 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.
E' 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 probabilita' 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 probabilita' 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 probabilita' 0.5 al risultato 1, probabilita' 0.2 al risultato 2 e probabilita' 0.3 al risultato x, e il risultato di ciascuna partita e' giudicato indipendente dai risultati delle altre (si veda il precedente Esempio 6).
(a) Qual e' la colonna piu' probabile? Quanto vale la sua probabilita'?
(b) Quanto vale la probabilita' del risultato $(1, 1, \mathsf{x}, 2, 1, 1, 1, 1, \mathsf{x}, \mathsf{x}, 1, 2, \mathsf{x})$?
(c) Qual e' la probabilita' che vi siano 7 risultati 1, 2 risultati 2 e 4 risultati x?

Esercizio 12.7. Un esperimento, che puo' dar luogo, con uguali probabilita' $p = \frac{1}{3}$ 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 probabilita' dell'evento $\{X_2 + X_3 = 6\}$
(b) Calcolare la probabilita' condizionata $P(\{X_2 = 3\}|\{X_1 = 4\})$.
(c) Calcolare la distribuzione di probabilita' 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 probabilita' p e poniamo $T = n - S$.
Che cosa si ottiene come distribuzione congiunta di S e T?

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	Υo anche Y	üpsilon
ϕ o anche φ	Φ	phi (fi)
χ	X	chi
ψ	Ψ	psi
ω	Ω	omega