Indice

CAPITOLO 1

INTRODUZIONE

SPAZI DI PROBABILITÀ E VARIABILI ALEATORIE

CAPITOLO 2

INTRODUZIONE

Il calcolo delle probabilità e la statistica costituiscono quella parte della matematica e della scienza più in generale che si occupa dei fenomeni casuali; questo fatto crea talvolta qualche difficoltà nel comprenderne scopi e metodi. Per tentare di ovviare a queste incomprensioni, noi iniziamo qui presentando alcuni problemi, tratti da situazioni concrete del mondo reale, che, opportunamente formalizzati e predisposti per un approccio scientifico alla loro trattazione, faranno da guida per lo sviluppo di quasi tutta la teoria che presenteremo. Il lettore è invitato a tentare di fornire qualche risposta, sia inizialmente che durante lo sviluppo della materia, nonché a riferirsi a questi problemi quando stia perdendo di vista il senso di quanto viene discusso. Tutto ciò renderà la teoria molto semplice e naturale.

Esempio 1. *Supponiamo di partecipare ad un gioco in cui viene lanciata, da un addetto, una moneta* 1000 *volte, e supponiamo che noi si vinca 1$ se esce testa e se ne perda 1 se esce croce. Vogliamo valutare la nostra reazione ai vari possibili valori di k, il numero di teste sui* 1000 *lanci. Quale reazione sarebbe ragionevole esprimere se k fosse* 510*? E* 492*? Oppure* 459, 423, 397, 354, 299, 212, 154, 22*?*

Esempio 2. *Su una scatola di componenti elettronici contenente 100 pezzi è indicato che ognuno ha tempo medio di funzionamento di 1000 ore. Come nell'esempio precedente, proviamo ad immaginare quale sarebbe una reazione ragionevole se il tempo totale di funzionamento dell'intera scatola fosse 99800, oppure 95400 oppure 90200 oppure 84000 oppure 72000 ore.*

Esempio 3. *Supponiamo che in una popolazione vi siano solo due possibili forme di un certo carattere, per semplificare diciamo capelli biondi o castani, determinati dalla trasmissione genetica che ci è ora nota, con biondo carattere recessivo. Se nella popolazione i biondi sono il* 30%, *su* 100 *figli di genitori entrambi castani, quanti ce ne aspettiamo di biondi?*

Problemi come questi si riferiscono ad esperimenti di cui non si riesce a determinare con certezza l'esito. Ci sono di solito varie o anche molte possibili alternative e le informazioni in nostro possesso non ci permettono di prevedere quale di queste si verificherà .

Non ci interessa discutere qui se questo sia solo un problema di mancanza di informazioni (come ad esempio nei problemi che si riferiscono alla meccanica classica) o se l'indeterminatezza sia intimamente connessa con la natura dell'esperimento (come si assume, per esempio, in meccanica quantistica o in una interpretazione soggettiva della materia qui esposta). Ci interessa soltanto che tale indeterminatezza renda per noi l'esperimento ad esito incerto. Chiameremo tali esperimenti casuali.

Poiché però spesso a noi interessa tentare qualche previsione dell'esito dell'esperimento descriveremo qui un metodo che è stato sviluppato a partire dal XVII secolo (e presumibilmente già nel mondo greco) per prepararci agli esiti più frequenti. L'idea principale è quella di introdurre una nuova quantità, a cui si dà comunemente il nome di probabilità, che appunto serva come misura di quanto spesso o quanto facilmente si realizza un certo esito di un esperimento casuale.

Naturalmente non importa che tale quantità esista realmente, così come non è molto importante sapere se la lunghezza di un tavolo esiste davvero, quest'ultima cosa essendo difficilmente accertabile anche per il fatto che alcuni atomi si attaccano e si staccano continuamente dalle estremità e che queste non sono affatto liscie quando osservate al microscopio. Quello che importa però , sia per le lunghezze che per le probabilità , è che queste quantità siano misurabili con relativa facilità in vari casi interessanti e che si sia poi in grado, sulla base di tali misurazioni, di dire qualcosa di utile per qualche problema di un certo rilievo.

In considerazione di queste finalità pratiche non deve sorprendere che, essendovi vari diversi ambiti di applicazione, vi siano definizioni di probabilità non uguali tra loro e che vi siano discussioni anche accese sui meriti delle varie definizioni; discussioni che spesso sfociano nella sterile diatriba di quale sia quella più 'vera'. La discussione filosofica è poi ulteriormente complicata dall'ovvia presenza di avvenimenti con esito non determinato a cui non è semplice associare una valutazione di probabilità, e dalla meno ovvia esistenza di altri il cui risultato è ben determinato, ma di complessità tale che conviene inventarsi una probabilità.

Noi presenteremo qui alcune di queste definizioni e ne discuteremo lo sviluppo della teoria elementare ed alcune applicazioni di rilievo.

La caratteristica principale della probabilità è che, al contrario della lunghezza che si misura attraverso la comparazione fisica con un oggetto campione, essa si misura attraverso la riflessione astratta, usando logica e calcolo matematico. Queste note espongono vari metodi di calcolo di probabilità e varie applicazioni dei risultati.

Essendo un prodotto di calcoli logici, la probabilità verrà misurata tramite numeri puri; spesso nel linguaggio comune essa viene riportata in frazione di 100, ossia in percentuale: così si parla di 30% o 2%. E' però più comodo matematicamente esprimere la probabilità in frazione di 1, indicando quindi $0,3 = 30\%$ oppure $0,02 = 2\%$, principalmente perchè la moltiplicazione di probabilità riesce così più comoda: il 2% del 30% è lo 0,6%, direttamente ottenibile da $0,02 \cdot 0,3 = 0,006$.

Si noti che anche le frequenze sono espresse tramite percentuali (come fatto nell'esempio 2), ma che si tratta per lo più di una coincidenza (così come, per esempio, sia pesi che lunghezze si esprimono tramite numeri decimali).

Avendo tuttavia deciso di misurare le probabilità con valori in $[0,1]$ si vede che per due casi estremi è possibile determinare subito il valore della probabilità. Agli eventi logicamente impossibili da realizzare è naturale assegnare probabilità 0, mentre a quelli certi va assegnata probabilità 1. A tutte le altre situazioni andrà assegnata una probabilità nell'intervallo $[0,1]$ chiuso (ciò signfica che potranno esserci altre situazioni che avranno probabilità 0, che saranno logicamente possibili ma senza la possibilità di realizzarsi). Ora cominciamo a vedere vari metodi per tale assegnazione.

Esercizio 1. *Calcolare il* 3% *del* 25%*.*

Esercizio 2. *Calcolare il* 90% *del* 20% *dell'*80%*.*

Esercizio 3. *Calcolare l'*80% *del* 120%*.*

Esercizio 4. *Calcolare, se possibile, la probabilità che una asserzione falsa venga correttamente ritenuta vera.*

Esercizio 5. *Calcolare, se possibile, la probabilità di un evento la cui probabilità sia uguale ad* 1 *meno la probabilità stessa.*

SPAZI DI PROBABILITÀ E VARIABILI ALEATORIE

CAPITOLO 3

PROBABILITÀ UNIFORMI

3.1. Spazi di probabilità uniformi

Iniziamo ora a definire questa nuova quantità, la probabilità, in modo che serva per alcune situazioni semplici, come ad esempio:

Esempio 4. *Nel lancio di una moneta calcolare la probabilità che venga testa.*

Esempio 5. *Nel lancio di un dado calcolare la probabilità che il dado mostri la faccia 3.*

In questi esempi dobbiamo naturalmente fare un piccolo sforzo di astrazione. Assumiamo che il risultato di un lancio sia necessariamente una faccia (non una moneta verticale o la sparizione del dado), e che prima del lancio vi sia stato un adeguato mescolamento. Cosa questo sia non è ben determinato, ma a noi interessa l'esito di tale mescolamento, e cioè che, per quanto ne sappiamo, ognuna delle facce si comporta in modo equivalente a tutte le altre. Ossia, se dobbiamo assegnare una probabilità ad una, dobbiamo assegnare la stessa probabilità alle altre. Avendo già deciso che la probabilità che qualcosa avvenga è 1 ne consegue che per queste situazioni è adeguata la prossima definizione. E' chiaro che il discorso esposto finora è euristico, ossia non rigoroso ma fatto cercando di interpretare la realtà esterna, mentre da ora in poi si inizia a fare matematica partendo da una definizione precisa e sviluppandone le conseguenze. Per decidere a quali situazioni si potrà applicare si ritorna a fare discorsi euristici: a tutte quelle situazioni in cui vi siano un numero finito di alternative equivalenti dal punto di vista probabilistico.

Visto che si parla di un numero finito di alternative conviene considerare un insieme finito ed adottare quindi la terminologia delle teoria degli insiemi.

Definizione 1. *(Spazi di probabilità uniformi). Sia S un insieme finito. I suoi sottinsiemi $A \subseteq S$ sono detti eventi, e la probabilità uniforme su S è una funzione P definita su ogni evento A da*

$$P(A) = \frac{|A|}{|S|},$$

avendo indicato con $|A|$ il numero di elementi dell'insieme A.

Si dice talvolta che queste probabilità sono definite come rapporto tra il numero di *casi favorevoli* ed il numero di *casi possibili.*

Inoltre, gli elementi di S sono anche detti eventi elementari.

Dalla definizione discendono alcune proprietà elementari:

LEMMA 1. *(i)* $P(\emptyset) = 0, P(S) = 1$
(ii) per ogni $A \subseteq S$, $0 \leq P(A) \leq 1$*;*
(iii) per ogni $s \in S$, $P(s) = \frac{1}{|S|}$*;*
(iv) per ogni $A \subseteq B \subseteq S$, $P(A) \leq P(B)$.

In questo modo si calcolano le probabilità di eventi relativamente ad un singolo lancio di una moneta, di un dado, all'estrazione di carte o di numeri nel lotto ecc.

In taluni casi è più semplice calcolare la probabilità della negazione di un evento A, ossia del suo complemento insiemistico $A^c = S \backslash A$. Si ha

COROLLARIO 1. *(I) Per ogni evento* $A \subseteq S$ *si ha* $P(A^c) = 1 - P(A)$*;*
(II) in generale, se $A \subseteq B \subseteq S$ *si ha che* $P(B \backslash A) = P(B) - P(A)$.

A volte è più agevole dedurre probabilità di certi eventi da altri per i quali la probabilità si deriva più facilmente. Come si vede dalle proprietà della cardinalità:

LEMMA 2. *Per ogni* $A, B \subseteq S$ *si ha*
(1) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(2) se $A \cap B = \emptyset$ *allora* $P(A \cup B) = P(A) + P(B)$
(1) se $A_i \cap A_j = \emptyset$ *allora* $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

ESEMPIO 6. *Calcoliamo la probabilità di uscita di un numero (diciamo il* 23*) su una certa ruota nel lotto (in cui si estraggono senza reinserimento* 5 *numeri da* 90*). Se* A *è l'evento che esce il* 23 *ed* A_i *è l'evento che il* 23 *esce all'i-sima estrazione, per* $i = 1, \ldots, 5$*, si ha che gli* A_i *sono disgiunti e* $P(A_i) = 1/90$ *(attenzione, qui si intende che il* 23 *esce alla i-sima prova senza sapere nulla delle altre, torneremo su questo calcolo), per cui*

$$P(A) = P(\cup_{i=1}^5 A_i) = \sum_{i=1}^5 P(A_i) = 5/90$$

Per la probabilità dell'unione di n eventi anche non disgiunti si ha:

LEMMA 3. *per ogni* $A_i \subseteq S$, $i = 1, \ldots, n$

$$P(\cup_{i=1}^n A_i) = \sum_{k=1}^n \sum_{\{i_1,\ldots,i_k\} \in I_n} (-1)^{k+1} P(A_{i_1} \cap \cdots \cap A_{i_k}), \tag{3.1}$$

ove $I_n = \{\{i_1, \ldots, i_k\} | i_j \in \{1, \ldots, n\} \text{ per ogni } j, \quad i_j \neq i_k \text{ per } j \neq k\}$.

Dimostrazione. Segue dalla parte (1) del Lemma 2 per induzione. Per $n = 2$ essa è equivalente infatti alla tesi. Supponiamo quindi valida la conclusione per ogni famiglia al più $n-1$ eventi. Di nuovo dalla parte (1) del Lemma 2 e dall'ipotesi di induzione si ha

$$\begin{aligned}
P(\cup_{i=1}^n A_i) &= P(\cup_{i=1}^{n-1} A_i \cup A_n) \\
&= P(\cup_{i=1}^{n-1} A_i) + P(A_n) - P(\cup_{i=1}^{n-1} A_i \cap A_n) \\
&= \sum_{k=1}^{n-1} \sum_{\{i_1,\ldots,i_k\}\in I_{n-1,k}} (-1)^{k+1} P(\cap_{j=1}^k A_{i_j}) + P(A_n) \\
&\quad - \sum_{k'=1}^{n-1} \sum_{\{i_1,\ldots,i_{k'}\}\in I_{n-1,k'}} (-1)^{k'+1} P(\cap_{j=1}^{k'} A_{i_j} \cap A_n) \\
&= \sum_{k=1}^{n} \sum_{\{i_1,\ldots,i_k\}\in I_n} (-1)^{k+1} P(A_{i_1} \cap \cdots \cap A_{i_k}).
\end{aligned}$$

L'ultima uguaglianza vale avendo posto $k' = k-1$, da cui $-(-1)^{k'+1} = (-1)^{k+1}$, in quanto i termini con $n \notin \{i_1,\ldots,i_k\}$ vengono dalla prima sommatoria, quelli con $k > 1$ ed $n \in \{i_1,\ldots,i_k\}$ vengono dalla seconda ed il termine con $k = 1$ ed $i_1 = n$ è $P(A_n)$. $\square$

Si noti che la dimostrazione precedente è basata unicamente sulla parte (1) del Lemma 2.

Esempio 7. *In 3 lanci di una moneta, se* $A = \{$*esce almeno una testa*$\}$ *e* $A_i = \{$*esce testa all'i-simo lancio*$\}$*, si ha* $A = \cup_{i=1}^3 A_i$*. Si può applicare la* (3.1) *e per questo basta osservare che per* $i, j \in \{1,\ldots,3\}$ *diversi tra loro*

$$P(A_i) = 4/8, \quad P(A_i \cap A_j) = 2/8 \text{ e } P(A_1 \cap A_2 \cap A_3) = 1/8$$

come si ottiene facilmente dalla scelta di un appropriato spazio di probabilità finito per ogni evento. Si ottiene quindi $P(A) = 7/8$.

Esempio 8. *Collocando a caso* n *palline numerate da* 1 *ad* n *in* n *urne anch'esse numerate, una per urna, calcoliamo la probabilità che nessuna pallina sia al posto giusto. Si può passare all'evento complementare ed usare la* (3.1)*. Osservando poi che se* A_i *è l'evento che l'i-sima pallina è al suo posto allora* $P(\cap_{i=1}^k A_i) = \frac{1}{n(n-1)\ldots(n-k+1)}$ *come descritto nel successivo paragrafo sul calcolo combinatorio, allora per la probabilità dell'evento considerato* B *è possibile dare una formula esplicita:*

$$P(B) = 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!}.$$

3.2. Calcolo combinatorio

La necessità di stabilire delle cardinalità di vari insiemi ha portato allo sviluppo del *calcolo combinatorio*, le cui formule principali sono le seguenti:

Il numero di campioni ordinati ossia delle k-ple ordinate con ripetizione da n elementi, dette disposizioni con ripetizione, è dato da

$$D_{n,k}^{(r)} = n^k.$$

Il numero dei campioni senza ripetizione ordinati ossia delle k-ple ordinate senza ripetizione da n elementi, dette disposizioni senza ripetizione, è dato da

$$D_{n,k} = n(n-1)\cdot \cdots \cdot (n-k+1) = (n)_k.$$

Il numero delle sottopopolazioni, ossia delle k-ple non ordinate senza ripetizione da n elementi, dette combinazioni senza ripetizione, è dato dal coefficiente binomiale

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Il numero delle sottopopolazioni con ripetizione, ossia delle k-ple non ordinate con ripetizione da n elementi, dette combinazioni con ripetizione, è dato da

$$C_{n,k}^{(r)} = \binom{n+k-1}{k-1}.$$

Il numero delle permutazioni di n elementi è dato da

$$P_n = D_{n,n} = n!.$$

Il numero di ripartizioni in sottopopolazioni di $k_1, k_2, \ldots, k_r$ elementi di un insieme di n elementi, con $k_1 + k_2 + \cdots + k_r = n$, è dato dal coefficiente multinomiale

$$C_{n,(k_1,k_2,\ldots,k_r)} = \frac{n!}{k_1!k_2!\ldots k_r!}.$$

3.3. Indipendenza

La teoria della probabilità esposta finora, basata sulla definizione di probabilità uniformi, era costituita essenzialmente di conteggi di cardinalità di insiemi, e, a parte la terminologia, non si discostava molto dalla teoria degli insiemi. C'è però un concetto intuitivo la cui traduzione nell'ambito della teoria le conferisce uno sviluppo automono. Si tratta dell'indipendenza che noi percepiamo tra vari eventi, ad esempio tra i risultati di lanci successivi di una moneta o di un dado (a patto che siano stati opportunamente mescolati tra un lancio e l'altro).

Per capire come inserire tale concetto all'interno della teoria consideriamo un esempio semplice: in due lanci successivi di una moneta il conteggio ci dice che la probabilità di due teste è 1/4, che risulta quindi uguale a 1/2 moltiplicato per 1/2. In altre situazioni che riteniamo indipendenti si verifica la stessa proprietà per cui è naturale porre la definizione seguente. Come al solito, queste erano riflessioni euristiche e da qui comincia la teoria.

DEFINIZIONE 2. *(i) due eventi* $A, B \subseteq S$ *si dicono indipendenti se*

$$P(A \cap B) = P(A)P(B); \tag{3.2}$$

(ii) n *eventi* $A_i \subseteq S$, $i = 1, \ldots, n$ *si dicono (collettivamente) indipendenti se per ogni sottofamiglia* $J \subseteq \{1, \ldots, n\}$ *di indici,*

$$P(\cap_{i \in J} A_i) = \prod_{i \in J} P(A_i).$$

ESEMPIO 9. *Nel lancio di un dado, se* A *è l'evento che esce un pari,* B *l'evento che esce un numero minore o uguale a* 2 *e* c *l'evento che esce un numero minore o uguale a* 3*, allora* $P(A) = 1/2$, $P(B) = 1/3$, $P(C) = 1/2$ *e* $P(A \cap B) = P(A \cap C) = 1/6$ *per cui* A *e* B *sono indipendenti e* A *e* C *non lo sono.*

Il concetto di indipendenza è però molto utile non quando si deve verificare l'indipendenza dalla definizione, come nell'ultimo esempio, ma quando l'indipendenza si deriva da qualche altra informazione e si utilizza la formula (3.2), ossia $P(A \cap B) = P(A)P(B)$, per calcolare $P(A \cap B)$ dagli altri due valori.

Questo alle volte comporta un possibile problema logico: possiamo usare la formula solo se sappiamo che i due eventi sono indipendenti, cosa che però dovremmo verificare dalla formula stessa. In realtà i casi più interessanti sono proprio quelli in cui non sarebbe semplice effettuare tale verifica, per cui in qualche modo si ragiona dicendo: io non so bene come sia fatta P perchè l'insieme S è troppo complicato, ma so che questi eventi sono indipendenti per cui la formula (3.2) deve valere, e conoscendo le probabilità dei due eventi A e B posso calcolare quella di $A \cap B$. La pratica permette di non fare errori applicando questa formula senza verificarla.

ESEMPIO 10. *Due lanci ripetuti di dado sono indipendenti, quindi se* A *indica l'uscita di due* 3 *e* A_i, $i = 1, 2$, *indica l'uscita del* 3 *all'i-simo dado, allora* $A = A_1 \cap A_2$ *e* $P(A) = P(A_1)P(A_2) = \frac{1}{36}$.

Questo caso è ancora facilmente verificabile, ma con più dadi questo diventa un po' più complesso.

ESEMPIO 11. *Anche* n *lanci ripetuti di dado sono collettivamente indipendenti, quindi se* A *indica l'uscita del* 3 *in tutti i dadi e* A_i, $i = 1, \ldots, n$, *indica l'uscita del* 3 *all'i-simo dado, allora* $P(A) = \prod_{i=1}^{n} P(A_i) = \frac{1}{6^n}$.

Vediamo ora che l'indipendenza collettiva di eventi implica l'indipendenza collettiva di parte degli eventi con i complementari degli altri. Questo è un esempio in cui l'indipendenza si deduce e non si verifica dalle definizioni.

TEOREMA 1. *Dati eventi* $A_1, \ldots, A_n \subseteq S$ *collettivamente indipendenti in uno spazio di probabilità* (S, P), *indicando con* $A_i^1 = A$ *e con* $A_i^0 = A^c$, *si ha che per ogni* $\alpha = (\alpha_1, \ldots, \alpha_n) \in \{0, 1\}^n$, *gli eventi* $A_1^{\alpha_1}, \ldots, A_n^{\alpha_n}$ *sono collettivamente indipendenti.*

DIMOSTRAZIONE. Una dimostrazione si può fare per induzione su n.

Per $n = 2$ verifichiamo prima che l'affermazione è vera per $\alpha = (1, 0)$, ossia per $A_1 = A_1^1$ e $A_2^c = A_2^0$. Si ha che, essendo $A_1 \cap A_2 \subseteq A_1$, vale

$$\begin{aligned} P(A_1 \cap A_2^c) &= P(A_1 \backslash A_1 \cap A_2) \\ &= P(A_1) - P(A_1 \cap A_2) \\ &= P(A_1) - P(A_1)P(A_2) \\ &= P(A_1)(1 - P(A_2)) = P(A_1)P(A_2^c) \end{aligned}$$

ove si è usato che $P(A_1 \cap A_2) = P(A_1)P(A_2)$ per la supposta indipendenza.

Per il resto della dimostrazione basta utilizzare più volte quanto ora verificato. Se scambiamo il ruolo di A_1 e A_2 otteniamo la tesi per $\alpha = (0, 1)$, e poi partendo dall'indipendenza di A_1 e A_2^c si ottiene quella per $\alpha = (0, 0)$.

Ora supponiamo che la tesi sia vera fino ad $n-1$ e verifichiamola per n. Nella definizione di indipendenza collettiva si considerano anche i sottinsiemi di indici e se $k \leq n-1$ e $\{i_1, \ldots, i_k\} = J \subseteq \{1, \ldots, n\}$ dalla indipendenza collettiva di $A_{i_1}, \ldots, A_{i_k}$ discende, per l'ipotesi di induzione su n anche quella di $A_{i_1}^{\alpha_{i_1}}, \ldots, A_{i_k}^{\alpha_{i_k}}$ per ogni $(\alpha_{i_1}, \ldots, \alpha_{i_k}) \in \{0, 1\}^k$. Quindi rimane solo da verificare la fattorizzazione per $A_1^{\alpha_1}, \ldots, A_n^{\alpha_n}$. Si può procedere con una seconda induzione sul numero m di zeri di α. Se $m = 0$ la fattorizzazione vale per ipotesi e supponiamo che valga quando ci sono al più $m-1$ zeri. Si consideri ora $\alpha = (\alpha_1, \ldots, \alpha_n)$ con m zeri e supponiamo, per semplicità di notazione e senza perdita di generalità , che $\alpha_n = 0$; si ha che la famiglia $A_1^{\alpha_1}, \ldots, A_{n-1}^{\alpha_{n-1}}, A_n^1$ è tale che il suo vettore α ha solo $m-1$ zeri e quindi per essa vale la fattorizzazione:

$$\begin{aligned} P(\cap_{i=1}^{n-1} A_i^{\alpha_i} \cap A_n^1) &= \prod_{i=1}^{n-1} P(A_i^{\alpha_i}) P(A_n^1) \\ &= P(\cap_{i=1}^{n-1} A_i^{\alpha_i}) P(A_n^1), \end{aligned}$$

ove la seconda uguaglianza è vera per l'ipotesi di induzione su n. Ma allora $\cap_{i=1}^{n-1} A_i^{\alpha_i}$ e A_n^1 sono indipendenti, per cui, per la verifica fatta per

$n = 2$, anche $\cap_{i=1}^{n-1} A_i^{\alpha_i}$ e A_n^0 sono indipendenti. Ne segue che anche per m zeri si ha:

$$\begin{aligned} P(\cap_{i=1}^{n} A_i^{\alpha_i}) &= P(\cap_{i=1}^{n-1} A_i^{\alpha_i} \cap A_n^1) \\ &= P(\cap_{i=1}^{n-1} A_i^{\alpha_i}) P(A_n^1) \\ &= \prod_{i=1}^{n-1} P(A_i^{\alpha_i}) P(A_n^n) \\ &= \prod_{i=1}^{n} P(A_i^{\alpha_i}), \end{aligned}$$

ove la terza uguaglianza segue dall'ipotesi di induzione su n.

□

Da qui non è difficile dedurre che anche per n eventi collettivamente indipendenti tutte le combinazioni di eventi o complementi sono collettivamente indipendenti.

Esempio 12. *Se A indica l'uscita di almeno una testa in 3 lanci di una moneta ed A_i l'uscita di testa all'i-simo lancio, allora gli A_i sono collettivamente indipendenti e $A = \cup_{i=1}^{3} A_i$; ma gli eventi non sono disgiunti e quindi il calcolo si complica. Tuttavia, $A^c = \cap_{i=1}^{3} A_i^c$ e dall'indipendenza collettiva si ha $P(A^c) = \prod_{i=1}^{3} P(A_i^c) = (\frac{1}{2})^3$, così che $P(A) = 1 - 1/8 = 7/8$.*

Facciamo una pausa per riassumere le regole che abbiamo visto per calcolare probabilità. Per calcolare la probabilità di un evento A si può provare a:

(1) contare gli elementi di A e di S;
(2) provare a passare a A^c;
(3) vedere A come unione, ossia $A = \cup_{i=1}^{n} B_i$ per certi B_i, e poi
(3_1) se B_i sono disgiunti si ha $P(A) = \sum_{i=1}^{n} P(B_i)$,
(3_2) se B_i non sono disgiunti ma ci sono solo due eventi (ossia $n = 2$) si ha $P(A) = P(B_1) + P(B_2) - P(B_1 \cap B_2)$, se invece $n > 2$ c'è una formula che non abbiamo descritto qui;
(4) vedere A come intersezione, ossia $A = \cap_{i=1}^{n} B_i$ per certi B_i, e poi
(4_1) se B_i sono collettivamente indipendenti si ha $P(A) = \prod_{i=1}^{n} P(B_i)$,
(4_2) se B_i non sono collettivamente indipendenti non abbiamo ancora una formula adeguata.

Esercizio 6. *In una vicenda che dette origine a molti di questi calcoli, un incallito giocatore, Chevalier De Méré, scrisse nel 1654 a Pascal chiedendogli se l'uscita di almeno un 6 in 4 lanci di dado avesse la stessa probabilità dell'uscita di almeno un doppio 6 in 24 lanci di due dadi. Si racconta che il De Méré avesse perso una grossa somma assumendo che in entrambi i casi si trattasse di 'due probabilità su tre' di vincere. Pascal rispose al De Méré, poi scrisse a sua volta a Fermat* *e questo fu l'inizio della teoria che stiamo presentando. Calcolare le probabilità di questi eventi.*

Con i metodi precedenti si può anche scrivere in forma esplicita la probabilità che escano esattamente k teste in n lanci di una moneta. Scriviamo ora l'espressione in forma ancora più generale.

ESEMPIO 13. *Supponiamo di effettuare n esperimenti indipendenti tali che in ciascuno la probabilità di successo sia un certo valore $p \in [0,1]$, ad esempio potrebbe essere $p = 1/6$ se per successo intendessimo l'uscita del 3 in un dado e così via. Se $\mathcal{P}(k,n,p)$ indica la probabilità di esattamente k successi su n prove indipendenti ognuna con probabilità di successo p ed A_i indica il successo alla i-sima prova si ha:*

$$\begin{aligned}
\mathcal{P}(k,n,p) &= P(\cup_{J\subseteq\{1,\dots,n\},|J|=k}(\cap_{j\in J}A_j \cap \cap_{j\in J^c}A_j^c)) \\
&= \sum_{J\subseteq\{1,\dots,n\},|J|=k} P(\cap_{j\in J}A_j \cap \cap_{j\in J^c}A_j^c) \\
&= \sum_{J\subseteq\{1,dots,n\},|J|=k} \prod_{j\in J} P(A_j) \prod_{j\in J^c} P(A_j^c) \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

Ci si riferisce a queste probabilità chiamandole distribuzione binomiale o di Bernoulli di parametri n e p od usando il simbolo $B(n,p)$.

Nel caso delle monete siamo interessati alla distribuzione di Bernoulli $B(n, 1/2)$. L'espressione ora ottenuta permette di fare un passo formale verso la soluzione del problema 1 posto all'inizio, ma ancora calcolare queste probabilità nel caso di 1000 monete non sarebbe agevole e non sapremmo nemmeno bene con cosa confrontare i valori così ottenuti.

Prima di analizzare meglio questo problema generalizziamo il concetto di probabilità.

3.4. Utilizzo della probabilità

Prima di passare alla generalizzazione facciamo una pausa di riflessione: abbiamo visto alcuni ragionamenti che conducono al calcolo di probabilità e forse è il momento di tentare una prima connessione tra i valori di tali probabilità e la realtà. Non abbiamo ancora abbastanza strumenti per una discussione approfondita, ma possiamo fare alcune osservazioni.

1) E' ragionevole aspettarsi che un evento che ha probabilità 1% si verifichi circa 1 volta su 100 esperimenti indipendenti. Dopo tutto la probabilità che si realizzi esattamente una volta, data dalla distribuzione di Bernoulli $B(100, \frac{1}{100})$, è

$$\mathcal{P}(1, 100, \frac{1}{100}) = 100\frac{1}{100^1}(\frac{99}{100})^{99} \equiv \frac{1}{e} \equiv 37\%$$

che quindi è abbastanza grande per poter procedere.

ESERCIZIO 7. *Se non vi sembra abbastanza grande stimate la probabilità che un evento di probabilità* 1% *si verifichi* $0, 1$ *o* 2 *volte su* 100 *prove indipendenti: interpretando la frase circa 1 volta come l'evento che l'evento originale si realizzi* $0, 1$ *o* 2 *volte vediamo che la probabilità di questo è una percentuale molto alta.*

Analogamente, eventi di probabilità $1/m$ si verificheranno circa 1 volta ogni m prove indipendenti (con alta probabilità).

Alla luce di questo è ragionevole affermare che se individuiamo un evento A *a priori*, ossia prima che sia effettuato l'esperimento che ne verificherà il realizzarsi o meno, e se la probabilità di A è circa $1/m$ con m grande ($m = 50, 100, 1000$), allora sarà sorprendente (tanto più quanto minore è la probabilità) vedere A realizzato già alla prima di queste prove.

2) D'altra parte se abbiamo 100 eventi disgiunti di probabilità ognuno 1% segue dal Lemma 2 che la probabilità dell'unione è 100%, ossia uno di questi accade di sicuro. Questo accade per esempio nella tombola: ogni numero ha probabilità $1/90$, ma in ogni estrazione viene estratto un numero.

Questo si può leggere così : *a posteriori*, ossia dopo che un esperimento è stato realizzato, possiamo selezionare eventi che avevano probabilità (a priori!) piccolissima e che si sono realizzati.

ESERCIZIO 8. *Calcolare la probabilità che 10 lanci successivi di un dado (quindi indipendenti) risultino in una successione data di facce, per esempio* (1234565432).

I calcoli a priori vengono definiti **probabilità** ed in questo ambito siamo riusciti a dare un senso ai nostri calcoli, mentre quelli a posteriori (in cui rientrano i calcoli relativi all'esempio 1 sulle monete) vengono definiti **statistica** e per questo la nostra analisi è ancora insufficiente.

OSSERVAZIONE 1. *Le riflessioni di questo paragrafo sembrano coincidere con una frase di Cicerone, che dice che gettando degli schizzi di colore a caso sul muro sarà molto facile osservare dei tratti che assomiglino ad una faccia, ma sarà assai difficile che questa possa essere quella della Venere di Milo.*

3.5. Stime di Probabilità e statistica

I risultati del capitolo precedente permettono di scrivere formalmente varie probabilità di interesse nel problema delle monete. Per esempio la probabilità di esattamente k teste su 1000 lanci di una moneta è

$$\begin{aligned}\mathcal{P}(k,1000,1/2) &= \binom{1000}{k}(\frac{1}{2})^{1000}\\ &= \frac{1000!}{k!(1000-k)!}.\end{aligned}$$

Con un calcolatore oggi è possibile ottenere una buona approssimazione di questo valore, ma nel '700 era necessario cercarne un'approssimazione analitica; oggi questa puo' comunque essere utile quando il numero di prove sia molto grande (10^7 è fuori dalla portata di programmi come MatLab) oppure per stime rapide quando non si abbia un computer a disposizione.

Il prossimo teorema introduce un'approssimazione analitica del fattoriale, detta di Stirling, esplicitata a meno di una costante che verrà determinata nel seguito.

Teorema 2. *Esiste una costante $c \in \mathbb{R}$ tale che per ogni $n \in \mathbb{N}$ si ha*

$$e^{\frac{1}{12(n+1)}} e^c n^{n+1/2} e^{-n} \le n! \le e^{\frac{1}{12n}} e^c n^{n+1/2} e^{-n}$$

per cui

$$\lim_{n\to\infty} \frac{n!}{e^c n^{n+1/2} e^{-n}} = 1.$$

Dimostrazione. Studiamo

$$d_n = \log(n!) - [(n+1/2)\log(n) - n];$$

si ha

$$d_n - d_{n+1} = (n+1/2)\log(\frac{n+1}{n}) - 1.$$

Da

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots$$

e

$$\log(\frac{1}{1-x}) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots,$$

si ha che

$$\log(\frac{1}{2}\frac{1+x}{1-x}) = x + \frac{x^3}{3} + \frac{x^5}{5} + \dots$$

Per cui, ponendo $\frac{n+1}{n} = \frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}$ si ha

$$\begin{aligned} d_n - d_{n+1} &= (n+1/2)(\frac{2}{2n+1} + \frac{2}{3(2n+1)^3} + \dots) - 1 \quad (3.3)\\ &= \frac{1}{3(2n+1)} + \frac{1}{5(2n+1)^4} + \dots \ge 0.\end{aligned}$$

Dalla (3.3) si ha che

$$\begin{aligned} d_n - d_{n+1} &\le \sum_{k=1} \left(\frac{1}{(2n+1)^2}\right)^k \\ &= \frac{1}{(2n+1)^2} \frac{1}{3(1 - \frac{1}{(2n+1)^2})} \\ &= \frac{1}{3(4n^2+4n)} \\ &= \frac{1}{12n^2+12n} \\ &= \frac{1}{12n} - \frac{1}{12(n+1)}. \end{aligned} \tag{3.4}$$

Pertanto la successione $d_n - \frac{1}{12n}$ è crescente in n. Poichè da (3.3) la successione d_n è decrescente in n, quindi limitata, si ha che $d_n - \frac{1}{12n}$ è anch'essa limitata, ed essendo crescente, ha limite: esiste $c \in \mathbb{R}$ tale che

$$\begin{aligned} \lim_{n\to\infty} d_n - \frac{1}{12n} &= \lim_{n\to\infty} d_n \\ &= \sup_n d_n - \frac{1}{12n} = c; \end{aligned}$$

inoltre, $d_n \le c + \frac{1}{12n}$ per ogni $n \in \mathbb{N}$.

D'altra parte, segue da da (3.3) che

$$\begin{aligned} d_n - d_{n+1} &\ge \frac{1}{3(2n+1)^2} \\ &= \frac{1}{12n^2+12n+3} \\ &\ge \frac{1}{12(n+1)} - \frac{1}{12(n+2)}. \end{aligned}$$

essendo $12(n+1)(n+2) \ge 12n^2 + 12n + 3$. Quindi $d_n - \frac{1}{12(n+1)}$ è decrescente e

$$\begin{aligned} c &= \lim_{n\to\infty} d_n - \frac{1}{12(n+1)} \\ &= \inf_n d_n - \frac{1}{12(n+1)} = \lim_{n\to\infty} d_n. \end{aligned}$$

Per cui $c + \frac{1}{12(n+1)} \le d_n$. Passando agli esponenziali di queste quantità si ottiene la tesi. □

Ora possiamo fare qualche stima più esplicita, eccetto che per la costante c.

ESEMPIO 14. *La probabilità di esattamente $n/2$ teste su n lanci di una moneta, assumendo n pari per semplicità, soddisfa:*

$$\begin{aligned}\mathcal{P}(n/2,n,1/2) &= \frac{n!}{((n/2)!)^2}\frac{1}{2^n}\\ &= \frac{e^c n^{n+1/2}e^{-n}}{e^{2c}(n/2)^n e^{-n}\frac{n}{2}}\frac{1}{2^n} = \frac{2}{\sqrt{n}e^c}\end{aligned}$$

e, tenuto conto degli errori,

$$\mathcal{P}(n/2,n,1/2) = \frac{2}{\sqrt{n}e^c}e^{S(n)}$$

con $|S(n) \leq \frac{1}{3(n+1)}$ *per* $n \geq 4$. *Il lettore è invitato a verificare questa stima.*

Non avendo ancora una stima per $\bar{c} = e^c$ possiamo principalmente concludere per ora che $\mathcal{P}(n/2,n,1/2)$ tende a 0 in n ed è dell'ordine di $1/\sqrt{n}$.

Per p generico si semplificano altrettanto bene le stime quando $k = np$ è intero:

$$\mathcal{P}(np,n,p) = \frac{n!}{(np)!((1-p)n)!}p^{np}(1-p)^{n(1-p)} \approx \frac{1}{\sqrt{n}e^c\sqrt{p(1-p)}}$$

Le stime per altri varori di k non si semplificano altrettanto bene, ma si può almeno confrontare i vari termini con quelli che abbiamo appena stimato, che per motivi che vedremo vengono definiti 'centrali'.

LEMMA 4. *Per ogni $n \in \mathbb{N}$ e per ogni $p \in [0,1]$ $\mathcal{P}(k,n,p)$ è strettamente crescente in k per $k < p(n+1)-1$, è strettamente decrescente per $k > p(n+1)-1$, così che*

$$\arg\max_k \mathcal{P}(k,n,p) = [p(n+1)-1, p(n+1)] \cap \mathbb{N}$$

DIMOSTRAZIONE. Si ha

$$\begin{aligned}\mathcal{P}(k+1,n,p) &= \binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}\\ &= \frac{n-k}{k+1}\frac{p}{1-p}\mathcal{P}(k,n,p)\end{aligned}$$

quindi la monotonia di $\mathcal{P}(k,n,p)$ dipende da $\frac{n-k}{k+1}\frac{p}{1-p}$. Le monotonie discendono dal fatto che $\frac{n-k}{k+1}\frac{p}{1-p} > 1$, e questo a sua volta vale se e solo se $p(n+1)- > k$. In particolare, se $p(n+1) \in \mathbb{N}$, $\mathcal{P}(p(n+1)-1,n,p) = \mathcal{P}(p(n+1),n,p)$ così che $\arg\max_k \mathcal{P}(k,n,p) = \{p(n+1)-1, p(n+1)\}$, come richiesto. □

ESEMPIO 15. *Nel caso delle monete l'andamento di $\mathcal{P}(k,n,1/2)$ in k dipende solo dai coefficienti binomiali;* $\arg\max_k \mathcal{P}(k,n,1/2)$ *dipende dalla parità di* $n = \{(n+1)/-1,(n+1)/2\}$

3.6. Teorema di De Moivre-Laplace

Abbiamo visto che il termine centrale, ossia più probabile, di una distribuzione binomiale relativa ad n prove vale circa $c/\sqrt{n}$. Poichè i termini vicini non saranno molto diversi con un numero di termini dell'ordine di $\sqrt{n}$ si otterrà una probabilità quasi piena. Questa osservazione è specificata molto meglio e resa rigorosa nel seguente teorema, che enunciamo e dimostriamo solo nel caso particolare di $p = 1/2$. Si utlizzerà la funzione gaussiana $e^{-x^2/2}$ definita per $x \in \ mathbbR$.

Teorema 3. **Teorema di De Moivre-Laplace per** $p = 1/2$. *Per ogni* $a_1, a_2 \in \mathbb{R}$, $0 \le a_1 \le a_2$ *si ha che*

$$Q_n(a_1, a_2) = \sum_{n/2+\frac{a_1}{2}\sqrt{n} \le j \le n/2+\frac{a_2}{2}\sqrt{n}} \mathcal{P}(j, n, 1/2)$$

soddisfa

$$\lim_{n\to\infty} Q_n(a_1, a_2) = \int_{a_1}^{a_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Dimostrazione. Per semplicità esponiamo la dimostrazione per n pari; se n è dispari la dimostrazione richiede solo modifiche irrilevanti.

Poniamo $n = 2\nu$ e, per ogni $k = 0, 1, 2, \dots$

$$\begin{aligned}
\alpha_k = \mathcal{P}(\nu + k, 2\nu, 1/2) &= \binom{2\nu}{\nu + k} \frac{1}{2^{2\nu}} \\
&= \frac{(2\nu)!}{(\nu + k)!(\nu - k)!} \frac{1}{2^{2\nu}} \\
&= \frac{(2\nu)!}{(\nu + k - 1)!(\nu - k + 1)!} \frac{1}{2^{2\nu}} \frac{\nu - k + 1}{\nu + k} \\
&= \alpha_{k-1} \frac{\nu - k + 1}{\nu + k} \\
&= \alpha_0 \frac{(\nu - k + 1)(\nu - k + 2) \dots \nu}{(\nu + k)(\nu + k - 1) \dots (\nu + 1)} \\
&= \alpha_0 \frac{(1 - \frac{k-1}{\nu})(1 - \frac{k-2}{\nu}) \dots 1}{(1 + \frac{k}{\nu})(1 + \frac{k-1}{\nu}) \dots (1 + \frac{1}{\nu})}.
\end{aligned}$$

Ricordiamo che

$$\log(1 + x) = x - x^2/2 + x^3/3 + \dots = x + R(x)$$

con

$$|R(x)| \le x^2/2 + x^3/3 + \dots \le 1/2 \sum_{k=2}^{\infty} x^k = \frac{x^2}{2(1 - x)} < x^2$$

per $2(1 - x) > 1$, ossia per $x < 1/2$.

Quindi $1 + x = e^{x+R(x)}$, per cui

$$\begin{aligned} \alpha_k &= \alpha_0 \frac{e^{-(\frac{k-1}{\nu}+\cdots+\frac{1}{\nu})}}{e^{(\frac{k}{\nu}+\cdots+\frac{1}{\nu})}} e^{\bar{R}(k)} \\ &= \alpha_0 e^{-\frac{k^2}{\nu}} e^{\bar{R}(k)} \\ &= \alpha_0 e^{-\frac{2k^2}{n}} e^{\bar{R}(k)}, \end{aligned} \tag{3.5}$$

con $|\bar{R}(k)| \leq 2kR(\frac{k}{\nu}) \leq 2\frac{k^3}{\nu^2} = 2\frac{2^3k^3}{n^2}$. Dalla formula di Stirling

$$\alpha_0 = \frac{2}{\sqrt{n}e^c} e^{S(n)}$$

con $|S(n)| \leq \frac{1}{3(n+1)}$.

Risulta quindi che $\alpha_k \approx c_1 e^{-c_2 k}$, con c_1 e c_2 costanti, così che possiamo stimarlo tramite l'integrale della funzione $c_1 e^{-c_2 x}$. Per far ciò si noti che la funzione $e^{-\frac{x^2}{2}}$ è decrescente per $x \geq 0$. Dalla monotonia discende che per $k \geq 1$:

$$\int_{k-1}^{k} e^{-\frac{x^2}{\nu}} dx \geq e^{-\frac{k^2}{\nu}} \geq \int_{k}^{k+1} e^{-\frac{x^2}{\nu}} dx$$

. Per cui

$$\begin{aligned} Q_n(a_1, a_2) &= \sum_{n/2+\frac{a_1}{2}\sqrt{n} \leq j \leq n/2+\frac{a_2}{2}\sqrt{n}} \mathcal{P}(j, n, 1/2) \\ &= \sum_{\frac{a_1}{2}\sqrt{n} \leq k \leq \frac{a_2}{2}\sqrt{n}} \mathcal{P}(n/2 + k, n, 1/2) \\ &\geq \frac{2}{\sqrt{n}} \int_{\lceil \frac{a_1}{2}\sqrt{n} \rceil}^{\lfloor \frac{a_2}{2}\sqrt{n} \rfloor + 1} e^{-c} e^{-\frac{2x^2}{n}} dx e^{S(n)} e^{-\max|\bar{R}(k)|} \end{aligned}$$

in cui $\lfloor t \rfloor$ indica la parte intera di t, $\lceil t \rceil = \lfloor t \rfloor + 1$, e il massimo è preso in k con $\frac{a_1}{2}\sqrt{n} \leq k \leq \frac{a_2}{2}\sqrt{n}$. Poichè l'integrando è positivo si può stimare restringendo ulteriormente l'intervallo di integrazione a $[a_1\sqrt{n} + 1, a_2\sqrt{n}]$ se $a_1\sqrt{n} + 1 \leq a_2\sqrt{n}$, oppure all'insieme vuoto. Per ottenere la funzione $e^{-y^2/2}$ poniamo $y = \frac{2x}{\sqrt{n}}$ ottenendo

$$\begin{aligned} Q_n(a_1, a_2) &\geq \left(\int_{a_1}^{a_2} e^{-c} e^{-\frac{y^2}{2}} dy - \int_{a_1}^{a_1+2/\sqrt{n}} e^{-c} e^{-\frac{y^2}{2}} dy \right) e^{-\max|\bar{R}(k)|+S(n)} \\ &\geq \left(\int_{a_1}^{a_2} e^{-c} e^{-\frac{y^2}{2}} dy - e^{-a_1^2} e^{-c} \frac{2}{\sqrt{n}} \right) e^{\frac{-a_2^3}{\sqrt{n}} - \frac{1}{3(n+1)}} \end{aligned}$$

essendo

$$\max|\bar{R}(k)| \leq \max_{\frac{a_1}{2}\sqrt{n} \leq k \leq \frac{a_2}{2}\sqrt{n}} \frac{2^3k^3}{n^2} = \frac{-a_2^3}{\sqrt{n}}.$$

Nella corrispondente stima dall'alto occorre separare il termine centrale, che non è dominato dall'integrale se $j = n/2$:

$$\begin{aligned} Q_n(a_1,a_2) &\le \left(\int_{\lfloor a_1\sqrt{n}\rfloor -1}^{\lfloor a_2\sqrt{n}\rfloor} e^{-c}e^{-\frac{x^2}{\nu}}e^{E([x])}dx+1\right)e^{-c}\frac{2}{\sqrt{n}}e^{-\frac{1}{3(n+1)}} \\ &\le \left(e^{\frac{a_2^3}{\sqrt{n}}}\int_{a_1-\frac{2}{\sqrt{n}}}^{a_2} e^{-c}e^{-\frac{y^2}{2}}dy+e^{-c}\frac{2}{\sqrt{n}}\right)e^{-\frac{1}{3(n+1)}}. \end{aligned}$$

Prendendo il limite per n che diverge si ottiene che

$$\lim_{n\to\infty} Q_n(a_1,a_2) = \int_{a_1}^{a_2} e^{-c}e^{-x^2/2}dx. \tag{3.6}$$

Rimane da determinare il valore di e^c. Si verificherà in modo probabilistico nel seguito, utilizzando il risultato (3.6) qui dimostrato, che $e^c = \sqrt{2\pi}$, da cui la tesi.

□

Osservazione 2. *Lo stesso risultato vale per $a_1 < a_2 < 0$ per la simmetria della funzione gaussiana e della distribuzione di Bernoulli per $p = 1/2$. Per cui, per l'additività dell'integrale rispetto al dominio di integrazione, il risultato di De Moivre Laplace vale per qualsiasi $a_1, a_2 \in \mathbb{R}$ con $a_1 \le a_2$. Poichè $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = 1$ segue per simmetria che $\int_0^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-c}e^{-x^2/2}dx = \frac{1}{2} = \sum_{n/2\le j}\mathcal{P}(j,n,1/2) + \frac{c}{\sqrt{n}}$ per qualche $c \in \mathbb{R}$. Quindi*

$$\begin{aligned} \lim_{n\to\infty} Q_n(a,+\infty) &= \lim_{n\to\infty} \sum_{n/2+a/2\sqrt{n}\le j} \mathcal{P}(j,n,1/2) \\ &= \lim_{n\to\infty} 1/2 - \frac{c}{\sqrt{n}} - \sum_{n/2\le j\le n/2+a/2\sqrt{n}} \mathcal{P}(j,n,1/2) \\ &= 1/2 - \int_0^a \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= \int_a^{+\infty} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \end{aligned}$$

Essendo valido questo ragionamento anche per valori negativi, si conclude che il Teorema di De Moivre Laplace vale per qualsiasi a_1, a_2 in $\bar{\mathbb{R}}$ esteso.

Osservazione 3. *Si noti che nella dimostrazione del teorema si è ottenuto non solo il limite richiesto, ma anche delle stime dall'alto e dal basso di $Q_n(a_1,a_2)$. Gli errori sono però solitamente piccoli, dell'ordine di $e^{\frac{a_2^3}{\sqrt{n}}}$ moltiplicativamente e $\frac{2}{\sqrt{2\pi n}}$ additivamente che in pratica è possibile trascurarli senza alterare sensibilmente l'approssimazione di Q. Pertanto abitualmente, occasionalmente senza prestare sufficiente cura, si approssima $Q_n(a_1,a_2)$ con $\int_{a_1}^{a_2} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx$.*

Questa tecnica ci permette quindi di approssimare i valori relativi alla probabilità che il numero di successi in prove indipendenti con probabilità di successo $1/2$ sia compresa in un certo intervallo di valori, che, come ora vedremo, può suggerire una risposta molto precisa al problema iniziale sulle monete.

Come si era visto, la probabilità di ogni singolo risultato è piccola se il numero di prove è grande, e quindi non permette di per sè di valutare la plausibilità di quel numero di teste. Invece la probabilità di un numero di teste minore o uguale ad un dato valore $\bar{k}$ cambia drasticamente, essendo attorno ad $1/2$ per $\bar{k}$ vicino ad $n/2$ e vicino a 0 per $\bar{k}$ molto piccolo. Quindi questa probabilità ci può permettere di distinguere tra i risultati plausibili e quelli che non lo sono.

Esempio 16. *Se si sono ottenute 460 teste su 1000 lanci di una moneta, la probabilità di ottenere un numero di testi minore o uguale a quello ottenuto è stimabile come segue. Se* S_{1000} *è il numero di teste ottenute in 1000 monete, vogliamo stimare*

$$P(S1000 \leq 460) = Q_n(-\infty, a)$$

per un certo a*. La determinazione di* a *va fatta imponendo* $1000/2 + \frac{a}{2}\sqrt{1000} = 460$*, ossia* $a = \frac{460-1000/2}{\sqrt{1000}/2} \approx -2,53$*. Ora si puo' ottenere dalle tavole della funzione gaussiana l'approssimazione*

$$P(S1000 \leq 460) \approx \int_{-\infty}^{-2,53} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0,006 = 0,6\%.$$

Se ne conclude che 460 *è un numero di teste assolutamente inaccettabile su* 1000 *lanci di una moneta e che già per questo numero di teste avremmo dovuto considerare l'esperimento come truccato!*

Osservazione 4. *E' chiaro che per decidere delle nostre reazioni in base ai calcoli precedenti dovremmo fissare delle soglie sotto le quali consideriamo una probabilità troppo piccola e quindi il risultato anomalo. Storicamente, allo scopo di scrivere delle tavole utilizzabili in molte situazioni, si sono considerate i valori del* 5% *come soglia per il dubbio sulla correttezza di un esperimento e dell'*1% *per la ragionevole assunzione che l'esperimento non è corretto*

Esercizio 9. *Verificare che su* 1000 *monete avremmo dovuto avere dubbi (soglia* 5%*) con meno di* 474 *teste e ragionevole certezza dell'anomalia del risultato (soglia* 1%*) con meno di* 463 *teste.*

CAPITOLO 4

PROBABILITÀ SU INSIEMI FINITI

4.1. Spazi di probabilità su insiemi finiti

Nel caso delle probabilità uniformi ci trovavamo di fronte ad alternative tutte equivalenti. Tuttavia negli calcoli successivi abbiamo preso un valore p qualsiasi ed abbiamo ricavato formule (come la distribuzione di Bernoulli) in dipendenza di questo p.

Viene quindi naturale di considerare delle probabilità che non vengano da conteggi di insiemi, ma siano semplicemente dei valori in $[0,1]$ soddisfacenti a certe regole, anche perchè così si possono fare modelli per situazioni in cui non ci sono elementi da contare (tipo la probabilità che un tiro faccia centro o che una misura ecceda di una certa frazione il valore vero).

Rimaniamo comunque per ora su un insieme finito e richiediamo per la probabilità che soddisfi alcune delle proprietà che abbiamo verificato essere vere nel caso uniforme. In particolare, si vede che la proprietà principale da cui derivano tutte le altre è la (2) del lemma 2, ossia quella secondo cui se $A \cap B = \emptyset$ allora $P(A \cup B) = P(A) + P(B)$. Assumiamola quindi come ipotesi:

Definizione 3. *Dato un insieme finito S si dice probabilità (finita) su S ogni funzione P definita sui sottinsiemi di S tale che:*
(1) $P(S) = 1$
(2) per ogni $A \subseteq S$, $P(A) \in [0,1]$;
(3) se $A \cap B = \emptyset$ allora $P(A \cup B) = P(A) + P(B)$.

Si può vedere che tutte le proprietà viste per le probabilità uniformi valgono anche per le probabilità finite. Così in particolare vale la formula della distribuzione di Bernoulli, adesso definita per ogni p anche non razionale.

Esercizio 10. *Un metodo per costruire probabilità finite è quello di partire dall'insieme finito S e da una funzione non negativa f su S. Si ottiene una probabilità ponendo:*
(a) per ogni $s \in S$, $P(S) = P(\{s\}) = \frac{f(s)}{\sum_{s\in S} f(s)}$
(b) per ogni evento A, $P(A) = \sum_{s\in A} P(s)$.
Verificare che P è una probabilità.

Osservazione 5. *Finora abbiamo visto quindi due definizioni: prima la probabilità come* **rapporto tra numero di casi favorevoli e**

possibili, *ed ora una definizione* **assiomatica** *di probabilità. Discuteremo di altre due definizioni nel proseguimento.*

4.2. Probabilità condizionate

In alcune situazioni si vuol calcolare la probabilità di un evento quando si sa che un altro si è già realizzato. Un buon sistema per trattare questa situazione è restringere l'ambito dei possibili risultati all'evento che si sa essere realizzato rendendone 1 la probabilità. Questo risulta naturale se l'evento già realizzato aveva probabilità positiva (ché altrimenti ci si immette in una considerazione forse non priva di senso, ma che richiede certamente un dettagliato chiarimento):

DEFINIZIONE 4. *Dato uno spazio di probabilità* (S, P) *e due eventi* $A, B \subseteq S$, *con* $P(B) \neq 0$, *si dice probablità condizionata di* A *dato* B *il valore*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{4.7}$$

ESEMPIO 17. *Nel lancio di un dado, se* $A = \{3\}$ *e* $B = \{dispari\}$ *(nello stesso lancio),* $P(A|B) = \frac{1/6}{1/2} = \frac{1}{3}$ *è la probabilità che sia uscito il* 3 *sapendo che è uscito un dispari.*

Come già accaduto per l'indipendenza, talvolta è conveniente 'dedurre' il valore della probabilità condizionata dal problema ed usare (4.7) per ricavare uno degli altri termini. Questo pone gli stessi problemi già discussi trattando dell'indipendenza: il valore della probabilità condizionata dovrebbe essere verificata dalla (4.7) prima di usarla. Tale verifica andrebbe fatta per esempio in astratto prima di usare l'espresssione in un caso concreto, ma in molte occasioni la verifica viene omessa e le probabilità condizionate vengono desunte direttamente dal problema. Il procedere senza una verifica teorica a volte crea difficoltà e rischia di condurre ad interpretazioni errate, ma una certa pratica evita comunque gli errori più grossolani.

ESEMPIO 18. *Estraendo senza reinserimento da un'urna contenente* 3 *palline bianche e* 5 *rosse, che indicheremo schematicamente come* $|3B \quad 2R|$, *ed indicando con* B_i *e* R_i *gli eventi che escano una pallina bianca o rossa, rispettivamente, alla i-sima estrazione si ha che* $P(B_1) = 3/8$, $P(R_2|B_1) = 5/7$ *da cui* $P(B_1 \cap R_2) = 18/56$.

Dalla conoscenza di tutte le probabilità condizionate a certi eventi è poi possibile risalire alle probabilità non condizionate degli eventi stessi. A tale scopo diciamo che una famiglia di insiemi $B_1, \ldots, B_n \subseteq S$ è una partizione di S se

(a) $B_i \cap B_j = \emptyset$ se i è diverso da j;
(b) $\cup_{i=1}^{n} B_i = S$.

Vale ora il seguente risultato, talvolta detto teorema delle probabilità totali o delle probabilità composte.

TEOREMA 4. *Se B_i, $i = 1, \ldots, n$, costituiscono una partizione di S con $P(B_i)$ diverso da 0 per ogni i, allora per ogni evento A si ha*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{4.8}$$

DIMOSTRAZIONE. Dalla definizione di probabilità condizionata si ha

$$\begin{aligned} \sum_{i=1}^{n} P(A|B_i)P(B_i) &= \sum_{i=1}^{n} \frac{P(A \cap B_i)}{P(B_i)} P(B_i) \\ &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= P(\cup_{i=1}^{n} A \cap B_i) = P(A) \end{aligned} \tag{4.9}$$

in cui le due ultime uguaglianze seguono, la prima dal fatto che i B_i sono disgiunti e quindi così sono gli $A \cap B_i$, e la seconda dal fatto che $\cup_{i=1}^{n} B_i = S$.

□

ESEMPIO 19. *Nella situazione e con le notazioni dell'ultimo esempio relativo alle estrazioni da $|3B \quad 2R|$ si ha:*

$$P(R_2) = P(R_2|B_1)P(B_1) + P(R_2|R_1)P(R_1) = \frac{5}{7} \cdot \frac{3}{8} + \frac{4}{7} \cdot \frac{5}{8} = \frac{5}{8}.$$

Si noti che $P(R_2) = P(R_1)$, come già in precedenza argomentato.

Questo teorema è molto utile, ma a volte bisogna saper trovare la partizione giusta a cui condizionare:

ESEMPIO 20. *In un gioco ci sono tre porte e dietro ad una sola vi è un premio. Un concorrente sceglie una porta senza aprirla, poi dal conduttore gli viene aperta una porta senza premio ed offerta la possibilità di cambiare la sua scelta (necessariamente nell'unica porta finora non menzionata). E' conveniente cambiare o la cosa è irrilevante in quanto comunque aveva scelto una porta a caso?*

Diciamo di giudicare sulla base della probabilità di trovare il premio, che vogliamo massima, e fissiamo due strategie: cambiare o non cambiare (al momento della scelta). Trovandoci ad aver scelto la prima porta, per calcolare le probabilità di trovare il premio la cosa migliore sembra essere scegliere la seguente partizione: se a è la porta scelta, siano A_a l'evento che il premio è in a ed A_a^c il suo complemento (che necessariamente formano una partizione dell'insieme dei possibili esiti, quali che essi siano). Ora, detto V l'evento che trovo il premio, si ha: $P(V) = P(V|A_a)P(A_a) + P(V|A_a^c)P(A_a^c)$. Naturalmente, visto che *la posizione del premio è casuale (o quantomeno ignota al concorrente)* $P(A_a) = 1/3$. *Nella strategia di cambiare,* $P(V|A_a) = 0$ *e* $P(V|A_a^c) = 1$ *e viceversa nella strategia di non cambiare, per cui quando si cambia* $P(V) = 2/3$ *e non cambiando* $P(V) = 1/3$. *Quindi conviene cambiare.*

Si può usare una combinazione del teorema delle probabilità totali con la definizione di probabilità condizionata per calcolare probabilità condizionate non direttamente evidenti.

ESERCIZIO 11. *Calcolare la probabilità che il secondo numero estratto nella tombola (quindi senza reinserimento) sia il 30 sapendo che il primo estratto è pari.*

4.3. Formula di Bayes

La formula presentata in questa sezione permette di tornare indietro, ossia calcolare la probabilità che avevano certi eventi già realizzati (ma del cui esito non abbiamo informazione).

ESEMPIO 21. *Supponiamo di avere due urne,* $|3B \quad 2R|$ *e* $|2B \quad 5R|$, *e di usare il seguente procedimento di estrazione: scelgo un'urna a caso e poi da questa estraggo una pallina a caso. Se la pallina è rossa, qual è la probabilità che l'urna estratta sia la prima?*

In questa situazione conosciamo le probabilità iniziali di scelta dell'urna e poi le probabilità condizionate di prendere una certa pallina scelta l'urna, ma appunto il problema chiede di tornare indietro.

TEOREMA 5. *Data una partizione Se* B_i, $i = 1, \dots, n$, *con* $P(B_i) > 0$, *ed un evento* A *sottinsieme di un insieme* S, *con* $P(A) \neq 0$, *si ha*

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} \tag{4.10}$$

DIMOSTRAZIONE. Dalla definizione di probabilità condizionata e dalla formula delle probabilità totali si ha

$$\frac{P(A|B_1)P(B_1)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} = \frac{P(A \cap B_1)}{P(A)} = P(B_1|A).$$

□

ESEMPIO 22. *Nell'esempio precedente siano* B_1 *l'evento che viene scelta l'urna* $|3B \quad 2R|$ *e* B_2 *l'evento che viene scelta l'altra urna;* B_1 *e* B_2 *formano una partizione. Se* A *è l'evento che è stata estratta una pallina rossa si risponde al problema determinando* $P(A|B_1)$. *E dalla formula di Bayes:*

$$P(A|B_1) = \frac{\frac{2}{5} \cdot \frac{1}{2}}{\frac{2}{5} \cdot \frac{1}{2} + \frac{5}{7} \cdot \frac{1}{2}} = \frac{14}{39}.$$

4.4. Alcuni calcoli di probabilità in genetica

Torniamo ora al secondo problema posto all'inizio di queste lezioni. La questione deve essere formulata in termini scientifici. La teoria rivelatasi più adeguata è quella genetica (che qui esponiamo in forma molto semplificata) in cui si assume:

1. ogni individuo ha, relativamente al carattere in oggetto, due alleli;
2. tali alleli possono essere o b per il carattere biondo o C per il carattere castano;
3. b è un allele recessivo, C dominante per cui il fenotipo (ossia la forma espressa del carattere da parte dell'individuo portatore di due alleli è sempre castano salvo quando i due alleli sono entrambi b);
4. (legge di Hardy-Weinberg) se p_b è la frequenza di alleli b in una popolazione (ossia il rapporto tra il numero di loci con allele b rispetto al numero totali di loci, quest'ultimo essendo due volte il numero di individui della popolazione) allora la frequenza di alleli C sarà $p_C = 1 - p_b$ e la frequenza di individui con alleli bb sarà $p_{bb} = p_b \cdot p_b$ e di quelli con alleli CC sarà $p_{CC} = p_C \cdot p_C$.

Sulla base di questi elementi, ed osservando che se un individuo è scelto a caso allora la probabilità di un certo evento coincide con la sua frequenza nella popolazione, è possibile determinare molte probabilità relativamente ai genotipi ed ai fenotipi di una popolazione in un dato momento.

Esempio 23. *La probabilità che un individuo sia eterozigote (ossia abbia due alleli diversi) è* $p_{bC} = 1 - p_{bb} - p_{CC} = 2p_b p_C$.

La probabilità che un individuo castano sia eterozigote è data dalla probabilità condizionata che l'individuo sia eterozigote dato che è castano: indicando con $\bar{C}$ *quest'ultimo evento, la probabilità è quindi*

$$P(bC|\bar{C}) = \frac{P(bC \cap \bar{C})}{P(\bar{C})} = \frac{2p_b p_C}{1 - p_b^2} = \frac{2p_b(1 - p_b)}{1 - p_b^2} = \frac{2p_b}{1 + p_b}.$$

Quest'ultima probabilità si può vedere come una applicazione, in un caso molto semplice, del teorema di Bayes in quanto è immediato desumere le probabilità di un certo fenotipo dato il genotipo (che sono peraltro 0 *o* 1*) e qui si richiede la probabilità condizionata inversa (esercizio).*

Tuttavia nel problema 2 dobbiamo anche prendere in considerazione la riproduzione in quanto siamo interessati a questioni di discendenza. Un'ipotesi ragionevole relativamente alla riproduzione è che

5. ogni figlio viene generato scegliendo due genitori scelti a caso (eventualmente lo stesso!) ed il suo genotipo è generato scegliendo a caso un allele da ogni genitore.

Questa ipotesi non è molto realistica per le popolazioni umane, ma sembra più ragionevole per animali inferiori e poi è molto semplice. Prima di procedere osserviamo che c'è però un problema di coerenza tra quest'ultima ipotesi e le precedenti (questo problema fu portato verso il 1920 al matematico Hardy dal biologo Punnet e dette in seguito luogo alla legge che porta il nome di Hardy): se ad una certa generazione valgono le ipotesi 1-4 e la riproduzione segue l'ipotesi 5 si manterranno le condizioni 1-4 anche alla generazione successiva?

Per questa verifica si usa il teorema delle probabilità totali. Se indichiamo con $P_{\alpha\beta}$, $M_{\alpha\beta}$ e $F_{\alpha\beta}$, gli eventi che il padre, la madre o il figlio rispettivamente, hanno genotipo $\alpha\beta$ si ha:

$$\begin{aligned} P(F_{bb}) &= \sum_{\alpha,\beta,\alpha',\beta'=b,C} P(F_{bb}|P_{\alpha,\beta}\cap M_{\alpha',\beta'})P(P_{\alpha,\beta}\cap M_{\alpha',\beta'}) \\ &= P(P_{b,b}\cap M_{b,b}) + \frac{1}{2}P(P_{b,C}\cap M_{b,b}) \\ &\qquad +\frac{1}{2}P(P_{b,b}\cap M_{b,C}) + \frac{1}{4}P(P_{b,C}\cap M_{b,C}) \\ &= p_b^4 + 2\frac{1}{2}2p_b^3(1-p_b) + \frac{1}{4}4p_b^2(1-p_b)^2 \\ &= p_b^2 = P(P_{bb}). \end{aligned} \tag{4.11}$$

La stessa cosa si può verificare per gli altri genotipi, per cui le condizioni 1-4 risultano in effetti stabili rispetto alla riproduzione modellizzata da 5.

Siamo ora in grado di fornire una risposta al problema 2. Se indichiamo con $F_{\bar{b}}$ ($F_{\bar{C}}$) l'evento che il figlio è biondo (o castano) ed analogamente denotiamo gli eventi che il padre o la madre hanno un certo fenotipo, una soluzione al problema viene data dal calcolo di

$$P(F_{\bar{b}}|P_{\bar{C}}\cap M_{\bar{C}}) = P(F_{bb}|P_{\bar{C}}\cap M_{\bar{C}}).$$

Dalla definizione di probabilità condizionate si ha:

$$\begin{aligned} P(F_{bb}|P_{\bar{C}}\cap M_{\bar{C}}) &= \frac{P(F_{bb}\cap P_{\bar{C}}\cap M_{\bar{C}})}{P(P_{\bar{C}}\cap M_{\bar{C}})} \\ &= \frac{\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb}\cap P_{\alpha,\beta}\cap M_{\alpha',\beta'})}{\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(P_{\alpha,\beta}\cap M_{\alpha',\beta'})}. \end{aligned}$$

L'indipendenza nella scelta dei genitori implica che

$$\sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(P_{\alpha,\beta}\cap M_{\alpha',\beta'}) = (1-p_b^2)^2.$$

Per il calcolo del numeratore possiamo utilizzare di nuovo la definizione di probabilità condizionata, osservando che solo una delle probabilità condizionate è diversa da 0:

$$
\begin{aligned}
& \sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb} \cap P_{\alpha,\beta} \cap M_{\alpha',\beta'}) \\
= & \sum_{\alpha,\beta,\alpha',\beta'=b,C \text{ ma non uguali a bb}} P(F_{bb}|P_{\alpha,\beta} \cap M_{\alpha',\beta'})P(P_{\alpha,\beta} \cap M_{\alpha',\beta'}) \\
= & \frac{1}{4}4(1-p_b)^2 p_b^2.
\end{aligned}
$$

Per cui si ha

$$
P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}) = \frac{(1-p_b)^2 p_b^2}{(1-p_b^2)^2} = \frac{p_b^2}{(1+p_b)^2}.
$$

Ora rimane solo da determinare p_b. In effetti, sappiamo che la frequanza di (fenotipi) biondi è il 30% e che la probabilità di un genotipo puro bb è p_b^2. Assumendo che questi due valori siano uguali si ha $p_b = \sqrt{0.3}$.

Per cui $P(F_{bb}|P_{\bar{C}} \cap M_{\bar{C}}) = \frac{0,3}{(1+\sqrt{0.3})^2} \simeq 0,125$.

Anche se era ovvio che la probabilità di un figlio biondo dovesse essere minore per due genitori castani rispetto a due qualunque genitori, quantificare tale riduzione è un risultato per nulla immediato.