Lo schema di Bernoulli (o delle prove indipendenti): un esempio di modello probabilistico applicabile a diversi contesti

Rita Giuliano (Pisa)

0. Introduzione.

È ormai acquisizione comune il fatto che uno degli scopi della Matematica è quello di costruire e studiare dei modelli per mezzo dei quali interpretare i fenomeni reali; accade poi di frequente che un modello introdotto per trattare qualche particolare situazione si riveli utilizzabile anche per altre, magari molto diverse dalla prima per quanto riguarda il contesto in cui nascono. Lo schema di Bernoulli, di cui ci occuperemo in queste pagine, è proprio uno di questi modelli.

Nell'esposizione che segue cercherò, per quanto possibile, di percorrere il cammino che ogni sistematizzazione matematica richiede, cioè non partirò dalla teoria, già confezionata, per arrivare alla pratica; tenterò invece di mettere in luce il fatto che ogni costruzione matematica, per quanto astratta, segue un'evoluzione, in cui pratica e teoria vanno dapprima di pari passo, fino al momento in cui i tempi diventano maturi per la cosiddetta "assiomatizzazione". Il Calcolo delle Probabilità, in particolare, nasce in Francia nel secolo 17-esimo, ma il momento della sistemazione teorica giunge ben più tardi (1933), ad opera del matematico russo Andrei Nikolaevich Kolmogorov (1903-1987).

Per tornare allo schema di Bernoulli, voglio qui osservare solo che, come si vedrà, la sua presentazione richiede strumenti matematici molto semplici, e dunque, dal punto di vista didattico, dovrebbe prestarsi bene per una trattazione in classe. Tuttavia su questo punto naturalmente l'ultima parola spetta non a me, ma ai Docenti di Matematica.

Termino questa breve introduzione con un ringraziamento ai Proff. G. Letta e M. Rossetti per la consulenza offertami, e ai Docenti di Materie Scientifiche dell'*Istituto Meucci–Galilei* in Massa Carrara, a cui questo intervento è destinato.

1. Alcuni esempi tratti dalla realtà.

Cominceremo con la descrizione di alcune situazioni, attinenti a contesti molto distanti tra loro, ma in realtà tutte riconducibili al modello probabilistico di cui ci vogliamo occupare.

1. (Giochi d'azzardo: lanci di monete). Si lancia 5 volte una moneta. È noto che la probabilità che essa dia "testa" in un lancio generico è pari a $p = 0.6$. Un giocatore con capitale iniziale nullo riceve dal banco 10 euro ogni volta che la moneta dà "testa", e paga invece al banco 5 euro ogni volta che essa dà "croce". Quanto vale la probabilità che dopo i 5 lanci il giocatore sia in credito con il banco?

2. (Giochi d'azzardo: la ruota della fortuna [4]). Questo è un gioco abbastanza popolare nelle fiere. Un giocatore scommette su uno dei numeri che vanno dall'1 al 6. Una ruota, numerata con gli stessi numeri, viene girata due volte. Se il numero su cui ha scommesso il giocatore appare i volte (con $i = 1, 2$), allora il giocatore vince i euro. Se invece il numero non appare su nessuna delle due ruote, allora egli perde 1 euro. Il gioco è equo per lo scommettitore? (ovvero, il capitale finale atteso è pari a 0?)

3. (L'ereditarietà dei caratteri secondo G. Mendel (1822-1884) [4]). Un particolare carattere (ad esempio il colore degli occhi) di una persona viene classificato sulla base di una coppia di geni; indichiamo con A (risp. a) il gene dominante (risp. recessivo). Dunque un individuo con la coppia di geni AA è puramente dominante (fenotipo A), uno con aa è puramente recessivo (fenotipo a), e uno con Aa è ibrido (fenotipo A). La coppia di geni dei figli è formata da un gene del padre e uno della madre. Se due genitori, ibridi relativamente ad un dato carattere, hanno 4 figli, qual è la probabilità che 3 di questi siano di fenotipo dominante?

4. (Controllo di qualità [4]). È noto che le viti prodotte da una certa fabbrica presentano un difetto con probabilità 0.01. La fabbrica vende le viti in confezioni di 10 e sostituisce i pacchetti che contengano una o più viti difettose. Qual è la percentuale dei pacchetti venduti che la fabbrica dovrà rimpiazzare?

5. (Un problema di decisione in scienze sociali [4]). Una giuria di un processo è formata da 12 persone. L'imputato viene condannato se almeno 9 dei 12 giurati lo ritiene colpevole. Supponiamo (in modo un

po' fittizio) che ogni giurato prenda la decisione corretta con probabilità 0.81. Qual è la probabilità che la giuria emetta un verdetto corretto se l'imputato è effettivamente colpevole?

6. (Affidabilità di sistemi [4]). Un sistema di comunicazione è formato da n componenti, ognuno dei quali funziona con probabilità pari a p. Il sistema, nella sua totalità, funziona se almeno la metà dei suoi componenti funziona. Per quali valori di p un sistema formato da 5 componenti ha maggiore probabilità di funzionare rispetto a uno formato da soli 3 componenti?

7. (Sperimentazione farmacologica). In un laboratorio si sperimenta un nuovo farmaco somministrandolo a 10 cavie. È noto (per esperienza precedente) che una generica cavia presenta reazione positiva al farmaco con probabilità 0.85. Si è deciso di ritenere efficace il farmaco (e dunque di metterlo in commercio) se almeno 8 cavie presenteranno reazione positiva. Qual è la probabilità che il farmaco venga messo in commercio?

8. (Operazione antimafia). In un'operazione di polizia, una pattuglia deve catturare 5 malviventi. Ciascun malvivente ha probabilità 0.67 di essere catturato (questo valore è stato calcolato in base ai risultati ottenuti dalla medesima pattuglia nel passato). Qual è la probabilità che al termine dell'operazione in corso al più un malvivente sia riuscito a sfuggire alla cattura?

9. (Manutenzione delle centraline ENEL). In un appartamento ciascuna delle 5 stanze è illuminata da una lampadina. È noto che un fulmine che si abbatta nelle vicinanze durante un temporale può fulminare una generica lampadina con probabilità p (questo valore dipende dalle qualità del sistema di controllo della vicina centralina Enel). Quanto deve valere (al massimo) p perché la probabilità che l'appartamento resti al buio non superi 0.01?

10. (Sicurezza aerea). Per un certo tipo di aereo quadrimotore, ogni motore ha la stessa probabilità di rompersi, pari a 0.04. Per ciascun motore di un dato tipo di bimotore la stessa probabilità vale invece 0.03. Ciascuno dei due tipi di aereo può arrivare a destinazione anche con solo la metà dei motori funzionante. È più affidabile un quadrimotore o un bimotore?

Nota. In molti degli esempi precedenti sono state fatte delle ipotesi semplificatrici e quindi certamente non realistiche, con lo scopo di mettere in risalto il nocciolo *vero* del problema, cioè per non essere costretti a tenere in considerazione problematiche che in questo momento sono per noi prive di interesse.

2. Lo schema generale.

Ognuna delle situazioni descritte sopra si può inquadrare nello schema generale seguente: in tutti i casi esaminati si è eseguita n volte una data *prova*, sempre la stessa tutte le volte; tale prova è, in ciascun esempio fatto, di un tipo abbastanza particolare: essa produce due (e due soli) possibili risultati, che chiameremo convenzionalmente *successo* e *insuccesso* e che indicheremo (per ragioni che saranno chiare in seguito, si veda l'osservazione (6.4)) rispettivamente con i simboli 1 e 0. Inoltre il *successo* (risp. l'*insuccesso*) si presenta con probabilità p (risp. $1-p$). Una prova di questo tipo viene comunemente chiamata *prova di Bernoulli* (dal nome del matematico svizzero Jakob Bernoulli (1654-1705)).

Un altro termine spesso usato come sinonimo di prova è *esperimento*. Dovrebbe essere chiaro però che si tratta comunque di una terminologia convenzionale, che non deve far pensare ad un esperimento necessariamente *di laboratorio*; nella lista di sopra l'unico vero esperimento di laboratorio è quello della situazione 7 (le cavie).

Torniamo alla descrizione dello schema. Un'ipotesi, diciamo così, "ineludibile" è quella che le varie prove si svolgano in condizioni di indipendenza, cioè (parlando con linguaggio per il momento solo intuitivo) tali che il risultato di una qualsiasi di esse non influenzi o non sia influenzato da quello di un'altra (o delle altre). A questo punto della nostra presentazione, non è affatto chiaro perché questa ipotesi debba essere fatta, né gli esempi di sopra sono di grande utilità in questo senso. Scioglieremo questo nodo tra poco, quando costruiremo il modello matematico rigoroso: durante la costruzione ci sarà un momento in cui, o ci rassegniamo a supporre l'indipendenza, oppure non riusciremo ad andare avanti.

La situazione generale sopra illustrata viene comunemente indicata con il nome di *schema delle prove ripetute*, o anche *schema di Bernoulli*. Nel titolo della presente esposizione ho preferito utilizzare il nome di *schema delle prove indipendenti* per sottolineare il fatto che le prove, oltre che tutte identiche, sono anche, appunto, tra loro indipendenti.

A titolo di esempio, discutiamo nei dettagli il caso dell'operazione antimafia (situazione 8); qui la prova in questione è la cattura di un generico malvivente. La (prova di) cattura del malvivente viene ripetuta $n = 5$ volte, e ogni cattura può risolversi in due modi: il malvivente viene catturato oppure no. Cosa chiamare *successo* e cosa *insuccesso* in questo caso?

Abbiamo detto sopra che anche queste due parole sono convenzionali (un poliziotto chiamerebbe *successo* la cattura, mentre un malvivente si esprimerebbe sicuramente in modo opposto); ovvero, i due termini devono essere considerati privi di qualsiasi connotazione, sia essa positiva o negativa; essi indicano solo l'uno o l'altro dei due esiti possibili, sono l'*acceso* (1) e lo *spento* (0) della situazione.

Dato però che un nome dobbiamo comunque darlo, mettiamoci dalla parte della polizia: diremo per convenzione che abbiamo avuto un successo (1) all'i esima prova se il malvivente i-esimo è stato catturato; diremo invece che abbiamo avuto un insuccesso (0) se il malvivente è riuscito a scappare; (ma, ripetiamo, niente ci impedirebbe, se necessario, di esprimerci "al negativo ", l'importante è prendere una decisione una volta per tutte).

Rimandiamo a fra poco le considerazioni che riguardano l'ipotesi di indipendenza, che, come abbiamo osservato sopra, non è facilmente giustificabile a questo punto della trattazione. Per terminare la discussione di questo esempio, non resta che osservare che in questo caso la probabilità di avere successo in una singola prova è ora la probabilità di riuscire a catturare un malvivente, e dunque abbiamo $p = 0.67$, $1 - p = 0.33$. (È quasi superfluo far notare che, se avessimo invertito i due termini (*successo* = non cattura, *insuccesso* = cattura), allora, per mantenere il simbolismo usato all'inizio, avremmo dovuto scrivere $p = 0.33$, e $1 - p = 0.67$).

Facciamo una breve digressione storica: il Calcolo delle Probabilità nasce in Francia nel secolo 17-esimo come strumento per risolvere problemi legati ai giochi d'azzardo (lanci di monete, dadi, giochi con le carte, eccetera). Famoso a questo proposito è il carteggio tra Blaise Pascal (1623-1662) e Pierre de Fermat (1601-1665), nel quale i due matematici risposero ad alcuni quesiti (sui giochi di dadi, appunto) che un incallito giocatore, il nobile Antoine Gombauld, cavaliere di Méré, proponeva a Pascal. (Pascal e Fermat hanno così inconsapevolmente posto le prime basi per la nascita del Calcolo delle Probabilità come disciplina matematica).

Questo spiega come mai, a proposito dello schema delle prove ripetute, l'esempio principe è tuttora quello dei lanci ripetuti di una moneta (situazione 1), tanto che spesso, invece di parlare di *prova*, *successo*, *insuccesso*, si usa dire anche, con identico significato, *lancio*, *Testa*, *Croce*.

Una piccola curiosità a margine: è forse istruttivo conoscere la visione che Pascal aveva della mente umana. In una lettera a Fermat, riferendosi al cavaliere di Méré, egli scrive:
Egli è molto intelligente, ma non è un matematico: questo, come tu sai, è un grande difetto.

Nel seguito di questa esposizione ci occuperemo di descrivere nei dettagli il procedimento per la costruzione matematica del modello sopra descritto. Per poter giungere a questo, tuttavia, sono necessarie alcune premesse generali.

3. Preliminari probabilistici: cos'è uno spazio probabilizzato?

Come abbiamo detto sopra, il Calcolo delle Probabilità nasce storicamente per risolvere problemi di giochi d'azzardo. Ben presto però ci si rende conto che le stesse tecniche possono essere utilizzate per trattare, più in generale, fenomeni dei quali non sia possibile avere sotto controllo tutti i fattori che ne determinano lo svolgimento, e dei quali, dunque, sia impossibile prevedere il risultato. Fenomeni di questo tipo vengono detti *casuali* o *aleatori* (dalla parola latina *alea* = *dado*: tutti ricordano la celebre frase di Cesare "alea iacta est..."), e, in lingua inglese, *random*. All'opposto dei fenomeni aleatori stanno i fenomeni *deterministici*, quelli cioè dei quali sono note tutte le condizioni iniziali e dunque è possibile stabilire lo svolgimento e l'esito. Un esempio di fenomeno deterministico è il moto di un corpo che si muova di moto rettilineo uniforme: supponendo che il corpo parta dall'origine e abbia velocità costante v, dopo un tempo (noto) t esso avrà percorso uno spazio $s = vt$ (calcolabile dunque a partire dai dati v e t). È evidente che in natura praticamente nessun fenomeno può essere considerato deterministico (lo stesso esempio fatto qui sopra deve essere considerato come un'astrazione ideale); è quindi altrettanto evidente che considerazioni di tipo "aleatorio" sono essenziali per lo studio di qualsiasi situazione reale, della natura più varia; le tecniche del Calcolo delle Probabilità si applicheranno allora in Biologia, Informatica, Sociologia eccetera.

Per giustificare le definizioni che daremo, cominciamo con qualche esempio semplice (ancora giochi d'azzardo).

(3.1) Esempio. Si lancia una moneta non truccata. Non siamo in grado di prevedere se essa darà "Testa" oppure "Croce", però qualcosa possiamo dire. Intanto elenchiamo tutti i possibili risultati di questo esperimento: sono evidentemente "Testa" e "Croce", che indicheremo convenzionalmente con i simboli 1 e 0. In formula scriviamo

$$\Omega = \{0, 1\},$$

e chiamiamo Ω *spazio campione*. I suoi elementi sono detti *eventi elementari*. Il generico evento elementare verrà indicato con il simbolo ω.
Tutti sono in grado di rispondere, per il momento in modo solo intuitivo, alle domande: "qual è la probabilità che esca Testa?" e "qual è la probabilità che esca Croce?" Ad esse chiunque risponderebbe 1/2: in simboli scriviamo

$$P(0) = P(1) = \frac{1}{2}.$$

Un secondo passo, cioè, consiste nell'associare ad ogni evento elementare un numero, la sua *probabilità*, che servirà a misurare, in qualche modo, il nostro *grado di fiducia* del verificarsi dell'evento considerato.

(3.2) Esempio. Si lanciano due monete equilibrate (non truccate). Cerchiamo di ripetere in questo secondo caso i passi dell'esempio precedente.
(i) Costruzione dello spazio campione Ω: si può scrivere (con evidente interpretazione dei simboli usati):

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\} = \{0, 1\}^2.$$

(ii) Probabilità di ogni evento elementare: poiché le monete sono entrambe equilibrate, non c'è motivo di ritenere che uno dei possibili risultati sia *più probabile* oppure *meno probabile* di un altro, dunque scriveremo

$$P((1, 1)) = P((1, 0)) = P((0, 1)) = P((0, 0)) = \frac{1}{4}.$$

È importante saper rispondere anche a domande un po' più complicate; ad esempio: qual è la probabilità che esca almeno una volta testa? Qual è la probabilità che esca esattamente una volta croce? Le risposte, molto semplici, sono rispettivamente 3/4 e 1/2; vediamo però un po' più in dettaglio quale ragionamento abbiamo fatto. Per il primo caso: gli eventi elementari che corrispondono al verificarsi dell'evento richiesto (cioè "esce almeno una volta testa") sono evidentemente $(1,1),(1,0),(0,1)$, e scriviamo

$$\{\text{esce almeno una volta testa}\} = \{(1,1),(1,0),(0,1)\} = A,$$

identificando così l'evento che ci interessa con un preciso sottoinsieme A di Ω. Per calcolare poi la probabilità, abbiamo sommato le probabilità degli eventi elementari $\omega \in A$, che sono note. In simboli

$$P(A) = \sum_{\omega \in A} P(\omega) = \frac{3}{4}.$$

Per la seconda domanda abbiamo

$$\{\text{esce esattamente una volta croce}\} = \{(1,0),(0,1)\} = B;$$

$$P(B) = \sum_{\omega \in B} P(\omega) = \frac{2}{4}.$$

Come risulta evidente da questo esempio, ogni *evento* può (e sarà) identificato con un ben preciso sottoinsieme di Ω. Inoltre, ad ogni evento abbiamo assegnato una probabilità. In altri termini, detta $\mathcal{A}$ la famiglia degli eventi, abbiamo costruito una "funzione" $P : \mathcal{A} \to \mathbb{R}^+$ (della quale fra poco studieremo le proprietà).
È appena il caso di far notare che la scrittura $P(\omega)$ usata per indicare la probabilità dell'evento elementare ω non è corretta (bisognerebbe scrivere invece $P(\{\omega\})$). Tuttavia continueremo ad utilizzarla per non appesantire le notazioni.

(3.3) Esempio. Qual è la probabilità che alla roulette (non truccata) esca un numero pari oppure un numero dispari superiore a 13? Si ha

$$\Omega = \{0,1,2,\ldots,36\} \qquad P(\omega) = \frac{1}{37};$$

poniamo poi

$$A = \{\text{ esce un numero pari}\} = \{0,2,4,\ldots,34,36\};$$

$$B = \{\,\text{esce un numero dispari superiore a } 13\} = \{15, 17, 19, \ldots, 33, 35\};$$

si ha

$$P(A) = \frac{19}{37}$$

e

$$P(B) = \frac{11}{37}.$$

Si chiede allora di calcolare $P(A \cup B)$ (cioè la probabilità che si verifichi A oppure B) e si ha

$$P(A \cup B) = \frac{19}{37} + \frac{11}{37} = \frac{30}{37}.$$

È chiaro che la somma è lecita perché i due eventi A e B sono disgiunti (o *incompatibili*).
La tecnica che abbiamo usato sopra (sommare la probabilità di eventi disgiunti per calcolare la probabilità dell'unione) è molto naturale per una funzione $P : \mathcal{A} \to \mathbb{R}^+$ "candidata" per essere una probabilità.

Ricapitolando, per modellizzare dal punto di vista matematico un fenomeno aleatorio dovremo:
(i) Costruire lo spazio campione Ω.
(ii) Stabilire una famiglia $\mathcal{A}$ di sottoinsiemi di Ω che rappresentino la famiglia degli eventi "interessanti" per il nostro esperimento.
(iii) Definire una funzione P su $\mathcal{A}$ e a valori in $\mathbb{R}^+$ che fornisca le nostre valutazioni di probabilità per i singoli eventi.

Siamo ora in grado di dare le definizioni precise:

(3.4) Definizione. Sia Ω un insieme. Una famiglia $\mathcal{A}$ di sottoinsiemi di Ω è una σ- *algebra* se verifica le seguenti proprietà:
(i) $\Omega \in \mathcal{A}$;
(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$;
(iii) Se $(A_n)_{n \in \mathbb{N}}$ è una successione di elementi di $\mathcal{A}$, allora $\bigcup_n A_n \in \mathcal{A}$.

(3.5) Osservazioni. (a) La famiglia $\mathcal{A}$ viene detta σ- *algebra degli eventi.* La proprietà $A \in \mathcal{A}$ si può esprimere a parole dicendo semplicemente : "A è un evento".

(b) Ω viene chiamato anche *evento certo.*

(c) Se A è un evento, A^c (che è un evento per l'assioma (ii)) si chiama *evento contrario di A*.

(d) Dalle proprietà (i) e (ii) segue facilmente che $\emptyset = \Omega^c \in \mathcal{A}$. Il sottoinsieme $\emptyset$ viene detto anche *evento impossibile*.

(e) L'assioma (iii) richiede in particolare che, se una famiglia finita di sottoinsiemi di Ω è *interessante* per l'esperimento che stiamo considerando, anche la loro unione lo sia (ciò si esprime dicendo che $\mathcal{A}$ è *stabile rispetto all'unione finita*). Ciò è del tutto ragionevole: non si vede perché, se ad esempio sono importanti il verificarsi di A e il verificarsi di B, non debba esserlo anche il verificarsi di uno dei due (A oppure B).

In realtà nell'assioma (iii) si pretende di più: precisamente si vuole che sia un evento anche l'unione di ogni successione numerabile di eventi (in altri termini, A è *stabile rispetto all'unione numerabile*). Il motivo è di natura tecnica: spesso alcuni sottoinsiemi interessanti di Ω si esprimono in modo molto naturale come unioni numerabili di eventi, ma non come unioni finite, e dunque con la sola proprietà di stabilità finita non sarebbe possibile considerarli eventi.

A noi capiterà nel seguito di dover usare la proprietà di stabilità numerabile, ma in quasi ogni situazione ci sarà sufficiente la sola proprietà di stabilità finita.

(f) Dagli assiomi (ii) e (iii) e dalle leggi di De Morgan discende facilmente la seguente proprietà (*stabilità di $\mathcal{A}$ rispetto all'intersezione numerabile*):

(iv) Se $(A_n)_{n\in\mathbb{N}}$ è una successione di elementi di $\mathcal{A}$, allora $\bigcap_n A_n \in \mathcal{A}$.

(3.6) Definizione. Siano Ω un insieme, $\mathcal{A}$ una σ-algebra di sottoinsiemi di Ω. Un'applicazione $P : \mathcal{A} \to \mathbb{R}^+$ si chiama *probabilità* (su $\mathcal{A}$) se verifica le due condizioni seguenti:
(i) $P(\Omega) = 1$;
(ii) Se $(A_n)_{n\in\mathbb{N}}$ è una successione di elementi di $\mathcal{A}$ due a due disgiunti (cioè tali che $\forall\, i \neq j$ si abbia $A_i \cap A_j = \emptyset$) risulta $P(\bigcap_n A_n) = \sum_n P(A_n)$.

(3.7) Osservazione. L'esempio (3.3) mostra che l'assioma (ii) è assolutamente ragionevole, almeno nel caso di una famiglia finita di eventi (e in tal caso si dice anche che P gode della proprietà di *additività finita*). L'assioma (ii) nella sua versione completa (detto di *σ-additività*) è dovuto alla necessità di assegnare una probabilità anche ad unioni numerabili di eventi.

(3.8) DEFINIZIONE. Siano Ω un insieme, $\mathcal{A}$ una σ-algebra di sottoinsiemi di Ω, P una probabilità (su $\mathcal{A}$). La terna $(\Omega, \mathcal{A}, P)$ si chiama *spazio di probabilità*, od anche *spazio probabilizzato*.

Ricapitolando, dunque, modellizzare un esperimento aleatorio significa costruire uno spazio di probabilità $(\Omega, \mathcal{A}, P)$ tale che

(i) Ω sia l'insieme di tutti i possibili risultati dell'esperimento;

(ii) $\mathcal{A}$ sia la famiglia degli eventi "interessanti" (dal punto di vista dell'esperimento in questione);

(iii) P sia la valutazione di probabilità per ciascun evento, assegnata in modo da rispettare il più possibile la situazione pratica che abbiamo davanti. È bene però tenere presente che tale valutazione è comunque soggettiva.

Nel prossimo paragrafo vedremo come gli assiomi imposti nelle definizioni permettano di effettuare il calcolo della probabilità di eventi di tipo particolare.

NOTA IMPORTANTE. Nel caso di insiemi Ω a cardinalità finita o numerabile (cioè con un numero di elementi finito o numerabile), tutti i sottoinsiemi di Ω sono considerati eventi, cioè si pone <u>sempre</u> $\mathcal{A} = \mathcal{P}(\Omega)$.
Tuttavia esistono situazioni più complicate nelle quali, per motivi di tipo matematico, non è possibile fare la stessa scelta. In questi casi dobbiamo contentarci di scegliere una σ-algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$. Ma queste considerazioni vanno oltre lo scopo di questa esposizione, e non ci soffermeremo oltre su questo punto.

4. Alcune conseguenze degli assiomi.

Sia $(\Omega, \mathcal{A}, P)$ uno spazio di probabilità, fissato nel seguito una volta per tutte.
Siano A e B due eventi. Si può allora scrivere

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

Gli eventi $(A \cap B)$ e $(A^c \cap B)$ sono evidentemente disgiunti (gli elementi di A non possono appartenere anche a A^c e viceversa); dunque, per la proprietà di additività di P abbiamo

$$(4.1) \qquad P(B) = P(A \cap B) + P(A^c \cap B).$$

Questa formula è già di per sé piuttosto importante nei calcoli: capita frequentemente di non saper calcolare direttamente la probabilità di un evento B, ma di saper "spezzare" B in due parti disgiunte (tramite un evento ausiliario A) la cui probabilità è facilmente calcolabile. La (4.1) ammette una generalizzazione al caso in cui B debba essere scomposto in più di 2 parti, e si chiama, per ovvi motivi, *formula della partizione dell'evento certo.*

La (4.1) ha alcune conseguenze importanti.

(i) Per $B = \Omega$ e per ogni $A \in \mathcal{A}$ la (4.1) diventa (ricordando l'assioma (3.6) (i)):

$$1 = P(A) + P(A^c),$$

ovvero

$$(4.2) \qquad P(A^c) = 1 - P(A).$$

La (4.2) permette dunque di calcolare la probabilità dell'evento contrario di A se è nota la probabilità di A, (ed è intuitivamente accettata da tutti).

(ii) Per $A \subseteq B$ la (4.1) dà

$$(4.3) \qquad P(B) = P(A) + P(A^c \cap B),$$

e quindi, poiché $P(A^c \cap B) \geq 0$, si ricava, per $A \subseteq B$,

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

Anche questa proprietà è del tutto naturale, e si chiama proprietà di *isotonia* della probabilità.

Una sua semplice conseguenza è il fatto che ogni evento ha probabilità inferiore o uguale a 1: infatti, se A è un evento, dalla ovvia relazione $A \subseteq \Omega$ segue

$$P(A) \leq P(\Omega) = 1.$$

(In altre parole, ogni funzione *probabilità* assume valori nell'intervallo $[0,1]$).

(iii) Se $A \subseteq B$, si usa scrivere

$$A^c \cap B = B \setminus A,$$

e la (4.3) diventa

$$P(B \setminus A) = P(B) - P(A).$$

(iv) Siano ora A e B due eventi qualsiasi (non necessariamente disgiunti). Possiamo scrivere

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

essendo i due eventi A e $A^c \cap B$ disgiunti, dalla proprietà di additività della probabilità e dalla (4.1) si deduce

$$P(A \cup B) = P(A) + P(A^c \cap B) = P(A) + P(B) - P(A \cap B) \tag{4.4}$$

Questa formula permette di calcolare la probabilità dell'unione di due eventi A e B quando siano note $P(A)$, $P(B)$ e $P(A \cap B)$.

La (4.4) è un caso particolare di una formula nota con il nome di *formula di inclusione-esclusione*, che si utilizza per il calcolo della probabilità dell'unione di una famiglia finita di eventi. Per completezza diamo (senza dimostrazione) la formula anche per il caso di tre eventi A, B e C:

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

A questo punto non dovrebbe essere difficile capire come le formule precedenti si generalizzano al caso di $n \geq 3$ eventi.

(v) Sia $(A_n)_n$ una successione di eventi. Poiché

$$(\cup_n A_n)^c = (\cap_n A_n^c),$$

dalla formula (4.2) si ricava

$$P(\cup_n A_n) = 1 - P(\cap_n A_n^c). \tag{4.5}$$

Questa formula è spesso utile perché tipicamente la probabilità di un'intersezione di eventi risulta più facile da calcolare della probabilità di un'unione.

(4.6) Esempio. Si lancia n volte una moneta equilibrata. Quanto vale la probabilità che la faccia "Testa" (contrassegnata convenzionalmente con il simbolo 1) esca almeno una volta nel corso degli n lanci?
Abbiamo già risposto a questa domanda per il caso $n = 2$ nell'esempio (3.2), dove abbiamo in pratica elencato tutti gli eventi elementari che rispondevano alla questione, sommando poi le loro probabilità. Nel caso generale ciò non è evidentemente possibile. Cominciamo allora con la costruzione dello spazio di probabilità che modellizza l'esperimento in questione. Analogamente al caso $n = 2$ già visto, lo spazio campione Ω sarà costituito da tutte le sequenze di lunghezza n composte dai simboli 0 e 1, e cioè, per usare una scrittura comoda

$$\Omega = \{0, 1\}^n.$$

Osserviamo che $\text{card}(\Omega) = 2^n$.
Anche in questo caso non c'è motivo di ritenere che una particolare sequenza sia più o meno probabile di un'altra, dato che la moneta è equilibrata, e dunque si pone

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

Poniamo ora, per $k = 1, 2, \ldots, n$

$$A_k = \{\text{esce}\,1\,\text{al}\,k - \text{esimo}\,\text{lancio}\}.$$

A noi interessa evidentemente calcolare $P(\cup_{k=1}^n A_k)$. Per la formula (4.5) si ha subito

$$P(\cup_{k=1}^n A_k) = 1 - P(\cap_{k=1}^n A_k^c).$$

D'altra parte è chiaro che

$$P(\cap_{k=1}^n A_k^c) = P(\{\text{esce}\,0\,\text{ad}\,\text{ogni}\,\text{lancio}\}) = \frac{1}{2^n},$$

(l'evento qui sopra è infatti costituito dall'unica sequenza di lunghezza n composta da soli simboli 0) e quindi la probabilità cercata vale

$$1 - \frac{1}{2^n}.$$

Tutte le proprietà viste fino a questo momento sono conseguenza della sola proprietà di additività finita della probabilità. Le ultime due, che enunciamo senza dimostrazione, utilizzano invece la $\sigma-$additività.

(vi) (*proprietà di passaggio al limite sulle successioni crescenti*). Sia $(A_n)_n$ una successione crescente di eventi (ciò significa che , $\forall n \in \mathbb{N}$, risulta $A_n \subseteq A_{n+1}$. Si ha allora

$$P(\cup_n A_n) = \lim_{n\to\infty} P(A_n)$$

(vii) (*proprietà di passaggio al limite sulle successioni decrescenti*). Sia $(A_n)_n$ una successione decrescente di eventi (ciò significa che , $\forall n \in \mathbb{N}$, risulta $A_n \supseteq A_{n+1}$. Si ha allora

$$P(\cap_n A_n) = \lim_{n\to\infty} P(A_n).$$

5. Probabilità condizionale e indipendenza.

Cominciamo al solito con un esempio.

(5.1) Esempio. Da un'urna contenente 10 palline numerate da 1 a 10 si esegue un'estrazione. Quanto vale la probabilità che la pallina estratta porti un numero ≤ 5?

Sia $\Omega = \{1, 2, 3, \ldots, 10\}$; poiché ogni pallina ha la stessa probabilità di essere estratta delle altre, per ogni evento elementare $\omega \in \Omega$ poniamo

$$P(\omega) = \frac{1}{\operatorname{card}\Omega} = \frac{1}{10}$$

Poiché

$$A = \{\text{esce un numero } \leq 5\} = \{1,2,3,4,5\},$$

avremo

$$P(A) = \frac{\text{card}A}{\text{card}\Omega} = \frac{5}{10}.$$

In altre parole, le informazioni a nostra disposizione ci hanno indotto a costruire lo spazio di probabilità sopra detto, e di conseguenza abbiamo ottenuto, come "grado di fiducia" del verificarsi di A, il numero 5/10.

Supponiamo adesso che un tizio che ha assistito all'estrazione ci abbia detto che è uscito un numero pari. Questa informazione supplementare ci induce a cambiare il nostro "grado di fiducia" del verificarsi di A? In caso affermativo, quale nuova valutazione di probabilità dovremo dare?
La risposta è molto semplice: adesso il nostro spazio campione sarà ridotto al sottoinsieme di Ω

$$B = \{\text{esce un numero pari}\} = \{2,4,6,8,10\};$$

Di nuovo per ogni evento elementare ω avremo

$$P(\omega) = \frac{1}{\text{card}\, B} = \frac{1}{5},$$

e, poiché due elementi di B sono ≤ 5, (cioè $\text{card}\,(A \cap B) = 2$), concludiamo che un numero inferiore o uguale a 5 esce con probabilità pari a

$$\frac{\text{card}\,(A \cap B)}{\text{card}\, B} = \frac{2}{5}.$$

Analizziamo un po' più in dettaglio il ragionamento fatto, per cercare di ricavarne il senso generale.
Possiamo scrivere

$$\frac{\text{card}\,(A \cap B)}{\text{card}\, B} = \frac{\text{card}\,(A \cap B)/\text{card}\,\Omega}{\text{card}\, B/\text{card}\,\Omega} = \frac{P(A \cap B)}{P(B)}.$$

Cioè la nostra valutazione del verificarsi di A *sapendo che si è verificato* B è stata

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

Questo esempio giustifica la definizione seguente

(5.2) DEFINIZIONE. Siano A, B due eventi, con $P(B) > 0$. Si chiama *probabilità condizionale di A, dato B* (oppure "sapendo che si è verificato B") la quantità

$$P(A|B) \doteq \frac{P(A \cap B)}{P(B)}. \tag{5.3}$$

(5.4) OSSERVAZIONE. Per quanto ovvio, si sottolinea il fatto che la scrittura $P(A|B)$ non rappresenta la probabilità dell'evento "$A|B$" (che non esiste!), ma solo un modo per indicare il rapporto che compare al secondo membro della (5.3).

Il nome di "probabilità" dato alla quantità $P(A|B)$ è giustificato dalla seguente proposizione:

(5.5) PROPOSIZIONE. Sia B un evento, con $P(B) > 0$. L'applicazione $Q : \mathcal{A} \to \mathbb{R}^+$ definita da $Q(A) = P(A|B)$ è una probabilità (nel senso della Definizione (3.6)).

Può capitare che l'informazione supplementare "si è verificato B" non cambi la nostra valutazione della probabilità del verificarsi di A.

(5.6) ESEMPIO. Si lancia due volte una moneta equilibrata. Poniamo

$$A = \{\text{esce testa al secondo lancio}\},$$

$$B = \{\text{esce testa al primo lancio}\}.$$

È immediato verificare che

$$P(A|B) = P(A) = \frac{1}{2}.$$

Diamo dunque la seguente definizione (che sarà però provvisoria):

(5.7) DEFINIZIONE. Siano A, B due eventi, con $P(B) > 0$. A e B si dicono *indipendenti (tra loro)* se

$$P(A|B) = P(A). \tag{5.8}$$

Usando la (5.3), la (5.8) si può scrivere nella forma

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

od anche, moltiplicando per $P(B)$,

$$P(A \cap B) = P(A)P(B). \tag{5.9}$$

La (5.9) è chiaramente simmetrica in A e B, e del resto lo è anche l'espressione *A e B sono indipendenti (tra loro)* usata nella definizione (5.7), mentre non è affatto evidente che lo sia anche la (5.8). Inoltre, mentre la (5.8) perde di significato se B è un evento di probabilità nulla, la (5.9) ha ancora senso. Essa risulta quindi più comoda come definizione di indipendenza; dunque sostituiremo la definizione provvisoria con la seguente:

(5.10) Definizione. Due eventi A e B si dicono *indipendenti (tra loro)* se vale la (5.9).

Sarà necessario estendere la definizione di indipendenza ad una famiglia qualsiasi di eventi. Consideriamo per iniziare il caso di tre eventi A, B e C. Per analogia con il caso di due eventi, potremmo pensare che sia corretto dire che essi sono indipendenti se

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{5.11}$$

Purtroppo questa condizione non garantisce che gli eventi siano anche "due a due"indipendenti, come ci dice invece l'intuizione (se ad esempio A, B e C sono tre eventi, con $C = \emptyset$, la relazione (5.11) è soddisfatta anche nel caso in cui A e B non siano indipendenti). La definizione ragionevole è allora la seguente

(5.12) Definizione. Tre eventi A, B e C si dicono *indipendenti* se valgono tutte le condizioni seguenti

$$P(A \cap B \cap C) = P(A)P(B)P(C); \tag{i}$$

$$P(A \cap B) = P(A)P(B); \tag{ii}$$

(iii) $$P(B \cap C) = P(B)P(C);$$

(iv) $$P(A \cap C) = P(A)P(C).$$

La definizione precedente si generalizza in modo ovvio ad una famiglia qualsiasi di eventi: essi si diranno indipendenti se per ogni sottofamiglia finita la probabilità dell'intersezione è uguale al prodotto delle singole probabilità; in termini precisi:

(5.13) DEFINIZIONE. Gli eventi $(A_i)_{i \in I}$ si dicono *indipendenti (fra loro)* se per ogni $k \in \mathbb{N}$ e per ogni successione finita $i_1, i_2 \ldots, i_k$ di indici di I, si ha

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

(5.14) OSSERVAZIONE. Siano A e B due eventi. Le affermazioni seguenti sono equivalenti

(a) A e B sono indipendenti;
(b) A e B^c sono indipendenti;
(c) A^c e B sono indipendenti;
(d) A^c e B^c sono indipendenti.

Dell'osservazione precedente si può dare una ovvia formulazione, solo formalmente più complicata, anche per il caso di una famiglia di $n > 2$ eventi.

L'ipotesi di indipendenza ci consentirà la costruzione dello schema di Bernoulli. Questa costruzione sarà l'oggetto del prossimo paragrafo.

6. La costruzione del modello. Il problema dell'indipendenza e il ruolo della Statistica.

Nel seguito, indicheremo con n ($n \in \mathbb{N}$) il numero di prove effettuate, e con p ($p \in (0,1)$) la probabilità di ottenere 1 (= *successo*) nella generica prova. Queste due quantità sono i *parametri* del nostro schema.
Come abbiamo già fatto nell'esempio (4.6), porremo

$$\Omega = \{0,1\}^n.$$

In questo caso, tuttavia, lo spazio che costruiremo non potrà assegnare la stessa probabilità ad ogni evento elementare (tranne che nel caso, già discusso in (4.6), in cui $p = 1/2$). Quale sarà allora la probabilità ragionevole da assegnare ad un generico sottoinsieme di Ω? Grazie alla formula

$$P(A) = \sum_{\omega \in A} P(\omega)$$

che è conseguenza ovvia dell'additività di P, basterà farlo per gli eventi elementari

$$\omega = (\omega_1, \omega_2, \ldots, \omega_n)$$

(con $\omega_i \in \{0, 1\}$). Per chiarezza tratteremo dapprima un caso numerico.

(6.1) ESEMPIO. Poniamo $n = 5$, e consideriamo la successione di risultati $\omega = (1, 1, 0, 1, 0)$.
Per $i = 1, \ldots, 5$ consideriamo gli eventi

$$A_i = \{\text{esce 1 alla prova } i - \text{esima}\}.$$

Evidentemente

$$\omega = A_1 \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c.$$

La scrittura precedente, pur corretta, non è di alcun interesse pratico, a meno di non saper calcolare in qualche modo $P(A_1 \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c)$, e in generale il calcolo della probabilità di una intersezione di eventi non è affatto un problema semplice. Lo è però **se gli eventi in questione sono indipendenti**, perché, come abbiamo visto in precedenza, in tal caso la probabilità della loro intersezione si può fattorizzare nel prodotto delle singole probabilità.

Ma cosa significa, nel nostro caso, dire che gli eventi $A_1, A_2, A_3^c, A_4, A_5^c$ sono indipendenti? Significa che, ad esempio, il fatto che alla quinta prova abbiamo un *insuccesso* (0) non cambia il nostro grado di fiducia del fatto che, alla prima prova abbiamo un *successo* (1), alla seconda ancora 1, e così via. Poiché questa considerazione deve valere non solo per l'evento elementare ω in questione, ma per ogni altro possibile risultato dell'esperimento, ciò (in linguaggio un po' improprio) si traduce nel fatto che le n prove sono eseguite in modo tale che il risultato di uno qualsiasi

di esse non influenzi le altre, ovvero **le n prove sono tra loro indipendenti**.
Introducendo dunque l' IPOTESI che le n prove siano tra loro indipendenti, possiamo andare avanti con i calcoli. Troviamo

$$P(\omega) = P(A_1 \cap A_2 \cap A_3^c \cap A_4 \cap A_5^c) = P(A_1)P(A_2)P(A_3^c)P(A_4)P(A_5^c).$$

A questo punto non ci resta altro da fare che *decidere* quali valori assegnare a $P(A_1)$, $P(A_2)$, $P(A_3^c)$, $P(A_4)$, $P(A_5^c)$: ad esempio, la quantità $P(A_1)$ deve essere la probabilità che la prima prova abbia avuto successo, e dunque la scelta naturale sarà

$$P(A_1) = p. \tag{6.2}$$

In modo analogo assegniamo gli altri valori:

$$P(A_2) = P(A_4) = p; \quad P(A_3^c) = P(A_5^c) = 1 - p. \tag{6.3}$$

Dovrebbe risultare chiaro che anche le relazioni (6.2) e (6.3) sono delle IPOTESI; si tratta della traduzione matematica rigorosa dell'espressione (fin qui usata in modo solo intuitivo): **le prove sono identiche fra loro e p è la probabilità di ottenere un successo in una singola prova.** Si conclude dunque che

$$P(\omega) = p^3(1-p)^2.$$

È facile capire che i numeri p e $1-p$ compaiono con esponenti 3 e 2 rispettivamente perché, nella sequenza ω considerata, 3 (risp. 2) è il numero di simboli uguali a 1 (risp. 0).

Il ragionamento dell'esempio è utilizzabile, evidentemente, anche nel caso generale; sia dunque $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ (con $\omega_i \in \{0, 1\}$) un elemento di $\Omega = \{0, 1\}^n$; osserviamo che le quantità

$$\sum_{i=1}^n \omega_i, \qquad n - \sum_{i=1}^n \omega_i$$

non sono altro che il numero di simboli uguali a 1 e il numero di simboli uguali a 0 che formano la sequenza ω. Pertanto avremo

$$P(\omega) = p^{(\sum_{i=1}^{n} \omega_i)}(1-p)^{(n-\sum_{i=1}^{n} \omega_i)}.$$

(6.4) Osservazione. La formula precedente chiarisce l'utilità dei simboli 1 e 0 per indicare il *successo* e l'*insuccesso* rispettivamente.

Discutiamo ora brevemente le ipotesi fatte.

Prima ipotesi: le prove sono identiche fra loro. È chiaro che si tratta di un'idealizzazione: nella pratica quasi mai le prove sono davvero tutte identiche fra loro; nel caso dell'operazione antimafia, ad esempio, ciò significherebbe che i 5 malviventi sono tutti egualmente "abili" a scappare, il che è evidentemente poco realistico. E anche in casi più "ripetitivi", come i lanci successivi di una moneta, si tratta comunque di un'astrazione: da un lancio all'altro cambia la situazione ambientale esterna (pressione, temperatura, eccetera); la moneta stessa cambia per gli urti ricevuti nei lanci precedenti, e così via.

Seconda ipotesi: le prove sono tra loro indipendenti. Deve essere ben chiaro che il solo fatto di essere in presenza di un fenomeno che consista in una serie di n prove ripetute *non autorizza* affatto ad usare lo schema che abbiamo costruito, a meno che non si sappia che le prove stesse siano state effettuate in condizioni di indipendenza, o non sia almeno plausibile che questo sia il caso. Purtroppo, nei casi pratici può capitare di non avere informazioni in questo senso. Riferendoci ancora all'operazione di polizia, l'ipotesi di indipendenza potrebbe essere ritenuta valida se ci fosse noto che i malviventi non sono in alcun modo in contatto tra loro (per esempio tramite telefoni cellulari). Ma, se i 5 fanno parte di una stessa banda, questo non sarà probabilmente il caso.

La domanda spontanea a questo punto è: ma se le ipotesi non sono valide, (o comunque non siamo in possesso di informazioni sufficienti per poterle ritenere tali), perché usare lo schema di Bernoulli?

Ciò che si fa in pratica, in tal caso, è *supporre* di essere effettivamente in presenza di prove identiche e fra loro indipendenti, sperando che questa sia un'ipotesi ragionevole, e affidando poi allo statistico il compito di verificare se i risultati ottenuti facendo i conti con lo schema di Bernoulli si accordano con la realtà. Tale è infatti uno dei ruoli della Statistica: controllare, appunto, se i risultati previsti dal Calcolo delle Probabilità sono in accordo con i risultati numerici effettivamente osservati; di più, spesso lo statistico, studiando le discrepanze fra il modello teorico e la realtà, è in grado di suggerire in quali punti il primo debba essere modificato per ottenere un miglior accordo con i dati reali. Dunque, nei casi incerti, lo schema di Bernoulli va considerato comunque un buon punto di partenza.

Per terminare questo paragrafo, illustriamo molto brevemente un esempio di prove ripetute certamente <u>non</u> indipendenti.

(6.5) Esempio. Il gioco del lotto. Nel gioco del lotto (supponiamo una sola ruota per semplicità), si estraggono 5 palline da un'urna che ne contiene 90, numerate da 1 a 90. Le successive estrazioni sono effettuate, come si suol dire, *senza rimpiazzo*, cioè ogni volta la pallina estratta viene gettata via (dopo aver preso nota del numero riportato su di essa, naturalmente). Per fissare le idee, supponiamo che un tizio abbia puntato sul 53.

Qui siamo in presenza di 5 prove ripetute (cioè identiche): mettendoci dalla parte del giocatore, diremo che la generica prova ha avuto successo se il 53 esce, (e ovviamente che ha avuto insuccesso nel caso contrario). La probabilità di ottenere il successo in ogni singola prova è evidentemente $p = 1/90$. Tuttavia, in questo caso è semplice capire che le 5 prove <u>non</u> possono essere ritenute indipendenti: se ad esempio la prima prova ha successo (cioè se il 53 esce alla prima estrazione), le prove successive daranno necessariamente un insuccesso (detto in termini più rigorosi: per esempio, la probabilità <u>condizionale</u> che il 53 esca nella seconda estrazione, <u>sapendo</u> che esso è già uscito alla prima, vale 0, mentre la probabilità <u>non condizionale</u> dello stesso evento vale 1/90).

7. La densità binomiale.

Per poter rispondere alle domande poste negli esempi del primo paragrafo, è necessario introdurre brevemente anche la densità binomiale.

(7.1) La densità binomiale. Siano $n \in \mathbb{N}$ e $p \in (0,1)$ due numeri fissati. Si chiama *densità binomiale* di parametri n e p (in simboli, $B(n,p)$) la funzione $f : \mathbb{R} \to \mathbb{R}$ così definita

$$f(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & k = 0, 1, 2, \dots, n \\ 0 & \text{altrimenti.} \end{cases}$$

Il nome assegnato a questa funzione deriva dalla cosiddetta *formula del binomio di Newton*, nel cui sviluppo appaiono i *coefficienti binomiali* $\binom{n}{k}$.

Una tipica situazione in cui la densità binomiale compare in modo naturale è proprio quella di uno schema di Bernoulli, come vedremo fra un attimo.

Sia $(\Omega, \mathcal{A}, P)$ lo schema delle prove ripetute indipendenti di parametri (noti) n e p. Ricordiamo che ciò significa che

$$\Omega = \{0, 1\}^n$$

ed inoltre, se $\omega = (\omega_1, \dots, \omega_n)$ è un evento elementare (con $\omega_i \in \{0, 1\}$ per ogni $i = 1, \dots, n$), si pone

$$P(\omega) = p^{(\sum_{i=1}^n \omega_i)} (1-p)^{(n - \sum_{i=1}^n \omega_i)}$$

Poniamo ora, per ogni $i = 1, \dots, n$

$$X_i(\omega) = \omega_i;$$

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = \sum_{i=1}^n \omega_i.$$

Le X_i e la X sono funzioni definite su Ω (in realtà sono qualcosa di più, ma non ci soffermeremo su questo punto). Diamone un'interpretazione concreta. Per fissare le idee, supponiamo che sia $n = 5$, e $\omega = (1, 1, 0, 1, 0)$. Ciò significa che, nello svolgimento delle 5 prove, la prima, la seconda e la quarta hanno dato "successo", la terza e la quinta "insuccesso".
Si ha subito

$$X_1(\omega) = X_2(\omega) = X_4(\omega) = 1, \; X_3(\omega) = X_5(\omega) = 0.$$

Inoltre

$$X(\omega) = \sum_{i=1}^n X_i(\omega) = 1 + 1 + 1 = 3.$$

La funzione X_i si chiama anche (variabile) *indicatrice* della prova $i-$esima (dice cosa è accaduto alla prova $i-$esima); la funzione X "conta" il numero di successi ottenuti nel corso di tutte le prove (anche da qui si capisce la comodità dei simboli 1 e 0: per avere il numero totale di successi, basta sommare i numeri che compaiono nella sequenza considerata).

Tornando al caso generale, ci interessa calcolare, per ogni $k \in \mathbb{R}$, la probabilità di ottenere esattamente k successi nel corso delle n prove, ovvero la probabilità dell'evento

$$\{\omega \in \Omega : X(\omega) = k\};$$

per semplificare le notazioni, questo evento viene indicato con la scrittura $\{X = k\}$ e la relativa probabilità con $P(X = k)$ (anziché con $P(\{\omega \in \Omega : X(\omega) = k\})$, come invece sarebbe corretto fare).

È evidente che la X assume i soli valori interi $k = 0, 1, 2, \ldots, n$. Pertanto le uniche probabilità $P(X = k)$ (eventualmente) non nulle sono quelle che corrispondono a tali valori. Sia dunque k un intero fissato, con $k \in \{0, 1, 2, \ldots, n\}$. L'evento $\{X = k\}$ è costituito dalle sequenze ω che contengono il simbolo "1 "esattamente k volte, cioè tali che

$$\sum_{i=1}^{n} \omega_i = k.$$

Pertanto, per calcolare la probabilità di $\{X = k\}$, basterà sommare le probabilità di tutte le sequenze del tipo sopra detto. Poiché esse hanno tutte la medesima probabilità, pari a

$$P(\omega) = p^{(\sum_{i=1}^{n} \omega_i)}(1-p)^{(n-\sum_{i=1}^{n} \omega_i)} = p^k(1-p)^{n-k},$$

per avere la somma basterà moltiplicare tale valore per il numero di sequenze ω appartenenti all'evento $\{X = k\}$. È facile capire che esso è esattamente il numero di combinazioni di n oggetti a k a k (è come se avessimo n caselle, k delle quali devono essere riempite con "1 ", mentre le restanti $n - k$ con "0 "). È noto che questo numero è uguale a $\binom{n}{k}$, e quindi otteniamo

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{7.2}$$

Si riconosce quindi la densità binomiale sopra definita, e la formula (7.2) si esprime a parole dicendo che X ammette densità binomiale di parametri n, p (in simboli, $X \sim B(n,p)$).

8. Come si usa lo schema di Bernoulli.

Per capire l'utilità dello schema di Bernoulli e per vedere come si fanno i calcoli, rispondiamo alle domande del paragrafo 1.

1. Sia X il numero di "Teste "ottenute nei 5 lanci. Allora X è il numero di successi in uno schema di Bernoulli di parametri $n = 5$, $p = 0.6$, dunque ha densità B(5, 0.6). Il capitale del giocatore al termine dei 5 lanci è pari a $10X - 5(5 - X) = 15X - 25$ (in euro). La domanda chiede di calcolare la probabilità che tale capitale sia strettamente positivo, cioè la probabilità dell'evento $\{\omega \in \Omega : 15X(\omega) - 25 > 0\}$ (che, al solito per brevità, indichiamo semplicemente con $\{15X - 25 > 0\}$. Analoghe scritture abbreviate, sia per gli eventi che per le loro probabilità, saranno utilizzate negli esempi successivi). Si vede facilmente che, se $X = 0$ oppure $X = 1$, il capitale è strettamente negativo, mentre per tutti gli altri valori che X può assumere (e cioé 2, 3, 4, 5) esso è strettamente positivo. Dunque si può scrivere

$$\{15X - 25 > 0\} = \bigcup_{k=2}^{5} \{X = k\};$$

Gli eventi $\{X = k\}$, al variare di $k = 0, 1, 2, 3, 4, 5$ sono due a due disgiunti (se ad esempio $X(\omega) = 1$, allora $X(\omega) \neq 2$, e così via); quindi, per la proprietà di additività della probabilità si ha, dalla formula della densità binomiale

$$P(15X - 25 > 0) = \sum_{k=2}^{5} P(X = k) = \sum_{k=2}^{5} \binom{5}{k} (0.6)^k \times (0.4)^{5-k} = 0.91296.$$

Alternativamente, con lo stesso tipo di ragionamento, si ha

$$\begin{aligned} &P(15X - 25 < 0) = P(X = 0) + P(X = 1) \\ &= \binom{5}{0} (0.6)^0 \times (0.4)^5 + \binom{5}{1} (0.6)^1 \times (0.4)^4 = 0.08704, \end{aligned}$$

da cui, per passaggio al complementare,

$$P(15X - 25 > 0) = 1 - 0.08704 = 0.91296.$$

NOTA. Nei calcoli che seguiranno sottintenderemo, per brevità, i ragionamenti che utilizzano l'additività della probabilità.

2. Per fissare le idee, supponiamo che il giocatore abbia scommesso sul numero 5, e sia X il numero di volte nelle quali il 5 appare sulle due ruote. Supponendo (raginevolmente) che esse funzionino in modo indipendente l'una dall'altra, $X \sim B(2, 1/6)$; la probabilità che il giocatore perda 1 euro è pari $P(X = 0) = \binom{2}{0}(1/6)^0 \times (5/6)^2 = 0.695$; la probabilità che il giocatore vinca 1 euro è pari a $P(X = 1) = \binom{2}{1}(1/6)^1 \times (5/6)^1 = 0.278$; infine la probabilità che il giocatore vinca 2 euro è pari a $P(X = 2) = \binom{2}{2}(1/6)^2 \times (5/6)^0 = 0.028$. Il valore medio della vincita è pertanto

$$-1 \times 0.695 + 1 \times 0.278 + 2 \times 0.028 = -0.361,$$

e il gioco non è equo, anzi, si può affermare che, alla lunga, il giocatore perde il 36% delle volte.

3. Supponiamo che ogni figlio abbia probabilità pari a 1/2 di ricevere ciascuno dei due geni (A o a) dal padre, e la stessa cosa accada per il gene ricevuto dalla madre. Supponiamo inoltre che il padre trasmetta il suo gene indipendentemente dalla madre. Ciò implica che un figlio sarà di genotipo AA, aa oppure Aa con probabilità rispettive 1/4, 1/4, 1/2. Dunque la probabilità che un figlio sia di fenotipo A sarà pari a 3/4. Se i fenotipi dei 4 figli si presentano in modo indipendente da un figlio all'altro, allora il numero X di figli di fenotipo A ha densità $B(4, 3/4)$, e quindi

$$P(X = 3) = \binom{4}{3}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 = \frac{27}{64} = 0.422.$$

A proposito delle ipotesi (in particolare l'indipendenza dei fenotipi e anche il fatto che ogni figlio abbia probabilità 1/2 di ricevere un gene da ognuno dei genitori), val la pena di osservare che un genetista avrebbe forti dubbi sulla loro ragionevolezza.

4. Supponiamo che le viti siano difettose o meno in modo indipendente l'una dall'altra. Se X è il numero di viti difettose in una confezione da 10, allora $X \sim B(10, 0.01)$. Quindi la probabilità cercata vale

$$1 - P(X=0) - P(X=1)$$
$$= 1 - \binom{10}{0}(0.01)^0 \times (0.99)^{10} - \binom{10}{1}(0.01)^1 \times (0.99)^9 = 0.004.$$

Dunque dovrà essere sostituito lo 0.4 % delle confezioni.

5. Se i giudici agiscono in modo indipendente l'uno dall'altro (per esempio se non possono scambiarsi opinioni prima di emettere la sentenza), il numero X di giudici che prende la decisione corretta (cioè ritiene l'imputato colpevole) ha densità $B(12, 0.81)$. L'imputato sarà (giustamente) condannato se $X \geq 9$, e la probabilità che questo accada vale

$$P(X \geq 9) = \sum_{k=9}^{12} P(X=k) = \sum_{k=9}^{12} \binom{12}{k}(0.81)^k \times (0.19)^{12-k} = 0.826.$$

6. Il numero di componenti funzionanti (X per il primo sistema, Y per il secondo) ha densità $B(n,p)$ ($n = 5$ per il primo sistema, $n = 3$ per il secondo); dunque, la probabilità che un sistema a 5 componenti sia funzionante vale

$$P(X \geq 3) = \binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p)^1 + p^5;$$

la stessa probabilità per un sistema a 3 componenti vale

$$P(Y \geq 2) = \binom{3}{2}p^2(1-p)^1 + p^3.$$

Quindi il sistema a 5 componenti risulta piú affidabile se

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3,$$

che semplificando dà $p > 1/2$.

7. Supponendo che le reazioni (positiva o negativa) dele 10 cavie siano l'una indipendente dall'altra, il numero X di cavie che presentano reazione positiva avrà una densità $B(10, 0.85)$. La probabilità che il farmaco venga messo in commercio è pari allora a

$$P(X \geq 8) = P(X=8) + P(X=9) + P(X=10)$$
$$= \binom{10}{8}(0.85)^8 \times (0.15)^2 + \binom{10}{9}(0.85)^9 \times (0.15)^1 +$$
$$+ \binom{10}{10}(0.85)^{10} \times (0.15)^0 = 0.8202.$$

L'ipotesi di indipendenza può essere ritenuta valida se le cavie sono state scelte in modo opportuno, ad esempio se fra di esse non ce ne sono alcune che hanno gli stessi genitori. In altre parole, dovrà essere cura dello sperimentatore scegliere gli animali "ad hoc", per evitare che le conclusioni risultino viziate in partenza.

8. Supponendo al solito l'indipendenza fra gli esiti delle varie catture, il numero X dei malviventi catturati al termine dell'operazione della polizia avrà densità $B(5, 0.67)$. La probabilità richiesta non è che

$$P(X \geq 4) = 1 - P(X \leq 1) = 1 - P(X=0) - P(X=1)$$
$$= 1 - (0.33)^5 - 5 \times (0.67) \times (0.33)^4 = 0.956.$$

9. Con l'usuale ipotesi di indipendenza, il numero X di lampadine fulminate ha densità $B(5, p)$. La probabilità che l'appartamento resti al buio è $P(X=5) = p^5$. Dunque dovrà essere

$$p^5 \leq 0.01,$$

ovvero

$$p \leq (0.01)^{0.2} = 0.398.$$

Osserviamo per inciso che l'ipotesi di indipendenza non sarebbe ragionevole se si sapesse che le lampadine dell'appartamento sono state collegate in serie; e forse non è ragionevole in assoluto perché normamente gli impianti elettrici delle abitazioni, pur non essendo costruiti in serie, sono costruiti in serie "a blocchi".

10. Supponendo che i vari motori abbiano funzionamenti indipendenti l'uno dall'altro (il che non è poi così assurdo), il numero X di motori rotti in un quadrimotore ha densità $B(4, 0.04)$, mentre lo stesso numero (che

indichiamo con Y) per il bimotore ha densità $B(2, 0.03)$. La probabilità per un quadrimotore di arrivare a destinazione è pari a

$$\begin{aligned}&P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)\\&= (0.04)^0 \times (0.96)^4 + 4 \times (0.04)^1 \times (0.96)^3 + 6 \times (0.04^2 \times (0.96)^2\\&= 0.9997.\end{aligned}$$

La stessa probabilità per un bimotore vale

$$P(Y \leq 1) = 1 - P(Y = 2) = 1 - (0.03)^2 = 0.9991.$$

Quindi è più affidabile un quadrimotore (sebbene di poco).

9. Un problema più difficile: perché e come modellizzare un numero infinito di prove (processo di Bernoulli). Il "paradosso della scimmia".

In questo paragrafo vedremo una situazione un po' più complicata: supponiamo di eseguire una *successione infinita* di prove di Bernoulli ripetute indipendenti, arrestandosi soltanto alla prima prova in cui si ottiene successo.

È chiaro che per modellizzare questo esperimento uno schema di Bernoulli del tipo fin qui studiato non può andare bene, perché esso presuppone di sapere fin dall'inizio quante prove si devono effettuare. Dunque avremo bisogno di uno *schema di Bernoulli ad infinite prove.* La costruzione non è semplice come la precedente (in particolare riposa su teoremi di teoria della misura niente affatto elementari, come il Teorema di Łomnicki-Ulam (1934)). La difficoltà principale sta in questo: per analogia con il caso finito, si prende come spazio campione Ω l'insieme di tutte le successioni infinite composte dai simboli 0 e 1; il problema sorge nel momento in cui si deve assegnare la probabilità a ciascun evento elementare, cioè ad ognuna di tali successioni. Il caso finito non ci aiuta, e in realtà non è difficile vedere che ogni probabilità P definita su Ω (cioè ogni funzione $P : \mathcal{A} \to \mathbb{R}$ che debba rispettare gli assiomi di una "probabilità") <u>deve</u> necessariamente assegnare <u>probabilità nulla</u> ad ognuna di tali successioni. Noi non entreremo ulteriormente nei dettagli. Ci limiteremo a fare una costruzione non rigorosa, mettendo però in luce i punti delicati della trattazione. Il nostro modello probabilistico sarà il seguente. Poniamo, per $i = 1, 2, 3, \ldots$

$$\omega_i = (\underbrace{0, 0, 0, \ldots}_{i-1 \text{volte}}, 1),$$

ed inoltre

$$\omega_\delta = (0, 0, 0, \ldots)$$

Il significato dei simboli qui sopra è abbastanza evidente. Notiamo solo che l'evento elementare ω_δ può essere espresso a parole dicendo "il successo non si ottiene mai". Porremo dunque

$$\Omega = \{\omega_\delta, \omega_1, \omega_2, \omega_3, \ldots\}$$

L'insieme Ω, questa volta non ha cardinalità finita, ma è comunque numerabile, e dunque anche in questo caso prenderemo $\mathcal{A} = \mathcal{P}(\Omega)$.
Dobbiamo ora definire la funzione P. Dallo schema delle prove indipendenti (di parametri k, p) sopra costruito, si deduce che, per ogni $k = 1, 2, \ldots$, si ha

$$P(\omega_k) = p(1-p)^{k-1};$$

di conseguenza

$$\begin{aligned} P(\omega_\delta) &= 1 - P(\cup_{k=1}^{\infty}\{\omega_k\}) = 1 - \sum_{k=1}^{\infty} P(\omega_k) \\ &= 1 - \sum_{k=1}^{\infty} p(1-p)^{k-1} = 1 - p \times \frac{1}{1-(1-p)} = 0. \end{aligned}$$

(la somma della "serie geometrica" $\sum_{k=0}^{\infty} x^k$ vale $1/(1-x)$ per $|x| < 1$).

Cioè: per quanto piccola (purché strettamente positiva) sia la probabilità p di ottenere "1" in un singolo lancio, l'eventualità che "1" non esca mai è "quasi impossibile" (cioè di probabilità nulla). Questo risultato è noto sotto il pittoresco nome di "paradosso della scimmia": una scimmia, battendo a caso sui tasti di una macchina da scrivere, prima o poi ne farà uscire l'intera *Divina Commedia.* Un nome più serio è "paradosso di Borel".

Vediamo in che cosa il ragionamento precedente non è rigoroso: per assegnare a $\{\omega_k\}$ la la probabilità "più ragionevole", (e cioè $p(1-p)^{k-1}$) abbiamo fatto ricorso allo schema delle prove indipendenti di parametri k, p. Ma cosa accade se dobbiamo calcolare la probabilità di $\{\omega_h\}$, con $h \neq k$? Lo schema in cui fare il calcolo dovrà essere questa volta quello di parametri h, p, dunque diverso dall'altro (nello schema di parametri k, p gli eventi elementari sono sequenze di lunghezza k, mentre in quello di parametri h, p sono sequenze di lunghezza $h \neq k$, dunque gli spazi campione sono diversi). In altre parole, siamo costretti a cambiare spazio di probabilità tutte le volte. Questo fatto non è rigorosamente giustificabile, a meno che non sia possibile guardare ciascuno di questi spazi di probabilità come *sottospazio* di un unico, "grande" spazio di probabilità che li *contenga* tutti (non ci soffermiamo ulteriormente sulla nozione di *sottospazio* e sul significato da dare al termine *contenere*). L'esistenza di questo unico spazio più grande $(\Omega, \mathcal{A}, P)$, che permette di dare un senso a tutto il nostro calcolo è appunto conseguenza del teorema sopra ricordato. Aggiungiamo solo che, se, per ogni intero n, si considera la variabile indicatrice dell'n-esima prova, e cioè la funzione (definita su Ω)

$$X_n(\omega) = \begin{cases} 1 & \text{se la } n\text{-esima coordinata di } \omega \text{ vale } 1 \\ 0 & \text{se la } n\text{-esima coordinata di } \omega \text{ vale } 0, \end{cases}$$

ovvero, come ci si esprime in genere, (e forse in modo più chiaro, per quanto non matematicamente rigoroso)

$$X_n(\omega) = \begin{cases} 1 & \text{se la } n\text{-esima prova ha avuto successo} \\ 0 & \text{se la } n\text{-esima prova ha avuto insuccesso}, \end{cases}$$

si ottiene, questa volta, una *successione infinita* di variabili, che descrive l'evolversi del fenomeno e che prende il nome di *processo di Bernoulli* (parlando in termini un po' vaghi, con il termine *processo* in Probabilità si intende la descrizione dell'evoluzione, generalmente nel tempo o nello spazio, del fenomeno aleatorio osservato. In questo caso il fenomeno sotto osservazione è la ripetizione delle infinite prove di Bernoulli).

Bibliografia

[1] G. Dall'Aglio, *Calcolo delle Probabilità*, Zanichelli, Bologna, (2000)

[2] R. Giuliano, *Elementi di Calcolo delle Probabilità*, ETS Editrice, Pisa (2002)

[3] G. Letta, *Probabilità Elementare*, Zanichelli, Bologna (1993)

[4] S. M. Ross, *Calcolo delle Probabilità*, Apogeo, Milano (2002)

Nota sulla bibliografia. La parte tecnica della presente esposizione si basa su [2]. Alcuni esempi sono tratti da [4]. Il testo [1], molto buono per quanto riguarda la trattazione matematica, riporta anche interessanti notizie storiche. Il testo [3] è di livello matematicamente più elevato, e permette quindi di farsi un'idea più precisa delle problematiche riguardanti l'assiomatizzazione della teoria.