CAPITOLO III: - PROBABILITA’

3.1 - Definizione della probabilità.

Riprendiamo in esame un esperimento casuale S. Come il Lettore certamente ricorderà, i suoi possibili risultati, ω, formano lo spazio dei campioni Ω; in esso abbiamo definito *evento* un qualsiasi sottoinsieme di Ω, proprio o improprio che sia. Tra tutti gli eventi, abbiamo definito come evento impossibile, $\emptyset$, il sottoinsieme vuoto di Ω, e, come evento certo, Ω stesso. Evidentemente, il primo non potrà mai verificarsi, il secondo si verificherà sempre, ossia ad ogni ripetizione dell’esperimento casuale S.

A parte questi, tutti gli altri eventi possono o meno realizzarsi, e sono detti *possibili*. E’ a questi che vogliamo dare una misura del grado di affidabilità concesso al loro verificarsi, vogliamo cioè esprimere la *fiducia* che abbiamo nel realizzarsi dell’evento stesso. Questa fiducia viene detta *probabilità* che l’evento si verifichi.

La probabilità sarà un numero, che andrà da un valore minimo, ragionevolmente assegnato all'evento impossibile, altrettanto ragionevolmente posto uguale a zero, ad un valore massimo dato all'evento certo, valore che riterremo finito e che, in base a considerazioni che saranno evidenti nel seguito più immediato, sarà assunto uguale ad uno. Dare una probabilità ad un dato evento dello spazio dei campioni dell’esperimento casuale S consiste nell'assegnare all'evento stesso un numero compreso tra zero ed uno, naturalmente rispettando regole ben precise che vedremo quasi subito.

Per semplificare le cose, restando nell’ambito di spazi dei campioni finiti, diremo che dare una *distribuzione di probabilità* consiste nell'assegnare una probabilità ad ogni *evento elementare*, cioè ad ogni sottoinsieme formato da un solo campione. Come vedremo tra breve, l'assegnazione di una probabilità a tutti gli eventi elementari consentirà di dare una probabilità ad ogni[1] evento nello spazio dei campioni. In assenza di ragioni contrarie, a tutti gli eventi elementari verrà data la medesima probabilità (si parla allora di una *distribuzione equiprobabile*); la cosa però non è di carattere generale come vediamo dal seguente esempio. Se lanciamo una normale moneta due volte, indicando con 0 l’uscita della croce e con 1 l’uscita della testa, lo spazio dei campioni sarà formato da quattro elementi, le coppie $\{(0,0),(0,1),(1,0),(1,1)\}$. Al contrario, nel caso di due monete indistinguibili lanciate assieme, lo spazio dei campioni si riduce, contando le teste, a $\{0,1,2\}$, e dunque a tre soli elementi. Nel primo caso sarà sensato dare

[1] In realtà la probabilità non è sempre definita su tutti i sottoinsiemi dello spazio dei campioni; noi però non insisteremo su questo aspetto e penseremo che ad ogni sottoinsieme $\mathbf{E}$ di $\mathbf{\Omega}$ sia possibile assegnare una probabilità.

ai quattro elementi dello spazio dei campioni la medesima probabilità, che sarà quindi pari ad un quarto; nel secondo caso sarebbe errato dare ad ognuno dei tre elementi la probabilità di un terzo, in quanto dovrebbe essere evidente che il risultato 1 comparirà con maggiore frequenza degli altri due avendo a disposizione per realizzarsi due situazioni elementari, sia pure per noi indistinguibili.

Prima di proseguire facciamo subito notare che quanto detto ora, relativo alla possibilità di dare una probabilità ad ogni singolo elemento dello spazio dei campioni, vale solo nel caso di spazi discreti; nel caso di spazi continui infatti non avrebbe senso cercare di dare una probabilità ad un singolo risultato, ma dovremo invece dare probabilità ad intervalli di risultati. Non avrebbe senso, nell'esempio citato in altre circostanze, relativo alla misura dell'altezza degli individui di una popolazione, cercare la probabilità di una statura esatta, quale potrebbe essere 172,235471 cm; avrà invece senso cercare di dare una probabilità all'intervallo di stature tra 172 e 173 cm.

L'assegnazione della probabilità ai singoli eventi, per noi elementari, deve essere *oggettiva*, e dunque legata all'evento, e non *soggettiva*, e cioè alle convinzioni, se non agli sbalzi di umore, di chi opera l'assegnazione. In altre parole, l'assegnazione di probabilità al medesimo evento fatta da operatori diversi dovrà risultare la medesima. Per esempio, tutti coloro che vogliono scommettere con il lancio di una moneta daranno probabilità pari ad $1/2$ a entrambe le facce. Non parleremo della probabilità del Milan di vincere lo scudetto, dal momento che ogni tifoso valuterà in modo diverso questa eventualità, seguendo più la sua passione sportiva che non la ragione.

Nell'assegnazione di probabilità dovrà sempre essere rispettato il cosiddetto *principio di coerenza*. Esso richiede che:

I. la probabilità dell'evento impossibile sia la minima, e precisamente sia nulla;

II. la probabilità dell'evento certo sia la massima, con tale massimo posto uguale ad 1;

III. se due eventi E_1 ed E_2 sono equivalenti (nel senso che il verificarsi dell'uno induce il verificarsi dell'altro e viceversa) i due eventi dovranno avere la stessa probabilità;

IV. se p è la probabilità assegnata ad un evento E, e q è quella assegnata alla sua negazione E', la somma delle due probabilità deve essere quella dell'evento certo, che è la massima, e cioè $P(E)+P(E')=p+q=1$, dal momento che il verificarsi dell'uno o dell'altro rappresenta la certezza.

Si vede la ragionevolezza di quanto richiesto, a parte l'assegnazione del valore uno al massimo di probabilità, forse non ancora giustificata. In particolare, la quarta condizione dice solamente che se riuscissimo a trovare chi scommette con noi sul verificarsi di un evento, e ci accordasse la vittoria sia quando l'evento si verifica, sia quando l'evento non si verifica, costui o ci vorrebbe bene o non potrebbe mai fare fortuna.

Ci sono due approcci classici alla definizione di probabilità, detti rispettivamente *a priori* ed *a posteriori*, che ora descriveremo. Per dire il vero, il primo approccio viene da qualche tempo fortemente criticato, lasciando così spazio solamente al secondo. Noi comunque preferiamo restare sulla vecchia strada, certi che le obiezioni che illustreremo al momento opportuno si possano controbattere, per lo meno nei casi di nostro interesse.

L'assegnazione *a priori* della probabilità presuppone la possibilità della determinazione sia del numero totale dei risultati possibili, cioè dei campioni (sono i *casi possibili*), sia del numero di questi per i quali l'evento al quale si vuole assegnare la probabilità risulta verificato (sono i *casi favorevoli*). In tali condizioni la probabilità sarà definita come il rapporto tra i casi favorevoli ed i casi possibili. Quindi ad esempio nel caso del lancio di un dado la probabilità dell'uscita di un numero pari sarà di un mezzo (tre casi favorevoli su sei possibili), quella dell'uscita di un numero compreso tra uno e sei sarà uno!

Nel caso in cui la stima del numero totale di campioni fosse difficile per non dire impossibile, così come quella del numero di campioni in corrispondenza ai quali l'evento E al quale si vuole assegnare la probabilità risulta verificato, non si potrebbe dare la definizione *a priori* della probabilità, e si dovrebbe quindi ricorrere alla definizione *a posteriori*. Per arrivare a questa sarà necessaria la ripetizione dell'esperimento S un *numero elevato di volte*, contando sia le ripetizioni effettuate che il numero di volte nelle quali l'evento E risulta verificato. La probabilità assegnata ad E *a posteriori* sarà allora il rapporto tra il numero di casi in cui l'evento si è realizzato (è questa la *frequenza* di E, tanto che l'approccio in questione viene anche detto *frequentistico*) ed il numero totale di prove (questo rapporto è la *frequenza relativa* di E). Chiariremo nel seguito che cosa si intenda per *numero elevato di ripetizioni*.

Come si vede, in entrambi i casi risulta giustificata la scelta di assumere il valore 1 come probabilità dell'evento certo, in quanto non sarà mai pensabile di avere un numero di casi favorevoli maggiore di quello dei casi possibili nella determinazione *a priori*, né, in quella *a posteriori*, di avere un numero di successi superiore al numero dei tentativi. Inoltre, è evidente che il principio di coerenza risulta verificato in entrambi i casi. Infatti abbiamo appena verificato che il valore massimo della probabilità deve essere 1, valore assegnato all'evento certo, e solo all'evento certo, ed è anche evidente che il valore minimo è zero, assegnato all'evento impossibile, e solo all'evento impossibile (non ci può infatti essere un numero di casi favorevoli o un numero di successi minore di 0); se poi due eventi sono equivalenti, il rispettivo numero di casi possibili o di successi deve essere uguale (altrimenti il primo potrebbe realizzarsi senza che si realizzi il secondo, e viceversa); infine, se f rappresenta il numero di casi favorevoli o di successi ottenuti su n casi possibili o tentativi fatti, ne risulta che $s = n - f$ è il numero dei casi sfavorevoli o dei tentativi falliti, di modo che la somma della probabilità di successo, $p = f/n$, con quella di insuccesso, $q = s/n$, deve dare 1.

L'obiezione che viene fatta alla definizione di probabilità come rapporto tra casi favorevoli e casi possibili è che, così facendo, affermiamo che non c'è motivo di credere che un campione debba realizzarsi con frequenza maggiore di quella degli altri: affermiamo cioè che tutti i campioni hanno la stessa probabilità, il che porta a definire la probabilità di un evento per mezzo della probabilità stessa. Il nostro sommesso parere è che questa obiezione sia, più o meno facilmente, aggirabile; ad ogni modo, Chi non ne fosse convinto potrà attenersi strettamente alla definizione frequentistica di probabilità.

Le due definizioni sono raccordate dal cosiddetto *principio o legge dei grandi numeri*. Tale principio afferma che:

- al crescere del numero di ripetizioni dell'esperimento **S** la frequenza dei vari eventi si stabilizza, cioè il rapporto tra successi e prove rimane sostanzialmente invariato (questa affermazione, se accettata, risponde anche al problema di che cosa significhi *grande* per il numero di ripetizioni delle prove: il numero di prove sarà considerato *grande* quando sia stata raggiunta tale stabilizzazione, sempre raggiungibile proprio in base al presente principio);
- a stabilizzazione ottenuta la frequenza del singolo evento deve coincidere con la probabilità calcolata *a priori* (ove naturalmente questo calcolo fosse possibile).

In altre parole il *principio (o legge) dei grandi numeri* dice che, per esempio, qualora un tradizionale dado a sei facce venisse lanciato un numero sufficientemente elevato di volte, ognuna delle sei facce comparirebbe in un sesto dei casi. Eventuali sistematici scostamenti da tale risultato sarebbero da attribuirsi a difetti strutturali del dado, quali una densità non uniforme nel suo volume, la presenza di spigoli non arrotondati allo stesso modo, ecc.

Entrambi gli approcci alla definizione di probabilità sono compatibili con un terzo approccio, che li riassume sotto forma di assiomi. La *definizione assiomatica* della probabilità consiste nell'affermare la validità dei tre assiomi seguenti:

1° la probabilità di un qualsiasi evento E deve essere un numero non negativo

$$P(E) \geq 0$$

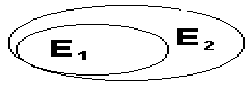
dove

$$P(E) = 0 \Leftrightarrow E = \emptyset;$$

2° $P(\Omega) = 1$;

3° $P(E_1 \vee E_2) = P(E_1) + P(E_2)$ se $E_1 \wedge E_2 = \emptyset$.

Questi assiomi sono chiaramente compatibili sia con i due approcci precedenti che con il *principio di coerenza*. Al proposito si noti che se nell'assioma 3° si pone $E_1 = E$ ed $E_2 = E'$, si ricava $P(E \vee E') = P(\Omega) = P(E) + P(E') = p + q = 1$, poiché $E \wedge E' = \emptyset$ e $E \vee E' = \Omega$; cosa che rende soddisfatta la richiesta IV del *principio di coerenza*. Anche la richiesta I di tale principio risulta immediatamente soddisfatta dal 1° assioma. La richiesta II viene solo parzialmente soddisfatta dal 2° assioma, in quanto questo garantisce che la probabilità dell'evento certo sia 1, ma non garantisce direttamente che tale valore sia il massimo. Per la verifica di ciò e per quella relativa alla richiesta III del principio di coerenza, che esige che eventi equivalenti abbiano la stessa probabilità, dobbiamo premettere la dimostrazione del fatto che, se un evento non impossibile, $E_1 \neq \emptyset$, implica l'evento E_2, la probabilità di E_2 è maggiore, o più correttamente, non è minore, della probabilità di E_1. Infatti, se il verificarsi di E_1 implica quello di E_2, nel diagramma di Venn avremo la situazione di figura. Scriviamo E_2 come $E_2 = E_1 \vee (E_2 \setminus E_1)$, cioè come somma logica di due eventi, il primo dei due e la differenza, con questo, del secondo, disgiunti dal momento che $E_1 \wedge (E_2 \setminus E_1) = E_1 \wedge (E_2 \wedge E_1') = E_2 \wedge (E_1 \wedge E_1') = \; = E_2 \wedge \emptyset$. Ne segue direttamente dal terzo assioma che



$$P(E_2) = P(E_1) + P(E_2 \setminus E_1),$$

da cui

$$P(E_2) - P(E_1) = P(E_2 \setminus E_1) \geq 0,$$

dove la diseguaglianza finale discende dal fatto che una probabilità non può essere negativa (1° assioma); l'eguaglianza vale solo nel caso in cui $P(E_2 \setminus E_1) = 0$, ossia**,** sempre in base al dettato del 1° assioma, solo nel caso in cui la differenza degli eventi è l'evento impossibile $\emptyset$; in questo caso non esistono campioni di E_2 che non siano al tempo stesso campioni di E_1; dal momento che per ipotesi tutti i campioni di E_1 sono anche campioni di E_2, le due condizioni $E_1 \square E_2$ ed $E_2 \setminus E_1 = \emptyset$ portano alla conclusione che $E_1 = E_2$.

Dunque, se E_1 implica E_2, la probabilità di E_2 non è minore della probabilità di E_1. Viceversa, se E_2 implica E_1, la probabilità di E_1 non è minore della probabilità di E_2. L'equivalenza degli eventi E_1 ed E_2 comporta l'implicazione nei due sensi, e dunque la validità di entrambe le diseguaglianze, in base alle quali concludiamo che la probabilità di E_1 e quella di E_2 devono dunque essere uguali, il che soddisfa la III richiesta del principio di coerenza.

Per quanto infine concerne l'impossibilità per un qualsiasi evento E di avere probabilità maggiore di 1 ci soccorre ancora il 3° assioma, applicato nuovamente ad un generico evento E ed alla sua negazione E':

$$P(E \vee E') = P(\Omega) = P(E) + P(E') \quad \Rightarrow \quad 1 = p + q \,.$$

Se per ipotesi assurda la probabilità di E fosse maggiore di 1, dalla uguaglianza precedente si ricaverebbe una probabilità negativa per E', in contrasto con il dettato del 1° assioma.

Il terzo assioma vale evidentemente anche per la somma di più eventi, sempre che tutti questi risultino a due a due incompatibili.

Allo scopo di fornire alcuni semplici esempi di tutto quanto detto si consideri sempre come esperimento S il lancio di un dado a sei facce che definisce quindi come spazio dei campioni Ω l'insieme formato dai primi sei numeri interi

$$\Omega = \{1, 2,.3, 4, 5, 6\} \;.$$

In Ω sia stata introdotta una distribuzione di probabilità uniforme sugli eventi elementari, che avranno allora tutti probabilità pari ad 1/6.

Es. 1 Come evento E si consideri l'uscita di un numero minore di quattro, E = {1, 2, 3}; evidentemente lo stesso evento E si può considerare come somma dei tre eventi elementari

$$E = \{1\} \vee \{2\} \vee \{3\},$$

tutti tra loro incompatibili, e dunque in base all'assioma 3° la probabilità dell'evento E è la somma delle probabilità dei tre eventi elementari (nel nostro caso in cui la distribuzione di probabilità è quella uniforme, la sua probabilità è pari ad un mezzo); rimane così verificata l'affermazione fatta a suo tempo che la distribuzione di probabilità fatta sull'insieme degli eventi elementari, cioè degli eventi composti di un solo campione, consente di ottenere la probabilità di qualsiasi altro evento dello spazio dei campioni.

Es. 2 L'evento E_1 sia l'uscita di un numero dispari: $E_1 = \{1,3,5\}$, con $P(E_1) = ½$, e l'evento E_2 sia l'uscita del numero 2, cioè $E_2 = \{2\}$ con $P(E_2) = 1/6$. Ne risulta che $E_1 \vee E_2 = \{1,2,3,5\}$ e $P(E_1 \vee E_2) = 2/3$, mentre $E_1 \wedge E_2 = \varnothing$; si verifica immediatamente che

$$P(E_1 \vee E_2) = P(E_1) + P(E_2)$$
$$2/3 = 1/2 + 1/6 .$$

Es. 3 L'evento E_1 sia lo stesso dell'esempio precedente, mentre l'evento E_2 sia costituito dall'uscita di un numero maggiore di 2: $E_2 = \{ 3, 4, 5, 6 \}$ con $P(E_2) = 2/3$; evidentemente $E_1 \vee E_2 = \{1,3,4,5,6\}$ e $P(E_1 \vee E_2) = 5/6$, mentre

$$P(E_1 \vee E_2) = P(E_1) + P(E_2) = 1/2 + 2/3 = 7/6,$$

cosa palesemente assurda! L'assurdo è dovuto al fatto che i due eventi non sono incompatibili, $E_1 \wedge E_2 \neq 0$, essendo $E_1 \wedge E_2 = \{3,5\}$. Occorrerà allora trovare in quale modo modificare il terzo assioma per poter tenere conto di queste situazioni tutt'altro che improbabili.

Abbiamo riconosciuto che il problema nasce dalla mancata incompatibilità dei due eventi addendi, cosa che comporta il fatto che la probabilità della parte comune ai due, ossia la probabilità del loro prodotto, viene contata due volte: la cosa da fare sarà dunque quella di correggere il terzo assioma per tenere conto di questa possibilità, ripresentandolo nella forma

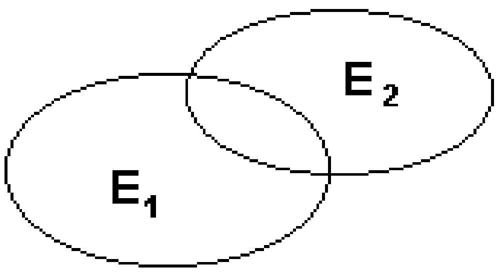


Fig. 2 Eventi non incompatibili

3°' $P(E_1 \vee E_2) = P(E_1) + P(E_2) - P(E_1 \wedge E_2)$.

Questa correzione è del tutto compatibile con la formulazione precedente del terzo assioma al quale la definizione attuale si riduce quando $E_1 \wedge E_2 = 0$, ossia quando i due eventi da sommare logicamente sono incompatibili.

3.2 - Probabilità condizionata (o subordinata).

Il problema da affrontare ora è quello di vedere se la probabilità del verificarsi di un certo evento E (che sarà detto *evento tesi*) risulti modificata, e se sì in quale modo, dal verificarsi di un altro evento H (detto *evento ipotesi*). Possono verificarsi due casi, e precisamente:

la probabilità di E rimane la stessa, ed allora diremo che E ed H sono tra loro *stocasticamente indipendenti*,

la probabilità di E risulta modificata; in questo caso diremo che E ed H sono *stocasticamente correlati*, e distingueremo ulteriormente, dicendo che

sono correlati in *modo positivo* se la probabilità di E subordinata ad H è maggiore di quella del solo E,

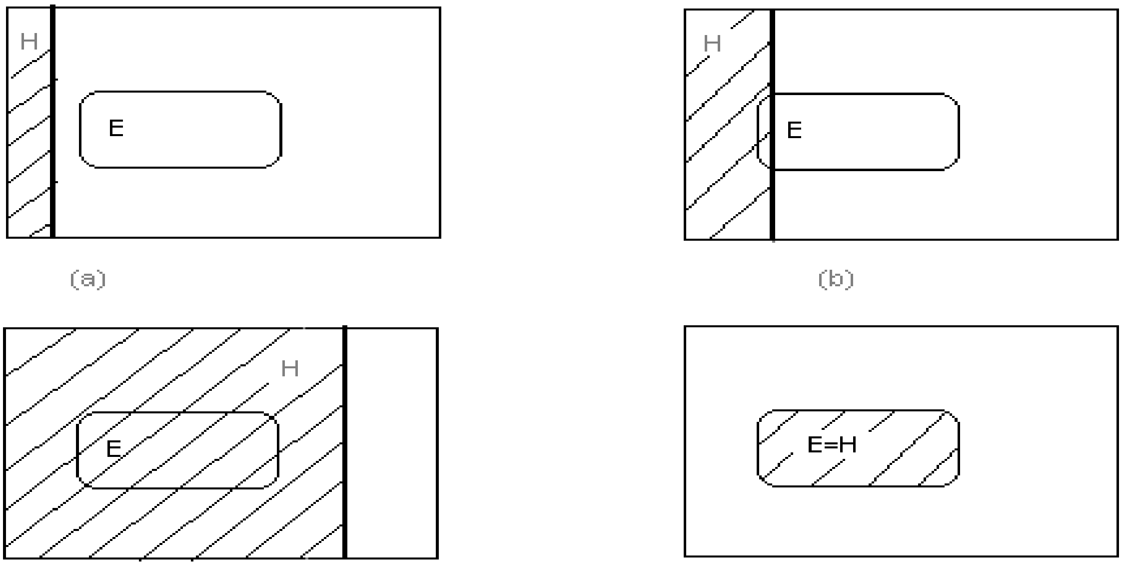
sono correlati in *modo negativo* nel caso contrario.

La probabilità condizionata, o subordinata, viene indicata come $P(E/H)$ (si legge: probabilità di E condizionata da –o subordinata ad- H).

E' evidente che l'evento ipotesi H potrà essere un qualunque evento dello spazio Ω, fatta eccezione per l'evento impossibile:

$$H \neq \emptyset \text{ e dunque } P(H)\neq 0;$$

infatti non avrebbe significato alcuno porsi la domanda se il *verificarsi dell'evento impossibile* abbia o meno influenza sul verificarsi dell'evento E (il *verificarsi dell'evento impossibile* è



(a) (b) (c) (d)

infatti una contraddizione in termini)!

Vediamo in quale modo il verificarsi dell'evento ipotesi H possa influenzare la probabilità del verificarsi dell'evento tesi E. Osserviamo come l'affermazione relativa al verificarsi di un evento ipotesi H non faccia altro che ridurre lo spazio dei campioni; allora, per la rideterminazione della probabilità di E, dallo spazio dei campioni devono essere tolti tutti i campioni che escludono il verificarsi dell'ipotesi H, e dunque i campioni della negazione di H. Ciò spiega come il verificarsi di H possa avere anche una influenza positiva sul verificarsi di E (possa cioè aumentarne la probabilità, fino a trasformarlo nell'evento certo nel caso in cui H ed E siano equivalenti o nel caso in cui H implichi E).

Esaminiamo la figura precedente, nella quale lo spazio dei campioni Ω è indicato da un rettangolo e, come di consueto, le rappresentazioni dei vari eventi sono in ragione della loro probabilità (nel senso che eventi di probabilità maggiore vengono rappresentati da un insieme più grande, in modo che il rapporto tra la misura dell'insieme E e la misura dello spazio dei campioni Ω fornisca un'idea grossolana della probabilità di E).

Nella figura, nella quale l'evento ipotesi è contraddistinto dal tratteggio, sono rappresentati quattro casi; il caso (a) indica un evento ipotesi H che non è compatibile con l'evento condizionato E: dunque in questo caso la probabilità di E condizionata da H è nulla; nel caso (b) E ed H hanno una piccola intersezione, ma è ragionevole pensare che il rapporto tra

E ed Ω (che indica la probabilità dell'evento E nell'esperimento S, indipendentemente dal verificarsi dell'ipotesi H) sia maggiore del rapporto tra il prodotto $E \wedge H$ ed H stesso: dunque il verificarsi di H influisce negativamente sulla probabilità di E; nel caso (c) siamo nella situazione opposta, in quanto il prodotto di E con H coincide con E stesso, e dunque il rapporto tra E e l'intero spazio dei campioni Ω è certamente minore del rapporto tra E ed H, il cui verificarsi influisce così positivamente sulla probabilità di E; nel caso (d) infine H coincide con E (ma sarebbe sufficiente che lo implicasse), e dunque il verificarsi di H comporta necessariamente il verificarsi di E che diviene così l'evento certo.

Proponiamo tre semplici esempi, tutti relativi allo spazio dei campioni generato dal lancio di un dado. L'evento tesi E è costituito dall'uscita dei numeri 1 o 2: $E = \{1,2\}$, di probabilità $P(E) = 1/3$, mentre cambia di volta in volta l'evento ipotesi H.

Es. 1 - H è dato dall'uscita di un numero dispari, cioè $H = \{1,3,5\}$; la probabilità di E condizionato da H è allora la probabilità che esca l'1 o il 2 quando è uscito un numero dispari: evidentemente va scartata la possibilità dell'uscita del 2, e dunque $P(E/H) = 1/3$, uguale alla probabilità di E: E ed H sono stocasticamente indipendenti.

Es. 2 - H è dato dall'uscita di un numero minore di 5, cioè $H = \{1,2,3,4\}$ }; la probabilità di E condizionato da H è allora la probabilità che esca l'1 o il 2 quando è uscito un numero minore di 5: dunque $P(E/H) = 1/2$, maggiore della probabilità di E: E ed H sono correlati positivamente.

Es. 3 - H è dato dall'uscita di un numero maggiore di 1, cioè $H = \{2,3,4,5,6\}$; la probabilità di E condizionato da H è allora la probabilità che esca l'1 o il 2 quando è uscito un numero maggiore di 1: dunque $P(E/H) = 1/5$, minore della probabilità di E: E ed H sono correlati negativamente.

Tutto ciò premesso, ci si presenta il problema di trovare una precisa espressione da dare alla probabilità condizionata, espressione che deve rispettare alcune esigenze, e precisamente le seguenti:

- la probabilità di E subordinata all'evento certo Ω deve coincidere con quella di E;

- la probabilità di E subordinata ad H deve essere nulla quando gli eventi E ed H sono incompatibili;
- la probabilità di E subordinata ad H deve essere 1 quando l'evento H induce l'evento E.

Queste considerazioni ci portano ad accettare come definizione della probabilità subordinata l'espressione seguente:

$$P(E/H) = \frac{P(E \wedge H)}{P(H)},$$

dove non dobbiamo preoccuparci che si possa avere $P(H) = 0$, poiché, per ipotesi, $H \neq \emptyset$.

Le tre esigenze riconosciute qui sopra sono certamente soddisfatte. Infatti:

- se $H = \Omega$ si ottiene P(E/Ω) = P(E ∧ Ω)/P(Ω) = P(E)/1;
- se $E \wedge H = \emptyset$ abbiamo $P(E \wedge H) = 0$, e dunque P(E/H) = 0;
- se $E \wedge H = H$, P(E/H) = P(E∧H)/P(H) = P(H)/P(H) = 1.

Riprendiamo in esame gli esempi proposti, ricordando che $E = \{1,2\}$, con $P(E) = 1/3$.

Es. 1 - $H = \{1,3,5\}$, $P(H) = 1/2$, $E \wedge H = \{1\}$, $P(E \wedge H) = 1/6$: $P(E/H) = \frac{1/6}{1/2} = \frac{1}{3} = P(E)$.

Es. 2 - $H = \{1,2,3,4\}$, $P(H) = 2/3$, $E \wedge H = \{1,2\}$, $P(E \wedge H) = 1/3$: $P(E/H) = \frac{1/3}{2/3} = \frac{1}{2} > P(E)$.

Es. 3 - $H = \{2,3,4,5,6\}$, $P(H) = 5/6$, $E \wedge H = \{2\}$, $P(E \wedge H) = 1/6$: $P(E/H) = \frac{1/6}{5/6} = \frac{1}{5} < P(E)$.

Abbiamo così ritrovato, applicando la definizione di probabilità condizionata, i risultati già ottenuti con il ragionamento diretto effettuato in precedenza.

Acquisita l’espressione della probabilità condizionata, possiamo dare una forma molto significativa al concetto di indipendenza stocastica, forma che talvolta viene proposta essa stessa come definizione di indipendenza tra eventi. Nella definizione originale l'indipendenza stocastica era indicata dall’eguaglianza tra la probabilità di un evento, E, e la probabilità dello stesso condizionata da un altro evento, H: era cioè $P(E) = P(E/H)$. Ricorrendo alla definizione di $P(E/H)$, abbiamo

$$P(E) = P(E/H) = \frac{P(E \wedge H)}{P(H)},$$

dalla quale

$$P(E \wedge H) = P(E)\,P(H).$$

Dunque, se due eventi sono stocasticamente indipendenti, la probabilità del loro prodotto è data dal prodotto delle rispettive probabilità.

Si noti che, in generale, né la probabilità di una somma logica di eventi coincide con la somma delle probabilità degli eventi stessi, né la probabilità di un prodotto logico di eventi coincide con il prodotto delle probabilità degli eventi stessi: per la somma, occorrerebbe la incompatibilità degli addendi, e per il prodotto l’indipendenza dei fattori.

3.3 Un primo esempio

Proponiamo ora un primo esempio di impiego della probabilità condizionata, più significativo di quelli, piuttosto banali, dei quali ci siamo serviti in precedenza.

Consideriamo come spazio dei campioni Ω una popolazione di individui suddivisa tra fumatori, F, e non fumatori, N. I sottoinsiemi, o eventi, F ed N sono tali da soddisfare le condizioni $F \vee N = \Omega$, $F \wedge N = \emptyset$ **,** dal momento che ogni individuo di Ω è fumatore o non è fumatore, e nessun individuo può considerarsi contemporaneamente fumatore e non fumatore.

Sia poi M l'evento: il generico individuo estratto dalla popolazione è affetto da una malattia che si può pensare connessa al fumo. Si conoscono le probabilità che un individuo sia o meno fumatore, appartenga cioè ad F o ad N (in realtà più che di probabilità si dovrebbe parlare di frequenze relative, rilevate statisticamente), nonché la probabilità che un fumatore sia malato e quella che un non fumatore sia malato: siano

$$P(F) = 0.40 \qquad\qquad P(M / F) = 0.25$$
$$P(N) = 0.60 \qquad\qquad P(M / N) = 0.08.$$

Si osservi come la somma delle probabilità dei fumatori e dei non fumatori dia 1, come doveva essere; e si osservi anche che tra i dati compaiono probabilità condizionate, probabilità cioè dell'essere malato essendo (o meno) un fumatore: nei due casi l'evento tesi è sempre '*essere malato*', M, e gli eventi ipotesi sono rispettivamente '*essere fumatore*', F, e '*non essere fumatore*', N. I dati proposti ci consentono di rispondere a domande del tipo: quale è la probabilità che un individuo scelto a caso dalla popolazione sia malato, cioè quale è $P(M)$?, quale è dunque la frequenza della malattia nell'intera popolazione?

Possiamo infatti scrivere la probabilità per il generico individuo di essere malato come somma della probabilità di avere scelto un fumatore malato e della probabilità di avere scelto un non fumatore malato:

$$M = M \wedge \Omega = M \wedge (F \vee N) = (M \wedge F) \vee (M \wedge N),$$

per cui

$$P(M) = P(M \wedge F) + P(M \wedge N).$$

dal momento che

$$(M \wedge F) \wedge (M \wedge N) = M \wedge F \wedge M \wedge N = (M \wedge M) \wedge (N \wedge F) = M \wedge \emptyset = \emptyset.$$

Per valutare le probabilità dei due prodotti logici possiamo fare ricorso alla definizione di probabilità condizionata; possiamo scrivere

$$P(M / F) = \frac{P(M \wedge F)}{P(F)},$$

da cui

$$P(M \wedge F) = P(M / F)P(F).$$

Allo stesso modo avremo

$$P(M \wedge N) = P(M / N)P(N),$$

e possiamo così concludere che

$$P(M) = P(M / F)P(F) + P(M / N)P(N) = 0{,}25 \cdot 0{,}40 + 0{,}08 \cdot 0{,}60 = 0{,}148 .$$

3.4 – Ulteriori esempi

Presentiamo ora altri esempi di un certo significato, al fine di rendere più chiari gli argomenti esposti ai punti precedenti, e non abbiamo di certo la pretesa di mettere in grado il Lettore di giungere autonomamente alle soluzioni. Chi dunque non fosse interessato a tale argomento potrebbe tranquillamente saltare quanto qui proposto.

3.4.1 – Lancio ripetuto di una moneta.

Consideriamo dapprima l'esperimento casuale S consistente nel lancio di una moneta ripetuto n volte. Lo spazio dei campioni Ω_n di questo nuovo esperimento si compone di 2^n elementi, ossia di tutte le possibili $n-ple$ ordinate che sono le disposizioni con ripetizione di *teste* e *croci*. L'evento tesi E sia quello costituito dall'uscita di n teste, la sua probabilità $P(E) = 1/2^n = 2^{-n}$, in quanto ci sarà un solo campione che lo realizza, ossia l'$n-pla$ costituita da sole teste. Come evento ipotesi consideriamo l'uscita di $n-1$ teste nelle prime $n-1$ ripetizioni, di probabilità $P(H) = 1/2^{n-1} = 2^{-(n-1)}$ (ci sono infatti in Ω due $n-ple$ che presentano $n-1$ teste nei primi $n-1$ posti, quella che al posto $n-esimo$ presenta ancora una testa, e realizza quindi anche l'esperimento tesi E, e quella che al posto $n-esimo$ presenta una croce). Ci si domanda se l'uscita di $n-1$ teste consecutive, e cioè il realizzarsi dell'evento ipotesi H, influenza la probabilità che nel lancio successivo, l'$n-esimo$, esca ancora testa, rendendo tale probabilità maggiore di $1/2$, o se invece questa rimane invariata. La risposta più logica, anche se non condivisa dalla maggior parte dei giocatori, è: NO, ossia la probabilità rimane $1/2$. Infatti, possiamo notare che $E \subset H$, e dunque $E \wedge H = E$; sfruttando la definizione di probabilità condizionata abbiamo

$$P(E/H) = \frac{P(E \wedge H)}{P(H)} = \frac{P(E)}{P(H)} = \frac{2^{-n}}{2^{-(n-1)}} = \frac{1}{2},$$

come doveva essere. Chi non si trovasse d'accordo con quanto ottenuto sarebbe caldamente sconsigliato dal frequentare un casinò per giocare p.e. sul rosso e nero.

3.4.2 – Somministrazione di un farmaco.

Come ulteriore esempio del concetto e del relativo impiego della probabilità condizionata consideriamo il caso di un farmaco la cui somministrazione provochi la indesiderata comparsa di due effetti collaterali, indicati rispettivamente con A e B. Le frequenze relative di questi, che assumeremo senz'altro come loro probabilità, siano $P(A) = 10^{-2}$, $P(B) = 10^{-1}$, mentre risulta essere $P(A' \wedge B') = 0{,}9$ la probabilità che non si verifichi nessuno dei due, cioè la probabilità dell'evento $A' \wedge B'$.

Ci si pone il problema della ricerca di un'eventuale correlazione tra gli effetti A e B, si vuole cioè verificare se la comparsa di B, effetto più probabile, abbia influenza sulla comparsa di A, effetto più raro (in rapporto di 1 a 10 con il precedente); dobbiamo dunque valutare la probabilità di A condizionato da B: $P(A/B) = \dfrac{P(A \wedge B)}{P(B)}$.

Il problema è così ricondotto alla valutazione della probabilità del verificarsi contemporaneo di entrambi gli effetti, $P(A \wedge B)$, che non compare tra i dati a disposizione ma è calcolabile in base agli stessi. Infatti, dal terzo assioma sulla probabilità, abbiamo

$$P(A \wedge B) = P(A) + P(B) - P(A \vee B) = P(A) + P(B) - [1 - P((A \vee B)')] =$$
$$= P(A) + P(B) - (1 - P(A' \wedge B')) = 0.01 + 0.1 - (1 - 0.9) = 0.01 = 10^{-2}.$$

Ne concludiamo che

$$P(A/B) = \frac{P(A \wedge B)}{P(B)} = \frac{10^{-2}}{10^{-1}} = 10^{-1},$$

con $P(A/B) > P(A)$: A e B sono dunque correlati positivamente, in quanto la comparsa dell'effetto B aumenta la probabilità della comparsa dell'effetto A (la moltiplica addirittura per 10!). In queste condizioni qualsiasi provvedimento atto a ridurre la comparsa di B automaticamente ridurrà la frequenza di A (o almeno, dovrebbe ridurla).

3.4.3 – Estrazioni successive.

Si voglia valutare la probabilità dell'evento, indicato come E, nel quale un numero prefissato, p.e. il 19, viene estratto da un'urna contenente 90 palline contraddistinte dagli interi da 1 a 90, entro le prime tre estrazioni, tenendo presente che ad ogni estrazione viene prelevato il numero estratto che non è più disponibile per le estrazioni successive. Definiti gli eventi

E_1: {il numero prescelto esce al primo tentativo},
E_2: {il numero prescelto esce al secondo tentativo},
E_3: {il numero prescelto esce al terzo tentativo},

possiamo assegnare loro le probabilità rispettive:

$$P(E_1) = \frac{1}{90}, \quad P(E_2) = \frac{1}{89}, \quad P(E_3) = \frac{1}{88},$$

dal momento che ad ogni successiva estrazione le residue possibilità di scelta diminuiscono di una unità. Il ragionamento precedente è senza dubbio corretto, ma si faccia attenzione a non spingersi oltre affermando che in tali condizioni la probabilità di E si possa ottenere direttamente dalla somma delle tre probabilità, e cioè che la probabilità dell'estrazione del 19 entro i primi tre tentativi sia

$$P(E) = P(E_1) + P(E_2) + P(E_3) = \frac{1}{90} + \frac{1}{89} + \frac{1}{88}.$$

Tale affermazione porterebbe infatti a conclusioni assurde, quali la determinazione di probabilità maggiori di uno! Basti pensare per semplicità ad un caso analogo a quello proposto nel quale le palline a disposizione siano solamente 3, p.e. quelle contraddistinte dai numeri 18, 19 e 20; è allora evidente che le tre estrazioni successive esauriscono tutte le possibilità, e dunque E diviene l'evento certo di probabilità 1 (in quanto, male che vada, alla terza estrazione il 19 esce). Applicando la formula suggerita, erroneamente, in precedenza, avremmo

$$P(E) = P(E_1) + P(E_2) + P(E_3) = \frac{1}{3} + \frac{1}{2} + \frac{1}{1} = \frac{2+3+6}{6} = \frac{11}{6}!!.$$

Il ragionamento corretto deve invece tenere conto del fatto che si deve considerare la probabilità di E_2 (ed eventualmente quella del successivo E_3) solo nel caso in cui il tentativo (o i tentativi) precedenti non hanno avuto successo. Dunque alla probabilità di E_1 dobbiamo sommare non quella dell'evento E_2, bensì quella dell'evento E_2 condizionato dal non verificarsi di E_1, cioè $P(E_2 / E_1')$, moltiplicata per la probabilità di questo, $P(E_1')$, ed ancora la probabilità di E_3 subordinata al verificarsi contemporaneo delle negazioni degli eventi E_1 ed E_2, $P(E_3 /(E_1' \wedge E_2'))$, moltiplicata per $P(E_1' \wedge E_2')$: dunque

$$P(E) = P(E_1) + P(E_2 / E_1')\,P(E_1') + P(E_3 /(E_1' \wedge E_2'))\,P(E_1' \wedge E_2')\,^2.$$

Ricordando la definizione di probabilità subordinata, la relazione precedente si riscrive come

$$P(E) = P(E_1) + P(E_2 \wedge E_1') + P(E_3 \wedge E_1' \wedge E_2')$$

(cosa del tutto naturale ove si pensi che l'espressione precedente esprime la somma della probabilità dell'uscita del 19 alla prima estrazione, $P(E_1)$, con quella dell'uscita dello stesso alla seconda estrazione, che presuppone la negazione di E_1, e cioè il verificarsi congiuntamente degli eventi E_2 ed E_1', $P(E_2 \wedge E_1')$, e con la probabilità dell'uscita al terzo tentativo, E_3, che presuppone la precedente negazione degli eventi E_1 ed E_2, e quindi il realizzarsi contemporaneo dell'evento E_3 e della negazione degli eventi E_1 ed E_2, cioè la probabilità $P(E_3 \wedge E_1' \wedge E_2')$: infatti potremo preoccuparci della probabilità di estrarre il 19 alla seconda o alla terza estrazione solo nel caso in cui tale estrazione non sia già avvenuta al primo o al secondo tentativo).

Per valutare le probabilità dei vari eventi coinvolti nell'espressione precedente ci basterà osservare che essi sono tutti stocasticamente indipendenti, in quanto per esempio il verificarsi di E_1', consente, ma di certo non assicura, il verificarsi di E_2 (se il numero cercato non è stato estratto la prima volta non è affatto certo che venga estratto la seconda!): dunque tutte le probabilità dei prodotti di eventi si possono sostituire con i prodotti delle probabilità:

$$P(E) = P(E_1) + P(E_2 \wedge E_1') + P(E_3 \wedge E_1' \wedge E_2') =$$
$$= P(E_1) + P(E_2)P(E_1') + P(E_3)P(E_1' \wedge E_2') = P(E_1) + P(E_2)P(E_1') + P(E_3)P(E_1')P(E_2')\ .$$

In definitiva

$$P(E) = \frac{1}{90} + \frac{1}{89}\frac{89}{90} + \frac{1}{88}\frac{88}{89}\frac{89}{90} = \frac{1}{90} + \frac{1}{90} + \frac{1}{90} = \frac{3}{90}\,.$$

Concludiamo la trattazione di questo esempio osservando che ogni successiva estrazione aumenta la probabilità di successo di E (uscita del numero prescelto in uno qualsiasi dei tentativi) nella stessa misura, che è di $1/90$. Dunque la ripetizione dell'estrazione per 90 volte (certo non di più perché non ci sono altri numeri a disposizione!) porta esattamente alla certezza del successo, come era intuitivo, e non ad assurde probabilità maggiori di 1.

Osserviamo infine che il risultato ottenuto, ossia $P(E) = 3/90$, si può giustificare anche

[2] Non avrebbe infatti senso considerare la probabilità dell'estrazione del 19 al secondo tentativo se il 19 fosse già uscito al primo.

con un ragionamento alternativo. Dal momento che il realizzarsi di E è garantito dalla presenza del numero 19 in una delle prime tre estrazioni, indipendentemente da quale essa sia, senza cioè dare importanza all'ordine, potremmo pensare di levare in un'unica estrazione tre palline, una qualsiasi delle quali potrebbe portare il numero cercato: i casi favorevoli per il realizzarsi di E sono dunque tre, sui novanta casi possibili.

3.4.4 – Estrazione di palline colorate.

Consideriamo, in quest'ultimo esempio, un'urna contenente un numero b di palline bianche ed un numero r di palline rosse; da tale urna si estraggono in successione due palline. Si chiede di determinare la probabilità che la seconda pallina estratta sia bianca indipendentemente dal colore della prima estratta.

Definiamo allora gli eventi

$$E_1 = \{\text{la prima estratta è bianca}\}$$
$$E_2 = \{\text{la prima estratta è rossa}\}.$$

Le rispettive probabilità sono ovviamente $P(E_1) = \dfrac{b}{b+r}$, $P(E_2) = \dfrac{r}{b+r}$. E' evidente che la prima estrazione abbassa il numero di palline inizialmente presenti nell'urna di una unità, portandole da $b+r$ a $b+r-1$; il realizzarsi di E_1 sarà causa della diminuzione delle palline bianche che passeranno da b a $b-1$, mentre il realizzarsi di E_2 farà diminuire le palline rosse da r a $r-1$. Se indichiamo con E l'evento: 'la seconda pallina è bianca', del quale appunto vogliamo valutare la probabilità, vediamo che la sua probabilità condizionata da E_1 è $P(E/E_1) = \dfrac{b-1}{b+r-1}$, dal momento che il verificarsi di E_1 ha lasciato $b-1$ palline bianche in un'urna che contiene complessivamente $b+r-1$, mentre la sua probabilità condizionata da E_2 è $P(E/E_2) = \dfrac{b}{b+r-1}$, perché il verificarsi di E_2 ha lasciato disponibili, nella stessa urna, tutte le b palline bianche inizialmente presenti**.** Dunque

$$P(E) = P(E/E_1)P(E_1) + P(E/E_2)P(E_2) = \frac{b-1}{b+r-1}\frac{b}{b+r} + \frac{b}{b+r-1}\frac{r}{b+r} =$$
$$= \frac{b(b+r-1)}{(b+r)(b+r-1)} = \frac{b}{b+r} \quad .$$

che è la stessa di E_1: dunque la probabilità dell'uscita di una pallina bianca alla seconda estrazione coincide con quella dell'uscita di una pallina bianca alla prima.

3.5 - Teorema di Bayes

Dalla definizione di probabilità subordinata si può ricavare, come del resto già fatto in precedenza, la probabilità del prodotto di due eventi:

$$P(E \wedge H) = P(E / H)\, P(H) .$$

Scambiando allora i ruoli di eventi ipotesi e tesi, il che comporta la sola ipotesi aggiuntiva che nemmeno l'evento E sia l'evento impossibile, otteniamo

$$P(H / E) = \frac{P(H \wedge E)}{P(E)}$$

e dunque

$$P(H \wedge E) = P(E \wedge H) = P(H / E)\, P(E) .$$

Dal confronto, si deduce $P(E / H)\, P(H) = P(H / E)\, P(E)$, o ancora

$$\frac{P(E / H)}{P(E)} = \frac{P(H / E)}{P(H)} .$$

Quella precedente costituisce una delle possibili espressioni, la più semplice, del notevole teorema di Bayes, dalla quale si ricava che, se gli eventi E ed H sono correlati stocasticamente, anche H ed E lo sono, cosa del tutto naturale, e le loro correlazioni sono uguali, coas questa meno intuitiva, dal momento che i rapporti tra probabilità condizionata e probabilità semplice sono gli stessi nei due casi.

Questo teorema consente anche di sostituire nelle applicazioni la probabilità di H condizionata da E a quella di E condizionata da H; è infatti sufficiente riscrivere il teorema nella ulteriore forma equivalente $P(H / E) = \frac{P(H)}{P(E)} P(E / H)$, risultato molto spesso utilizzato nella soluzione di vari problemi.

Come immediata applicazione del teorema di Bayes, riprendiamo in esame un esempio precedente, relativo ad una popolazione di individui, ripartita in fumatori e non fumatori, nel quale si era calcolata la probabilità che un individuo, scelto a caso, sia affetto da un morbo riconducibile al fumo, evento indicato allora con M, con probabilità $P(M)$. Poniamoci adesso la domanda: dato che un individuo sia malato, quale è la probabilità che esso sia un fumatore? cioè, quale è la probabilità di essere un fumatore condizionata dall'essere malato, $P(F / M)$? Applicando il teorema di Bayes abbiamo

$$P(F / M) = \frac{P(F)}{P(M)} P(M / F) = \frac{0,4}{0,148} 0,25 = 0,676 .$$

Come esempio conclusivo, per altro non elementare, consideriamo il caso di tre armadi identici con due cassetti ciascuno; in ognuno dei sei cassetti sia stata messa una moneta d'argento od una moneta d'oro, facendo in modo che in un armadio ci siano due monete d'oro, in uno due d'argento e nel terzo una moneta d'oro ed una d'argento. Si supponga di avere aperto a caso uno dei due cassetti di uno dei tre armadi, e di averci trovata una moneta d'oro. Si domanda quale è la probabilità che l'armadio prescelto sia quello con due monete d'oro, ossia la probabilità che aprendo l'altro cassetto del medesimo armadio ci si trovi un'altra moneta d'oro.

Non si ceda al ragionamento: l'aver trovato una moneta d'oro esclude l'armadio con due monete d'argento, e dunque la probabilità che l'armadio sia quello con due monete d'oro sarà di uno su due armadi possibili, cioè 1/2! Ragionando in termini corretti, definiamo tre eventi:

E_1 : è stato scelto l'armadio con due monete d'oro;
E_2 : è stato scelto l'armadio con una moneta d'oro ed una moneta d'argento;
E_3 : è stato scelto l'armadio con due monete d'argento;

ed ancora l'evento A : è stata trovata una moneta d'oro al primo tentativo.

E' immediato riconoscere che le probabilità degli eventi E_i sono tutte uguali, dal momento che gli armadi sono indistinguibili, e dunque $P(E_i) = 1/3$ per $i = 1, 2, 3$. Possiamo anche determinare le probabilità di A condizionate dall'aver scelto rispettivamente il primo, il secondo od il terzo armadio:

$P(A/E_1) = 1$: nel primo armadio ci sono solo monete d'oro;
$P(A/E_2) = 1/2$: nel primo armadio c'è una sola moneta d'oro;
$P(A/E_3) = 0$: nel primo armadio non ci sono monete d'oro.

Possiamo così determinare la probabilità dell'evento A, come somma della sua probabilità condizionata da E_1 moltiplicata la probabilità di E_1, della sua probabilità condizionata da E_2 per la probabilità di E_2 e della sua probabilità condizionata da E_3 moltiplicata per la probabilità di E_3: dunque

$$P(A) = P(A/E_1)P(E_1) + P(A/E_2)P(E_2) + P(A/E_{31})P(E_3) =$$
$$= 1\cdot\frac{1}{3} + \frac{1}{2}\frac{1}{3} + 0\frac{1}{3} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2},$$

e quindi la probabilità di A, cioè dell'uscita al primo tentativo di una moneta d'oro, è di 1/2, come ci si poteva aspettare pensando che si è trovata una delle tre monete d'oro su un totale di sei monete. Conoscere $P(A)$ ci consente di considerare A stesso come evento ipotesi e di determinare in conseguenza la probabilità di E_1, scelta dell'armadio contenente due monete d'oro, condizionata da A, cioè dall'aver trovato subito una moneta d'oro; infatti abbiamo:

$$P(E_1 / A) = \frac{P(A / E_1)}{P(A)} P(E_1) = \frac{1}{1/2} \frac{1}{3} = \frac{2}{3} .$$

Dunque la probabilità cercata è pari a due terzi e non ad un mezzo come forse ci si poteva attendere. Veda il Lettore di giustificare tale conclusione.[3]

[3] Si potrebbe ragionare a questo modo. Il verificarsi di A esclude la scelta dell'armadio con due monete d'argento e, tra i due restanti, quello con due monete d'oro ha il doppio di probabilità dell'altro: dunque la probabilità che anche il secondo cassetto contenga una moneta d'oro è di $2/3$, e di $1/3$ quella contraria.