Elementi di statistica bayesiana parametrica

Francesco Bertolino

October 7, 2019

Contents

Chapter 1

Le concezioni della probabilità

> *...But, as to what probability is and how it is connected with statistics, there as seldom been such complete disagreement and breakdown of comunication since Tower of Babel.* (Leonard J. Savage, 1954)[1]
> *Quot homines, tot sententiae* (Publio Terenzio Afro)

Parlando in modo generico, si può dire che il ragionamento induttivo o per induzione è un processo che ha lo scopo di apprendere dalla esperienza. Di trarre, nelle condizioni date e per quanto è possibile, conclusioni generali da esperienze e/o da osservazioni particolari.

Piû specificamente, il nostro interesse, va a quei ragionamenti induttivi che richiedono il decisivo intervento del Calcolo delle Probabilità, (d'ora in poi in sigla *CdP*). Come precisò de Finetti, il ragionamento induttivo

> ...[è il] *ragionamento che precisa il senso ed il modo in cui si fanno, ed è giustificato fare, delle previsioni, in termini di probabilità, basandosi sull'esperienza, e precisamente - in particolare - sull'osservazione della frequenza dei successi in un numero (possibilmente grande) di casi "analoghi" a quello (o quelli) di cui ci interessa prevedere il risultato. In termini piú generali, si tratta*

[1] "Quanto a che cosa si debba intendere per probabilità e come essa sia collegata con la statistica, di rado vi è stato un disaccordo cosí completo ed una tale disgregazione comunicativa dai tempi della Torre di Babele." ; L. J. Savage, *The Foundation of Statistics.* Wiley & Sons, NY, 1954.

> *di vedere il modo in cui "la" (o "le") probabilità in questione vengono modificate in seguito all'acquisizione di ulteriori informazioni.*[2]

Secondo l'impostazione qui adottata, il ruolo che il ragionamento induttivo affida alla teoria consiste nell'indicare le procedure formali (ben inteso in base ai *dati* passati) allo scopo di aggiornare le nostre opinioni iniziali sul collettivo in studio e le nostre valutazioni di probabilità di eventi futuri in base ai *dati* che via via sopraggiungono.

Ma accettare l'idea che l'apprendere dalla realtà è un processo di revisione continua delle opinioni del soggetto, presentandosi nuovi *dati*, il ragionamento induttivo è possibile solo adottando la concezione soggettiva della probabilità. Nel qual caso tutto si riduce nel richiamare i concetti di coerenza, di dipendenza subordinata, di scambiabilità e nell'applicare il teorema delle probabilità composte ed il teorema di Bayes che ne consegue.

Affatto estranea a questa dispensa è l'analisi critica delle basi filosofiche e dei temi fondativi del ragionamento per induzione. Nessuno spazio è riservato agli sviluppi storici della concezione soggettiva, né a questioni generali e pur importanti, del tipo
(*a*) quali sono le condizioni generali (minime) che rendono possibile il ragionamento induttivo e, posto che esse si possano stabilire, quali sono le procedure induttive ammissibili?
(*b*) il ragionamento deduttivo, che trae conclusioni particolari da principî generali procedendo entro i limiti della logica del certo, può tornare utile (e se si in qualche forma) nelle procedure induttive, *etc.*?

Ma prima di entrare nel merito del ragionamento per induzione giova riprendere certi concetti di teoria della probabilità, in larga misura già noti al lettore, visti ora in ottica soggettivista.

Iniziamo dunque richiamando le principali concezioni della probabilità.

1.1 Definizioni di probabilità

Probabilities are states of mind and not states of nature.
(Leonard J. Savage)[3]

È un fatto, come fa osservare L. J. Savage, che in matematica non vi siano concetti espressi in modi tanto differenti (ed opposti) quanto i concetti di

[2] Dalla voce "Decisione", vol. IV *Enciclopedia* Einaudi, 1978. p. 421-484.

[3] *Le probabilità sono stati della mente e non stati di natura.*

probabilità e di induzione statistica[(4)].

Ed è complicato spiegare come dalle "varie definizioni" di probabilità siano seguiti "varî approcci inferenziali". E dunque l'approccio frequentista (in ottica Fisher, in ottica Neyman-Pearson-Wald, in sigla *NPW*, in ottica supportista o neo-fisheriana) e l'approccio bayesiano.

Iniziamo dunque con le principali definizioni di probabilità che discendono da concezioni e punti di vista difficilmente conciliabili. Procedendo in modo per forza di cose schematico, la dispensa considera solo le definizioni razionale, empirica, asintotica, dette oggettive; e la definizione soggettiva (o personale).[(5)]

Definizione 1.1.1. (**Razionale** o **classica**)
La probabilità dell'evento E è il rapporto tra il numero $n(E)$ dei casi favorevoli al verificarsi di E ed n_T, numero totale dei casi, purché tutti *ugualmente probabili.*

$$p(E) \overset{def}{=} \frac{n(E)}{n_T} \,. \lhd$$

Definizione 1.1.2. (**Empirica**)
Data una serie di prove ripetute un *gran* numero N di volte, sempre nelle stesse condizioni, nel corso della quale l'evento E si è manifestato n_E volte, la proporzione empirica di successo $f_N(E) = \dfrac{n_E}{N}$ è *presso a poco* uguale alla sua probabilità

$$p(E) \overset{def}{=} f_N(E) \,.$$

Di norma, l'approssimazione migliora col numero delle prove.[(6)] $\lhd$

Definizione 1.1.3. (**Asintotica** o **frequentista**)
Data una sequenza indefinita di prove tutte condotte nelle stesse condizioni, si definisce probabilità dell'evento E, il limite della proporzione empirica di successo $f_N(E)$ al divergere del numero N delle prove

$$p(E) \overset{def}{=} \lim_{n\to\infty} f_N(E) \,. \lhd$$

Definizione 1.1.4. (**Soggettiva**)
La probabilità di un evento E è la quota $p(E)$ che un individuo reputa di dover pagare ad un banco per ricevere "1" ovvero "0" verificandosi o non

---

[4)] Di certo che non è semplice spiegare allo studente che, scorrendo un libro di *CdP*, scopre che in letteratura circolano "diverse definizioni" di probabilità. Piú difficile è persuaderlo del fatto che il *CdP* è, a pieno titolo, un capitolo nella matematica.

[5)] In questa dispensa, non si dà conto della concezione logicista, dovuta a J. M. Keynes (1921), e della concezione neopositivista dovuta a R. Carnap (1950).

[6)] La definizione **1.1.2** è nota come legge (o postulato) empirico del caso.

verificandosi E. Le valutazioni di probabilità degli eventi devono rispondere ai pricipî di equità e coerenza. [7] $\lhd$

Segue la definizione di *principio* di *equità* e *coerenza*.

Definizione 1.1.5. Una scommessa risponde ai pricipî di
- equità allora che il ruolo di banco e giocatore sono scambiabili in ogni momento del gioco e sempre alle stesse condizioni;
- coerenza se non vi sono combinazioni di scommesse che consentano (sia al banco che al giocatore) di realizzare perdite o vincite certe. $\lhd$

La definizione razionale, che si ispira al mondo delle lotterie e dei giochi d'azzardo, prevede che le alternative *equiprobabili* siano in numero finito e note in partenza, trova applicazione in tutti quei campi (ad es., nei sondaggi statistici, nella genetica, nella meccanica delle particelle, *etc.*) nei quali lo sperimentatore reputa "ragionevole" l'*ipotesi di equiprobabilità* di tutti gli eventi *elementari* coinvolti, favorevoli o sfavorevoli che siano.

Ipotesi volta per volta giustificata da argomenti di simmetria (di dadi e monete) e/o dal principio di ragione non sufficiente[8].

A nessuno sfugge la circolarità logica della definizione **1.1.1** che, per definire la probabilità, fa uso del concetto di *equiprobabilità*. Concetto del tutto incomprensibile a quanti non sanno (ancora) cosa sia la probabilità.[9] Osserva J. H. Poincaré (*La Science et l'hypothèse*, 1902)

> "[Con la definizione *razionale*] *siamo costretti a definire il probabile dal probabile. Come possiamo sapere se due casi sono ugualmente probabili? Sarà per convenzione?*"

La definizione **1.1.1** non si estende ai casi in cui l'*equiprobabilità* degli eventi è irragionevole.

In tal caso e quando si è in presenza di sequenze di eventi da "prove ripetute di uno stesso fenomeno", si *deve passare* alla definizione empirica **1.1.2**. La quale, come chiunque può verificare, contiene espressioni labili,

---

7) La definizione soggettiva fu data, per la prima volta, da Thomas Bayes (1762-1761), "*La probabilità di un evento è il rapporto tra il valore al quale una aspettativa che dipende dall'accadere di quell'evento deve essere calcolata ed il valore che ciò che si attende assume una volta che l'evento si è verificato*". Essa comparve nell'articolo Essay Towards solving a problem in the doctrine of chances. (1763, *post.*)

8) Tali concetti sono lasciati all'intuizione del lettore. Il principio di *ragione non sufficiente*, noto sia pure in implicito da Galileo e definito (in una larga accezione filosofica) da Leibniz, apparve in un saggio di P. S. Laplace (1812, 1814) sulla probabilità. Ad esso J. M. Keynes (1921) oppose i principi di indifferenza e di irrelevanza.

9) Piú avanti si mostra come la definizione classica possa, a certe condizioni, essere "recuperata" in ottica soggettiva.

imprecise, vaghe, tipiche di quelle situazioni in cui si vuol creare un ponte o (peggio) una coincidenza tra un oggetto empirico ed un oggetto astratto, tra la frequenza relativa e la probabilità.

A giudizio dei fautori della concezione empirica, è cruciale che le prove che via via si succedono siano
(i) distinte e/o separate, condizione che ne assicurerebbe l'*indipendenza*;
(ii) ripetibili, si torni alla locuzione "*prove ripetute ... sempre nelle stesse condizioni*", condizione che garantirebbe l'*equiprobabilità* delle prove e la *stabilità* della proporzione empirica $f_N(E)$, al crescere del numero N.

Fatto che se da una parte evidenzia il carattere tautologico della definizione stessa dall'altra esclude alla radice ogni valutazione di probabilità, di qualsiasi evento singolo.

Giova, in proposito, citare J. F. Bertrand (*Calcul des Probabilités*, 1889.)

> *Qual'è la probabilità che domani piova? Non esiste. Non perché essa varia da un giorno all'altro con lo stato del cielo e la direzione dei venti; ma perché in nessuna circostanza essa non ha valore obiettivo, la stessa per tutti quelli che la valutano senza ingannarsi. ... Un uomo ha quaranta anni, qual'è la probabilità che egli viva per dieci anni? ...*

Un empirista, persuaso che l'*oggettività* della probabilità poggi sulla *perfetta ripetibilià dell'esperimento*, che valore assegna alle probabilità di eventi unici? Saprebbe valutare, per conto di una compagnia assicurativa, la probabilità dell'evento "*nel 2020 il signor Leo Bianchi muore*"?

Nota 1.1.1. Per tradizione, le probabilità razionali sono dette *a priori*, in quanto calcolabili in via pre-sperimentale. A differenza delle probabilità empiriche, dette *a posteriori*, cioè ottenute dopo aver fatto esperimenti.

Con gli anni è venuta meno l'esigenza di mantenere una distinzione (di *significato*) tra probabilità razionali e probabilità empiriche in quanto, da piú parti, giudicata inutile e/o fuorviante:
- inutile perché priva di conseguenze circa le proprietà della probabilità;
- fuorviante perché interferisce con i concetti di probabilità *a priori* e *a posteriori* che entrano nel teorema di Bayes aventi differente significato;
- inutile e fuorviante perché le probabilità nascono dalla incertezza di chi guarda e valuta gli eventi e non dalla loro (supposta) natura. $\triangleleft$

Nota 1.1.2. Le definizioni classica **1.1.1** ed empirica **1.1.2**, prive del necessario *statuto di definizione*, possono essere riguardate (al piú) come delle *pseudo-definizioni* e, come tali, accolte come *utili regole di calcolo.* $\triangleleft$

Quanto alla definizione asintotica **1.1.3**, balza all'occhio che il limite, per $N \to \infty$, che vi appare non ha il senso esatto che l'analisi matematica richiede ad un limite.(10) Diversamente dalla definizione empirica, la definizione asintotica non è neppure una regola di stima della probabilità, non potendosi concepire una regola che per essere applicata richiede un numero infinito di prove condotte, tutte, nelle stesse condizioni sperimentali(11). Per dirla con de Finetti, "*L'idea di una successione infinita di prove è puramente fittizia. Il numero di prove effettivamente realizzate sarà sempre finito*".

In certi manuali di *CdP* di indirizzo frequentista si tenta di giustificare la definizione **1.1.3** invocando il *teorema di Bernoulli* il quale, con ovvio significato dei simboli, stabilisce che

$$\lim_{N\to\infty} \mathbb{P}\Big\{ \,\big| \, p(E) - f_N(E) \, \big| < \varepsilon \Big\} \; = \; 1 \, ,$$

e tutto ciò senza avvedersi di due vistosi aspetti

(i) se si ignora il significato ed il valore di $p(E)$ il teorema non ha senso;

(ii) è illogico dimostrare mediante teoremi (*!*) ciò che prima è stato assunto come definizione.

Affatto differente è la concezione soggettiva, adottata in queste dispense, secondo la quale le valutazioni di probabilità esprimono il grado di fiducia (*degree of belief*) del soggetto circa l'accadere dell'evento E.

Strumento privilegiato (ma non unico) per fissare la probabilità di un evento aleatorio E qualsiasi, singolo o ripetibile che sia, è la scommessa in cui il giocatore paga al banco il valore certo p, la quota di scommessa, a fronte di un premio incerto unitario che il banco si impegna ad erogare al giocatore solo accadendo E. Probabilità dunque come equivalente certo di una somma aleatoria unitaria.(12)

Allo scopo di impedire arbitrarie valutazioni di probabilità, occorre far intervenire i principî di equità e coerenza, secondo la definizione **1.1.5**.

Il principio di *equità* nel pretendere la simmetria di banco e giocatore non evita (da solo) il pericolo di quote di scommessa stravaganti, illogiche ed arbitrarie, ancorché accettate da entrambe le parti. Pericolo che può essere

10) R. von Mises (1928), principale fautore della probabilità-limite, reputava che il "*carattere empirico*" del limite presente nella definizione **1.1.3** non costituisse un ostacolo insuperabile e che bastasse aggiungere la locuzione "*per definizione*".

11) Stando ai fautori della concezione asintotica, non è lecito parlare di probabilità neppure dopo $n = 10^3$, $n = 10^4$,... prove.

12) Ben inteso, le parole *scommessa*, *banco*, *giocatore*, *premio*, *danaro*, *etc.*, non indicano entità matematiche, bensí oggetti empirici extralogici, ben noti al soggetto, a lui necessarî per ragionare intorno alla probabilità.

impedito solo facendo intervenire il principio di *coerenza*, sul quale si tornerà nel prossimo §**1.2**.

Nota 1.1.3. Chi reputa che le probabilità non sono stati di natura, ma esistono solo come *pensiero* (cioè sono stati della mente), è tenuto a sostenere che, nella sua essenza, la probabilità ha un carattere *relativo*.[13]
Relativo alle competenze, alla psicologia del soggetto, alle informazioni possedute dal soggetto chiamato ad tradurle in probabilità coerenti ed al momento in cui si compie la valutazione.[14] Per dirla con D.V. Lindley.

> "... *La probabilità, a differenza della lunghezza di un tavolo, non ha una propria esistenza indipendente dai soggetti da cui è valutata. La probabilità esprime piuttosto una relazione tra un soggetto ed il mondo che egli osserva. Alcuni considerano questo fatto negativo. Al contrario, la natura soggettiva della probabilità è un grande vantaggio: essa infatti descrive lo stato reale di un soggetto che osserva il mondo e non, come accade in altre scienze, un mondo separato dagli osservatori che vivono in esso.*"[15]

È affatto *naturale* che il soggetto possa modificare, in presenza di nuovi fatti e nuove evidenze sperimentali, le sue valutazioni di probabilià. $\triangleleft$

Giova osservare che la definizione razionale **1.1.1** può essere *recuperata*, come si è detto poco sopra, in ottica soggettiva in forma di teorema.

Sia $E = \{E_1, E_2, \ldots, E_K\}$ una partizione dell'evento certo, sia $B = \{E_{k_1}, E_{k_2}, \ldots, E_{k_m}\}$ un suo sottoinsieme costituito da m eventi.

Se il soggetto, ad es. per ragioni di "*simmetria*", è propenso a valutare equiprobabili gli eventi della partizione A, cioè $\mathbb{P}(E_k) = \dfrac{1}{card\{A\}}$, $\forall k$, deve dire che $\mathbb{P}(B) = \dfrac{card\{B\}}{card\{A\}} = \dfrac{m}{K}$. Tale conclusione, che ricorre in tante situazioni pratiche e non solo nel caso di lotterie e di lanci di dadi, è espressa dal teorema.

***Teorema* 1.1.1.** *Se il soggetto reputa ugualmente probabili gli eventi di una partizione* (*finita*) *costituita da* $K \geq 2$ *eventi, allora* (*per lui*) *la probabilità della unione di* m *di essi è* $\dfrac{m}{K}$. $\triangleleft$

---

[13] Fuori dell'ambienti soggettivisti, negli anni '30 del secolo passato, tale idea era condivisa anche dai logicisti J. M Keynes e H. Jeffreys, e da qualche oggettivista. In questi ultimi trenta anni sono pochi gli autori che non accettano il carattere relativo della probabilità.

[14] Se C_0 indica l'insieme delle conoscenze del soggetto, al momento della valutazione, il simbolo $\mathbb{P}(E|C_0)$ sarebbe preferibile a $\mathbb{P}(E)$.

[15] Dennis V. Lindley, *Making Decisions*, Wiley, 1971.

Nota 1.1.4. La portata pratica del teorema **1.1.1** è notevole. Sussistendo le condizioni del teorema, le valutazioni combinatorie della probabilità (le cosidette techiche e/o *regole di calcolo*) trovano, tra i soggettivisti, piena giustificazione. $\triangleleft$

Esempio 1.1.1. Un'urna contiene due biglie che possono essere entrambe nere, entrambe bianche, una nera e l'altra bianca. Calcolare la probabilità dell'evento $E =$"*le biglie sono di diverso colore*".

Cosí come è formulato il problema non ha risposta. Se, ad esempio, si è adottata la regola di lanciare (per due volte) una moneta regolare e di inserire ad ogni lancio nell'urna una biglia nera o bianca a seconda che sia apparsa T o C, allora $\mathbb{P}(E) = \frac{1}{2}$.

Se invece si è deciso di formare tre gruppi di biglie (N, N), (N, B) e (B, B) e di riversare nell'urna il gruppo sorteggiato, allora $\mathbb{P}(E) = \frac{1}{3}$.

Non conoscendo il procedimento di riempimento dell'urna è possibile stabilire la probabilità "oggettiva" di E? $\triangleleft$

Il lettore faccia attenzione al fatto che la (*pseudo*-)definizione **1.1.1** ed il teorema **1.1.1** "raccontano" cose affatto differenti.

Infatti, mentre per i fautori della concezione razionale la *equiprobabilità* è un fatto *oggettivo* che precede il concetto di probabilità, che si vorrebbe definire, il teorema **1.1.1**, banalmente, si limita a trarre le ovvie conclusioni dal giudizio di *equiprobabilità* esplicitamente espresso dal soggetto. Soggetto che (oltre tutto) è già in possesso del significato di probabilità.

Per contro, per i soggettivisti tutti gli eventi, singoli o ripetibili che siano, sono comunque probabilizzabili.

Le procedure induttive che fanno uso esclusivo di probabilità oggettive rientrano nella statistica oggettiva o frequentista; prende nome di statistica soggettiva o bayesiana o neo-bayesiana, l'insieme delle procedure che fanno uso (solo) di probabilità soggettive.

1.2 Definizioni ed assiomi del *CdP*

In questa sezione si intende mostrare che da ognuna delle definizioni di probabilità (razionale, empirica, asintotica e soggettiva) sono deducibili i cosí detti assiomi centrali del *CdP*. Ovvero gli assiomi di non negatività, di normalizzazione, di addittività della probabilità, nonché il teorema delle probabilità composte.

Per chi accetta la definizione razionale, per cui la probabilità dell'evento E è il rapporto fra il numero $n(E) \geq 0$ dei casi favorevoli ad E ed il numero dei casi totali $n_T \in \mathbb{N}$ tutti ugualmente possibili, la limitazione $0 \leq p(E) \leq 1$ non è che una ovvia conseguenza.

Per dedurre l'assioma di addittività è sufficiente considerare gli eventi incompatibili A e B, i cui casi favorevoli sono $n(A)$ ed $n(B)$ e l'evento $A \cup B$ i cui casi favorevoli sono $n(A \cup B) = n(A) + n(B)$. Donde

$$p(A \cup B) \;=\; \frac{n(A) + n(B)}{n_T} \;=\; \frac{n(A)}{n_T} + \frac{n(B)}{n_T} \;=\; p(A) + p(B)\,.$$

Allo scopo di dedurre il teorema delle probabilità composte, è necessario considerare gli eventi A, B e $A \cap B$ i cui casi favorevoli sono $n(A) \in \mathbb{N}$, $n(B) \in \mathbb{N}$ ed $n(A \cap B) \in \mathbb{N}_{\{0\}}$, con $n(A \cap B) \leq \min\big\{n(A), n(B)\big\}$. Donde

$$p(A \cap B) \;=\; \frac{n(A \cap B)}{n_T} \;=\; \frac{n(A)}{n_T} \cdot \frac{n(A \cap B)}{n(A)} \;=\; p(A) \cdot p(B|A)\,,$$

donde la definizione di *probabilità condizionata* $p(B|A) = \dfrac{p(A \cap B)}{p(A)}$.

Dedurre gli assiomi centrali, dalle definizioni empirica **1.1.2** ed asintotica **1.1.3**, richiede procedimenti analoghi. Il compito è lasciato al lettore.

Affatto differente è l'approccio soggettivista che fa discendere gli assiomi centrali del *CdP* adottando nulla piú che il principio di coerenza.

Se $p \cdot S$ è la puntata (il prezzo *equo* di scommessa) relativa all'evento E, che dà diritto al giocatore di riceve dal banco il premio S, verificandosi E, il guadagno aleatorio $\mathcal{G}(\cdot)$ del giocatore risulta (si ponga $S > 0$)

$$\begin{cases} \mathcal{G}(E) \;=\; S - p \cdot S \;=\; (1-p)S & \text{se si verifica E,} \\ \mathcal{G}(\bar{E}) \;=\; -p \cdot S & \text{se } E \text{ non si verifica.} \end{cases}$$

Fissando $p < 0$, donde $1 - p > 1$, i guadagni del giocatore sarebbero sempre positivi; per contro, fissando $p > 1$, donde $1 - p < 0$, il giocatore sarebbe sempre in perdita. Se dunque si vuole che il giuoco sia coerente è necessario che $0 \leq p \leq 1$; condizione che è anche sufficiente. Infatti, facendo il prodotto dei guadagni si ha

$$\mathcal{G}(E) \cdot \mathcal{G}(\bar{E}) \;=\; -p(1-p) \cdot S^2\,,$$

da cui si evince che i guadagni $\mathcal{G}(E)$ e $\mathcal{G}(\bar{E})$ sono di segno differente solo valendo la condizione $0 \leq p \leq 1$.(16)

16) E ciò anche se banco e giocatore avessero fissato $S < 0$.

Il principio di coerenza si applica anche agli eventi certo ed impossibile; infatti, il guadagno $\mathcal{G}(\Omega) = (1-p)S$ è nullo *sse* $p = 1$; mentre, la perdita $\mathcal{G}(\bar{\Omega}) = \mathcal{G}(\emptyset) = -pS$ è nulla *sse* $p = 0$.

Lo stesso principio impedisce che banco e giocatore possano per uno stesso evento E assegnare due differenti quote di scommessa, ad es. $p_1 \neq p_2$.

Ove ciò venisse consentito, si potrebbe scommettere su E con la quota p_1 e premio S_1 e, sempre su E, con la quota p_2 e premio S_2. I guadagni relativi agli eventi E ed $\bar{E}$ risulterebbero

$$\begin{cases} \mathcal{G}(E) &= (1-p_1)S_1 + (1-p_2)S_2\,, \\ \mathcal{G}(\bar{E}) &= -p_1S_1 - p_2S_2\,, \end{cases}$$

ovvero, dopo opportuno riordino,

$$\begin{pmatrix} 1-p_1 & 1-p_2 \\ -p_1 & -p_2 \end{pmatrix} \cdot \begin{pmatrix} S_1 \\ S_2 \end{pmatrix} = \begin{pmatrix} \mathcal{G}(E) \\ \mathcal{G}(\bar{E}) \end{pmatrix},$$

avente determinante $\Delta = p_1 - p_2$. Ma se $\Delta \neq 0$, allora è sempre possibile determinare l'*opportuna* combinazione di premi $\big\{S_1, S_2\big\}$ che rende i guadagni $\big\{\mathcal{G}(E), \mathcal{G}(\bar{E})\big\}$ tutti positivi o tutti negativi.

Ancora invocando il principio di coerenza, si può dedurre l'assioma di addittività e poi dimostrare il teorema delle probabilità totali.

Si assumano per semplicità gli eventi $\{E_1, E_2, E_3\}$, che formano una partizione di Ω. Se $\{p_1, p_2, p_3\}$ sono le rispettive probabilità fissate da banco e giocatore e se $\{S_1, S_2, S_3\}$ sono le quote messe in palio per ciascun evento, segue che l'importo complessivo della scommessa è $\sum_{i=1}^{3} p_iS_i$.

I guadagni associati a ciascuno degli eventi E_j producono il sistema

$$\mathcal{G}(E_j) = S_j - \sum_{i=1}^{3} p_iS_i\,,\; j = 1, 2, 3\,,$$

ovvero

$$\begin{pmatrix} 1-p_1 & -p_2 & -p_3 \\ -p_1 & 1-p_2 & -p_3 \\ -p_1 & -p_2 & 1-p_3 \end{pmatrix} \cdot \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} \mathcal{G}(E_1) \\ \mathcal{G}(E_2) \\ \mathcal{G}(E_3) \end{pmatrix},$$

con determinante $\Delta = 1 - (p_1 + p_2 + p_3)$; da cui segue che le probabilità fissate da banco e giocatore sono coerenti solo se $p_1 + p_2 + p_3 = 1$.

Qualora fosse $p_1 + p_2 + p_3 \lesseqgtr 1$, è sempre possibile stabilire una certa terna $\{S_1, S_2, S_3\}$ tali da aversi guadagni o perdite certe.

Con poca fatica si dimostra che la condizione $p_1 + p_2 + p_3 = 1$, oltre ad essere necessaria, è pure sufficiente.

Accettando dunque la definizione **1.1.4**, si deduce l'assioma di addittività, e si dimostra il teorema delle probabità totali.

***Teorema* 1.2.1.** *Se* $\{E_1, E_2, \ldots, E_K\}$, *con* $K \geq 2$, *sono eventi due a due incompatibili, allora*

$$\mathbb{P}\left(\bigcap_{j=1}^{K} E_j\right) = \sum_{j=1}^{K} \mathbb{P}(E_j) \ . \lhd$$

Applicando la definizione **1.1.4** ad un opportuno sistema di eventi e di puntate, si ritrova il teorema delle probabilità composte. Prima di procedere, giova ricordare che

• l'evento subordinato $E|H$ è l'evento che: (a) è vero se sono veri E ed H; (b) è falso se E è falso ed H è vero; (c) è privo di valore logico se H è falso (nel qual caso la scommessa è nulla ed il banco restituisce le puntate);

• per i fautori della concezione soggettiva le probabilità $\mathbb{P}(E)$ e $\mathbb{P}(E|H)$ sono valutazioni dell'evento E effettuate in differenti contesti informativi ed entrambe valide.

Siano p_H, $p_{E\cap H}$ e $p_{E|H}$ le quote di scommessa unitaria che un soggetto coerente assegna al verificarsi di H, $E \cap H$ e $E|H$. Siano $p_H S_1$, $p_{E\cap H} S_2$ e $p_{E|H} S_3$ le puntate per ricevere S_1, S_2 e S_3, verificandosi gli eventi H, $E \cap H$ e $E|H$ rispettivamente.

Agli eventi $\bar{H}$, $E \cap H$ e $\bar{E} \cap H$, costituenti una partizione dell'evento certo, corrispondono i guadagni $\mathcal{G}_{\bar{H}}$, $\mathcal{G}_{E\cap H}$ e $\mathcal{G}_{\bar{E}\cap H}$, ordinati in forma di sistema

$$\begin{cases} \mathcal{G}_{\bar{H}} &= p_H S_1 - p_{E\cap H} S_2 \,, \\ \mathcal{G}_{E\cap H} &= S_1 + S_2 + S_3 - \big(p_H S_1 + p_{E\cap H} S_2 + p_{E|H} S_3\big) \,, \\ \mathcal{G}_{\bar{E}\cap H} &= S_1 - \big(p_H S_1 + p_{E\cap H} S_2 + p_{E|H} S_3\big) \,, \end{cases} \tag{1.1}$$

ovvero nella forma

$$\begin{pmatrix} -p_H & -p_{E\cap H} & 0 \\ \big(1 - p_H\big) & \big(1 - p_{E\cap H}\big) & \big(1 - p_{E|H}\big) \\ \big(1 - p_H\big) & p_{E\cap H} & p_{E|H} \end{pmatrix} \cdot \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} \mathcal{G}_{\bar{H}} \\ \mathcal{G}_{E\cap H} \\ \mathcal{G}_{\bar{E}\cap H} \end{pmatrix} ,$$

avente determinante $\Delta = p_H \cdot p_{E|H} - p_{E\cap H}$. Se fosse $\Delta \neq 0$, allora sarebbe possibile trovare una *opportuna* terna $\big\{S_1, S_2, S_3\big\}$ capace di produrre guadagni

$\{\mathcal{G}_{\bar{H}}, \mathcal{G}_{E\cap H}, \mathcal{G}_{\bar{E}\cap H}\}$ tutti positivi o tutti negativi. Pericolo che può essere evitato solo adottando la *condizione di coerenza* $p_{E\cap H} = p_H \cdot p_{E|H}$.

Allo scopo di provare che la condizione di coerenza, oltre che necessaria, è anche *sufficiente* si consideri l'espressione del guadagno medio

$$\begin{aligned}\bar{\mathcal{G}} &= (1 - p_H)\,\mathcal{G}_{\bar{H}} + p_{E\cap H}\,\mathcal{G}_{E\cap H} + (1 - p_{E|H})\,\mathcal{G}_{\bar{E}\cap H} \\ &= (1 - p_H)\,\mathcal{G}_{\bar{H}} + p_H \cdot p_{E|H}\,\mathcal{G}_{E\cap H} + (1 - p_{E|H})\,\mathcal{G}_{\bar{E}\cap H}\,,\end{aligned}$$

che diviene, facendo intervenire le (1.1), $\bar{\mathcal{G}} = \big(p_H \cdot p_{E|H} - p_{E\cap H}\big) \cdot S_3$.

Per scongiurare combinazioni di scommesse a vincita sempre positiva/negativa e cioè $\bar{\mathcal{G}} \gtrless 0$, è sufficiente la condizione $p_{E\cap H} = p_H \cdot p_{E|H}$.

I ragionamenti che hanno preceduto costituiscono la dimostrazione del teorema delle probabilità composte.

***Teorema* 1.2.2.** *Le valutazioni di probabilità* $\mathbb{P}(H)$, $\mathbb{P}(E\cap H)$, $\mathbb{P}(E \mid H)$ *sono coerenti sse* $\mathbb{P}(E\cap H) = \mathbb{P}(H)\cdot\mathbb{P}(E \mid H)$ *. Se poi* $\mathbb{P}(H) > 0$*, allora* $\mathbb{P}(E \mid H) = \dfrac{\mathbb{P}(E\cap H)}{\mathbb{P}(H)}$. $\lhd$

Alla stessa maniera è possibile dimostrare la formula di decomposizione, di cruciale importanza nel *CdP*. Ci limitiamo ad un breve richiamo.

Sia $\{H_1, H_2, \ldots, H_k\}$ una partizione di Ω, sia $A \subset \Omega$ un evento compatibile con almeno uno degli H_i. Si dimostra che la probabilità $\mathbb{P}(A)$ è data dalla combinazione lineare convessa delle probabilità condizionate $\mathbb{P}(A|H_i)$ di *pesi* $\mathbb{P}(H_i) > 0$ e con $\sum_{i=1}^{k} \mathbb{P}(H_i) = 1$, detta formula di decomposizione

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(H_i)\cdot\mathbb{P}(A|H_i)\,. [17] \tag{1.2}$$

Sempre accettando il principio di coerenza, il teorema di Bayes non è che una conseguenza (se si preferisce un corollario) del Teorema **1.2.2** delle probabilità composte.

Per chi è fautore della concezione soggettivista, applicare teoremi e proprietà del *CdP* non correttamente, non solo comporta una violazione del principio di coerenza, ma offre ad ogni giocatore abile nel creare sistemi di scommesse, la possibilità di assicurarsi guadagni certi.

In tal modo inteso, il *CdP* "*non è allora se non la teoria matematica che insegna ad essere coerenti*"[18].

---

[17] Qualche autore tiene a precisare che la $\mathbb{P}(A)$ è la media delle $\mathbb{P}(A|H_i)$ pesate con le $\mathbb{P}(H_i)$. Donde la proprietà di internalità $\min_i \{\mathbb{P}(A|H_i)\} \leq \mathbb{P}(A) \leq \max_i \{\mathbb{P}(A|H_i)\}$ detta anche proprietà conglomerativa.

[18] B. de Finetti, in *Fundamenta Mathematicae*, **17**, 1931, pp. 298-329.

1.2.1 Note sul sistema assiomatico di Kolmogorov

Risale al 1933, e si deve ad Andrei N. Kolmogorov, il piú accettato sistema assiomatico-deduttivo, completo e non contraddittorio, del *CdP*, costituito da cinque assiomi e dalla definizione di probabilità condizionata. La sua semplicità fu motivo non ultimo del suo successo.

Partendo dal fatto che nessun sistema assiomatico possa basarsi su definizioni "deboli" e "controverse", Kolmogorov propose di tenere fuori dal sistema assiomatico ogni definizione di probabilità, troncando alla radice tutte le controversie ad esse legate(19).

Oggi i cinque assiomi sono presentati, per lo piú, col seguente ordine

(i) gli eventi $A, B, C, \dots,$ sottoinsiemi di Ω, costituiscono una classe additiva $\mathcal{A}$;

(ii) $\forall A \in \mathcal{A},\ \mathbb{P}(A) \geq 0\,;$

(iii) $\mathbb{P}(\Omega) = 1\,;$

(iv) $\forall A, B \in \mathcal{A},$ t.c. $A \cap B = \emptyset \ \Rightarrow\ \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)\,;$

(v) se $\{A_n, n = 1, 2, \dots\}$ è una successione decrescente di eventi di $\mathcal{A}$, t.c. $\lim_{n\to\infty} A_n = \emptyset,\ \Rightarrow\ \lim_{n\to\infty} \mathbb{P}(A_n) = 0.$

La terna $\big(\Omega, \mathcal{A}, \mathbb{P}\big)$, che raccoglie lo spazio Ω, la classe additiva $\mathcal{A}$, la misura di probabilità $\mathbb{P}$ sugli eventi di Ω è comunemente detta spazio di probabilità.

Nel 1933, tanto gli assiomi centrali $(ii) \div (iv)$ (cioè il nucleo essenziale del sistema assiomatico) che la definizione di probabilità condizionata non erano di certo una novità. Ed era noto (lo si è visto nel precedente §**1.2**) come essi fossero *deducibili* dalle varie definizioni di probabilità.

Le novità introdotte da Kolmogorov furono

• di assimilare, con l'assioma (i), (l'assioma della adittività semplice) gli eventi, sia *singoli* che *ripetibili*, ad insiemi muniti (almeno) di algebra di Boole;

• di non fornire indicazioni sul modo di calcolare le probabilità; precisando solo che le probabilità, in quanto funzioni di insieme, sono *misure secondo Radon-Nikodym* o, caso particolare, *secondo Stieltjes-Lebesgue*;

---

19) È assai curioso il fatto che, in certi manuali di *CdP*, il sistema assiomatico sia indicato come *definizione assiomatica di Kolmogorov* essendo ben noto il suo proposito di escludere dal sistema qualsiasi definizione di probabilità.

• di dotare, mediante l'assioma di continuità (v), (l'assioma della adittività infinita o completa) le classi numerabili di eventi di σ-algebra; fatto che consente di estendere il teorema delle probabilità totali a tali classi;
• di introdurre la probabilità condizionata, "in forma di definizione", ovvero $\mathbb{P}(E|H) \overset{def}{=} \dfrac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}$, purché $\mathbb{P}(H) > 0$. [20]

Col sistema che porta il suo nome, Kolmogorov fece entrare il *CdP* nella teoria della misura capitolo, a pieno titolo, della matematica.[21]

Si osservi che i fautori delle concezioni razionale, empirica ed asintotica, non potendo mai concepire *infinità* di eventi distinti alternativi non avrebbero, *sensu strictu*, necessità dell'assioma (v). Ciò nondimeno non si trova chi, fra detti fautori, dichiari di volerne farne a meno.

Differente è il punto di vista soggettivo che mettendo al centro delle valutazioni di probabilità il principio di *coerenza* giunge alla *addittività*, come conseguenza: addittività *finita* se gli eventi su cui stiamo lavorando sono in numero finito, addittività *numerabile* (o *completa*) se gli eventi costituiscono una infinità numerabile.

In sintesi si può dire che se le probabilità fissate dal soggetto sono coerenti allora gli assiomi di Kolmogorov sono soddisfatti. E *viceversa*.

Una menzione (per forza di cose) fuggevole merita il postulato (v) da cui possono provenire inconvenienti persino in esempi/casi non complicati.

Si pensi al fatto che la probabilità non può essere, allo stesso tempo, infinitamente additiva e uniformemente distribuita su tutti gli eventi. (Ad es. su tutti i punti di $\mathbb{N}_0$.) Per non parlare degli inconvenienti legati agli spazi aventi la potenza del continuo.

Va segnalata infine, per completezza, la presenza di una (autorevole) minoranza di soggettivisti che rifiuta l'assioma (v) non reputandolo né necessario né pienamente giustificato; non andando, tuttavia, oltre applicazioni limitate, particolari e talora dubbie.

Nella dispensa, che non entrerà nel merito delle controversie fondazionali legate a tale rifiuto, si adotterà, per ragioni di convenienza matematica, la base assiomatica di Kolmogorov.

---

20) Se ogni definizione di probabilità è esclusa dal sistema assiomatico, il teorema delle probabilità composte e la probabilità condizionata possono entrarvi solo come definizione autonoma. In certi manuali la formula della probabilità subordinata è detta, inopinatamente, "definizione di Kolmogorov di probabilità subordinata".

21) Il che non impedì nuovi interrogativi (e critiche). La $\sigma-$addittività è proprio necessaria? La probabilità è una misura d'insieme? Eventi ed insiemi "coincidono"? Ad es., l'evento $E =$"*domani Caio arriverà puntuale*" in che modo può entrare in una qualche algebra di insiemi? *Etc.*

1.2.2 Il valore atteso da un punto di vista bayesiano

Sia $\{E_j\}_{j=1}^n$ una $n-$pla di eventi necessari ed incompatibili, siano $\{x_j\}_{j=1}^n$ i premi messi in palio su ciascuno di essi, siano $\{p_j\}_{j=1}^n$ le rispettive valutazioni (coerenti) di probabilità di un certo soggetto. Sia $\mathbf{1}_{E_j}$ la v.a. indicatore di evento che vale 1 se E_j si verifica e 0 se non si verifica.[22]

Si pone il problema di stabilire l'ammontare certo $m = \mathcal{P}(X)$ che un giocatore che accetta di scommettere su tutti gli n eventi E_j alle condizioni indicate, deve pagare al banco per partecipare alla lotteria, sapendo che dal banco riceverà il premio aleatorio $X = x_1\mathbf{1}_{E_1} + x_2\mathbf{1}_{E_2} + \cdots + x_n\mathbf{1}_{E_n}$. Da cui segue l'espressione del guadagno aleatorio dello scommettitore

$$G = x_1\mathbf{1}_{E_1} + x_2\mathbf{1}_{E_2} + \cdots + x_n\mathbf{1}_{E_n} - m\,,$$

per cui, accadendo E_h, il guadagno (in segno) del giocatore è

$$g_h = x_h - m\,, \quad \forall h = 1, 2, \ldots, n\,,$$

donde moltiplicando entrambi i membri per p_h e sommando $\forall h$ si ottiene

$$\sum_{j=1}^n g_j \cdot p_j = \sum_{j=1}^n x_j \cdot p_j - m \gtreqless 0\,.$$

Il giocatore $\big[$il banco$\big]$ è favorito se $\sum_{j=1}^n x_j \cdot p_j > m$ $\Big[\sum_{j=1}^n x_j \cdot p_j < m\Big]$. Il gioco è equo e coerente *sse* la quota di scommessa è $m = \sum_{j=1}^n x_j \cdot p_j$.

Il valore certo $m = \mathcal{P}(X)$ prende il nome di valore atteso o previsione o speranza matematica o aspettazione[23] o valor medio o equivalente certo della v.a. X.

Quando gli eventi su cui scommettere sono $\{E, \bar{E}\}$, con $\mathbb{P}(E) = \theta$ e $\mathbb{P}(\bar{E}) = 1 - \theta$, ed il premio aleatorio è $X = 0 \cdot \mathbf{1}_{\bar{E}} + 1 \cdot \mathbf{1}_E = \mathbf{1}_E$, ovvero quando la v.a. X assume solo i valori $\{0, 1\}$, si ha $\mathcal{P}(X) = \mathcal{P}(\mathbf{1}_E) = 0 \cdot (1-\theta) + 1 \cdot \theta = \theta = \mathbb{P}(E)$.

A parole: la probabilità di un evento è il valore atteso di una scommessa a due sole alternative con premio 1.

In base alla impostazione seguita, il valor medio può essere dunque concepito come *estensione* della probabilità ad una scommessa a piú alternative.

---

[22] Si osservi che per "costruzione" è $\sum_{j=1}^n \mathbf{1}_{E_j} = 1$ e $\sum_{j=1}^n p_j = 1$.

[23] Si deve a C. Huygens, in una corrispondenza del 1666, l'impiego del termine latino expectatio per indicare il valore di un giuoco. Tale termine accolto dagli autori inglesi fu reso con expectation, tutt'ora in uso.

Per tale motivo B. de Finetti, che per primo evidenziò questo aspetto della probabilità, adottò lo stesso termine di previsione (e lo stesso simbolo) tanto per la *probabilità* che per la *speranza matematica.*

Per concludere si può dire che, dal punto di vista bayesiano, la speranza matematica è l'estensione diretta della nozione di probabilità soggettiva.

Esempio 1.2.1. Sia $(Y_1, Y_2, \ldots, Y_n)$, con $Y_i \in \{0, 1\}$, un successione di n v.a.*iid* e $\mathbb{P}(Y_i = 1) = \theta$. Sia $X = \mathbf{1}_{Y_1} + \mathbf{1}_{Y_2} + \cdots + \mathbf{1}_{Y_n}$ il numero aleatorio di successi nelle n prove. La previsione di X è

$$\begin{aligned} \mathcal{P}(X) &= \mathcal{P}(\mathbf{1}_{Y_1} + \mathbf{1}_{Y_2} + \cdots + \mathbf{1}_{Y_n}) \\ &= \mathcal{P}(\mathbf{1}_{Y_1}) + \mathcal{P}(\mathbf{1}_{Y_2}) + \cdots + \mathcal{P}(\mathbf{1}_{Y_n}) \\ &= n \cdot \theta \; = \; E(X)\,, \end{aligned}$$

Lo stesso risultato si poteva conseguire con altra procedura. Poiché $X \sim Bin(x|\theta, n)$, $X \in \{0, 1, \ldots, n\}$, ne consegue che $E(X) = n\theta = \mathcal{P}(X)$. La speranza matematica come equivalente certo, espresso in £, che il giocatore deve pagare al banco per partecipare ad un gioco in cui il banco consegna al giocatore 1£ ad ogni successo ottenuto nel corso di n prove. $\triangleleft$

Essendo $\mathcal{P}(X)$ una combinazione lineare convessa, ne consegue che, se gli eventi su cui scommettere costituiscono una partizione finita, allora la previsione esiste sempre finita. Inoltre

• $x_{(1)} \le \mathcal{P}(X) \le x_{(n)}$, con $x_{(1)} = \min\{x_1, \ldots, x_n\}$ e $x_{(n)} = \max\{x_1, \ldots, x_n\}$, (proprietà della internalità)[24];

• $\mathcal{P}(X)$ è il baricentro di un insieme unitario di n masse puntiformi $\{p_j\}_{j=1}^n$ distribuite nei punti $\{x_j\}_{j=1}^n$ di un asse;

Esempio 1.2.2. Sia X il punto che si realizza lanciando un dado regolare a 6 facce. La previsione

$$\mathcal{P}(X) = \mathcal{P}(1 \cdot \mathbf{1}_1 + 2 \cdot \mathbf{1}_2 + \cdots + 6 \cdot \mathbf{1}_6) = \sum_{x=1}^{6} x \cdot \mathcal{P}(\mathbf{1}_x) = \tfrac{1}{6} \cdot \sum_{x=1}^{6} x = \tfrac{7}{2}£.$$

è la quota di scommessa certa per partecipare ad un giuoco in cui il giocatore riceve dal banco la quota aleatoria X£. $\triangleleft$

Esempio 1.2.3. Da una massa di n biglie numerate da 1 a n se ne estrae una a caso; sia X n.a. che si realizza. Si calcoli la quota di scommessa in un giuoco in cui il banco è tenuto a pagare al giocatore X^2£. Si ha

[24] Vale simultaneamente il segno di uguale *sse* $x_1 = x_2 = \cdots = x_n$.

$$\mathcal{P}\big(X^2\big) = \mathcal{P}\big(1^2 \cdot \mathbf{1}_1 + 2^2 \cdot \mathbf{1}_2 + \cdots + n^2 \cdot \mathbf{1}_n\big) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{(n+1)(2n+1)}{6}.$$

Ad es., per $n = 10$ si ha $\mathcal{P}(X^2) = 38.5£$. $\lhd$

L'estensione della nozione di scommessa coerente a partizioni costituite da infiniti eventi $\big\{E_j\big\}_{j=1}^{\infty}$ aventi probabilità $\big\{p_j\big\}_{j=1}^{\infty}$, comporta qualche cautela.

Per le operazioni finora eseguite è stata sufficiente la proprietà della additività finita. In presenza di collezioni infinite di eventi (e dunque di partizioni infinite) occorre adottare la addittività completa.

Siano $\big\{x_j\big\}_{j=1}^{\infty}$ i premi associati agli eventi, sia $X = \sum_{j=1}^{\infty} x_j \mathbf{1}_{E_j}$ il premio aleatorio. Segue la definizione.

Definizione 1.2.1. Si definisce valore atteso, o speranza matematica, o valor medio, o aspettazione, o previsione, o semplicemente media della v.a. discreta X, il valore certo

$$E(X) \;=\; \sum_{j=1}^{\infty} x_j \cdot p_j\,, \tag{1.3}$$

purché esista finito. $\lhd$

È noto, dal teorema di H. Lebesgue, che condizione *n.s.* affinché la previsione della v.a. X esista finita è che $E\big(|X|\big) = \sum_{j=1}^{\infty} |x_j| \cdot p_j < \infty$. Teorema che torna utile in presenza di v.a. per cui è $E(X) = \infty$. È legittimo, in tali casi, concepire una quota di scommessa?

Dagli esempi appena proposti si ricava l'idea che la speranza matematica del guadagno possa essere *sempre* usata come valutazione complessiva ed equa delle conseguenze aleatorie. L'esempio che segue pone un serio limite a questa certezza. Si consideri dunque l'importante esempio.

Esempio 1.2.4. (*Il paradosso di San Pietroburgo.*)
Si immagini una sequenza indefinita di lanci indipendenti di una moneta regolare con esito $\{T, C\}$ ed un giocatore che riceve dal banco $2£$ se T esce al primo colpo, $4£$ se T esce solo al secondo colpo, $\ldots$, $2^x£$ se T esce solo all'$x-$esimo colpo.

Poiché 2^X è la somma versata dal banco al giocatore, con probabilità $X \sim q(x) = \big(\frac{1}{2}\big)^x$, $\forall x = 1, 2, \ldots$, ne consegue che l'equivalente certo che il giocatore deve dare al banco per partecipare al giuoco è

$$\mathcal{P}\big(2^X\big) \;=\; \sum_{x\in\mathbb{N}} 2^x \cdot q(x) \;=\; \sum_{x\in\mathbb{N}} 2^x \cdot \big(\tfrac{1}{2}\big)^x \;=\; \sum_{x\in\mathbb{N}} 1 \;=\; \infty\,.$$

È evidente che il valore ∞ non può essere il *prezzo equo* da pagare al banco per partecipare ad un giuoco il cui ricavo è comunque finito. D'altra parte siamo in presenza di una scommessa in cui "si vince sempre"...

In sostanza noi ci troviamo di fronte ad infinite scommesse eque, in cui, $\forall x \in \mathbb{N}$ il giocatore paga $1\pounds$ per ricevere il premio $2^x \pounds$, con probabilità 2^{-x}. Ma se l'addittività delle scommesse è valida, ed è equo pagare $n\pounds$ per partecipare ad n scommesse, allora non vi è somma che sia sufficiente per partecipare a tutte le infinite scommesse. $\lhd$

Nota 1.2.1. Nella memoria nella quale Daniel Bernoulli propose il suo celebre paradosso (1738) si sostenne che andasse criticata alla radice l'idea stessa di equità. E con essa la pretesa di ragionare in termini di *danaro in senso assoluto*. E che, per contro, occorresse spostare l'attenzione sull'utilità attesa, cioè su quanto la vincita rappresenta per il giocatore.

Ovvero, per dirla con le parole del coevo Gabriel Cramer: "*I matematici stimano il denaro in ragione della sua quantità, mentre un uomo di buon senso lo stima in proporzione all'uso che può farne.*"

L'idea di utilità attesa divenne di centrale importanza nella teoria delle decisioni, nella teoria economica, nel campo delle assicurazioni. Le importanti considerazioni critiche che ne seguirono, misero seri limiti al principio stesso di equità e con esso la possibilità di scambio (in ogni istante del giuoco) tra giocatore e banco.

Scambio realistico in giuochi "fra amici e parenti" cioè con premi e perdite di "piccolo importo" e e comunque senza tenere conto della propensione (o meno) al rischio di chi scommette. Affatto irrealistico quando il giuoco si svolge, ad esempio, tra assicuratore ed assicurato.

La critica di D. Bernoulli, accolta *in toto* dalla comunità scientifica (e dai soggettivisti in particolare) fu il primo (e decisivo) passo verso lo sviluppo della teoria delle decisioni in condizioni di incertezza, nella teoria della utilità e del rischio. $\lhd$

Chapter 2

Probabilità condizionata

Le probabilità sono sempre condizionate.
(Bruno de Finetti, 1970)[(1)]

I concetti di dipendenza/indipendenza stocastica (subordinata o no) sono regolati in forma di definizione da semplici (e precise) relazioni formali che, di per sé, non presentano particolari difficoltà.

Le sorprese nascono quando, persino in certi casi elementari, si assiste a conclusioni controintuitive se non paradossali. Ed è proprio qui che meglio si comprende la nozione di dipendenza/indipendenza stocastica.

Presupposto imprescindibile di ogni ragionamento induttivo è la possibilità di rivedere, mediante formali procedure, le personali valutazioni di probabilità alla luce di nuovi esperimenti, di nuovi fatti.

Come si è detto nella nota **1.1.3**, le valutazioni di probabilità dipendono dalle conoscenze del soggetto nel momento in cui esse si fanno. Permodoché subentrando *altri fatti* e/o *altre evidenze sperimentali* e/o *altre informazioni*, vecchie o recenti che siano, le valutazioni del soggetto non sono (e non possono essere) piú le stesse.

Se $\mathbb{P}(E)$ è la probabilità che il soggetto assegna ad E, è naturale che, verificandosi l'evento H, o qualunque altra circostanza, egli avverta l'esigenza di riconsiderare le precedenti valutazioni e a giudicare l'evento E come piú probabile oppure meno probabile, o infine a lasciare le cose come si trovano. In simboli $\mathbb{P}(E \mid H) \lesseqgtr \mathbb{P}(E)$. Ma in quali situazioni ed in che modo e un *nuovo fatto* può essere informativo?

Infine un cenno è riservato a certe strutture di dipendenza di cui occorre tenere conto nella induzione.

---

1) Bruno de Finetti, *Teoria delle Probabilità*. Vol. I, Einaudi Ed., Torino, 1970.

2.1 Indipendenza/dipendenza stocastica

> *. . . la condizione di indipendenza viene spesso sottintesa e ammessa per valida quando non lo è affatto.*
> *. . . in certo senso è l'indipendenza stocastica che costituisce il caso-limite piuttosto idealizzato e la dipendenza il caso normale, anziché il contrario . . .*
>
> (Bruno de Finetti, 1970)[(2)]

Iniziamo con la definizione di indipendenza stocastica tra due eventi.

Definizione 2.1.1. Gli eventi $(A, B) \subset \Omega$, a probabilità positiva, sono detti stocasticamente indipendenti (o indipendenti e basta se non ci sono equivoci) allora che l'accadere dell'evento B non modifica la valutazione di probabilità dell'evento A, e *viceversa*. Formalmente se

$$\mathbb{P}(A \mid B) \;=\; \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(B \mid A) \;=\; \mathbb{P}(B)\,. \triangleleft \tag{2.1}$$

È facile provare che se una delle (2.1) è vera lo è pure l'altra, cosí come sono vere le $\mathbb{P}(A|\bar{B}) = \mathbb{P}(A)$, $\mathbb{P}(B|\bar{A}) = \mathbb{P}(B)$, *etc.* Ed ancora, la (2.1) è vera *sse* è vera la fattorizzazione

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A) \cdot \mathbb{P}(B)\,. \tag{2.2}$$

Una facile estensione delle (2.1): dati gli eventi (A, B, C), se A è indipendente da $B \cap C$, allora

$$\mathbb{P}(A) \;=\; \mathbb{P}(A|B, C) \;=\; \mathbb{P}(A|\bar{B}, C) \;=\; \mathbb{P}(A|B, \bar{C}) \;=\; \mathbb{P}(A|\bar{B}, \bar{C})\,.$$

Se le (2.1) e (2.2) non sono vere, se cioè tra A e B vi è dipendenza, vale la definizione che segue.

Definizione 2.1.2. Gli eventi A e B sono positivamente o negativamente correlati a seconda che

$$\mathbb{P}(A \mid B) \begin{Bmatrix} > \\ < \end{Bmatrix} \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(A \cap B) \begin{Bmatrix} > \\ < \end{Bmatrix} \mathbb{P}(A) \cdot \mathbb{P}(B)\,. \triangleleft \tag{2.3}$$

Provata l'equivalenza delle condizioni (2.3), è poi facile verificare che $\mathbb{P}(A \mid B) > \mathbb{P}(A) \;\; \Leftrightarrow \;\; \mathbb{P}(B \mid A) > \mathbb{P}(B)$, *etc.*

[2)] Bruno de Finetti, *Teoria delle Probabilità*. Vol. I, Einaudi Ed., Torino, 1970.

Nota 2.1.1. In certi manuali di *CdP* si evince che la dipendenza stocastica abbia un carattere cronologico e poggi sull'idea di "causa". Gli argomenti sono: se *prima* accade A è probabile che *poi* accadrà B, l'accadere di A favorisce (in base a qualche potere o proprietà) l'accadere di B, *etc.*

Non si nega l'esistenza di eventi che accadono in tempi differenti e che sono stocasticamente dipendenti. Si nega che fra tempi e dipendenza stocastica vi sia un nesso logico. È sufficiente notare che gli eventi che entrano nelle relazioni di indipendenza (2.1) e (2.2) e di dipendenza (2.3) sono simmetrici, quale che ne sia la temporalità.

Se proprio non si vuole rinunciare all'idea di "causa", si può, per lo meno in certune situazioni, pensare che una stessa "causa" che abbia agito su tutti gli eventi in gioco senza che tra essi vi sia dipendenza.(3) Per approfondimenti si vada piú avanti all'esempio **2.2.1**. $\triangleleft$

Mentre l'incopatibilità degli eventi è una proprietà della logica formale, l'indipendenza è una proprietà della probabilità e come tale è soggettiva.

L'esempio che segue mostra come l'indipendenza possa dipendere dalla *precisione* con cui il soggetto fissa le probabilità.

Esempio 2.1.1. Data una famiglia con almeno due figli e gli eventi

$$\begin{aligned} M_j &= \text{“}\textit{il j-mo figlio è M}\text{”}\,, \\ F_j &= \text{“}\textit{il j-mo figlio è F}\text{”}\,, \\ D &= \text{“}\textit{i primi due figli sono di differente sesso}\text{”}\,, \end{aligned}$$

dire se (M_1, D) sono stocasticamente indipendenti ovvero se positivamente o negativamente correlati, nell'ipotesi che il sesso del secondo figlio non dipenda dal sesso del primo e che $\mathbb{P}(M_j) = p_M$ e $\mathbb{P}(F_j) = p_F = 1 - p_M$, $\forall j$.

Si ha $\mathbb{P}(M_1 \cap D) = \mathbb{P}(M_1 \cap F_2) = \mathbb{P}(M_1) \cdot \mathbb{P}(F_2) = p_M(1 - p_M)$. Inoltre, data la scomposizione $D = \big(M_1 \cap F_2\big) \cup \big(F_1 \cap M_2\big)$ si ha

$$\begin{aligned} \mathbb{P}(D) &= \mathbb{P}\big(M_1 \cap F_2\big) + \mathbb{P}\big(F_1 \cap M_2\big) \\ &= \mathbb{P}(M_1) \cdot \mathbb{P}(F_2) + \mathbb{P}(F_1) \cdot \mathbb{P}(M_2) \\ &= p_M \cdot (1 - p_M) + (1 - p_M) \cdot p_M \\ &= 2 \cdot p_M(1 - p_M) \end{aligned}$$

donde $\mathbb{P}(M_1) \cdot \mathbb{P}(D) = 2p_M^2(1 - p_M)$. Riassumendo, gli eventi (M_1, D) sono indipendenti *sse* $p_M(1 - p_M) = 2p_M^2(1 - p_M)$. Se dunque il soggetto fissa $p_M = p_F = \frac{1}{2}$, allora $\mathbb{P}(M_1 \cap D) = \mathbb{P}(M_1) \cdot \mathbb{P}(D) = \frac{1}{4}$ e gli eventi (M_1, D)

(3) Ad es., con un inverno freddo (la "causa") crescono le probabilità di morte della popolazione anziana. Ciò nonostante le morti di due anziani sono tra loro indipendenti.

sono indipendenti. Gli eventi (M_1, D) sono, invece, dipendenti se il soggetto, consultate le tavole dell'ISTAT, con maggior cura, fissa $\mathbb{P}(M_j) = 0.511^{(4)}$. Poiché $\mathbb{P}(M_1|D) = \dfrac{\mathbb{P}(M_1 \cap D)}{\mathbb{P}(D)} = \dfrac{p_M(1-p_M)}{2 \cdot p_M(1-p_M)} = \frac{1}{2}$ e $\mathbb{P}(M_1) = p_M \geq \frac{1}{2}$, ci fa concludere che M_1 e D sono negativamente correlati. $\lhd$

Le "controintuitive" conclusioni dell'esempio **2.1.1** mostrano che l'indipendenza è una proprietà legata alla probabilità (che è soggettiva) e non degli eventi in sé. Il concetto è ribadito dall'esempio che segue.

Esempio 2.1.2. Un soggetto, soppesata una moneta, fissa $\mathbb{P}(C) = p$ e $\mathbb{P}(T) = q = 1 - p$. La moneta è lanciata tre volte. Dire se gli eventi

$$\begin{aligned} A &= \text{“\textit{nei tre lanci appaiono sia } } T \text{ \textit{che} } C\text{”}\,, \\ B &= \text{“\textit{al piú si verifica una} } T\text{”}\,, \end{aligned}$$

sono indipendenti. Tenuto conto che i tre lanci possono generare 8 distinti risultati, si ha

$$\mathbb{P}(A) \;=\; 3pq\,, \quad \mathbb{P}(B) \;=\; q^2(q+3p)\,, \quad \mathbb{P}(A \cap B) \;=\; 3pq^2\,.$$

Gli eventi A e B sono indipendenti *sse* $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ ovvero

$$3pq^2 \;=\; 3pq \cdot q^2(q+3p)\,,$$

donde $2q^2 - 3q + 1 = 0$. La quale è verificata *sse* $p = q = \frac{1}{2}$, cioè *sse* il soggetto valuta la moneta bilanciata. $\big($Il lettore verifichi che se $p < \frac{1}{2}$ $\left[p > \frac{1}{2}\right]$ allora A e B sono negativamente [positivamente] correlati.$\big)$ $\lhd$

Quando si passa ad una terna di eventi (E_1, E_2, E_3), la condizione

$$\mathbb{P}(E_1 \cap E_2 \cap E_3) \;=\; \mathbb{P}(E_1) \cdot \mathbb{P}(E_2) \cdot \mathbb{P}(E_3), \tag{2.4}$$

non è piú sufficiente per dire che gli eventi sono indipendenti. È necessario che per ogni coppia di indici valgano le fattorizzazioni

$$\mathbb{P}(E_i \cap E_j) \;=\; \mathbb{P}(E_i) \cdot \mathbb{P}(E_j)\,, \quad \forall i \neq j\,. \tag{2.5}$$

È facile provare che se gli eventi della terna (E_1, E_2, E_3) sono indipendenti, allora ciascuno dei tre è indipendente dagli altri due, e *viceversa*. Dunque per ogni permutazione degli indici valgono le condizioni

[4)] Nel 2001 l'ISTAT forní, per la Sardegna, il valore $p_M = 0.511$ (*XIV* Censimento della Popolazione Italiana, (2001). Tale valore si discosta di poco dal dato nazionale.

$$\mathbb{P}(E_i \mid E_j, E_k) = \mathbb{P}(E_i)\,, \quad \forall i \neq j \neq k\,. \tag{2.6}$$

Nel caso in cui una sola delle (2.6) non fosse vera, gli eventi (E_1, E_2, E_3) non sarebbero indipendenti. Si consideri in proposito l'esempio.

Esempio 2.1.3. Si lanciano tre monete regolari che producono 8 possibili sequenze equiprobabili costituite da T e C. Dire se gli eventi

$$\begin{aligned} E_1 &= \text{“}\textit{la sequenza contiene almeno due } T\text{”}\,, \\ E_2 &= \text{“}\textit{la sequenza ha un numero pari di } T\text{”}\,, \\ E_3 &= \text{“}\textit{la sequenza inizia con } C\text{”}\,, \end{aligned}$$

sono stocasticamente indipendenti.

Sebbene $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \mathbb{P}(E_3) = \frac{1}{2}$ e $\mathbb{P}(E_1, E_2, E_3) = \frac{1}{8}$, gli eventi (E_1, E_2, E_3) non sono indipendenti. Risulta infatti che

(i) facendo intervenire le (2.5) si ha

$$\mathbb{P}(E_1, E_2) = \tfrac{3}{8}\,, \quad \mathbb{P}(E_1, E_3) = \tfrac{1}{8}\,, \quad \mathbb{P}(E_2, E_3) = \tfrac{2}{8}\,;$$

(ii) o anche, applicando le (2.6), si ha

$$\begin{aligned} \mathbb{P}(E_1|E_2, E_3) &= \frac{\mathbb{P}(E_1, E_2, E_3)}{\mathbb{P}(E_2, E_3)} = \tfrac{1/8}{2/8} = \tfrac{1}{2} = \mathbb{P}(E_1)\,, \\ \mathbb{P}(E_2|E_1, E_3) &= \frac{\mathbb{P}(E_1, E_2, E_3)}{\mathbb{P}(E_1, E_3)} = \tfrac{1/8}{1/8} = 1 \neq \mathbb{P}(E_2)\,, \\ \mathbb{P}(E_3|E_1, E_2) &= \frac{\mathbb{P}(E_1, E_2, E_3)}{\mathbb{P}(E_1, E_2)} = \tfrac{1/8}{3/8} = \tfrac{1}{3} \neq \mathbb{P}(E_3)\,. \ \lhd \end{aligned}$$

Ma può pure accadere che, valendo le (2.5), non valga la fattorizzazione (2.4) e dunque che gli eventi non sono indipendenti. Si consideri l'esempio.

Esempio 2.1.4. Lanciati due dadi regolari, si considerano gli eventi

$$\begin{aligned} P_j &= \text{“}\textit{il dado } j\textit{−mo mostra una faccia pari}\text{”}\,,\ j \in \{1, 2\} \\ D &= \text{“}\textit{la somma dei punti è dispari}\text{”}\,. \end{aligned}$$

È facile vedere che $\mathbb{P}(P_1) = \mathbb{P}(P_2) = \mathbb{P}(D) = \frac{1}{2}$ e che

$$\mathbb{P}(P_1, P_2) = \mathbb{P}(P_1, D) = \mathbb{P}(P_2, D) = \tfrac{1}{4} \quad \text{e che} \quad \mathbb{P}(P_1, P_2, D) = 0\,.$$

Dunque: gli eventi (P_1, P_2, D) sono indipendenti se presi due alla volta. Non lo sono piú se presi in blocco. $\lhd$

L'esempio che segue tratta il caso prove senza restituzione da urne finite a composizione nota.

Esempio 2.1.5. Una certa urna contiene N biglie n delle quali azzurre (A) ed $N-n$ bianche (B), con N ed n noti. Estraendo le biglie senza restituzione, sia $E_i =$"*la i-ma biglia estratta è di color A*".

Si calcoli la probabilità di E_2, quando non è noto l'esito della estrazione precedente e si stabilisca se gli eventi (E_1, E_2) sono indipendenti.

Poiché $E_2 = (E_1 \cap E_2) \cap (\bar{E}_1 \cup E_2)$, la formula (1.2) porge

$$\begin{aligned}
\mathbb{P}(E_2) &= \mathbb{P}(E_1 \cap E_2) + \mathbb{P}(\bar{E}_1 \cup E_2) \\
&= \mathbb{P}(E_1) \cdot \mathbb{P}(E_2 \mid E_1) + \mathbb{P}(\bar{E}_1) \cdot \mathbb{P}(E_2 \mid \bar{E}_1) \\
&= \frac{n}{N} \cdot \frac{n-1}{N-1} + \Big(1 - \frac{n}{N}\Big) \cdot \frac{n}{N-1} = \frac{n}{N}\,,
\end{aligned}$$

da cui si evince l'equiprobabilità degli eventi E_1 ed E_2. Per mostrare che gli eventi E_1 ed E_2 non sono indipendenti, è sufficiente considerare che

$$\mathbb{P}(E_1 \cap E_2) = \frac{n}{N} \cdot \frac{n-1}{N-1} \neq \mathbb{P}(E_1) \cdot \mathbb{P}(E_2) = \Big(\frac{n}{N}\Big)^2,$$

eventi che divengono indipendenti per $(N, n) \to \infty$, con $\dfrac{n}{N} \to \theta$. Il fatto che

$$\mathbb{P}(E_2 \mid E_1) = \frac{n-1}{N-1} < \frac{n}{N} = \mathbb{P}(E_2)\,,$$

mostra che E_1 ed E_2 sono correlati negativamente.

I risultati dell'esempio si estendono facilmente a successioni di $k > 2$ eventi. E si ha

$$\mathbb{P}(E_{k+1}) = \frac{n}{N}\,, \quad \mathbb{P}(E_k \mid E_j) = \frac{n-1}{N-1} < \frac{n}{N} = \mathbb{P}(E_k)\,. \ \triangleleft$$

Segue la definizione di indipendenza stocastica riferita ad una $n-$pla di eventi qualsiasi.

Definizione 2.1.3. Gli eventi $\big(A_1, A_2, \ldots, A_n\big)$ sono stocasticamente indipendenti *sse* per qualunque $k-$pla di indici $\{1 \leq h_1 < h_2 < \cdots < h_k \leq n\}$, con $k \in \{2, \ldots, n\}$, vale la fattorizzazione

$$\mathbb{P}\Big(\bigcap_{i=1}^{k} A_{h_i}\Big) = \prod_{i=1}^{k} \mathbb{P}(A_{h_i})\,. \quad \triangleleft \tag{2.7}$$

La definizione di indipendenza stocastica di n eventi poteva essere data, in modo affatto equivalente, ma in forma piú complicata, usando la probabilità condizionata. Il teorema che segue fissa solo delle condizioni necessarie

***Teorema* 2.1.1.** *Condizione necessaria (ma non sufficiente) affinché gli eventi* $(A_1, A_2, \ldots, A_n)$ *siano stocasticamente indipendenti è che ciascuno di essi sia indipendente dalla* $(n-1)-$*pla degli eventi rimanenti. Ovvero*

$$\mathbb{P}(A_j \mid A_1, \ldots, A_{j-1}, A_{j+1}, \ldots A_n) \;=\; \mathbb{P}(A_j)\,, \quad \forall j = 1, 2, \ldots, n. \quad \triangleleft \qquad (2.8)$$

Nota 2.1.2. Le precedenti definizioni disegnano il quadro formale della indipendenza stocastica; meno semplice è spiegarne il significato. In breve
− l'indipendenza stocastica dipende delle informazioni e dalle opinioni del soggetto; essa è la conseguenza (forse la piú importante) della concezione soggettiva adottata;
− è impossibile stabilire l'indipendenza stocastica *solo* in base della descrizione fisica dell'esperimento; e magari con la stessa *oggettiva* certezza con cui si stabilisce, ad esempio, l'incompatibilità e l'indipendenza logica.(5) $\triangleleft$

Al di là di ogni altra considerazione, l'esempio che segue mostra ancora che l'indipendenza stocastica è una proprietà della probabilità.

Esempio 2.1.6. L'urna U_1 contiene 4 biglie di color bianca, rossa, verde, nera (B, R, V, N); l'urna U_2, 3 biglie di color B, R, V. Da una di esse, sono estratte due biglie con restituzione. Stabilire se gli eventi $H = B \cup R$ e $K = R \cup V$ sono indipendenti quando l'urna di provenienza è la U_1 oppure la U_2. Poiché
$[\mathsf{U}_1]$ $\quad \mathbb{P}(B) = \mathbb{P}(R) = \mathbb{P}(V) = \mathbb{P}(N) = \frac{1}{4}$ e $\mathbb{P}(H) = \mathbb{P}(K) = \frac{1}{2}$; e dunque

$$\mathbb{P}(H \cap K) \;=\; \mathbb{P}(R) \;=\; \tfrac{1}{4} \quad \text{e} \quad \mathbb{P}(H) \cdot \mathbb{P}(K) \;=\; \tfrac{1}{4}\,;$$

$[\mathsf{U}_2]$ $\quad \mathbb{P}(B) = \mathbb{P}(R) = \mathbb{P}(N) = \frac{1}{3}$ e $\mathbb{P}(H) = \mathbb{P}(K) = \frac{2}{3}$; e dunque

$$\mathbb{P}(H \cap K) \;=\; \mathbb{P}(R) = \tfrac{1}{3} \;\neq\; \mathbb{P}(H) \cdot \mathbb{P}(K) \;=\; \tfrac{4}{9}\,,$$

diciamo che H e K sono indipendenti o dipendenti a seconda che l'urna di provenienza sia U_1 oppure U_2.

---

5) Giova ribadire il fatto che lncompatibilità e lndipendenza logica sono proprietà *degli* eventi (come tali esse prescindono da ogni valutazione di probabilità e rientrano nella logica del certo) e che l'indipendenza stocastica non è una *proprietà fisica degli eventi*, è bensí una proprietà che dipende dalle valutazioni soggettive di probabilità.

Il fatto che, dal punto di vista fisico, i due casi siano del tutto simili, rende il caso paradossale. La spiegazione del paradosso sta (tutta) nella valutazione "oggettiva" (si osservi: "oggettiva") delle probabilità degli eventi elementari (B, R, V, N) e (B, R, V), relativi alle urne U_1 e U_2.[6] $\lhd$

2.1.1 Prove fisicamente separate

L'espressione "*prove ripetute con probabilità costante ma incognita*" che spesso ricorre tra i fautori della concezione frequentista della probabilità, nasce dall'errata idea secondo cui "poiché le estrazioni ripetute da un'urna prove sono *fisicamente separate*, *ergo* sono indipendenti". Dunque, l'indipendenza stocastica come proprietà fisica delle prove[7].

Tipici casi di eventi separati considerati indipendenti sarebbero le prove ripetute di uno stesso fenomeno quali

• gli esiti di prove condotte su pezzi di un certo lotto, indipendenti per molti esperti di controllo di qualità;

• le misure ripetute di uno stesso oggetto effettuate nelle stesse condizioni sperimentali, (*specie*) se condotte in luoghi, tempi, laboratori differenti.

• il caso di un unico esperimento che si articola in distinte sotto prove (esempio tipico è dato dal lancio in blocco di k monete che producono k eventi T o C, in cui noi ignoriamo del tutto come le monete, nei loro reciproci urti, si influenzino tra loro).

Per mostrare che l'indipendenza ha poco da spartire con le prove separate, è sufficiente tornare all'esempio **2.1.6**, in cui si mostra che ci possono essere eventi non separati (gli eventi H e K, urna U_1) che sono indipendenti.

Si supponga dunque di avere un'urna U che contiene biglie di colore A e B, secondo le proporzioni θ e $1-\theta$. Con θ nota a Tizio ed ignota a Caio.

Da essa si prelevano n biglie con restituzione, con $n \geq 2$. Sia X_i l'indicatore d'evento relativo alla $i-$ma prova: $X_i = 1$ se la $i-$ma biglia estratta è di color A e $X_i = 0$ se è di color B.

Posto che in $n = 10$ prove si siano verificati solo insuccessi, cioè $X_i = 0$, $\forall i = 1, 2, \ldots, n$, si chiede, a Tizio e Caio, di valutare la probabilità che in una ulteriore prova si abbia $X_{n+1} = 1$.

---

[6] L'Esempio **2.1.6** può essere (ri-)formulato, in modo equivalente, assumendo un'unica urna contenente le biglie (B, R, V, N) e due differenti soggetti. Il primo che valuta $\mathbb{P}(B) = \mathbb{P}(R) = \mathbb{P}(V) = \mathbb{P}(N) = \frac{1}{4}$; il secondo $\mathbb{P}(B) = \mathbb{P}(R) = \mathbb{P}(V) = \frac{1}{3}$ e $\mathbb{P}(N) = 0$.

[7] Ma oggi c'è qualcuno che reputi che vi sia un legame tra gli eventi fisicamente separati $P =$ "*peste a Milano*(1630)" e $C =$"*congiuntura di Giove con Saturno*"?

Tizio che conosce θ afferma che gli eventi $X_i = x_i$ sono indipendenti ed identicamente distribuiti (in sigla *iid*) subordinatamente a θ fatto che gli consente di scrivere

$$\mathbb{P}\big(X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}\,,\, \theta\big) \;=\; \mathbb{P}\big(X_n = x_n \mid \theta\big)\,, \qquad (2.9)$$

o l'equivalente condizione

$$\mathbb{P}\big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \theta\big) \;=\; \prod_{i=1}^{n} \mathbb{P}\big(X_i = x_i \mid \theta\big)\;, \qquad (2.10)$$

se cioè è noto il valore di θ è del tutto ininfluente conoscere gli esiti delle prime n prove per stabilire la probabilità dell'evento $X_{n+1} = x_{n+1}$.

Tizio giustifica l'indipendenza degli eventi $\big\{(X_i = x_i), i = 1, 2, \ldots, n\big\}$, dicendo che, essendo le prove con resa, la composizione dell'urna U non è modificata in alcun modo da precedenti prove. Fatto che consente di dire che $X_i = x_i$ e $X_j = x_j$ sono, $\forall i \neq j = 1, 2, \ldots, n$, equiprobabili.

Differente e (di certo) piú interessante è il caso in cui la composizione θ dell'urna U non è nota. (O è parzialmente nota.)

Coloro i quali definiscono la probabilità in base alla frequenza (cioè gli empiristi) *sono tenuti* a dire che gli eventi $X_i = x_i$ sono:

- *indipendenti*, in quanto riferentesi a prove *fisicamente separate*, per cui sarebbe priva di valore qualsiasi idea che un soggetto potrebbe farsi su U nel corso delle prove;
- *equiprobabili*, dovendo la probabilità riguardare ogni singola prova;

asserzioni riassunte nella già citata oscura formula "*prove equiprobabili con probabilità θ costante ma incognita*".

Per contro, dal punto di vista soggettivistico il fatto stesso che ad ogni prova (via via che le prove si succedono) il soggetto *modifichi* le sue personali valutazioni di probabilità indica che le prove siano tra loro dipendenti e che *non* sia lecito scrivere

$$\mathbb{P}\big(X_{n+1} = x_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n\big) \;=\; \mathbb{P}\big(X_{n+1} = x_{n+1}\big)\,,$$

$$\mathbb{P}\big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\big) \;=\; \prod_{i=1}^{n} \mathbb{P}\big(X_i = x_i\big)\,,$$

il tutto a voler trascurare il fatto che se le prove fossero stocasticamente indipendenti nessun esperimento sarebbe informativo.

Quanto ha preceduto suggerisce di rivisitare le le nozioni di indipendenza stocastica condizionata e di scambiabilità stocastica, centrali nel ragionamento induttivo.

2.2 Dipendenza condizionata e scambiabilità

In certo senso, il concetto piú importante della teoria soggettivistica è quello di "eventi scambiabili". (Kyburg e Smoker, Wiley, 1964)

Come si è visto al §**2.1.1** può accadere il caso che due eventi A e B siano stocasticamente indipendenti, subordinatamente ad una certa ipotesi H, ma che lo siano piú se H non è nota. Segue dunque la definizione.

Definizione 2.2.1. Si dice che tra gli eventi A e B vi è indipendenza stocastica condizionata (o subordinata) all'evento H, allora che

$$\mathbb{P}(A|H,B) = \mathbb{P}(A|H) \quad \Leftrightarrow \quad \mathbb{P}(B|H,A) = \mathbb{P}(B|H)\,. \quad \triangleleft \tag{2.11}$$

I teoremi che seguono (semplice da provare il primo, meno semplice il secondo) forniscono proprietà dell'indipendenza stocastica subordinata.

***Teorema* 2.2.1.** *Se vale la* (2.11)*, allora*

$$(i) \quad \mathbb{P}(A|H,\bar{B}) = \mathbb{P}(A|H) \quad \Leftrightarrow \quad \mathbb{P}(\bar{A}|H,B) = \mathbb{P}(\bar{A}|H)\,,$$
$$(ii) \quad \mathbb{P}(A \cap B \mid H) = \mathbb{P}(A \mid H) \cdot \mathbb{P}(B \mid H)\,. \triangleleft$$

***Teorema* 2.2.2.** *L'indipendenza degli eventi A e B implica l'indipendenza subordinata a qualsivoglia evento H a probabilità positiva*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \;\Rightarrow\; \mathbb{P}(A \cap B|H) = \mathbb{P}(A|H) \cdot \mathbb{P}(B|H)\,. \triangleleft$$

L'esempio seguente è una applicazione del teorema **2.2.2** e che richiama concetti esposti nella nota **2.1.1**.

Esempio 2.2.1. (*B. de Finetti, 1970.*) Il banco lancia un dado regolare a sei facce e realizza il punto H. Leo vince se (con lo stesso dado) realizza un punto che supera H; cosí pure fa Ugo. Dato H, gli eventi $L =$ "*Leo vince*" e $U =$ "*Ugo vince*", sono indipendenti. *Ergo* $\mathbb{P}(L \cap U|H) = \mathbb{P}(L|H) \cdot \mathbb{P}(U|H)$. La tavola **2.1** riporta tali probabilità condizionate.

H	*1*	*2*	*3*	*4*	*5*	*6*
$\mathbb{P}(L\|H)=\mathbb{P}(U\|H)$	5/6	4/6	3/6	2/6	1/6	0
$\mathbb{P}(L\cap U\|H)$	25/36	16/36	9/36	4/36	1/36	0

Tavola **2.1** - *Probabilità condizionate.*

Per provare che l'indipendenza subordinata non implica l'indipendenza è sufficiente notare che $\mathbb{P}(L)=\mathbb{P}(U)=\sum_{H=1}^{6}\mathbb{P}(H)\cdot\mathbb{P}\big(L|H\big)=\frac{1}{6}\sum_{H=1}^{6}\mathbb{P}\big(L|H\big)=\frac{5}{12}$ e che $\mathbb{P}(L\cap U)=\sum_{H=1}^{6}\mathbb{P}(H)\cdot\mathbb{P}\big(E\cap U|H\big)=\frac{110}{432}>\mathbb{P}(L)\cdot\mathbb{P}(U)=\big(\frac{5}{12}\big)^2$. $\lhd$

Diffidente nei confronti delle *facili* ipotesi di indipendenza, per lo piú assunte da *default*, de Finetti avanzò l'idea che per la dipendenza stocastica (condizionata o meno) si potesse parlare di dipendenza in senso diretto ed in senso indiretto

− *diretto* quando un evento modifica (in modo significativo ai fini delle valutazioni di probabilità) le circostanze in cui si verifica un altro evento;

− *indiretto* quando si manifesta un fatto H che influisce su tutta una collezione di eventi "separati" $(E_1, E_2, \ldots, E_k)$.

Per quest'ultimo, tipico è il caso (ben noto agli attuari) di un inverno rigido che influisce sulla probabilità di morte di soggetti che non hanno tra loro né relazioni né vicinanze. L'esempio seguente che tratta un tale caso presenta evidenti analogie con l'esempio **2.2.1**.

Esempio 2.2.2. In una nota città, nella quale vivono gli anziani a e b, gli inverni sono, nel 25(%), nel 50(%) nel 25(%) dei casi, o molto rigidi (H_1), o nella norma (H_2), o tiepidi (H_3). Sia A $[B]$ l'evento "*a $[b]$ muore nel corso del prossimo inverno*". Siano $\mathbb{P}(A|H_j)$ e $\mathbb{P}(B|H_j)$ le probabilità di morte subordinate all'ipotesi H_j, $j=1,2,3$; vedi tavola **2.2**. Supposta l'indipendenza condizionata di A e B, cioè $\mathbb{P}(A\cap B|H_j)=\mathbb{P}(A|H_j)\cdot\mathbb{P}(B|H_j)$, dire se A e B sono indipendenti.

H_j	$\mathbb{P}(H_j)$	$\mathbb{P}(A\|H_j)$	$\mathbb{P}(B\|H_j)$	$\mathbb{P}(A\cap B\|H_j)$
H_1	0.25	0.30	0.20	.0600
H_2	0.50	0.12	0.10	.0120
H_3	0.25	0.05	0.03	.0015

Tavola **2.2** - *Probabilità condizionate.*

Donde

$$\begin{aligned}
\mathbb{P}(A) &= \sum_{j=1}^{3} \mathbb{P}(H_j)\cdot\mathbb{P}\big(A|H_j\big) = 0.1475\,,\\
\mathbb{P}(B) &= \sum_{j=1}^{3} \mathbb{P}(H_j)\cdot\mathbb{P}\big(G|H_j\big) = 0.1075\,,
\end{aligned}$$

Il fatto che

$$\mathbb{P}(A\cap B) = \sum_{j=1}^{3} \mathbb{P}(H_j)\cdot\mathbb{P}\big(A\cap B|H_j\big) \;=\; 0.0214 > \mathbb{P}(A)\cdot\mathbb{P}(B) = 0.0159\,,$$

indica che l'indipendenza subordinata non implica l'indipendenza. $\triangleleft$

Diversamente da quanto l'intuito suggerirebbe *non è vero* che

$$\mathbb{P}(A|H,B) \;=\; \mathbb{P}(A|H) \;\Leftrightarrow\; \mathbb{P}(A|\bar{H},B) \;=\; \mathbb{P}(A|\bar{H})\,.$$

L'affermazione sarà provata a mezzo di un esempio elementare.

Esempio 2.2.3. Sono dati due dadi a sei facce; regolare il primo (ipotesi H) e non bilanciato il secondo (ipotesi $\bar{H}$). Si considerino gli eventi

$$\begin{aligned}
E_1 &= \text{“}\textit{appare il “5” per } 8 \textit{ volte”}\,,\\
E_2 &= \text{“}\textit{appare il “5” al } 9^o \textit{ lancio”}\,.
\end{aligned}$$

Si ha $\mathbb{P}(E_2|H,E_1) \;=\; \mathbb{P}(E_2|H)$ e $\mathbb{P}(E_2|\bar{H},E_1) \;>\; \mathbb{P}(E_2|\bar{H})$. $\triangleleft$

Di “norma” le probabilità di eventi tipici dei giochi d'azzardo sono valutate nell'ipotesi di onestà del gioco, (monete, dadi, carte, *etc.*, regolari). Cadendo tale condizione, per i soggettivisti, cade pure l'ipotesi di indipendenza delle prove. E questo è un (altro) punto su cui oggettivisti e soggettivisti hanno opinioni differenti.

Nel caso di prove ripetute da urne di composizione incognita, i primi assumono ancora l'ipotesi di indipendenza delle prove, i secondi al contrario, reputano che ad ogni prova le valutazioni di probabilità dipendano dagli esiti delle prove che hanno preceduto e che le prove siano scambiabili[(8)]. La nozione di scambiabilità stocastica sarà data a mezzo di esempi. Ecco il primo.

[8)] Del concetto di *scambiabilità*, quale concetto chiave nel ragionamento per induzione, parlò B. de Finetti in occasione di un ciclo di conferenze tenute presso l'*Institut Henri Poincaré* (Parigi, 1935). B. de Finetti, che nell'occasione aveva parlato di *eventi equivalenti*, fu indotto ad adottare la piú appropriata locuzione di *eventi scambiabili* su suggerimento di M. Fréchet, presente alle conferenze.

Esempio 2.2.4. L'urna U contiene un numero noto N di biglie; Z di esse, Z non noto, sono azzurre (A, i successi) ed $N-Z$ bianche (B, gli insuccessi).

Si osservi che il n.a. $Z \in \mathcal{Z} = \{0, 1, \ldots, N\}$ è un evento *decidibile*, cioè fattualmente verificabile. Le possibili composizioni di U sono esprimibili in forma di ipotesi aleatorie

$$H_z = \text{“}Z = z \textit{ delle } N \textit{ biglie contenute in } \mathsf{U} \textit{ sono } A\text{”} , \quad Z \in \mathcal{Z} ,$$

le quali costituiscono la partizione $\mathcal{H} = \{H_0, H_1, \ldots, H_N\}$[9].

Si supponga che il soggetto, che ignora la composizione di U, abbia, sugli eventi $H_z \in \mathcal{H}$, informazioni parziali espresse dalla distribuzione coerente di probabiltà $\{p_0^0, p_1^0, \ldots, p_N^0\}$; cioè $p_z^0 = \mathbb{P}(H_z) \geq 0$, $\forall z$, con $\sum_{z=0}^N p_z^0 = 1$. Siano

$$E(Z) = \sum_{z=0}^{N} z \cdot p_z^0 \ , \quad E(Z^2) = \sum_{z=0}^{N} z^2 \cdot p_z^0 \, ,$$

i primi due momenti *a priori* del n.a. Z.

Da U sono estratte delle biglie con resa, sia $(X_1, X_2, \ldots)$ la successione aleatoria degli indicatori d'evento, dove $X_i = 0$ ovvero $X_i = 1$ a seconda che l'esito della $i-$ma prova sia un insuccesso od un successo.

Subordinatamente a $Z = z$, cioè all'ipotesi H_z, i n.a. X_i sono *iid* con $X_i|z \sim Bern\big(\cdot \,|\, \frac{1}{N} z\big)$, $\forall z$ e, $\forall i \neq j$ e $\forall (x, x') \in \{0,1\} \times \{0,1\}$, si ha

$$\begin{aligned} \mathbb{P}\big(X_i = x \mid H_z\big) &= \Big(\frac{z}{N}\Big)^x \Big(1 - \frac{z}{N}\Big)^{1-x} \\ \mathbb{P}\big(X_i = x \, , \, X_j = x' \mid H_z\big) &= \mathbb{P}\big(X_i = x \mid H_z\big) \cdot \mathbb{P}\big(X_j = x' \mid H_z\big) \\ &= \Big(\frac{z}{N}\Big)^{x+x'} \Big(1 - \frac{z}{N}\Big)^{2-x-x'} . \end{aligned} \tag{2.12}$$

Se *viceversa* la composizione dell'urna non è nota, allora i n.a. X_i
(a) sono identicamente distribuiti;
(b) *non* sono stocasticamente indipendenti.

Per provare l'affermazione (a) si riprenda la (3.4)(i) e si consideri che

$$(X_i = x) \ = \ \Big(\bigcup_{z=0}^{N} H_z\Big) \cap (X_i = x) \ = \ \bigcup_{z=0}^{N} (X_i = x) \cap H_z \, ,$$

[9] Giova ricordare che l'insieme di eventi $\mathcal{H}$ costituisce una partizione dell'evento certo Ω se: (i) $\cup_{j=0}^N H_j = \Omega$ e se (ii) $\forall i \neq j$ si ha $H_i \cap H_j = \emptyset$.
Si osservi poi che, anzichè considerare le Z, si sarebbe potuto ragionare in termini di proporzione aleatoria, cioè $\theta_Z = \frac{1}{N} \cdot Z$, con $\theta_Z \in \big\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\big\}$.

donde

$$\mathbb{P}\big(X_i = x\big) \;=\; \mathbb{P}\left(\sum_{z=0}^{N}(X_i = x) \cap H_z\right) \;= \sum_{z=0}^{N} \mathbb{P}(X_i = x | H_z) \cdot p_z^0 \;, \qquad (2.13)$$

ma poiché

$$\mathbb{P}(X_i = 1) \;=\; \sum_{z=0}^{N} \frac{1}{N} \cdot z \cdot p_z^0 \;=\; \frac{1}{N} \cdot E(Z)\,,$$

la (2.13) diviene

$$\begin{aligned} \mathbb{P}(X_i = x) \;&=\; \sum_{z=0}^{N} \left(\frac{z}{N}\right)^x \left(1 - \frac{z}{N}\right)^{1-x} \cdot p_z^0 \\ &=\; \left(\tfrac{1}{N} \cdot E(Z)\right)^x \left(1 - \tfrac{1}{N} \cdot E(Z)\right)^{1-x} \;, \end{aligned} \qquad (2.14)$$

la quale dipende solo dalla distribuzione $\{p_0^0, p_1^0, \ldots, p_N^0\}$ ed è vera $\forall i = 1, 2, \ldots$. Fatto, quest'ultimo, che prova l'affermazione $(a)^{(10)}$.

Quanto poi all'affermazione (b), si osservi che subordinatamente ad H_z le v.a. (X_i, X_j) sono indipendenti, come indica la (3.4)(ii), esse non lo sono piú quando la composizione dell'urna U non è nota (o è parzialmente nota); $\forall (x, x') \in \{0, 1\} \times \{0, 1\}$, si ha infatti

$$\begin{aligned} \mathbb{P}\big(X_i = x, X_j = x'\big) \;&=\; \mathbb{P}\left(\sum_{z=0}^{N}(X_i = x, X_j = x') \cap H_z\right) \\ &=\; \sum_{z=0}^{N} \mathbb{P}\big(X_i = x, X_j = x' \mid H_z\big) \cdot \mathbb{P}\big(H_z\big) \\ &=\; \sum_{z=0}^{N} \left(\frac{z}{N}\right)^{x+x'} \left(1 - \frac{z}{N}\right)^{2-x-x'} \cdot p_z^0\,, \end{aligned} \qquad (2.15)$$

donde, con un po' di pazienza,

---

10) Dalla (2.14) si deduce poi che $E(X_i) = E(X_i^2) = \frac{1}{N}E(Z)$ e dunque $Var(Z_i) = \frac{1}{N}E(Z) \cdot \big(1 - \frac{1}{N}E(Z)\big)$.

$$\mathbb{P}\big(X_i = x, X_j = x'\big) \;=\; \begin{cases} 1 - \dfrac{2}{N}E(Z) + \dfrac{1}{N^2}E(Z^2) & (x = x' = 0) \\ \dfrac{1}{N}E(Z) - \dfrac{1}{N^2}E(Z^2) & (x = 1 - x') \\ \dfrac{1}{N^2}E(Z^2) & (x = x' = 1) \end{cases} .$$

Il fatto che $\forall(x, x')$ risulti

$$\mathbb{P}(X_i = x) \cdot \mathbb{P}(X_j = x') \;\neq\; \mathbb{P}(X_i = x, X_j = x')\ , \text{[11]}$$

prova la dipendenza delle v.a. X_i ed X_j.[12]

Quale che sia la distribuzione $\{p_0^0, p_1^0, \ldots, p_N^0\}$ non degenere, gli eventi $(X_i = 1, X_j = 1)$, cosí come gli eventi $(X_i = 0, X_j = 0)$, sono, *correlati positivamente*. Gli eventi $(X_i = 0, X_j = 1)$, al contrario, e $(X_i = 1, X_j = 0)$ *che* sono *negativamente* correlati. Si ha infatti

$$\begin{aligned} \mathbb{P}(X_i = 1 \mid X_j = 1) \;&=\; \frac{\mathbb{P}(X_i = 1 \cap X_j = 1)}{\mathbb{P}(X_j = 1)} \;=\; \frac{1}{N} \cdot \frac{E(Z^2)}{E(Z)} \\ &> \; \mathbb{P}(X_i = 1) \;=\; \frac{1}{N} \cdot E(Z)\,, \end{aligned}$$

$$\begin{aligned} \mathbb{P}(X_i = 0 \mid X_j = 1) \;&=\; \frac{\mathbb{P}(X_i = 0 \cap X_j = 1)}{\mathbb{P}(X_j = 1)} \;=\; 1 - \frac{1}{N} \cdot \frac{E(Z^2)}{E(Z)} \\ &< \; \mathbb{P}(X_i = 0) \;=\; 1 - \frac{1}{N} \cdot E(Z)\,. \end{aligned}$$

Il ragionamento non cambia se le estrazioni (le prove) da U fossero ad es. esaustive. Saprebbe il lettore verificare l'affermazione? $\triangleleft$

Estendere la (2.15) agli $h \geq 2$ eventi $\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big)$. è facile. Posto $s = \sum_{i=1}^h x_i$, si ha infatti

---

11) Ad es.: $\mathbb{P}(X_i = 1) \cdot \mathbb{P}(X_j = 1) = \big(\frac{1}{N}E(Z)\big)^2 \neq \mathbb{P}(X_i = 1, X_j = 1) = N^{-2} \cdot E(Z^2)$.

12) Giova osservare che se un altro soggetto (in base a sue informazioni) è in condizione di affermare che $p_{z_0}^0 = 1$ e $p_z^0 = 0$, $\forall z \neq z_0$, ovvero che la distribuzione di Z è degenere in $Z = z_0$, gli eventi $(X_i = x, X_j = x')$ sono (per lui) indipendenti; si veda la (2.15).

$$\begin{aligned}\mathbb{P}\big(X_1 = x_1, \ldots, X_h = x_h\big) &= \mathbb{P}\left(\sum_{z=0}^{N} \bigcap_{i=1}^{h} (X_i = x_i) \cap H_z\right) \\ &= \sum_{z=0}^{N} \mathbb{P}\Big(\bigcap_{i=1}^{h} (X_i = x_i) \,\big|\, H_z\Big) \cdot \mathbb{P}\big(H_z\big) \qquad (2.16) \\ &= \sum_{z=0}^{N} \left(\frac{z}{N}\right)^s \left(1 - \frac{z}{N}\right)^{h-s} \cdot p_z^0 \,,\end{aligned}$$

la quale evidenzia che
- gli h eventi $\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big)$ sono equiprobabili ma non sono stocasticamente indipendenti;
- la $\mathbb{P}\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big)$ dipende da h e dal numero di successi $s = \sum_{i=1}^{h} x_i$ e *non* dall'ordine delle prove, potendosi solo dire che

$$\begin{aligned}&\mathbb{P}\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big) = \\ &\qquad\qquad = \mathbb{P}\big(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \ldots, X_{i_h} = x_{i_h}\big)\,, \qquad (2.17)\end{aligned}$$

dove $\{i_1, i_2, \ldots, i_h\}$ è una qualsiasi permutazione degli indici $\{1, 2, \ldots, h\}$.

Se vale la condizione (2.16) o la equivalente condizione (2.17), gli eventi $\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big)$ sono detti scambiabili.

La condizione di scambiabilità (2.17), valida per il n.a. $(0-1)$, o dicotomici, non si limita a tali n.a. Si vada al prossimo paragrafo.

Come si ribadirà alla fine del prossimo capitolo, se i n.a. dicotomici $\big(X_1, X_2, \ldots, X_h\big)$ sono scambiabili allora $\forall h$, e vale perciò la (2.17), la probabilità $\mathbb{P}\big(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h\big)$ dipende *solamente* da h e dal numero dei successi nelle h prove $s = \sum_{i=1}^{h} x_i$. Inoltre

$$\mathbb{P}\big(S = s \mid h\big) = \binom{h}{s} \cdot \mathbb{P}\big(X_{i_1} = x_{i_1}, X_{i_2} = x_{i_2}, \ldots, X_{i_h} = x_{i_h}\big)\,.$$

Si dimostra infine che se gli eventi $(X_1 = x_1, X_2 = x_2, \ldots, X_h = x_h)$ sono indipendenti allora sono anche scambiabili. Non vale il *viceversa*.

Nota 2.2.1. (*Spiegare la scambiabilità*)
Si supponga di essere in presenza delle urne U_1 e U_2 contenenti biglie color A (successi) e B (insuccessi) secondo le proporzioni note $\theta_1 \neq \theta_2$. Caio, una

volta scelta una delle due urne secondo la lotteria $P(\mathsf{U}_j) = p_j > 0$, $j = 1, 2$, preleva da essa, senza dirmi quale delle due sia n biglie con restituzione.

Per Caio le prove sono *iid* e, quale che sia il risultato conseguito, la probabilità del futuro successo non risentirà dell'esperimento realizzato. Né alcun risultato modificherà lo stato delle sue conoscenze.

Per me, *viceversa*, le prove sono solo *id* e *scambiabili* e da esse ho tutto da apprendere al fine di: (i) fare previsioni sull'esito della futura prova (ii) risalire all'urna sorteggiata.

Il ragionamento per induzione secondo l'impostazione di Bruno de Finetti si riduce, come -in piú riprese- da lui stesso dichiarato (e come si vedrà) alla applicazione del teorema delle probabilità composte e delle probabilità totali; il problema (i) rimanda ai modelli predittivi, il problema (ii) richiama il teorema di Bayes. $\triangleleft$

Al pari dell'indipendenza stocastica, anche la scambiabilità è una proprietà che dipende dalle valutazioni di probabilità del soggetto. E si rifletta sul fatto che le valutazioni di probabilità di eventi futuri in base ad eventi passati è possibile solo quando si riesce a stabilire un *legame logico* di dipendenza fra eventi passati ed eventi futuri.

Sulla natura di tale legame logico o, per meglio dire, dei varî tipi di legami che possono sussistere tra gli eventi di successioni di prove, si tornerà fra poco. Nell'Esempio **2.2.4** appena discusso il legame che appare piú appropriato è la scambiabilità stocastica.

Come si vedrà fra breve, il concetto di scambiabilità, che riguarda le piú differenti prove dalle piú disparate popolazioni, è cruciale nel ragionamento induttivo in ottica soggettiva.

2.2.1 Cenni sulle strutture di dipendenza

Il soggetto che è chiamato ad esaminare le relazioni fra fenomeni ed a costruire l'*adeguato* modello probabilistico, deve aver un'idea delle varie forme di dipendenza/indipendenza fra eventi.

Dato un processo di osservazione $\big(X_n\big)_{n\geq 1} = \big(X_1, X_2, \ldots, X_n \ldots\big)$, dire se i n.a. X_i sono indipendenti, oppure scambiabili, oppure periodici, *etc.*, è possibile solo se si guarda agli aspetti fisici (oggettivi) del processo.

Ne discende che le possibili forme di dipendenza/indipendenza fra eventi che il soggetto (caso per caso) può osservare vadano assunte come ipotesi e mai essere postulate.

Iniziamo col dire che tutto ciò che si può dire è che il processo è noto allora che è nota la f.r. congiunta

$$F_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;=\; \mathbb{P}(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n)$$

da cui discendono tutte le possibili f.r. marginali e condizionate. Studiare una successione “in generale” non è né semplice, né utile. Piú interessante è lo studio delle relazioni di dipendenza che il ricercatore ipotizza.

Seguono le ipotesi di dipendenza a cui piú spesso si ricorre in statistica.

▶ *Ipotesi di indipendenza ed ipotesi di bernoullinità.*
L'ipotesi di indipendenza comporta che

$$F_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;=\; \prod_{j=1}^{n} F_j(x_j)\ ,$$

la bernoullianità richiede, anche, l'equidistribuzione degli eventi, donde

$$F_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;=\; \prod_{j=1}^{n} F(x_j)\ .$$

▶ *Ipotesi di scambiabilità.*
Con tale ipotesi risulta

$$F_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;=\; F_{i_1,i_2,\ldots,i_n}(x_1, x_2, \ldots, x_n)\,,$$

dove $\{i_1, i_2, \ldots, i_n\}$ è una qualsiasi permutazione degli indici $\{1, 2, \ldots, n\}$.

▶ *Ipotesi di markovianità.*
In caso di processo markoviano omogeneo del 1^o ordine si ha

$$F_{0,1,2,\ldots,n}(x_0, x_1, x_2, \ldots, x_n) \;=\; F_0(x_0) \cdot \prod_{j=1}^{n} F_{j|j-1}(x_j \mid x_{j-1})\ .$$

Giova ricordare una nota proprietà delle successioni markoviane.

Nota 2.2.2. Se i $k+1$ eventi $(A_0, A_1, \ldots, A_k)$ sono implicati uno nell'altro, cioè $A_0 \supset A_1 \supset \cdots \supset A_k$, allora essi sono dotati di struttura markoviana. Tenuto conto che se $A \supset B$, allora $\mathbb{P}(H|A, B) = \mathbb{P}(H|B)$, si ha

$$\begin{aligned}
\mathbb{P}(A_k) &= \mathbb{P}(A_0, A_1, \ldots, A_k) \\
&= \mathbb{P}(A_0) \cdot \mathbb{P}(A_1|A_0) \cdots \mathbb{P}(A_k|A_0, \ldots, A_{k-1}) \\
&= \mathbb{P}(A_0) \cdot \prod_{j=1}^{k} \mathbb{P}(A_j|A_{j-1})\ .
\end{aligned}$$

Un processo di alternativa, o $(0-1)$, che si realizza, ad esempio, lanciando ripetutamente una moneta, è markoviano? $\lhd$

Vi sono infine importanti processi per i quali è realistico formulare altre ipotesi: ad es. di *stazionarietà*, di *periodicità*, *etc.*

Pur di notevole interesse in statistica, tali ipotesi non saranno trattate in questa dispensa.

Chapter 3

Aggiornare le probabilità

Il teorema di Bayes costituisce la chiave di volta e il concetto informatore di ogni attività costruttiva del pensiero.
(Bruno de Finetti, 1970)[(1)]

Da circa un secolo a questa parte si è affermata l'abitudine di attribuire a Bayes qualunque analisi statistica e/o ragionamento induttivo il quale

− rifiuta l'idea di identificare la probabilità con la frequenza;
− accetta la definizione soggettiva di probabilità **1.1.4**;
− prescrive di assegnare una probabilità (sia pur provvisoria) a tutti gli eventi incerti che entrano nell'analisi;
− aggiorna le probabilità assegnate mediante il teorema di Bayes.

La dimostrazione del teorema di Bayes, impropriamente detto teorema delle cause, è molto semplice essendo conseguenza immediata dei teoremi delle probabilità composte e di scomposizione, o di disintegrazione.

L'applicazione del teorema è semplice e spesso *routinaria*. Difficoltà, di natura numerica, possono sorgere in certe applicazioni.

3.1 Il teorema di Bayes

Si consideri una $k-$pla di eventi (detti anche cause, o anche ipotesi, o anche stati di natura) $\mathcal{H} = \{H_1, H_2, \ldots, H_k\}$, $k \geq 2$, costituenti una partizione dell'evento certo Ω. Sulle ipotesi H_j sono date le probabilità ***a priori*** o iniziali (cioè presperimentali) $\{\mathbb{P}(H_1), \mathbb{P}(H_2), \ldots, \mathbb{P}(H_k)\}$.

Se è una evidenza sperimentale compatibile con almeno una delle ipotesi della partizione $\mathcal{H}$, e dunque $E \subset \Omega$ con $E \neq \emptyset$, è possibile scrivere la formula

---
[1)] Bruno de Finetti, *Teoria delle Probabilità.* Vol. I, Einaudi Ed., Torino, 1970.

di scomposizione, data dalla combinazione lineare convessa

$$\mathbb{P}(E) = \sum_{j=1}^{k} \mathbb{P}(H_j) \cdot \mathbb{P}(E|H_j)\,, \tag{3.1}$$

che risponde alla condizione di conglomerabilità

$$\min_j \left\{\mathbb{P}(E|H_j)\right\} \;\le\; \mathbb{P}(E) \;\le\; \max_j \left\{\mathbb{P}(E|H_j)\right\},$$

e dunque alla proprietà di internalità della media.

Segue l'enunciato del teorema di Bayes (1702-1761).

***Teorema* 3.1.1.** (*Thomas Bayes*, 1763, *post.*)
Data la partizione $\mathcal{H} = \{H_j, j = 1, 2, \ldots, k\}$ di Ω, con $\mathbb{P}(H_j) > 0$, $\forall j$, e l'evento $E \subset \Omega$, con $\mathbb{P}(E) > 0$, compatibile con almeno uno degli eventi H_j, si ha

$$\mathbb{P}(H_j \mid E) \;=\; \mathbb{P}(H_j) \cdot \frac{\mathbb{P}(E \mid H_j)}{\mathbb{P}(E)} \;=\; \frac{\mathbb{P}(H_j) \cdot \mathbb{P}(E \mid H_j)}{\displaystyle\sum_{i=1}^{k} \mathbb{P}(H_i) \cdot \mathbb{P}(E \mid H_i)}\,. \tag{3.2}$$

Dim. Uguagliando i secondi membri delle espresioni

$$\mathbb{P}(E \cap H_j) = \mathbb{P}(E) \cdot \mathbb{P}(H_j|E) \quad \text{e} \quad \mathbb{P}(H_j \cap E) = \mathbb{P}(H_j) \cdot \mathbb{P}(E|H_j)$$

si ha $\mathbb{P}(H_j|E) = \mathbb{P}(H_j) \cdot \dfrac{\mathbb{P}(E|H_j)}{\mathbb{P}(E)}$. Richiamando la formula di scomposizione (3.1), il teorema è dimostrato. $\lhd$

Il teorema di Bayes è lo strumento logico per aggiornare, le probabilità *a priori* $\mathbb{P}(H_j)$ con le probabilità $\mathbb{P}(H_j|E)$, dette probabilità *a posteriori* alla luce dell'esperimento osservato E. Le $\mathbb{P}(E|H_j)$ sono dette verosimiglianze delle ipotesi H_j dato E.

La probabilità $\mathbb{P}(E)$, a denominatore della (3.2), è una quantità di normalizzazione che non dipende piú dalle ipotesi H_j. Tale è il motivo per cui il teorema di Bayes è anche espresso nella forma

$$\mathbb{P}(H_j|E) \;\stackrel{c}{=}\; \mathbb{P}(H_j) \cdot \mathbb{P}(E|H_j)\,.^{(2)} \tag{3.3}$$

---

[2] Ad avviso di chi scrive, il simbolo "$\stackrel{c}{=}$" che vuole significare "*uguale a meno di una certa costante moltiplicativa positiva*" è, preferibile al simbolo "$\propto$" correntemente usato.

Se $\alpha = \dfrac{\mathbb{P}(E \mid H_j)}{\mathbb{P}(E)} = 1$ allora E e H_j sono indipendenti ed il verificarsi di E non modifica la valutazione *a priori* di H_j. Per contro, se $\alpha > 1$ $[\alpha < 1]$ si dice che l'evidenza E rafforza [indebolisce] l'ipotesi H_j.

Nota 3.1.1. Accettando l'assioma (v) della addittività completa, è possibile trattare partizioni che prevedono infinite ipotesi $\mathcal{H} = (H_j)_{j \geqslant 2}$. La formula di scomposizione (3.1) risulta, in tal caso. $\mathbb{P}(E) = \sum_{j=1}^{\infty} \mathbb{P}(H_j) \cdot \mathbb{P}(E|H_j)$. $\triangleleft$

Con i casi proposti si tenterà di mostrare perché il teorema di Bayes è lo strumento *par excellence* capace di aggiornare (con palese facilità d'uso) le probabilità delle ipotesi H_j alla luce di *nuovi fatti*.

Si osservino, in particolare, i casi di ipotesi deboli *a priori*, cioè ritenute in partenza poco probabili, le quali alla luce di certe evidenze empiriche cioè *a posteriori*, si rivelano piú probabili, e *viceversa*.

3.1.1 Applicazioni elementari del teorema di Bayes

Nelle applicazioni che seguono, sia le probabilità *a priori* che le verosimiglianze, che entrano nella espressione del teorema di Bayes, sono state valutate "*oggettivamente*", ricorrendo alle definizioni di probabilità **1.1.1** e **1.1.2**, razionale ed empirica.

Piú avanti le applicazioni del teorema tratteranno, in buona misura, casi in cui le probabilità *a priori* sono state fissate soggettivamente.

Iniziamo con un caso ben noto presso gli epidemiologi.

Esempio 3.1.1.
Un ragazzo su 10000 di una certa popolazione giovane è affetto dalla malattia M. Si sottopongono gli alunni a *screening* di massa. Il *test*, essendo poco preciso, diagnostica "falsi positivi" e "falsi negativi" con probabilità pari a 0.04 e 0.02. Calcolare la probabilità che un alunno con diagnosi positiva sia, in realtà, sano.

Stato di natura	$(-)$	$(+)$
Sano	0.96	0.04
Malato	0.02	0.98

Tav.**3.1** - *Verosimiglianze.*

Si indichi con M $[S]$ l'evento "l'alunno è positivo [negativo]" e con $(+)$ $\big[(-)\big]$ "l'alunno è dichiarato positivo [negativo]". In tavola **3.1** sono riportate le verosimiglianze, cioè le probabilità di "falso positivo" $\mathbb{P}\big\{(+)|S\big\} = 0.04$

e di "falso negativo" $\mathbb{P}\big\{(-)|M\big\} = 0.02$. Essendo le probabilità *a priori* $\mathbb{P}(S) = 0.9999$, $\mathbb{P}(M) = 0.0001$, la formula di Bayes porge

$$\begin{aligned}\mathbb{P}\{M \mid (+)\} &= \frac{\mathbb{P}(M)\cdot\mathbb{P}\big\{(+)\mid M\big\}}{\mathbb{P}(M)\cdot\mathbb{P}\big\{(+)\mid M\big\}+\mathbb{P}(S)\cdot\mathbb{P}\big\{(+)\mid S\big\}}\\ &= \frac{0.0001\cdot 0.98}{0.0001\cdot 0.98 + 0.9999\cdot 0.04} = 0.00244\,,\end{aligned}$$

donde $\mathbb{P}\big\{S \mid (+)\big\} = 1-0.00244 = 0.99756$. *Ergo* la diagnosi $(+)$ *non* implica che l'alunno sia malato.

Il teorema di Bayes dice solo che se *a priori* $\mathbb{P}(M) = 0.0001$, la probabilità *a posteriori*, pur cresciuta di $k \cong 25$ volte è ancora bassa. $\lhd$

Nota 3.1.2. Nell'esempio appena considerato, cosí come in tanti altri casi, taluno è portato a confondere verosimiglianze, le $\mathbb{P}(E|H_j)$, e probabilità *a posteriori*, le $\mathbb{P}(H_j|E)$. Nel nostro caso vi è la fallace propensione a ritenere che se il *test* medico (si badi: applicato ai soli alunni ammalati) mette in evidenza la malattia nel 98 (%) dei casi, allora è molto probabile che chi è diagnosticato $(+)$ sia ammalato.

L'altro piú grave errore sta nel dimenticare che il 4 (%) degli alunni sani (la maggioranza degli alunni) è diagnosticato $(+)$. Ovvero che i falsi positivi superano di gran lunga i falsi negativi.[3]

Per chi commette i due errori, la discrepanza (numerica) fra la verosimiglianza $\mathbb{P}\big\{(+)\big|M\big\} = 0.98$ e la probabilità *a posteriori* $\mathbb{P}\big\{M\big|(+)\big\} = 0.00244$, ottenuta col teorema di Bayes, è un mistero misterioso. $\lhd$

L'esempio che segue mostra un caso in cui le probabilità *a priori* sono calcolate applicando la definizione razionale di probabilità; fatto che per gli oggettivisti è la sola legittimazione all'uso del teorema di Bayes.

L'esempio evidenzia l'errore che si commette quando si trascurano le informazioni disponibili, nel nostro caso le probabilità *a priori*, e si traggono inferenze (e conclusioni) usando solo le verosimiglianze.

Esempio 3.1.2. Si lancia per 8 volte una moneta regolare. Se per 8 volte appare la faccia T, nell'urna U si depongono 20 biglie color A. Diversamente, si mettono in U 10 biglie color A e 10 color B.

Sulla composizione di U, Leo ignaro dell'esito dei lanci fa delle ipotesi

[3] Non molto tempo fa, il nostro Parlamento proibí la *cannabis* sentenziando (nero su bianco) che "se il 99.9 (%) degli eroinomani (in passato) ha fatto uso di *spinelli*, allora gli attuali *spinellatori* presto o tardi arriveranno all'eroina". Esercizio letterario: si ripeta il "ragionamento" sostituendo la parola *spinello* con la parola *caramella*.

$$\begin{cases} H_1 & = \text{“}\mathsf{U}\ \textit{contiene solo biglie color } A\text{”}, \\ H_2 & = \text{“}\mathsf{U}\ \textit{contiene } 10\ \textit{biglie } A\ \textit{e}\ 10\ B\text{”}. \end{cases}$$

Da U sono estratte 5 biglie con restituzione e tutte sono di color A. Dato tale risultato, lo si indichi con E, si calcoli la probabilità delle ipotesi H_j.

Tenuto conto che le probabilità *a priori* sono

$$\mathbb{P}(H_1) = 2^{-8} = \tfrac{1}{256} \qquad \text{e} \qquad \mathbb{P}(H_2) = 1 - 2^{-8} = \tfrac{255}{256}\,.$$

che le verosimiglianze sono

$$\mathbb{P}\big(E \mid H_1\big) = 1 \qquad \text{e} \qquad \mathbb{P}\big(E \mid H_2\big) = 2^{-5} = \tfrac{1}{32}\,,$$

e che $\mathbb{P}\big(E\big) = \sum_{j=1}^{2} \mathbb{P}(H_j) \cdot \mathbb{P}\big(E \mid H_j\big) = \frac{1}{2^8} \cdot 1 + \big(1 - \frac{1}{2^8}\big) \cdot \frac{1}{2^5} \cong 0.035$, si ottiene $\mathbb{P}\big(H_1 \mid E\big) = 0.1115$ e $\mathbb{P}\big(H_2 \mid E\big) = 0.8885$. $\triangleleft$

Coloro i quali basano il ragionamento induttivo solo sulla verosimiglianza devono propendere per l'ipotesi H_1. Se, piú correttamente, si tiene conto (anche) del meccanismo di costruzione di U (che fa parte integrante delle conoscenze in possesso del soggetto) si deve accettare l'ipotesi H_2.

Esempio 3.1.3.

Ivo, Pio e Ugo hanno in tasca rispettivamente 18, 10 e 8 monete, 11, 7 e 4 delle quali coniate in Italia e le altre all'Estero. Ad uno dei tre un ladro ruba 5 monete 2 delle quali estere. Si calcoli la probabilità che la vittima sia Ivo. (*Traccia.*) Per le informazioni in nostro possesso, non siamo autorizzati a pensare che, per il ladro, vi sia stata una particolare “preferenza” per uno dei tre. La descrizione... dell'esperimento, suggerisce di riguardare ciascuna delle tre tasche come... un'urna esaustiva.

Siano H_1, H_2, H_3, gli eventi “il derubato è o Ivo o Pio o Ugo”. Sia $E =$ “*delle* $t = 5$ *monete rubate* $x = 2$ *sono estere*”.

Dal testo si ricava che le verosimiglianze delle ipotesi H_j , dato E, sono

$$\mathbb{P}(E \mid H_j) = Hyperg(x \mid m_j, n_j, t)\,,\ j = 1, 2, 3\,,$$

dove (m_j, n_j) sono le monete coniate in Italia e all'estero, possedute dal $j-$mo amico prima del furto e con $x = 2$ e $t = 5$. *Ergo*

$$\text{(Ivo)} \qquad Hyperg(2 \mid m_1, n_1, 5) = \frac{\binom{11}{2} \cdot \binom{7}{3}}{\binom{18}{5}} = 0.2247\,,$$

$$\text{(Pio)} \qquad Hyperg(2 \mid m_2, n_2, 5) \;=\; \frac{\binom{7}{2}\cdot\binom{3}{3}}{\binom{10}{5}} \;=\; 0.0833\,,$$

$$\text{(Ugo)} \qquad Hyperg(2 \mid m_3, n_3, 5) \;=\; \frac{\binom{4}{2}\cdot\binom{4}{3}}{\binom{8}{5}} \;=\; 0.4286\,.$$

La tav. **3.2** riporta: le probabilità *a priori* delle ipotesi H_j, le verosimiglianze, le probabilità *a posteriori* delle H_j, avendo osservato E . $\lhd$

ipotesi	H_j	H_1	H_2	H_3	*tot.*
p. *a priori*	$\mathbb{P}(H_j)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1.
verosimiglianze	$\mathbb{P}(E\|H_j)$	.2247	.0833	0.4286	-
	$\mathbb{P}(H_j)\cdot\mathbb{P}(E\|H_j)$	$\frac{1}{3}$.2247	$\frac{1}{3}$.0833	$\frac{1}{3}$.4286	$\frac{1}{3}$.7366
p. *a posteriori*	$\mathbb{P}(H_j\|E)$	.3050	.1131	.5818	.9999

Tav. **3.2** - *Tavola dei calcoli - esempio dei tre amici.*

Esempio 3.1.4. (Corrado Gini, 1936, *XIV E. F.*)
Italo non ha piú avuto notizie del suo migliore amico Scipio da quando, un anno fa, partì volontario per la guerra d'Africa. Si sa che, dopo un anno di guerra, l'1 (%) dei soldati non è piú in vita, che solo il 70 (%) è solito scrivere lettere, che il 29 (%) dei soldati è distratto. Si sa anche che il 20 (%) della posta inviata, data la precarî dei trasporti, va perduta. Italo si domanda se per caso l'amico non sia deceduto.

Sulle ipotesi $H_1 =$"*Scipio è morto*", $H_2 =$"*Scipio scrive*" e $H_2 =$"*Scipio è distratto*", Italo adotta come probabilità *a priori* le proporzioni $\{0.01, 0.70, 0.29\}$. Le verosimiglianze delle ipotesi H_j, dato $N =$"*da un anno Italo non riceve posta*", sono $\mathbb{P}(N|H_1) = 1$, $\mathbb{P}(N|H_2) = 0.20$ e $\mathbb{P}(N|H_3) = 1$.

La tavola **3.3** riporta: le probabilità *a priori* delle H_j, le verosimiglianze, le probabilità *a posteriori* delle ipotesi H_j, dato N. Da notare che, dato N, a rafforzarsi sensibilmente è l'ipotesi H_3 e non l'ipotesi H_1. $\lhd$

ipotesi	H_j	H_1	H_2	H_3	*tot.*
prob. *a priori*	$\mathbb{P}(H_j)$	0.01	0.70	0.29	1.
verosimiglianze	$\mathbb{P}(N\|H_j)$	1.	0.20	1.	-
	$\mathbb{P}(H_j)\cdot\mathbb{P}(N\|H_j)$	0.01	0.14	0.29	0.44
prob. *a posteriori*	$\mathbb{P}(H_j\|N)$	0.0227	0.3182	0.6591	1.

Tav.**3.3** - *Tavola dei calcoli per l'esempio di C. Gini.*

Nota 3.1.3. L'esempio **3.1.4** pur fittizio riflette ansie ed atteggiamenti di certo *grande* pubblico. Cosa accade in certe famiglie quando un giovane rincasa con qualche ritardo? o un parente non si fa vivo da tempo? *etc.* Anche in questi casi il paradigma bayesiano mostra tutta la sua utilità. $\triangleleft$

Esempio 3.1.5. (*Paradosso di Bertrand*)
Nella scatola S_1 ci sono due monete d'oro (A), nella scatola S_2 ci sono una moneta A e l'altra di bronzo (B), in S_3 due monete B.

Da una delle tre scatole scelta a caso si preleva sempre a caso una moneta. Posto che questa sia A, si calcoli la probabilità che anche l'altra sia A.

Si ha $\mathbb{P}(S_1) = \mathbb{P}(S_2) = \mathbb{P}(S_3) = \frac{1}{3}$. Se $E =$"*la moneta prelevata è* A", è agevole verificare che $\mathbb{P}(E|S_1) = 1$, $\mathbb{P}(E|S_2) = \frac{1}{2}$, $\mathbb{P}(E|S_3) = 0$.

Poiché la moneta nascosta è A *sse* la scatola sorteggiata è la S_1, la formula di Bayes porge

$$\begin{aligned}\mathbb{P}(S_1|E) &= \frac{\mathbb{P}(S_1)\cdot\mathbb{P}(E|S_1)}{\mathbb{P}(S_1)\cdot\mathbb{P}(E|S_1)+\mathbb{P}(S_2)\cdot\mathbb{P}(E|S_2)+\mathbb{P}(S_3)\cdot\mathbb{P}(E|S_3)}\,,\\ &= \frac{\frac{1}{3}\cdot 1}{\frac{1}{3}\cdot 1+\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\cdot 0} = \frac{2}{3}\,.\end{aligned}$$

Volendo evitare l'uso del teorema di Bayes, si può pensare di nominare le monete $\{A_1, A_2, A_3\}$ e $\{B_1, B_2, B_3\}$ ed osservare che se si verifica l'evento E, allora restano i casi equiprobabili (A_1, A_2), (A_2, A_1) e (A_3, B_1), due soli dei quali favorevoli. Donde la *non* intuitiva risposta $\frac{2}{3}$. $\triangleleft$

Nota 3.1.4. Allo scopo di mettere in guardia i lettori dalle risposte *svelte*, *ovvie* e ... *sbagliate* J. Bertrand costruì numerosi esempi a "risposta controintuitiva", facendo cosí, del paradosso, il suo genere preferito.

Se A è la moneta sorteggiata, allora va esclusa la scatola S_3; e siccome le scatole restanti sono equiprobabili, cioè $\mathbb{P}(S_1) = \mathbb{P}(S_2)$, la risposta *svelta/ovvia* è ... "$\frac{1}{2}$"; quella che piú ricorre presso il *grande pubblico*[4]. $\triangleleft$

Gli esempi che hanno preceduto evidenziano la capacità del teorema di Bayes di aggiornare, alla luce di nuove evidenze sperimentali, le probabilità fissate *a priori* dal soggetto. Cosí inteso, il meccanismo bayesiano è strumento insostituibile nel ragionamento per induzione.

Allo scopo di mostrare, sia pure *in nuce*, tale capacitià si torni all'esempio **2.2.4**, già discusso nel precedente capitolo, che si riferisce ad una popolazione finita, cioè un'urna costituita da un numero finito (e noto) di biglie. Si faccia attenzione al ruolo della scambiabilità.

---

[4] Nel mondo televisivo dell'America negli anni ′50, il paradosso (con altro nome e varî adattamenti) ebbe un grande successo di ascolti.

3.1.2 Una applicazione riassuntiva

Nell'esempio **2.2.4**, si è considerata un'urna U contenente N biglie, Z delle quali di colore A (i successi) e $N - Z$ le restanti di colore B (gli insuccessi). Il soggetto, in base a sue informazioni ed opinioni, valuta equiprobabili le ipotesi $H_z \in \mathcal{H}$, con $z \in \mathcal{Z} = \{0, 1, \ldots, N\}$; cioè $p_z^0 = \mathbb{P}(H_z) = \dfrac{1}{N+1}$. Sia $\{p_0^0, p_1^0, \ldots, p_N^0\}$ la distribuzione *a priori* delle H_z.

Si mostra in che modo la regola di Bayes aggiorni le nostre conoscenze su U dato il risultato (n, s) "*s successi in n prove*", quando le prove sono bernoulliane, caso (c_b), o esaustive, caso (c_h).

Esempio 3.1.6. La verosimiglianza di H_z è $\mathbb{P}_b\big(e \mid H_z\big) = Bin\big(s|\frac{z}{N}, n\big)$, nel caso (c_b); mentre, nel caso (c_h), è $\mathbb{P}_h\big(e \mid H_z\big) = Hyp\big(s|z, N-z, n\big)$; *ergo*

$$
\begin{aligned}
(c_b) \quad \mathbb{P}_b\big(e \mid H_z\big) &= \binom{n}{s} \cdot \left(\frac{z}{N}\right)^s \left(1 - \frac{z}{N}\right)^{n-s}, \quad z \in \mathcal{Z}\,, \\
(c_h) \quad \mathbb{P}_h\big(e \mid H_z\big) &= \frac{\dbinom{z}{s} \cdot \dbinom{N-z}{n-s}}{\dbinom{N}{z}}, \quad z \in \mathcal{Z}_h\,,
\end{aligned}
\tag{3.4}
$$

dove $\mathcal{Z}_h = \{s, s+1, \ldots, N-n+s\}$. Grazie al il teorema di Bayes si calcolano le leggi *a posteriori* delle ipotesi H_z. Nei due casi (c_b) e (c_h) si ha

$$
\begin{aligned}
(c_b) \quad p_z^b = \mathbb{P}_b(H_z \mid e) &\overset{c}{=} \mathbb{P}(H_z) \cdot \mathbb{P}_b(e \mid H_z) \\
&\overset{c}{=} \frac{\left(\frac{z}{N}\right)^s \left(1 - \frac{z}{N}\right)^{n-s}}{\sum_{u=0}^{N} \left(\frac{u}{N}\right)^s \left(1 - \frac{u}{N}\right)^{n-s}}, \quad z \in \mathcal{Z}\,,
\end{aligned}
\tag{3.5}
$$

$$
\begin{aligned}
(c_h) \quad p_z^h = \mathbb{P}_h(H_z \mid e) &\overset{c}{=} \mathbb{P}(H_z) \cdot \mathbb{P}_h(e \mid H_z) \\
&\overset{c}{=} \frac{\dbinom{z}{s} \cdot \dbinom{N-z}{n-s}}{\sum_{j=s}^{N-n+s} \dbinom{j}{s} \cdot \dbinom{N-j}{n-s}}, \quad z \in \mathcal{Z}_h\,.
\end{aligned}
\tag{3.6}
$$

(*Una applicazione*). Se l'urna U è costituita da $N = 40$ biglie, *A priori* si ha $E(Z) = 20$, $sd(Z) = 11.83$ e $\mathbb{P}(X_i = 1) = \frac{1}{2}$. Si supponga ora che un primo esperimento fornisca $e \equiv (4, 1)$ e, proseguendo, si abbia $e \equiv (8, 2)$.

- Se $e \equiv (4,1)$ si si hanno le leggi *a posteriori* $\mathbb{P}_b(H_z|e)$, caso (c_b), e $\mathbb{P}_h(H_z|e)$, caso (c_b). Vedi figure **3.1** $[A]$ e $[C]$ rispettivamente.

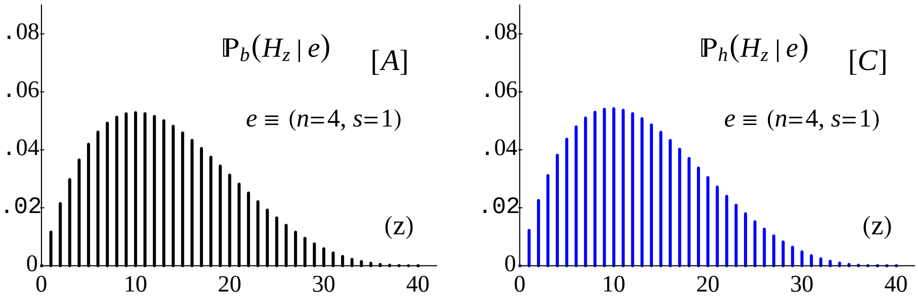

Figura 3.1: - *Probabilità a posteriori delle ipotesi H_z, casi (c_b) e (c_b).*

- Se $e \equiv (8,2)$ le leggi *a posteriori* $\mathbb{P}_b(H_z|e)$, caso (c_b), e $\mathbb{P}_h(H_z|e)$, caso (c_h) sono riportate in figura **3.2** $[B]$ e $[D]$ rispettivamente.

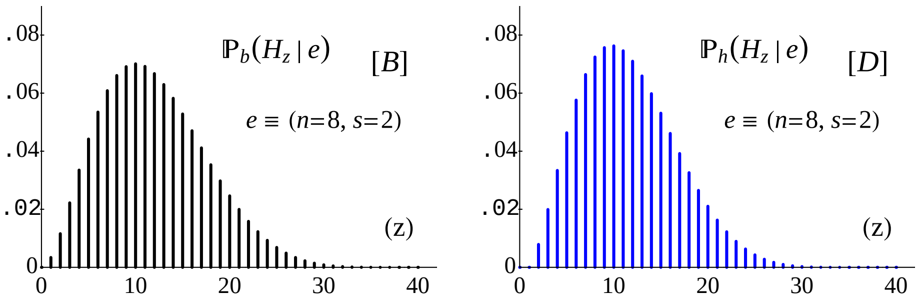

Figura 3.2: - *Probabilità a posteriori delle ipotesi H_z nei casi $[B]$ e $[D]$.*

Come emerge dalle figure **3.1** e **3.2** e come indica la tavola **3.4**, al crescere della taglia n: (i) la media *a posteriori* $E_*(Z|e)$ (il simbolo "$*$" significa "b" o "h" a seconda che le prove siano con o senza restituzione) tende a stabilizzarsi sul valore $\dfrac{s}{n} \times N = 10$, (ii) la *sd a posteriori* $sd_*(Z|e)$ si riduce, (iii) la legge *a posteriori* della v.a. Z (o delle H_z) tende a concentrarsi, come indicano le probabilità $\mathbb{P}_*(Z \in C|e)$, con $C = \{6, 7, \ldots, 19\}$.

La probabilità $\mathbb{P}(X = x)$, data dalla (2.14), è aggiornabile alla luce dell'esperimento e. Con la formula di scomposizione (3.1) si ha infatti

$$\begin{aligned}
\mathbb{P}_*(X_i = x \mid e) &= \sum_{z=0}^{N} \mathbb{P}(X_i = x|H_z) \cdot \mathbb{P}_*(H_z|e) \\
&= \sum_{z=0}^{N} \left(\frac{z}{N}\right)^x \left(1 - \frac{z}{N}\right)^{1-x} \cdot p_z^* \qquad (3.7) \\
&= \left(\frac{1}{N} \cdot E_*(Z|e)\right)^x \left(1 - \frac{1}{N} \cdot E_*(Z|e)\right)^{1-x} .
\end{aligned}$$

La tavola riporta infine la probabilità $\mathbb{P}_*(X_i = 1|e)$ nei casi (c_b) e (c_h).

Al lettore è lasciato l'onere di spiegare le ragioni per cui

▶ legge *a posteriori* a verosimiglianza sono proporzionali se la legge *a priori* delle ipotesi è uniforme;

▶ risulta $\mathbb{P}_b(Z = 0|e) = \mathbb{P}_b(Z = N|e) = 0$ e $\mathbb{P}_h(Z \notin \mathcal{Z}_h|e) = 0$;

▶ al crescere della consistenza dell'urna (cioè al crescere di N) le differenze tra la (3.5) e la (3.6) tendono a ridursi(5);

nonché di calcolare

▶ la probabilità congiunta *a posteriori* $\mathbb{P}_*(X_{n+1} = x', X_{n+2} = x''|\, e)$ e la covarianza $Cov_*(X_{n+1}, X_{n+2}|\, e)$.

	$(c_b)\,[A]$	$(c_b)\,[B]$	$(c_h)\,[C]$	$(c_h)\,[D]$
$e \equiv (n, s)$	*(4,1)*	*(8,2)*	*(4,1)*	*(8,2)*
$E_*(Z|e)$	13.35	12.00	13.00	11.60
$sd_*(Z|e)$	7.12	5.53	6.93	5.07
$\mathbb{P}_*(Z \in C|e)$	0.699	0.831	0.714	0.870
$\mathbb{P}_*(X = 1|e)$	0.334	0.325	0.300	0.290

Tav. **3.4** - *Media e sd per i campioni ed il tipo di prove considerate.*

Meno semplici da spiegare sono i motivi per i quali, al crescere di n,

▶ la legge *a posteriori* tende a concentrarsi su una sola delle ipotesi H_z;

▶ le future osservazioni *id* e *scambiabili* $(X_{n+1}, X_{n+2}, \dots)$ tendono a divenire, al crescere di n, *iid*. ◁

Nota 3.1.5. Da osservare che è proprio grazie alla legge *a posteriori* (3.5), o (3.6), che è possibile calcolare la (3.7) detta probabilità (o legge) predittiva dell'evento $X = x$ data l'osservazione e. Di fatto con la (3.7) il soggetto "tiene conto" delle sue incertezze circa la composizione dell'urna. ◁

(5) Come mostra la Tavola **2.3** le leggi *a posteriori* $\mathbb{P}_b(H_z|e)$ e $\mathbb{P}_h(H_z|e)$ sono assai simili se $e \equiv (4, 1)$. Vi sono delle differenze quando $e \equiv (8, 2)$. *Coeteris paribus*, la $\mathbb{P}_b(H_z|e)$ si presenta un poco piú dispersa della $\mathbb{P}_h(H_z|e)$.

Il caso appena discusso non è che un esempio (elementare) di ragionamento per induzione. Esso solleva questioni, di sicuro rilievo teorico e pratico, che meritano di essere approfonditi.

3.2 Verso il ragionamento induttivo

La nozione di probabilità che abbiamo descritto è senza dubbio la piú vicina a quella dell'"uomo della strada", anzi è proprio quella che viene impiegata nei giudizi pratici di tutti i giorni. Perché la scienza dovrebbe ripudiarla?
(Bruno de Finetti, 1931)

Come è noto, prende il nome di induzione il procedimento che a partire dalla acquisizione di evidenze empiriche particolari (fatti, situazioni, osservazioni, cose) conduce alla conoscenza di una *verità* generale.

Teorie e leggi che si incontrano nelle scienze sperimentali sono nate, senza eccezioni, dal ragionamento per induzione. Il quale non può fondarsi sulla logica del certo, la logica classica, bensí sulla logica del probabile. La teoria della probabilità soggettiva.

Il processo induttivo che consente, a partire da fatti osservati, di aggiornare ed accrescere le nostre conoscenze su certe caratteristiche di un dato universo richiede l'esame degli *oggetti* coinvolti nel teorema di Bayes. L'esame cioè

(1) del modello campionario inteso come "macchina" capace di generare dati: passate osservazioni, le x_{obs}, e nuove osservazioni[6];

(2) della legge *a priori* delle ipotesi, le probabilità $\mathbb{P}(H_j)$, e/o del parametro di interesse θ del modello[7], la distribuzione $G_0(\theta)$, che riassume le informazioni presperimentali del soggetto;

(3) della verosimiglianza delle ipotesi H_j (o del parametro) dato x_{obs}[8].

Su tali oggetti, si tornerà ampiamente nei prossimi capitoli.

Nella letteratura statistica soggettivista è prevalsa l'idea di organizzare il

---

[6] In letteratura vi è l'abitudine di classificare i modelli in parametrici, semiparametrici e non parametrici. Al primo gruppo appartengono i modelli definiti (o indicizzati) da un certo parametro (o carattere o ipotesi) $\theta \in \Theta$. Il secondo gruppo prevede (ad esempio) i modelli *kernel*, *etc.* Nel terzo gruppo i modelli definiti in base a certe proprietà: ad esempio i modelli unimodali, i modelli a supporto positivo, *etc.*

[7] Si osservi che nell'esempio **3.1.6** il parametro del modello è il numero $Z \in \mathcal{Z}$.

[8] Per funzione di verosimiglianza del parametro θ, dato x_{obs}, si intende la probabilità di osservare x_{obs} al variare di θ.

ragionamento induttivo, e le conseguenti procedure del paradigma bayesiano, secondo le due differenti impostazioni: ipotetica e predittiva.

(i) L'impostazione ipotetica che ha l'obiettivo di determinare, mediante il teorema di Bayes, la legge *a posteriori* del parametro $\theta \in \Theta$, e cioè dell'*ipotesi*, una volta osservato x_{obs}; tale impostazione considera particolari aspetti di θ che dànno luogo ai (classici) problemi della

▶ stima puntuale la quale prevede la ricerca di una stima $\hat{\theta} \in \Theta$, del parametro θ, una volta che si è specificato un opportuno criterio;

▶ stima per intervalli che si occupa della costruzione di un insieme $\Theta_0 \subset \Theta$ che, con "alta" probabilità contiene il "vero" valore del parametro θ;

▶ scelta tra ipotesi (o *test* di ipotesi) che prevede il confronto e la scelta fra ipotesi, ad es. $\big\{H_0 \,:\, \theta \in \Theta_0 \; vs \; H_1 \,:\, \theta \in \Theta_1\big\}$.

(ii) L'impostazione predittiva (o previsiva) la quale, in base ad osservazioni *passate* x_{obs} da una certa popolazione, intende stabilire la legge del *futuro* risultato $Y = y$ ottenibile da esperimenti sulla *stessa* popolazione.

La distinzione fra impostazione ipotetica ed impostazione predittiva, tra loro *praticamente* equivalenti, e la stessa parola "*impostazione*", saranno via via precisate nel corso della dispensa.

Negli esempi finora presentati e discussi, sono state considerate ipotesi decidibili e come tali "pacificamente" probabilizzabili.

La sezione che segue cercherà di trattare le ipotesi non decidibili, le quali comportano delle difficoltà.

3.3 Indipendenza e scambiabilità

Si indichi con $\big(X_n\big)_{n\geq 1} = \big(X_1, X_2, \ldots, X_n, \ldots\big)$ una successione di n.a. provenienti da uno stesso fenomeno e costituenti un processo, che è noto allora che è nota $\forall n \geq 1$ la successione delle f.r.

$$Q_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;=\; \mathbb{P}\big(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\big)\,,$$

detta legge del processo.

Se il soggetto ipotizza l'indipendenza del processo dicendo che esso è costituito da n.a. indipendenti, nessuna inferenza sarà mai possibile.

Postulare l'indipendenza stocastica dei n.a. impedisce (*sensu strictu*) ogni aggiornamento delle probabilità, quale che siano le ulteriori informazioni che si rendano disponibili. E, sopra tutto, non vi è possibilità alcuna di far uso degli eventi osservati per valutare probabilità di eventi futuri.

È solo assumendo l'ipotesi (e lo schema) di scambiabilità che è possibile utilizzare le *osservazioni passate* per inferire su eventi futuri.

Si torni al processo bernoulliano, vedi §**2.2.1**, che produce osservazioni *iid*. Ovvero in cui le X_i sono stocasticamente indipendenti dato il parametro ϕ e sono *identicamente distribuite*, o *somiglianti*, in sigla *id.* Hanno cioè stessa f.r. cioè, $\forall i$, $X_i \sim F(\cdot|\phi)$, $\phi \in \Phi$.

Se il processo è bernoulliano la *f.r.* congiunta è

$$F_{1,2,\ldots,n}\big(x_1, x_2, \ldots, x_n \mid \phi\big) \;=\; \prod_{i=1}^{n} F(x_i|\phi)\,. \tag{3.8}$$

È immediato osservare che ogni mistura di processi bernoulliani di peso $H(\phi)$ qualsiasi è un processo scambiabile. E dunque

$$\begin{aligned} Q_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \;&=\; \int_\Phi F_{1,2,\ldots,n}\big(x_1, x_2, \ldots, x_n|\phi\big) \cdot \mathrm{d}H(\phi) \\ &=\; \int_\Theta \prod_{i=1}^{n} F(x_i|\phi) \cdot \mathrm{d}H(\phi)\,, \end{aligned} \tag{3.9}$$

da cui segue che, solo nel caso in cui la $H(\phi)$ risulti degenere, ad es. nel punto ϕ_0, la *f.r.* $Q_{1,2,\ldots,n}$, $\forall n$, si ridurrebbe al prodotto delle *f.r.* marginali.

Si pone il quesito: ogni processo scambiabile ed illimitato di alternativa può sempre essere rappresentato come mistura, di peso opportuno, di un processo bernoulliano?

La risposta è data dal fondamentale *teorema di caratterizzazione* di de Finetti, che fissa il collegamento fra i due processi.

3.3.1 Il teorema di rappresentazione

Si consideri il caso speciale di processo bernoulliano di alternativa, detto anche $(0-1)$, o dicotomico, o di indicatori di evento, *etc.*, nel quale $X_i \in \{0, 1\}$, $\mathbb{P}(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, con $\theta \in [0, 1]$. La (3.8) diviene

$$\mathbb{P}\big(X_1 = x_1, \ldots, X_n = x_n \mid \theta\big) \;=\; \prod_{i=1}^{n} \mathbb{P}(X_i = x_i|\theta) \;=\; \theta^s(1-\theta)^{n-s}\,, \tag{3.10}$$

dove $s = \sum_{i=1}^{n} x_i$. Segue dunque il teorema di rappresentazione.

***Teorema* 3.3.1.** (*B. de Finetti*, 1937)
Sia $\big(X_n\big)_{n\geq 1}$ *una successione indefinita e scambiabile di n.a.* $(0-1)$*, sia* $\big(P_n\big)_{n\geq 1}$*, con* $P_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$*, la corrispondente successione delle proporzioni di successo. Allora*
(i) *la legge della* $n-$*pla* $(X_1, X_2, \ldots, X_n)$ *è,* $\forall n$*,* rappresentabile *nella forma*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \;=\; \int_0^1 \theta^s (1-\theta)^{n-s} \cdot \mathrm{d}H(\theta)\;, \qquad (3.11)$$

(ii) *la funzione peso* $H(\cdot)$ *è la f.r. limite*

$$H(\theta) \;=\; \lim_{n\to\infty} \mathbb{P}\big(P_n < \theta\big)\,. \lhd \qquad (3.12)$$

La dimostrazione del teorema è laboriosa. Per tale ragione è omessa.

Con le (3.11) e (3.12) il teorema di rappresentazione afferma che

- tutti e soli i processi di alternativa scambiabili ed illimitati sono misture di processi bernoulliani dicotomici;
- assegnati il processo bernoulliano (3.10) e la funzione peso $H(\theta)$, la legge del processo scambiabile è determinato; *viceversa*, nota che sia la legge del processo scambiabile, si può sempre risalire al peso $H(\theta)$.

Pertanto se il processo scambiabile è noto, allora è definito il *parametro aleatorio* θ assieme alla sua legge (3.12).

L'esempio che segue mostra, noto il processo, come risalire ad $H(\theta)$.

Esempio 3.3.1. È nota la legge del processo scambiabile $\big(X_i\big)_{i\geq 1}$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \;=\; \frac{\big(\sum_{i=1}^{n} x_i\big)! \cdot \big(n - \sum_{i=1}^{n} x_i\big)!}{(n+1)!}\;,\; X_i \in \{0,1\}\,.$$

Determinare la funzione peso $H(\theta)$ tale che, $\forall n \in \mathbb{N}$, risulti

$$\frac{\big(\sum_{i=1}^{n} x_i\big)! \cdot \big(n - \sum_{i=1}^{n} x_i\big)!}{(n+1)!} \;=\; \int_0^1 \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i} \cdot \mathrm{d}H(\theta)\,.$$

La legge assegnata è invariante per permutazioni delle X_i, fatto che ne certifica la scambiabilità e che, in forza del teorema **3.3.1**, implica l'esistenza di una funzione peso $H(\theta)$. La quale, nel caso in esempio, è calcolabile sfruttando note proprietà delle funzioni caratteristiche (*f.c.*).

Poiché la v.a. $S_n = \displaystyle\sum_{i=1}^{n} X_i$ segue la legge beta-binomiale

$$\mathbb{P}(S_n = s) \;=\; \binom{n}{s} \cdot \frac{(s)! \cdot (n-s)!}{(n+1)!}\,,$$

ne consegue, vedi appendice, che le *f.c.* delle v.a. S_n e $P_n = \frac{1}{n} S_n$ sono

$$\begin{aligned}
H_{S_n}(t) &= \sum_{s=0}^{n} e^{its} \cdot \binom{n}{s} \cdot \frac{(s)! \cdot (n-s)!}{(n+1)!} \\
&= \frac{1}{n+1} \sum_{s=0}^{n} e^{its} \;=\; \frac{1}{n+1} \cdot \frac{\exp\{it(n+1)\}}{1 - e^{it}}\,, \\
H_{P_n}(t) &= H_{S_n}\!\left(\tfrac{1}{n} t\right) \\
&= \frac{1}{n+1} \cdot \frac{\exp\left\{it \frac{n+1}{n}\right\}}{1 - \exp\left\{it \frac{1}{n}\right\}} \;\; \underset{n \to \infty}{\longrightarrow} \;\; \frac{e^{it} - 1}{t}\,,
\end{aligned}$$

la quale risulta essere la *f.c.* della legge uniforme. La funzione peso (3.12) assume dunque la forma $P_n \xrightarrow{\mathcal{L}} \theta \sim unif(\cdot|0,1)$. $\lhd$

In presenza di un processo di n.a. $(0-1)$ relativi ad uno stesso fenomeno, il soggetto che conosce la probabilità di successo θ, afferma che (per lui) le osservazioni sono *iid*. Per contro, il soggetto che (di fronte allo stesso fenomeno) ha incertezze circa il valore di θ può solo dire che le (stesse) osservazioni sono scambiabili. Restando con ciò confermato, ed ancora una volta, il carattere relativo della dipendenza, indipendenza, scambiabilità stocastica.

Il teorema **3.3.1**, nell'esprimere il collegamento concettuale fra le ipotesi di bernoullianità e di scambiabilità, mette in relazione, *in certo senso*, le opinioni dei due soggetti[(9)].

3.3.2 Urne, monete ed ipotesi

Iniziamo a descrivere due semplici (ed usuali) esperimenti *isoformi* nei quali è ragionevole assumere la condizione di scambiabilità delle osservazioni.

Caso [1]. Da un'urna contenente biglie azzurre (A) e bianche (B) secondo la composizione $\theta \in [0,1]$ ignota, sono estratte n biglie con restituzione; s sono risultate essere di color A ed $n-s$ di color B. Dato il risultato, qual'è la probabilità che in una ulteriore prova sia estratta una biglia color A?

Caso [2]. Si lanci per n volte una moneta che forse è regolare. Sia $\theta \in [0,1]$ la pobabilità non nota che appaia testa (T). Siano s ed $n-s$ le volte in cui

[9)] Per dirla (molto) alla buona: le X_i dentro l'integrale (3.11) sono *iid*, mentre le stesse X_i fuori l'integrale sono scambiabili.

è uscita T e C. Con tale risultato, qual'è la probabilità che in un ulteriore colpo si abbia T?

Tenuto conto che i due descritti processi $(0-1)$, subordinatamente a θ, sono bernoulliani, il lettore dica se sono *isoformi* pure i ragionamenti necessari per arrivare alla risposta.

Naturalmente nulla impedisce a chiunque di formulare ipotesi circa la composizione dell'urna e circa la probabilità dell'evento T. Ad esempio si possono considerare, in entrambi i *casi* [1] e [2], le ipotesi $\big(\Theta_1, \Theta_2, \ldots, \Theta_k\big)$ costituenti una partizione dell'intervallo $[0,1]$. Meno scontato è stabilire se sia lecito, in entrambi i *casi* [1] e [2], probabilizzare tali ipotesi.

Nel *caso* [1] la composizione dell'urna è un evento incerto e decidibile e, come tale, probabilizzabile potendosi sempre, almeno in linea di principio, contare le biglie contenute nell'urna.

Nel *caso* [2], al contrario, dall'esame accurato delle "imperfezioni" della moneta, non v'è alcuna possibilità di introdurre ipotesi aventi significato fisico e dunque di stabilire oggettivamente quale sia l'*ipotesi vera*.

La giustificazione (forse) piú persuasiva per probabilizzare tali ipotesi e per "dare senso" a tale operazione, è che, replicando un *gran numero di volte* i lanci della moneta, si arriva all'*ipotesi vera*.

Giustificazione che tuttavia manca di concretezza, non essendo data, se non in astratto, la possibilità di fare *illimitati lanci* della moneta.

Se dunque si vuole inferire sull'evento $X_{n+1} = x_{n+1}$, avendo osservato $E_n = \big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\big)$, non si può calcolare la probabilità condizionata $\mathbb{P}\big(X_{n+1} = x_{n+1} \mid E_n\big)$ ricorrendo alla formula (3.7), valida solo nel caso in cui le ipotesi sono decidibili.

Posto che $\mathbb{P}\big(E_n\big) > 0$, $\forall n$, è possibile aggirare la difficoltà esprimendo tale probabilità condizionata come rapporto fra probabilità congiunte

$$\mathbb{P}\big(X_{n+1} = x_{n+1} \mid E_n\big) \;=\; \frac{\mathbb{P}\big(E_n, X_{n+1} = x_{n+1}\big)}{\mathbb{P}\big(E_n\big)}\,, \tag{3.13}$$

calcolabili se il soggetto già conosce oppure è capace (in qualche modo) di assegnare la legge del processo $(0-1)$ illimitato $\big(X_i\big)_{i\geq 1}$. Si noti che la (3.13) è valida in entrambi i *casi* [1] e [2] e che se il processo è scambiabile allora siamo nelle condizioni richieste dal teorema **3.3.1**. *Ergo*: solo grazie alla scambiabilità urne e monete sono "*trattabili allo stesso modo*".

Se dunque, subordinatamente a θ si ha $\mathbb{P}(X_i = x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $\forall i$, e $\mathbb{P}(E_n|\theta) = \theta^s(1-\theta)^{n-s}$, dove $s = \sum_{i=1}^n x_i$, e se il processo è scambiabile, siamo autorizzati a dire che esiste una funzione peso $H(\theta)$, espressa dalla

(3.12), tale che $\forall n$ risulta $\mathbb{P}(E_n) = \int_0^1 \theta^s(1-\theta)^{n-s} \cdot \mathrm{d}H(\theta)$.

Grazie al teorema di rappresentazione, l'impostazione predittiva risulta equivalente, sia pure sotto opportune condizioni, allo schema ordinario basato sul concetto di esperimento statistico, trattato nel prossimo capitolo.

Da dire che i casi [1] e [2], urne e monete, nella loro semplicità, non sempre segnalano le difficoltà che si incontrano nella pratica.

Già il passaggio dall'urna descritta nel *caso* [1], ad urne (fisicamente) piú complicate comporta qualche delicato problema di metodo, piú o meno superabile. Si pensi alle *urne* contenenti N *oggetti*, Z dei quali dotati di una certa caratteristica, con N e Z *praticamente* indeterminabili.

Casi ben noti, le specie di bacteri che popolano uno stagno; le molecole costituenti una miscela gassosa . . . In presenza di popolazioni *non identificate* è (sempre) lecito parlare di ipotesi decidibili?

Quanto al *caso* [2], sono decidibili le ipotesi: "per Tizio affetto da quella certa patologia, il farmaco F è efficace all'85 (%)", "la difettosità di una certa macchina operatrice è $\theta = 0.06$"? È lecito confrontare le difettosità di due macchine operatrici?

Nota 3.3.1. La portata gnoseologica del teorema **3.3.1** è notevole. Già dagli anni '50, il teorema (che ha attirato l'attenzione di probabilisti ed epistemoligi) si è arricchito di contributi e sviluppi che hanno consentito di passare dai casi elementari di *urne e monete* a casi piú generali. $\triangleleft$

3.3.3 Scambiabilità e schemi di urne (*)

Per motivi di semplicità, si assume che le urne siano costituite da biglie di due soli colori, azzurro A e bianco B.

Il lettore proverà facilmente che le urne ipergeometiche (n prove senza restituzione da un'urna avente N_A biglie A ed N_B biglie B) è scambiabile.

Si mostra ora che i processi da urne di Polya, che rispondono cioè allo schema di Polya o di contagio, sono ancora scambiabili.[10]

L'esempio che segue considera un'urna di Polya con $r = +1$.

Esempio 3.3.2. (*Urne di Polya e scambiabilità*)
Se $r = +1$ il verificarsi di A (e simmetricamente di B) rende piú probabile il suo ripetersi. Pertanto, posto $N = N_A + N_B$, la probabilità di estrarre prima k biglie color A e poi $n - k$ biglie color B risulta

---

[10] Giova ricordare che lo schema di Polya prevede che, eseguita l'estrazione e la restituzione della biglia, si mettono in essa r biglie dello stesso colore della biglia sorteggiata. Per $r = -1$ si ha l'urna ipergeometrica, per $r = 0$ l'urna bernoulliana.

$$\frac{N_A}{N} \cdot \frac{N_A+1}{N+1} \cdots \frac{N_A+k-1}{N+k-1} \cdot \frac{N_B}{N+k} \cdot \frac{N_B+1}{N+k+1} \cdots \frac{N_B+n-k-1}{N+n-1} \ .$$

Poiché la probabilità di estrarre k biglie color A e $n-k$ biglie color B non dipende né dall'ordine dei successi e degli insuccessi(11), né dal valore di r, segue che il processo da urne di Polya è scambiabile. $\lhd$

L'esempio che segue mostra due processi non scambiabili. In entrambi i casi le probabilità di osservare k successi e $n-k$ insuccessi dipendono oltre che da (k,n) anche dall'ordine con cui gli eventi si succedono.

Esempio 3.3.3. (*Urne non scambiabili*)
(i) È data una successione di urne di composizione $\{\theta_1, \theta_2, \ldots, \theta_n, \ldots\}$ da ciascuna delle quali è estratta una biglia (*schema di Poisson*). Per provare che il processo non è scambiabile, è sufficiente il confronto tra la probabilità che le prime k biglie siano A e le altre $n-k$ siano B *vs* la probabilità che le prime $n-k$ siano B e le altre k siano A

$$\prod_{i=1}^{k} \theta_i \times \prod_{i=k+1}^{n} (1-\theta_i) \neq \prod_{i=1}^{n-k} (1-\theta_i) \times \prod_{i=n-k+1}^{n} \theta_i \ .$$

(ii) È data un'urna tale che se la biglia sorteggiata è A la si restituisce, mentre se è di color B se ne rendono due. Il confronto fra la probabilità di estrarre prima k biglie A e poi $n-k$ biglie B *vs* la probabilità di estrarre prima $n-k$ biglie B e poi k biglie A porge la disuguaglianza

$$\left(\frac{N_A}{N}\right)^k \cdot \frac{N_B}{N} \cdots \frac{N_B+n-k-1}{N+n-k-1} > \frac{N_B}{N} \cdots \frac{N_B+n-k-1}{N+n-k-1} \left(\frac{N_A}{N+n-k-1}\right)^k \ .$$

$\lhd$

11) Scambiando l'ordine di successi ed insuccessi la successione dei denominatori resta identica, mentre i numeratori, *sic et simpliciter*, cambiano ordine.