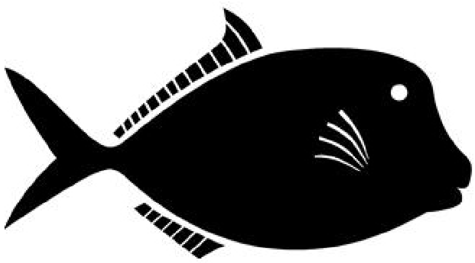


CHI DORME NON PIGLIA PESCI.
UN PERCORSO DIDATTICO SUL TEOREMA DI BAYES

Michele Impedovo

Riassunto Quanto segue è una personale proposta didattica di avvicinamento allo strumento più importante per aggiornare le stime di probabilità via via che si ottengono nuove informazioni: il teorema di Bayes. In particolare il lavoro cerca di mettere in luce l'effettiva potenza del teorema di Bayes, applicato nella sua forma più semplice ed espressiva: la probabilità di un evento condizionato è proporzionale alla sua verosimiglianza.

Introduzione La probabilità è molto di più di un argomento nel curriculum: è uno strumento di cultura, è un modo di pensare, di guardare il mondo, di capire le sue leggi, che ci sono rimaste celate per secoli semplicemente perché le guardavamo con ingenui occhi deterministici, perché non coglievamo l'aspetto aleatorio, casuale della natura. Se guardiamo per esempio alla teoria cinetica dei gas, dobbiamo accettare che i miracoli possano esistere (può accadere che tutte le molecole si addensino in una piccola porzione di spazio): è solo che hanno probabilità zero di accadere. Un pezzo di legno può sprigionare una forza magnetica; è sufficiente che una parte significativa delle orbite dei suoi elettroni si orientino più o meno nella stessa direzione: ma questo ha probabilità zero di accadere.

Michele Impedovo Università Bocconi di Milano michele.impedovo@unibocconi.it

Forse proprio noi insegnanti di matematica abbiamo una qualche avversione per il pensiero aleatorio? Pretendiamo di essere precisi e rigorosi su minuzie e casi particolarissimi, su eccezioni eccezionali, e ci rifiutiamo di esprimere valutazioni probabilistiche, ci limitiamo ai due estremi, vero o falso.

Per esempio, se un numero naturale (maggiore di 2) è primo allora è dispari. D'accordo. E se è dispari, qual è la probabilità che sia primo?

Qual è la probabilità che un'equazione di secondo grado a coefficienti interi abbia soluzioni razionali?

Quali sono le probabilità che un sistema lineare sia determinato, o indeterminato, o impossibile?

L'orbita di una cometa catturata dalla forza gravitazionale del Sole è necessariamente una conica. D'accordo, ma è una parabola con probabilità 0.

Chi dorme non piglia pesci. D'accordo. Ma se non ho pigliato pesci ho dormito?

LA PROBABILITÀ È SEMPRE CONDIZIONATA

Secondo gli assiomi di Kolmogorov la probabilità di un evento A è una *misura* dell'insieme A *normalizzata* rispetto a un insieme-universo Ω[1], che rappresenta l'evento certo e ha probabilità 1. L'universo Ω circoscrive le informazioni in nostro possesso: uno e uno solo dei risultati $\omega \in \Omega$ si verificherà, dunque non ci interessa (lo riteniamo trascurabile) alcun risultato che stia al di fuori di Ω.

Anche nel caso semplice del solito dado, quando diciamo che Ω è l'insieme

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

affermiamo che non riteniamo degni di osservazione altri casi quali: il dado rimane in equilibrio su un vertice, o su uno spigolo, oppure cade nel tombino, oppure si distrugge (era di materiale fragilissimo), oppure si deforma irreparabilmente (era di pongo), oppure viene mangiato al volo dal

[1] Chiamato sovente *spazio campionario*: è un insieme, non uno spazio, e non è un campione (contiene tutti i possibili risultati di un esperimento aleatorio). Su alcuni testi Ω viene chiamato *insieme dei risultati*, anche se “risultato” evoca ciò che è già accaduto.

mio gatto (era una crocchetta di salmone), …; ipotizzare $\Omega = \{1,2,3,4,5,6\}$ è dunque il risultato di diverse ipotesi (o, se si preferisce, di stratificazioni successive del contratto didattico) ormai tacitamente ammesse. Quando fissiamo Ω stiamo già usando l'informazione (o l'ipotesi) che altri risultati non si verificheranno, che li riteniamo al di fuori dei risultati possibili.

Questo accade per qualsiasi esperimento aleatorio.

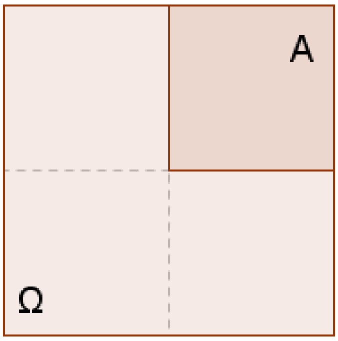
La probabilità Pr(A) di un evento $A \subseteq \Omega$ è perciò sempre condizionata: in prima battuta è condizionata dalla scelta di Ω, che rappresenta in qualche modo le nostre conoscenze preliminari, dalle quali ricaviamo ipotesi (che consideriamo irrinunciabili e stabili) sui possibili risultati dell'esperimento. Quando scriviamo

$$\Pr(A)$$

sottintendiamo in realtà la probabilità di A *subordinata* a Ω, oppure *condizionata* da Ω: dovremmo scrivere

$$\Pr(A|\Omega)$$

ESEMPIO. Utilizziamo, come paradigma di misura di probabilità, l'*area*. È questo un esempio molto significativo di rappresentazione semiotica della probabilità[2]; supponiamo che Ω sia una figura piana, per semplicità un quadrato, e che A sia un suo sottoinsieme; per esempio:



[2] La probabilità è in definitiva la misura di un insieme A normalizzata rispetto ad un insieme Ω: "vedere" la probabilità come misura (lunghezza, area, volume, …) dovrebbe affiancare almeno con pari dignità il conteggio del numero di elementi di A rispetto a quelli di Ω nel cosiddetto approccio "classico".

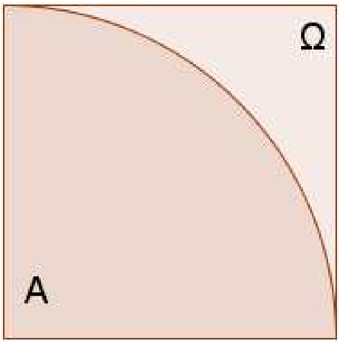
Immaginiamo ora di scegliere un punto a caso in Ω. È ragionevole ipotizzare che la probabilità che tale punto appartenga ad A sia proporzionale all'area di A, anzi, che sia proprio

$$\Pr(A) = \frac{\text{area}(A)}{\text{area}(\Omega)}$$

In questo caso Pr(A) = 1/4 = 0.25 = 25%.

Naturalmente si potrebbero usare in modo analogo le lunghezze o i volumi, tuttavia nell'esperienza personale mi è sembrato che l'area dia un contributo più solido alla costruzione del concetto, forse perché le porzioni di piano sono più facilmente visualizzabili rispetto a porzioni di linee o di volume. In fondo l’area “vive” naturalmente sul piano del foglio o della lavagna e anche sullo schermo di un computer.

Tra l'altro, un esempio ben noto di utilizzo dell'area come misura di probabilità è quello in cui A è il quarto di cerchio inscritto nel quadrato.



In questo caso

$$\Pr(A) = \pi/4 \approx 78.5\%$$

ed è facile simulare l'evento A con un foglio elettronico. Si scelgono, mediante il comando

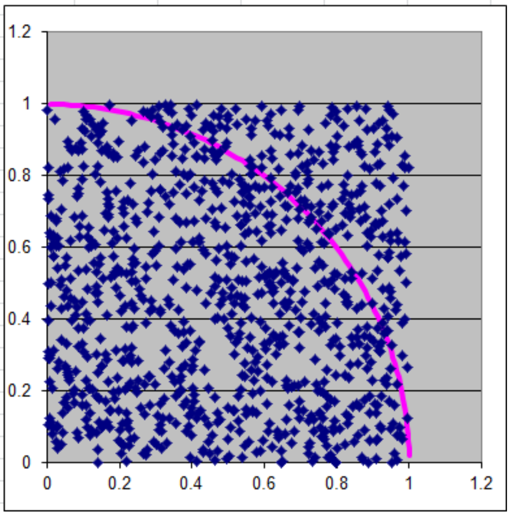
=CASUALE()

le coordinate (x, y) di un punto a caso nel quadrato di vertici opposti (0,0) e (1,1) e si stabilisce se appartiene ad A controllando se risulta

$$x^2 + y^2 < 1$$

Si ripete l'esperimento con un numero "grande" N di punti e si conta quanti sono, diciamo F, quelli che stanno in A. La frequenza relativa di successo F/N può essere considerata una stima della probabilità Pr(A). Nella figura seguente abbiamo ottenuto, con N = 1000, F = 792.

x	y	x^2+y^2	dentro	frequenza	pigreco
0.790679	0.086642	0.63268	1	792	3.168
0.68684	0.808373	1.125217	0		
0.29025	0.803535	0.729913	1		
0.305334	0.963557	1.021671	0		
0.317169	0.087336	0.108224	1		
0.017086	0.575453	0.331438	1		
0.763931	0.065036	0.58782	1		
0.841421	0.697829	1.194955	0		
0.461084	0.865394	0.961505	1		
0.670241	0.933783	1.321174	0		
0.661331	0.231392	0.490901	1		
0.356714	0.187088	0.162247	1		
0.559346	0.827478	0.997587	1		
0.626459	0.478485	0.621399	1		
0.822453	0.514641	0.941284	1		
0.610571	0.089316	0.380774	1		
0.956006	0.109928	0.926031	1		
0.328615	0.999314	1.106616	0		
0.01928	0.314483	0.099271	1		
0.755311	0.721782	1.091463	0		
0.142174	0.252893	0.084168	1		
0.823818	0.694957	1.161641	0		
0.73007	0.130778	0.550106	1		
0.962603	0.426627	1.108615	0		



Possiamo utilizzare la frequenza relativa

$$F/N = 0.792$$

per stimare Pr(A) e dunque stimare π:

$$\Pr(A) = \pi/4 \approx F/N = 0.792$$

da cui

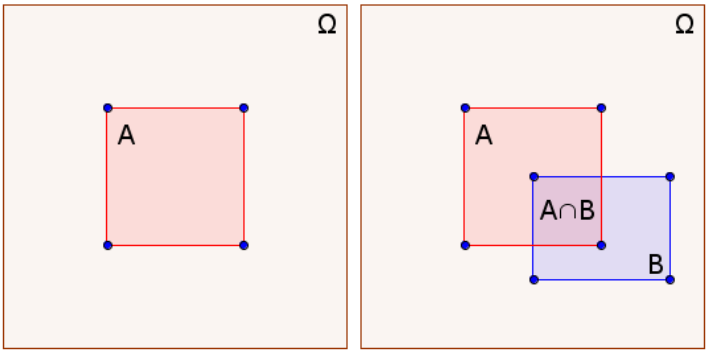
$$\pi \approx 3.168.$$

Supponiamo ora che le nostre informazioni si rafforzino: veniamo a sapere (o ipotizziamo) che i risultati possibili dell'esperimento stanno in un sottoinsieme (non vuoto) di Ω, diciamo B. Allora B è il nostro nuovo universo, non ci interessa nessun risultato che stia al di fuori di B; di conseguenza la probabilità dell'evento A, che valutavamo prima rispetto a Ω, andrà aggiornata rispetto a B.

Poiché alla luce delle nostre nuove informazioni non ci importa nulla di ciò che sta al di fuori di B, ciò che interessa di A è solo la parte che sta in

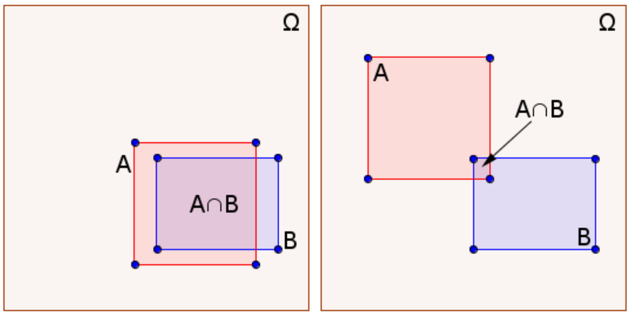
B, cioè A∩B. Da qui la definizione di *probabilità di* A *condizionata* (o *subordinata*) *a* B:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

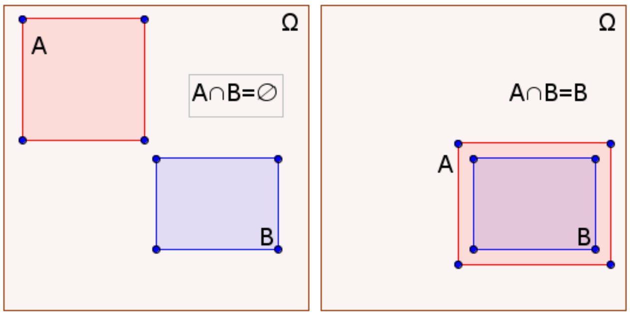


Si immagini, data l'informazione B, di *muovere* A: cambia l'area di A∩B e dunque cambia Pr(A|B), come mostrano le figure seguenti.

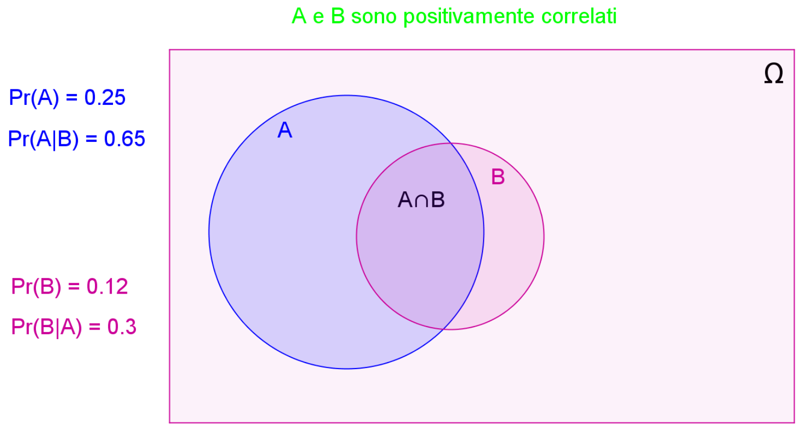
Pr(A|B) "grande" Pr(A|B) "piccola"



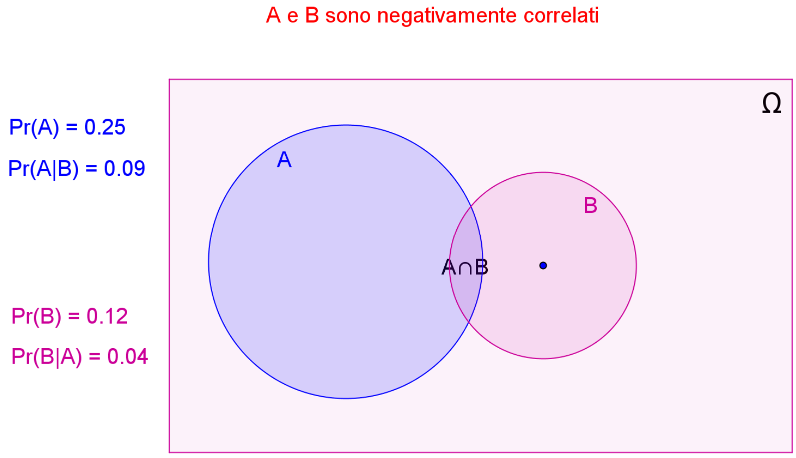
Pr(A|B) = 0 Pr(A|B) = 1



Se Pr(A|B)>Pr(A), il che accade se e solo se Pr(B|A)>Pr(B),[3] si dice che A e B sono *positivamente correlati*. La figura seguente mostra un esempio.



Se Pr(A|B) < Pr(A) (il che accade se e solo se Pr(B|A) < Pr(B)) si dice che A e B sono *negativamente correlati*.



ESEMPIO. Utilizziamo il comando

=CASUALE()

---

[3] Risulta Pr(A|B)=Pr(A) se e solo se Pr(B|A)=Pr(B). Infatti Pr(A|B)=Pr(A) ⇔ Pr(A∩B)/Pr(B)=Pr(A) ⇔ Pr(B∩A)/Pr(A)=Pr(B) ⇔ Pr(B|A)=Pr(B). Nello stesso modo si ottiene che Pr(A|B)>Pr(A) se e solo se Pr(B|A)>Pr(B) e Pr(A|B)<Pr(A) se e solo se Pr(B|A)<Pr(B).

di Excel, che simula la generazione di un numero reale casuale uniformemente distribuito (cioè senza alcuna preferenza) in $\Omega = [0,1]$; possiamo ragionevolmente assumere che la probabilità che il punto scelto appartenga a un intervallo $[a, b] \subseteq \Omega$ sia

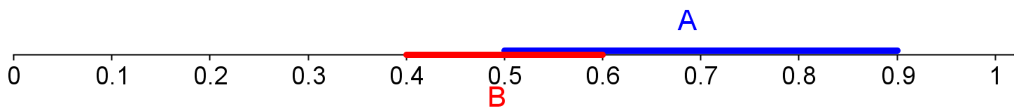
$$\Pr([a, \quad b]) = b - a$$

Consideriamo per esempio $A = [0.5, 0.9]$. Risulta

$$\Pr(A) = 0.9 - 0.5 = 0.4 = 40\%$$

Supponiamo ora di venire in possesso della seguente informazione: il numero scelto appartiene all'intervallo

$$B = [0.4, 0.6]$$



B, la cui probabilità è 0.2, diventa il nuovo Ω. Cambia la nostra stima di probabilità di A? Cioè: l'informazione B modifica la valutazione di probabilità di A?

Se supponiamo che B sia vero, allora dobbiamo ignorare tutto ciò che è fuori di B: di A ci resta solo $A \cap B$, cioè l'intervallo

$$A \cap B = [0.5, 0.6]$$

Se indichiamo con $\Pr(A|B)$ la "nuova" probabilità di A sotto l'ipotesi che B sia vero, allora

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{0.1}{0.2} = 50\%.$$

Come si vede la probabilità di A è passata dal valore iniziale $\Pr(A) = 0.4$ (prima dell'informazione B) al valore finale $\Pr(A|B) = 0.5$ (dopo l'informazione B): A e B sono positivamente correlati.

La stessa cosa accade se rovesciamo il condizionamento:

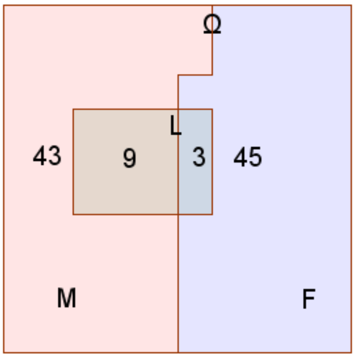
$$\Pr(B|A) = \frac{\Pr(B \cap A)}{\Pr(A)} = \frac{0.1}{0.4} = 0.25 > \Pr(B) = 0.2.$$

PROBABILITÀ CONDIZIONATA: TABELLE E GRAFI

Consideriamo il seguente problema.

In una popolazione di 100 persone (Ω), 12 sono mancini (L): se prendo un individuo a caso in Ω, la probabilità che sia mancino è dunque $Pr(L) = Pr(L|\Omega) = 12\%$. Ora vengo a sapere che l'individuo scelto è maschio (M): cambia la mia valutazione di probabilità? In altri termini, quanto vale $Pr(L|M)$?

Bisognerebbe sapere come si distribuisce il mancinismo tra maschi e femmine. Ecco la statistica completa, mostrata attraverso un diagramma di Venn che permette di "vedere" i 4 sottoinsiemi (e le corrispondenti frequenze assolute[4]) in cui si divide la "popolazione" Ω quando se ne considerano due caratteristiche: $M \cap L$ (sono 9), $M \cap R$ (43), $F \cap L$ (3), $F \cap R$ (45)[5].



La nuova popolazione-universo è M (52 unità): i mancini sono 9 su 52, dunque la nuova informazione porta a rivalutare la probabilità che l'individuo estratto sia mancino, da 12% a $9/52 \approx 17.3\%$:

$$Pr(L) = 12\% \rightarrow Pr(L|M) = 17.3\%$$

La nuova informazione fa aumentare la stima di probabilità dell'evento L: l'informazione "essere maschio" è *positivamente correlata* con "essere mancino".

---

[4] Oppure le frequenze relative: si arriva alle stesse conclusioni quasi con lo stesso processo cognitivo.

[5] Naturalmente non è necessario usare 4 simboli differenti: è sufficiente indicare R come il complementare L^C di L e M come il complementare (!) di F.

Risulta invece

$$Pr(L|F) = 3/48 = 6.25\% < Pr(L)$$

L'informazione "essere femmina" è *negativamente correlata* con "essere mancino"[6].

Le stesse informazioni possono essere sintetizzate mediante una tabella a doppia entrata, che riporti in ogni cella la probabilità che una persona a caso appartenga all'intersezione delle rispettive popolazioni.

	M	F
L	0.09	0.03
R	0.43	0.45

Dunque:

$$Pr(L \cap M) = 9\% \qquad Pr(L \cap F) = 3\%$$
$$Pr(R \cap M) = 43\% \qquad Pr(R \cap F) = 45\%$$

Se sommiamo per righe e per colonne otteniamo le cosiddette *distribuzioni marginali* dei due caratteri, cioè le probabilità dei quattro eventi L, R, M, F.

	M	F	**tot.**
L	0.09	0.03	**0.12**
R	0.43	0.45	**0.88**
tot.	**0.52**	**0.48**	**1**

Dunque:

$$Pr(M) = 52\% \qquad Pr(L) = 12\%$$

$$Pr(F) = 48\% \qquad Pr(R) = 88\%$$

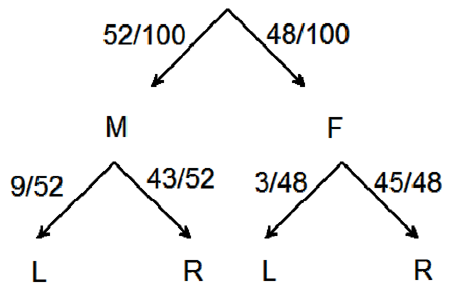
In questo contesto la probabilità condizionata è interpretata come rapporto di probabilità, proprio come vuole la definizione:

$$Pr(L|M) = Pr(L \cap M)/Pr(M) = 0.09/0.52 \approx 17.3\%$$

$$Pr(L|F) = Pr(L \cap F)/Pr(F) = 0.03/0.48 \approx 6.3\%$$

6 Si può dimostrare in generale che $Pr(A|B)>Pr(A)$ se e solo se $Pr(A|B^C)<Pr(A)$.

Un altro modo per visualizzare e analizzare il problema è quello di utilizzare un *grafo*.



Il grafo mostra che un individuo scelto a caso nella popolazione iniziale è maschio con probabilità

$$\Pr(M) \quad = \quad 0.52$$

Se ora scegliamo a caso un individuo nella popolazione M (che costituisce, come sottoinsieme di Ω, la nuova informazione), questo sarà mancino con probabilità 9/52 (e "non mancino" con probabilità 43/52). Dunque le probabilità sulla seconda riga del grafo sono le <u>probabilità condizionate</u>

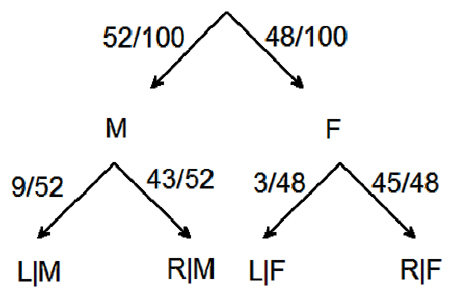
$$\Pr(L|M) = 9/52$$

$$\Pr(R|M) = 43/52$$

$$\Pr(L|F) = 3/48$$

$$\Pr(R|F) = 45/48$$

Il grafo dovrebbe essere più correttamente rappresentato nel seguente modo.



Qual è la probabilità che una persona a caso sia maschio e mancino?

Ricordiamoci che la definizione di probabilità condizionata mette in relazione il condizionamento con l'intersezione:

$$\Pr(L|M) = \frac{\Pr(L \cap M)}{\Pr(M)} \quad \Leftrightarrow \quad \Pr(L\cap M) = \Pr(L|M)\Pr(M)$$

Questo significa che dobbiamo moltiplicare le probabilità (condizionate) lungo i rami per ottenere le probabilità delle intersezioni. Nel nostro esempio risulta

$$\Pr(L|M)\Pr(M) = \Pr(L\cap M) = \frac{9}{52}\cdot\frac{52}{100} = \frac{9}{100}$$
$$\Pr(R|M)\Pr(M) = \Pr(R\cap M) = \frac{43}{52}\cdot\frac{52}{100} = \frac{43}{100}$$
$$\Pr(L|F)\Pr(F) = \Pr(L\cap F) = \frac{3}{48}\cdot\frac{48}{100} = \frac{3}{100}$$
$$\Pr(R|F)\Pr(F) = \Pr(R\cap F) = \frac{45}{48}\cdot\frac{48}{100} = \frac{45}{100}$$

Possiamo ora completare il grafo.

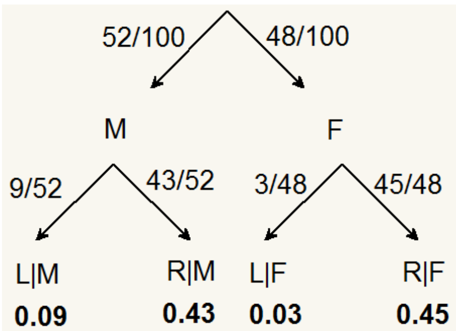


Tabella e grafo veicolano dunque informazioni equivalenti ma forniscono immagini diverse (e complementari) del problema, mettendo in evidenza rispettivamente le probabilità dell'intersezione e le probabilità condizionate.

INDIPENDENZA DI EVENTI

Può accadere che la probabilità di A, dopo aver acquisito l'informazione B, rimanga invariata. Quest'ultimo caso è particolarmente importante perché segnala il fatto che l'informazione B non muta la stima sulla probabilità di A; risulta cioè

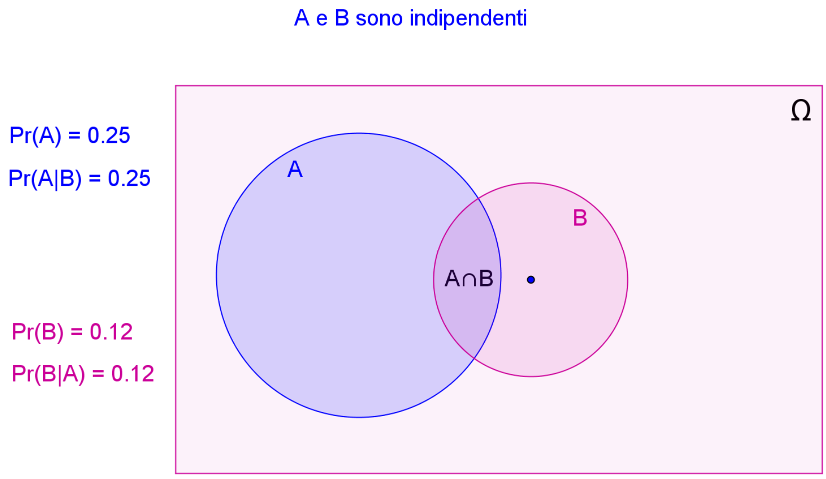
$$\Pr(A|B) = \Pr(A)$$

In questo caso si dice che A è *indipendente da* B.

Poiché si dimostra facilmente che

$$\Pr(A|B) = \Pr(A) \text{ se e solo se } \Pr(B|A) = P(B)$$

si dice semplicemente che A e B sono *indipendenti*.



È straordinario essere riusciti a catturare razionalmente il concetto di indipendenza tra due eventi, che può essere assai sfuggente. Per esempio è intuitivo ritenere che nel lancio di due dadi il risultato D_2 sul secondo dado sia indipendente dal risultato D_1 sul primo:

$$\Pr(D_2 = a|D_1 = b) = \Pr(D_2 = a)$$

Ma spesso l'intuizione non è sufficiente; sempre restando in tema di dadi, considerate gli esempi seguenti. Lanciando due dadi regolari, la probabilità che la somma D_1+D_2 delle facce sia 7 è 6/36 = 1/6 (1+6, 2+5, 3+4, 4+3, 5+2, 6+1): se il primo dado si ferma sul 4, cambia la mia stima di probabilità? Se $D_1 = 4$ allora necessariamente D_2 deve dare 3:

$$\Pr(D_1+D_2 = 7|D_1 = 4) = 1/6$$

Dunque gli eventi $D_1+D_2=7$ e $D_1=4$ sono indipendenti.

1. Analizziamo ora il caso $D_1+D_2 = 6$: la probabilità che la somma delle facce sia 6 è 5/36 (1+5, 2+4, 3+3, 4+1, 5+1); se il primo dado si è fermato sul 4 allora il secondo deve necessariamente essere 2:

   $$\Pr(D_1+D_2 = 6|D_1 = 4) = 1/6$$

   Dunque gli eventi $D_1+D_2 = 6$ e $D_1 = 4$ non sono indipendenti, ma sono (debolmente) correlati positivamente: se so che il primo dado si è fermato sul 4, il mio grado di fiducia nel fatto che la somma sia 6 aumenta, passando da 5/36 a 1/6.

2. Analizziamo ora il caso $D_1+D_2 = 2$: la probabilità che la somma delle facce sia 2 è 1/36 (solo 1+1); se il primo dado si è fermato sul 4 allora la probabilità che la somma sia 2 è nulla:

   $$\Pr(D_1+D_2 = 2|D_1 = 4) = 0$$

   Dunque gli eventi $D_1+D_2 = 2$ e $D_1 = 4$ non sono indipendenti, ma sono (fortemente) correlati negativamente.

Una definizione alternativa di indipendenza, che viene presentata su molti libri di testo, è la seguente: gli eventi A e B sono indipendenti se

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

Questa definizione è logicamente equivalente a quella precedente, ma non cognitivamente equivalente. Ha infatti il pregio (a differenza della prima, che sfrutta la probabilità condizionata) di essere simmetrica rispetto ai due eventi, ma ha il difetto di essere meno ricca dal punto di vista semantico. In un percorso didattico sceglierei senz'altro la prima definizione e mostrerei la seconda come conseguenza. Infatti da

$$\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)}$$

ricaviamo (se $\Pr(B) \neq 0$) la relazione equivalente

$$\Pr(A\cap B) = \Pr(A|B)\Pr(B)$$

e questa uguaglianza si riduce a

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

nel caso in cui sia $\Pr(A|B) = \Pr(A)$, cioè A e B siano indipendenti.

Insomma, mi sembra molto più efficace dal punto di vista didattico immaginare l'indipendenza tra A e B in questo modo (che suona ricco di suggestioni):

"sapere che B è vero non muta la stima di probabilità di A"

piuttosto che in questo modo (che suona dimessamente sintattico):

"la probabilità dell'intersezione è uguale al prodotto delle probabilità".

INDIPENDENZA: TABELLE E GRAFI

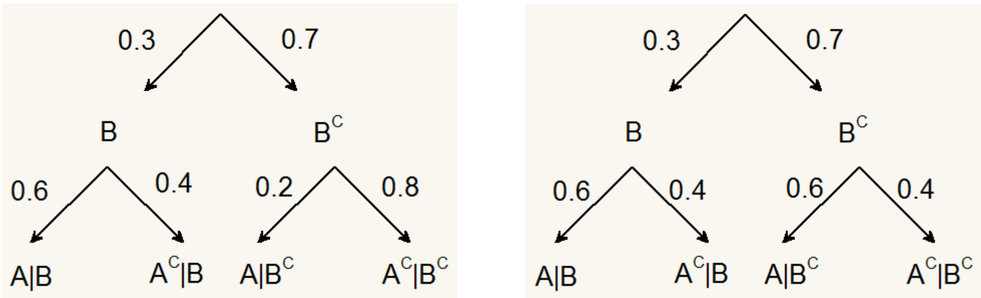
Per quanto abbiamo appena detto, in una tabella è immediato riconoscere eventi indipendenti: ogni cella deve contenere il prodotto delle corrispondenti probabilità marginali:

	B	non B	**tot.**
A	ab	a(1-b)	**a**
non A	(1-a)b	(1-a)(1-b)	**1-a**
tot.	**b**	**1-b**	**1**

Nell'esempio precedente, ferme restando le distribuzioni marginali, mancinismo e sesso sarebbero indipendenti se e solo se la tabella fosse la seguente:

	M	F	**tot.**
L	0.0624	0.0576	**0.12**
R	0.4576	0.4224	**0.88**
tot.	**0.52**	**0.48**	**1**

In un grafo invece non è immediato riconoscere eventi indipendenti. Consideriamo ad esempio i due grafi seguenti.



Nel primo A e B non sono indipendenti: infatti

$$\Pr(A|B) = 0.6$$

mentre

$$\begin{aligned}\Pr(A) &= \Pr((A\cap B)\cup(A\cap B^C))\\ &= \Pr(A\cap B)+\Pr(A\cap B^C)\\ &= \Pr(A|B)\Pr(B)+\Pr(A|B^C)\Pr(B^C)\\ &= 0.6\cdot 0.3+0.2\cdot 0.7\\ &= 0.18+0.14 = 0.32\end{aligned}$$

Nel secondo invece A e B sono indipendenti: infatti

$$\Pr(A|B) = 0.6$$

e

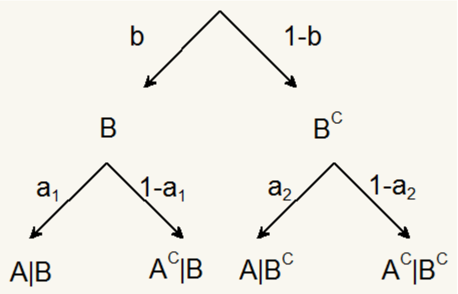
$$\begin{aligned}\Pr(A) &= 0.6\cdot 0.3+0.6\cdot 0.7\\ &= 0.18+0.42 = 0.6\end{aligned}$$

In generale si dimostra che A e B sono indipendenti se e solo se

$$\Pr(A|B) = \Pr(A|B^C)$$

il che è del tutto sensato: se la probabilità di A non è influenzata dal verificarsi di B allora non è influenzata nemmeno dal non verificarsi di B.

Dunque in un grafo del tipo seguente



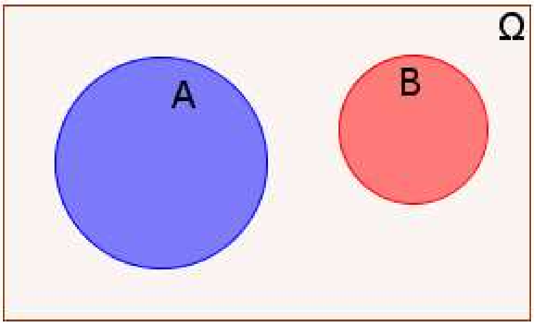
A e B sono indipendenti se e solo se

$$a_1 = a_2$$

il che mette in luce che se A e B sono indipendenti lo sono anche A e B^C, A^C e B, A^C e B^C.

EVENTI INDIPENDENTI E INCOMPATIBILI

Una tipica misconcezione è quella che confonde eventi *indipendenti* con eventi *incompatibili* (vedi per esempio Barra 2005). Due eventi A e B si dicono incompatibili se $A \cap B = \varnothing$, cioè se il verificarsi dell'uno comporta l'impossibilità dell'altro. È evidente che se A e B (con $Pr(A) > 0$ e $Pr(B) > 0$) sono incompatibili



allora non possono essere indipendenti: se so che si è verificato B, la mia stima di probabilità su A muta radicalmente, perché diventa nulla: certamente non può essersi verificato A. Due eventi incompatibili sono sempre negativamente correlati:

$$Pr(A|B) = Pr(A \cap B)/Pr(B) = 0 < Pr(A)$$

Nell'esempio precedente con i dadi, i due eventi $D_1+D_2 = 2$ (probabilità iniziale = 1/36) e $D_1 = 4$ sono evidentemente incompatibili: se sul primo dado è uscito 4 allora sono certo che la somma dei due dadi non può dare 2: e dunque i due eventi non sono indipendenti.

ANCORA SUI GRAFI

Molti problemi di probabilità si risolvono in modo naturale mediante un grafo.

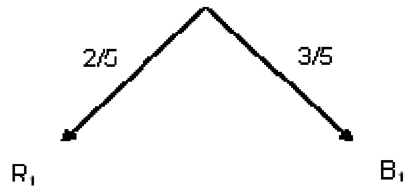
Consideriamo il seguente classico esempio.

Un'urna contiene 2 palline rosse e 3 blu. Se ne pescano 2 (non ha importanza se tutte e due insieme o una dopo l'altra, il risultato è lo stesso: si dice che vengono estratte senza reimmissione). È più probabile che siano dello stesso colore o di colore diverso?

Consideriamo l'esperimento aleatorio che consiste nell'estrazione della prima pallina. Il risultato può essere una pallina rossa (indichiamolo con R_1) o una pallina blu (B_1)[7]; le probabilità sono:

$$Pr(R_1) = 2/5 \qquad Pr(B_1) = 3/5$$

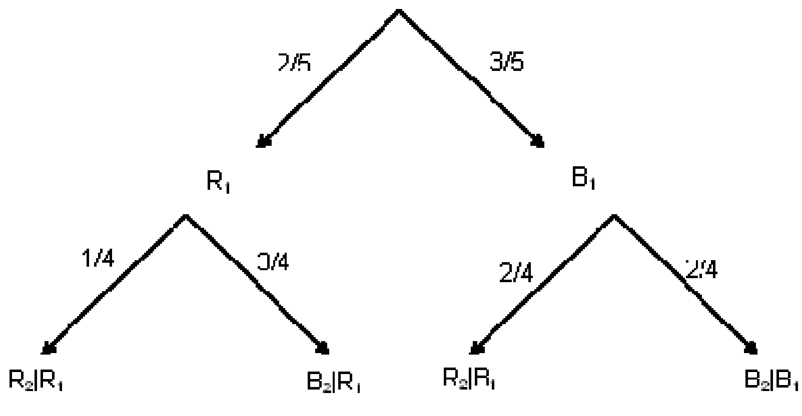
Il grafo è il seguente:



Ora procediamo alla seconda estrazione, che può dare come esito R_2 oppure B_2: da entrambi i nodi R_1 e B_1 si dipartono altri due rami. Quali sono le probabilità lungo questi 4 nuovi rami? Il fatto è che:

- le probabilità sono cambiate rispetto a quelle assegnate a R_1 e B_1, perché è cambiata la composizione dell'urna rispetto alla prima estrazione, c'è una pallina in meno;
- ai due rami che partono dal nodo R_1 dobbiamo assegnare probabilità diverse rispetto a quelle sui due rami che partono da B_1, perché a seconda che il primo risultato sia stato R_1 oppure B_1, le composizioni dell'urna sono differenti; per esempio se la prima pallina è rossa allora nell'urna sono rimaste 1 pallina rossa e 3 blu; la probabilità di R_2 condizionata da R_1 è dunque 1/4.

Un grafo che tenga conto di tutto ciò è il seguente.



[7] Ho scritto R_1 e B_1 per segnalare il fatto che si tratta della prima pallina estratta.

Qual è la probabilità di pescare due palline rosse? Se traduciamo "due palline rosse" in termini di eventi, ciò che vogliamo calcolare è

$$\Pr(R_1 \cap R_2) = \Pr(R_2|R_1)\Pr(R_1)$$

Risulta

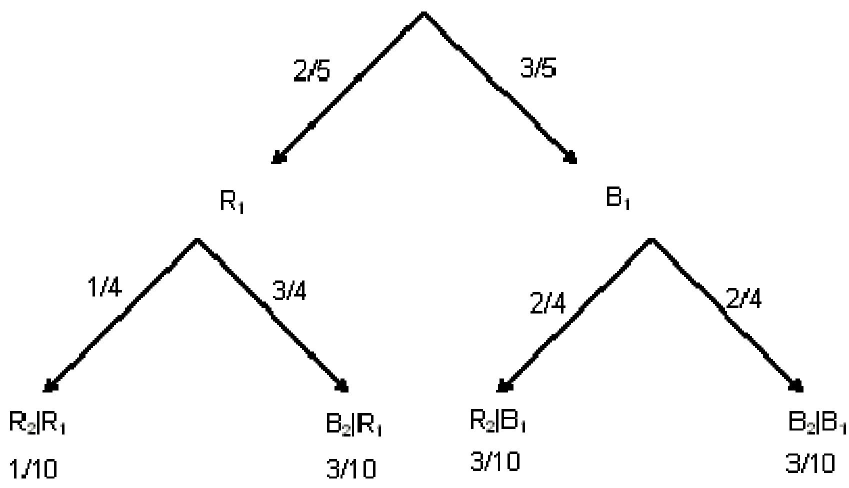
$$\Pr(R_2 \cap R_1) = \Pr(R_2|R_1)\Pr(R_1) = \frac{1}{4}\cdot\frac{2}{5} = \frac{1}{10}$$

$$\Pr(R_2 \cap B_1) = \Pr(R_2|B_1)\Pr(B_1) = \frac{3}{5}\cdot\frac{2}{4} = \frac{3}{10}$$

$$\Pr(B_2 \cap R_1) = \Pr(B_2|R_1)\Pr(R_1) = \frac{2}{5}\cdot\frac{3}{4} = \frac{3}{10}$$

$$\Pr(R_2 \cap B_1) = \Pr(R_2|B_1)\Pr(B_1) = \frac{2}{4}\cdot\frac{3}{5} = \frac{3}{10}$$

Possiamo ora completare il grafo.



Conclusione:

$$\Pr(R_1 \cap R_2) = 10\%$$
$$\Pr(B_1 \cap B_2) = 30\%$$

La probabilità che le palline siano dello stesso colore è dunque

$$\Pr((R_1 \cap R_2) \cup (B_1 \cap B_2)) = 40\%$$

Inoltre

$$\Pr(R_1 \cap B_2) = \Pr(B_1 \cap R_2) = 30\%$$

La probabilità che le palline siano di colore diverso è dunque

$$\Pr((R_1 \cap B_2) \cup (B_1 \cap R_2)) = 60\%$$

Dal punto di vista didattico, per giustificare il prodotto

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B)$$

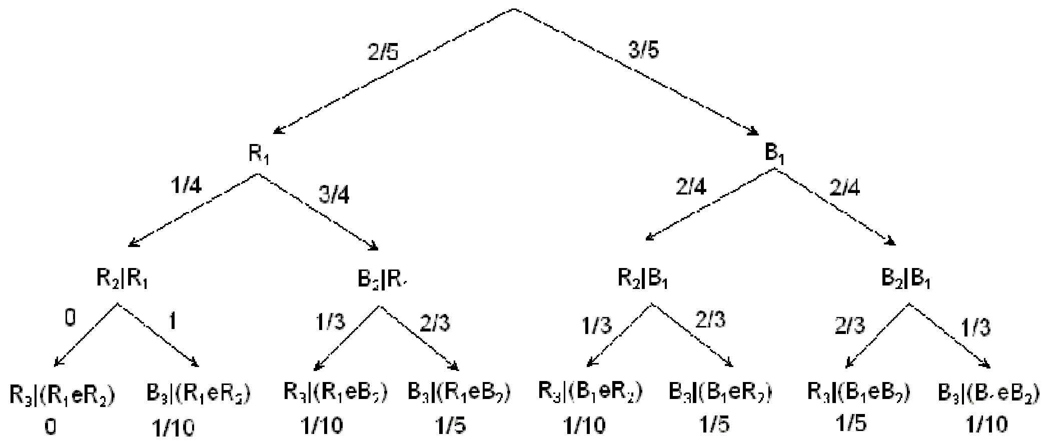
può essere utile ricorrere ad un grafo di frequenze, anziché di probabilità. Il fatto che le probabilità lungo i rami si moltiplichino può essere mostrato in modo convincente seguendo la strada "frequentista". Supponiamo di avere 100 urne tutte con la stessa composizione: 2 palline rosse e 3 palline blu. Da ciascuna di queste 100 urne peschiamo la prima pallina. Ci aspettiamo che in 2/5 delle estrazioni, cioè in 40 urne, esca una pallina rossa e nelle altre 60 una pallina blu. Procediamo ora all'estrazione della seconda pallina in ciascuna delle 100 urne.

Nelle 40 da cui abbiamo estratto una pallina rossa è rimasta una pallina rossa e 3 blu: ci aspettiamo di pescare una pallina rossa in 1/4 delle 40 urne (cioè 10 urne) e una pallina blu nelle rimanenti 30 urne (3/4 di 40). Dunque il numero di urne da cui ci aspettiamo di pescare 2 palline rosse è 1/4 di 2/5 di 100 urne, cioè

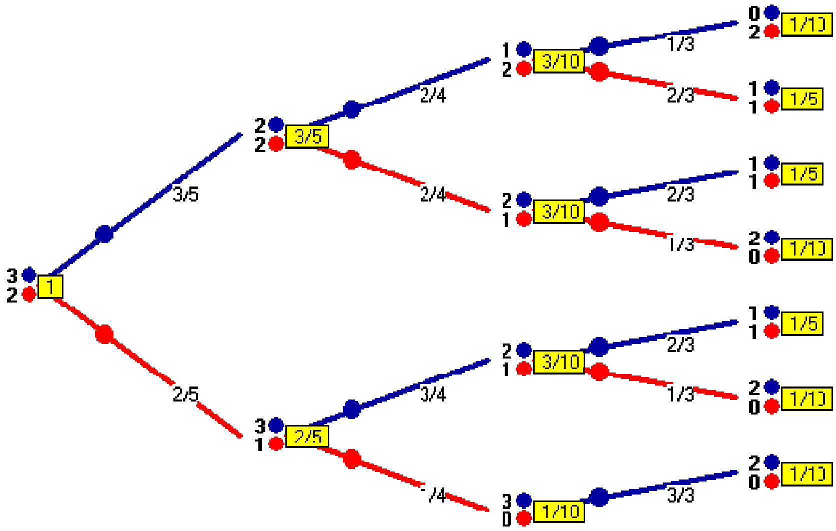
$$\frac{1}{4}\cdot\frac{2}{5}\cdot 100 = \frac{1}{10}\cdot 100 = 10$$

In definitiva ci aspettiamo di pescare 2 palline rosse in 1/10 delle urne, il che è quanto dire che la probabilità di pescare 2 palline rosse è 1/10, cioè il prodotto tra 1/4 e 2/5. Nello stesso modo si ricavano le probabilità lungo gli altri tre rami.

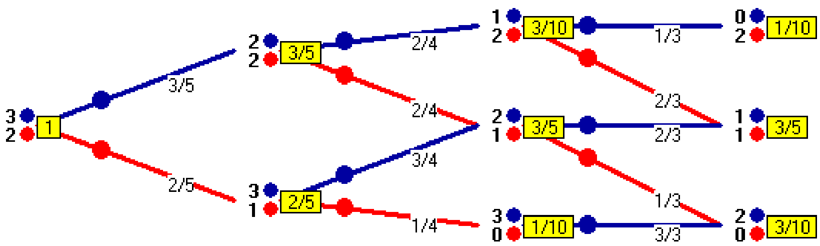
Naturalmente possiamo continuare estraendo una terza pallina. Il grafo si aggiorna nel seguente modo.



C'è un ottimo software che costruisce grafi di probabilità relativi al problema dell'urna; si chiama VuStat, e lo trovate online all'indirizzo www.vusoft2.nl/VuStat.htm (è lo stesso che distribuisce Graphic Calculus, il software a cui ha collaborato David Tall).



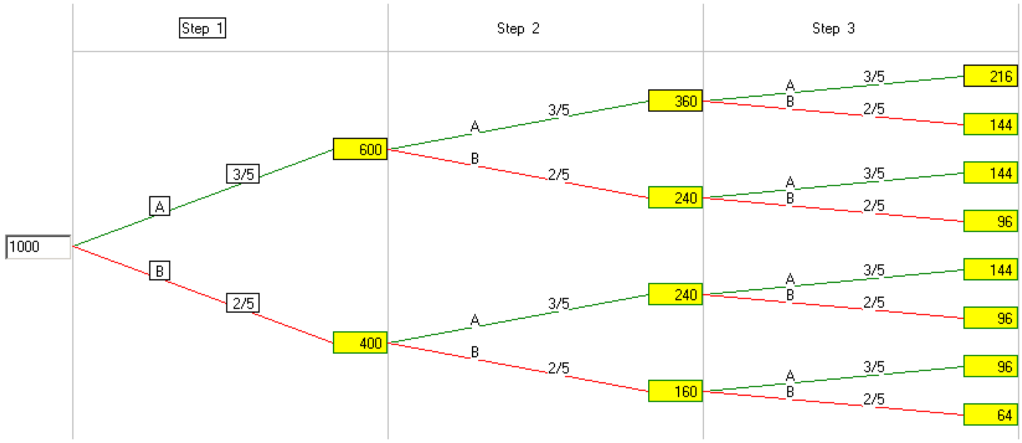
Fino ad ora abbiamo separato i rami successivi, prendendo nota dell'ordine dei colori estratti. Se siamo interessati solo al numero dei colori estratti e non all'ordine, si combinano i rami successivi, come mostra il grafo seguente.



ESEMPIO. *Un giocatore di basket centra il tiro libero con probabilità 60%. Se tira tre volte, qual è la probabilità di fare 0, 1, 2, 3 centri?*

Il problema può essere riformulato in termini di urna: *Un'urna contiene 40 palline rosse e 60 blu. Se ne pescano 3 (questa volta con reimmissione, in modo tale che le probabilità restino sempre 2/5 e 3/5). Quali sono le probabilità di pescarne 0, 1, 2, 3 blu?*

Un grafo (sempre costruito con VuStat) che risolve il problema potrebbe essere il seguente.



Come si vede si parte da una "popolazione" (nel caso del basket da un numero di tiri) di 1000 elementi. Assumiamo che le frequenze relative siano uguali alle probabilità. Allora questi 1000 tiri si trasformeranno in centri nei 3/5 delle volte, cioè in 600 tiri e ci sarà errore i 2/5 delle volte, cioè in 400 tiri. Ora tiro di nuovo: sui 600 centri del primo tiro, ne azzeccherò i 3/5 (cioè 360, i 3/5 dei 3/5 di 1000); di questi 360 doppi successi otterrò al terzo tiro i 3/5 dei centri e quindi in totale 216 centri:

$$216 = \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} \cdot 1000$$

Lo stesso ragionamento per tutti gli altri rami.

IL TEOREMA DI BAYES

Bayes' theorem: it is the chief rule involved
in the process of learning from experience.
Harold Jeffreys, 1931

Le nozioni di probabilità condizionata e di indipendenza tra eventi sono cruciali per l'intera teoria della probabilità, in particolare per l'impostazione soggettivista, per la quale la probabilità di un evento si aggiorna in funzione delle informazioni che l'osservatore via via raccoglie.

Rivediamo il meccanismo di aggiornamento:

- ho una certa stima della probabilità di un evento A (probabilità *iniziale*, oppure *a priori*);
- vengo in possesso dell'informazione B (si è verificato l'evento B);
- la probabilità *a posteriori* (o *finale*) di A, una volta acquisita l'informazione B, è

$$\Pr(A\,|\,B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Il cosiddetto *teorema di Bayes* (Thomas Bayes, 1702-1761) è enunciato e dimostrato in *An Essay towards solving a Problem in the Doctrine of Chances*, pubblicato postumo nel 1763. Esso è talmente semplice che qualcuno si rifiuta di chiamarlo *teorema* e preferisce *regola di Bayes*; è una diretta conseguenza del conflitto tra la simmetria dell'intersezione

$$A \cap B = B \cap A$$

e la non simmetria del condizionamento:

$$\Pr(A|B) \neq \Pr(B|A)$$

Risulta infatti:

$$\Pr(A \cap B) = \Pr(A|B)\Pr(B)$$
$$\Pr(B \cap A) = \Pr(B|A)\Pr(A)$$

dunque

$$\Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$$

da cui

Teorema di Bayes: $\Pr(A|B) = \dfrac{\Pr(B|A)}{\Pr(B)}\Pr(A)$

Questa relazione si può leggere in diversi modi: innanzitutto collega direttamente la *probabilità iniziale* Pr(A) alla *probabilità finale* Pr(A|B) mediante il *fattore di aggiornamento* $k = \Pr(B|A)/\Pr(B)$:

$$\Pr(A|B) = k \cdot \Pr(A)$$

$k > 1$	A e B sono positivamente correlati
$k = 1$	A e B sono indipendenti
$0 < k < 1$	A e B sono negativamente correlati

Oppure: la probabilità finale Pr(A|B) è il prodotto della probabilità iniziale Pr(A) per la *verosimiglianza* (*likelihood*) Pr(B|A), il tutto normalizzato mediante Pr(B): fissate la probabilità iniziale Pr(A) e la probabilità della nuova informazione Pr(B), la probabilità finale Pr(A|B) è *direttamente proporzionale* alla verosimiglianza:

$$\Pr(A|B) \propto \Pr(B|A)$$

Questo è il nocciolo del teorema di Bayes: tanto più un evento è verosimile, tanto più è probabile. Attenzione: non è un commento da bar, è un teorema.

La probabilità Pr(B|A) prende il nome di *verosimiglianza di A rispetto a B*. Questo termine viene da lontano e si riferisce all'antica disputa filosofica tra ciò che è *vero* e ciò che è *verosimile*.

Se un numero naturale (maggiore di 2) è primo allora è dispari, e questo è *vero*.

Formalizziamo; posto

P = insieme dei numeri primi (maggiori di 2)

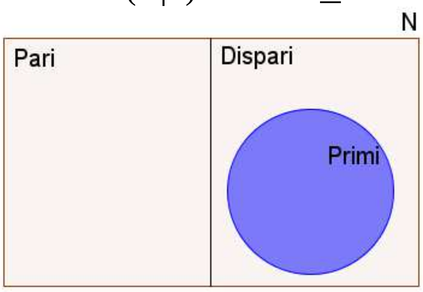
D = insieme dei numeri dispari

allora risulta

$$\Pr(D|P) = 1$$

Dal punto di vista insiemistico:

$$\Pr(D|P) = 1 \Leftrightarrow P \subseteq D$$



Se un numero naturale (maggiore di 2) è dispari allora non è detto che sia primo, però … sapere che è dispari rende più *verosimile* che sia primo; la conoscenza del fatto che *n* sia dispari (rispetto a non sapere nulla) fa aumentare il nostro grado di fiducia che *n* sia primo; di quanto?

Poiché Pr(D|P)=1 e poiché possiamo assumere Pr(D)=1/2, risulta:

$$\Pr(P|D) = \Pr(D|P)\Pr(P)/\Pr(D) = 2\Pr(P)$$

La probabilità che *n* sia primo, se si acquisisce l'informazione che *n* è dispari, raddoppia.

CHI DORME NON PIGLIA PESCI

Se dormo non piglio pesci, e questo è *vero*. Se non ho pigliato pesci ho dormito? Questo è "solo" *verosimile*.

La matematica tradizionalmente si è sempre sottratta, altezzosa, al problema di quantificare la verosimiglianza di un evento, limitandosi a sostenere che dal teorema $A \Rightarrow B$ non è possibile in generale concludere *alcunché* sulla implicazione inversa $B \Rightarrow A$. Non è vero, qualcosa è possibile concludere, anche se non in termini di verità (1) o falsità (0), ma in termini di probabilità (tra 0 e 1).

Consideriamo gli eventi

A = ho dormito
B = non ho preso pesci

Si tratta del prototipo popolare del processo deduttivo: se A allora necessariamente B.

Bene, e se B? Posso concludere qualcosa su A?

Se superiamo il fossato storico che ha impedito per secoli alla matematica di pronunciarsi sull'implicazione inversa, e alla probabilità di imporsi come strumento di lettura del mondo, possiamo finalmente dire:

- se non hai preso pesci è *verosimile* che tu abbia dormito

oppure

- se non hai preso pesci c'è una certa probabilità che tu abbia dormito

oppure

- se non hai preso pesci sono più propenso a pensare che tu abbia dormito.

Insomma, se non hai preso pesci (B) allora o hai dormito (A) oppure sei rimasto sveglio (A^C); le probabilità (condizionate da B) di A e di A^C, per il teorema di Bayes, sono date da:

$$\Pr(A|B) = \frac{\Pr(B|A)}{\Pr(B)}\Pr(A) \qquad \Pr(A^C|B) = \frac{\Pr(B|A^C)}{\Pr(B)}\Pr(A^C)$$

Poiché entrambe si ottengono dividendo per Pr(B), possiamo in prima battuta tralasciare il calcolo di Pr(B) e fornire le stesse informazioni in termini di *proporzionalità* anziché di uguaglianza:

$$\Pr(A|B) \propto \Pr(B|A)\Pr(A)$$

$$\Pr(A^C|B) \propto \Pr(B|A^C)\Pr(A^C)$$

Inoltre, se non abbiamo informazioni sul fatto che A si sia verificato oppure no (ho dormito oppure no), possiamo supporre inizialmente[8] $\Pr(A) = \Pr(A^C) = 0.5$ e dunque in definitiva

$$\Pr(A|B) \propto \Pr(B|A)$$

$$\Pr(A^C|B) \propto \Pr(B|A^C)$$

Ritroviamo il fatto davvero notevole:

<u>La probabilità finale Pr(A|B) è proporzionale alla verosimiglianza Pr(B|A).</u>

Tanto maggiore è la verosimiglianza (cioè la probabilità che si verifichi B in presenza di A), tanto più probabile è che si sia verificato A in presenza di B.

Indichiamo con $x = \Pr(B|A^C)$ la probabilità di non prendere pesci (B) nonostante si sia trascorsa la giornata a pescare con impegno (A^C). Risulta:

$$\Pr(A|B) \propto \Pr(B|A) = 1$$

$$\Pr(A^C|B) \propto \Pr(B|A^C) = x$$

In sostanza le due probabilità Pr(A|B) e $\Pr(A^C|B)$ sono proporzionali ai due numeri 1 e x: poiché risulta $x \le 1$, possiamo già dedurre che

$$\Pr(A|B) \ge \Pr(A^C|B)$$

cioè se non hai preso pesci, è più probabile che tu abbia dormito piuttosto che tu abbia passato la giornata a pescare con impegno.

[8] È questa una situazione tipica: in assenza di informazioni che lascino supporre il contrario, si assume inizialmente che tutte le ipotesi alternative siano ugualmente probabili (*distribuzione uniforme*). È quanto si dovrebbe fare anche in ambito giudiziario (vedi Peccati 2002): prima di iniziare il processo l'imputato dovrebbe essere considerato innocente o colpevole con uguali probabilità 1/2 (in fondo è stato rinviato a giudizio).

E siccome

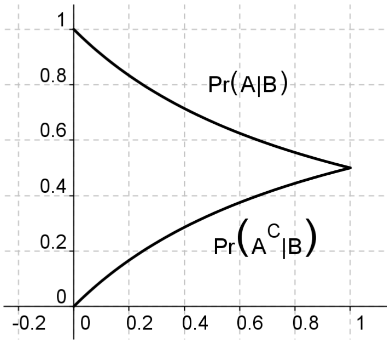
$$\Pr(A|B)+\Pr(A^C|B) = 1$$

(se non hai preso pesci (B): o hai dormito (A) oppure sei rimasto sveglio (A^C)), allora stiamo cercando due numeri proporzionali a 1 e x di somma 1; è sufficiente normalizzare, quindi dividere per $1+x$, per ottenere:

$$\Pr(A|B) = 1/(1+x)$$
$$\Pr(A^C|B) = x/(1+x)$$

Decidete voi quale potrebbe essere un valore ragionevole per $x = \Pr(B|A^C)$: 0? 0.1? 0.5? 0.9? 1?

Il grafico seguente mostra l'andamento di $\Pr(A|B)$ e di $\Pr(A^C|B)$ in funzione di x.



Qualunque valore scegliate per x la conclusione è sempre la stessa: se non si è preso pesci (B) la probabilità di aver dormito (A|B) non è mai minore (e in media notevolmente maggiore) della probabilità di essere rimasto sveglio (A^C|B). Altro che "non si può dire nulla"![9]

Qualunque punto di vista vogliate usare, si tratta di una conclusione forte e non banale, che non possedevo prima di applicare il teorema di Bayes. La morale è questa: possiamo arricchire le nostre conoscenze non solo di verità assolute, ma anche (per dirla con il linguaggio di Galileo) di convincenti ragionevolezze, di sensate probabilità.

[9] Addirittura se $x = 0$, cioè se possedete l'informazione che chi passa la giornata a pescare almeno un'acciughina la porta a casa, allora è sicuro che chi non piglia pesci ha dormito. L'unico caso in cui non possiamo dire nulla è quello ($x = 1$) in cui è sicuro che chi passa la giornata a pescare con impegno non porta comunque a casa nulla (perché ad esempio sta pescando in un bacino privo di pesci). Ma questo è proprio il caso banale, in cui non ha senso porsi il problema.

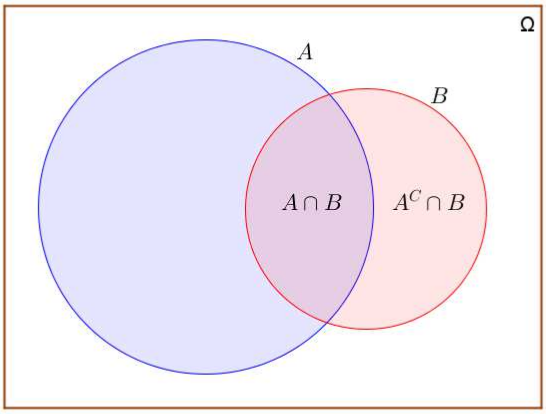
IL TEOREMA DI BAYES COMPLETO

Il teorema di Bayes viene anche indicato come il teorema che fornisce le *probabilità delle cause*, nel senso che tra diverse ipotesi alternative che possono in qualche modo "spiegare" un certo evento, è in grado di assegnare a ciascuna la corrispondente probabilità, e quindi fornire indicazioni preziose su quale sia l'ordine di rilevanza delle ipotesi prese in considerazione.

Nell'esempio precedente, partendo dall'informazione B (non ho preso pesci) abbiamo visto che l'ipotesi A (ho dormito) è sempre da preferire all'ipotesi A^C (ho pescato tutto il giorno con impegno), cioè l'evento B si spiega con maggior fiducia con l'ipotesi A piuttosto che con l'ipotesi A^C.

Come si è visto, per ottenere le due probabilità Pr(A|B) e Pr(A^C|B) siamo riusciti, mediante la proporzionalità, a non calcolare il denominatore comune Pr(B). Si tratta di una scorciatoia notevole, perché il calcolo di Pr(B) può essere laborioso; B può verificarsi sia in presenza di A che di A^C.

La figura seguente



mostra che qualunque siano A e B (purché Pr(A) > 0 e Pr(B) > 0) risulta

$$B = (A \cap B) \cup (A^C \cap B)$$

e poiché $A \cap B$ e $A^C \cap B$ sono disgiunti risulta

$$Pr(B) = Pr(A \cap B) + Pr(A^C \cap B)$$

cioè

$$Pr(B) = Pr(B|A)Pr(A) + Pr(B|A^C)Pr(A^C)$$

Il teorema di Bayes, applicato alle due ipotesi alternative A e A^C, può essere allora scritto in forma completa così:

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A)+\Pr(B|A^C)\Pr(A^C)}$$

Se le ipotesi alternative sono n, cioè se A_1, A_2, …, A_n costituiscono una partizione di Ω, allora abbiamo il teorema di Bayes nella forma più generale:

Teorema di Bayes. $$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B|A_1)\Pr(A_1)+\ldots+\Pr(B|A_n)\Pr(A_n)}$$

ALCUNI ESEMPI

1. Quattro monete sono regolari e una è truccata (ha Testa su entrambi i lati). Si sceglie una moneta a caso, la si lancia 5 volte, ed esce sempre Testa. Qual è la probabilità che la moneta scelta sia quella truccata?

Prima di lanciare la moneta scelta non ho alcuna informazione su quale sia quella truccata, dunque assumiamo che l'evento

A = "la moneta scelta è quella truccata"

abbia una probabilità iniziale

$$\Pr(A) = 1/5 = 20\%$$

Ora lancio la moneta (eseguo un esperimento); l'evento

B = "è uscito 5 volte Testa"

modifica (radicalmente) la valutazione di probabilità; se è uscito 5 volte Testa, il mio grado di fiducia sul fatto che la moneta scelta sia quella truccata è ragionevolmente aumentato. Come tener conto delle nuove informazioni fornite dall'esperimento? Teorema di Bayes!

$\Pr(B|A)$ è la probabilità di ottenere 5 teste consecutive con la moneta truccata: è 1.

$\Pr(B|A^C)$ è la probabilità di ottenere 5 teste consecutive con una delle 4 monete regolari: è $1/2^5$.

Dunque la verosimiglianza di A è 32 volte maggiore della verosimiglianza di A^C.

Usando la proporzionalità (questa volta le probabilità iniziali delle due ipotesi alternative A e A^C sono diverse: Pr(A) = 1/5, Pr(A^C) = 4/5) risulta:

$$\Pr(A|B) \propto \Pr(B|A)\Pr(A) = 1\cdot\frac{1}{5}$$

$$\Pr(A^C|B) \propto \Pr(B|A^C)\Pr(A^C) = \frac{1}{2^5}\frac{4}{5} = \frac{1}{40}$$

Dunque le probabilità Pr(A|B) e Pr(A^C|B) sono proporzionali ai numeri 1/5 e 1/40 o, se si preferisce, ai numeri 8 e 1. Dunque

$$\Pr(A|B) = \frac{8}{1+8} \approx 89\% \qquad \Pr(A^C|B) = \frac{1}{1+8} \approx 11\%$$

2. Uno studente deve rispondere a una domanda d'esame a risposta multipla, con una sola risposta giusta. Se risponde correttamente qual è la probabilità che conosca davvero l'argomento e che non abbia risposto a caso?

Sia

A = "lo studente conosce la risposta giusta"
B = "lo studente risponde correttamente"

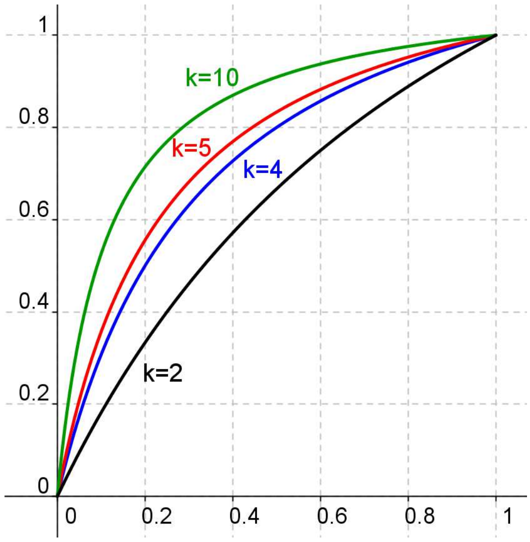
È ovvio supporre che $A \Rightarrow B$, cioè se lo studente conosce la risposta allora risponde correttamente con probabilità 1: Pr(B|A) = 1. Noi vogliamo conoscere invece Pr(A|B).

Se lo studente non conosce la risposta, risponde a caso e l'azzecca con probabilità $1/k$, se k è il numero di risposte del test. Assumiamo che la probabilità iniziale $\Pr(A) = x$ sia una misura della preparazione dello studente in quell'esame. Risulta allora:

$$\Pr(A|B) \propto \Pr(B|A)\Pr(A) = 1\cdot x$$
$$\Pr(A^C|B) \propto \Pr(B|A^C)\Pr(A^C) = \frac{1}{k}(1-x)$$

Dunque, normalizzando:

$$\Pr(A|B) = \frac{x}{x+\frac{1}{k}(1-x)} = \frac{kx}{1+(k-1)x}$$



La figura mostra i grafici di queste funzioni per diversi valori di k. Per esempio:

- se la preparazione è scadente ($x = 0.4$), con $k = 2$ possibili risposte la probabilità che lo studente conosca davvero la risposta giusta è circa del 57%; con $k = 10$ possibili risposte è dell'87%;
- se la preparazione è appena sufficiente ($x = 0.6$) allora con k =2 la probabilità che lo studente conosca davvero la risposta giusta è 75%; con $k = 10$ è 94%;
- se la preparazione è eccellente ($x = 0.9$) allora con $k = 2$ la probabilità che lo studente conosca davvero la risposta giusta è 95%; con $k = 10$ è 99%;

UN PROBLEMA CLASSICO: IL TEST CLINICO

Vediamo un'applicazione classica, di notevole portata sociale.

Un test diagnostico per l'accertamento di una malattia non è mai perfetto. C'è una certa percentuale di falsi negativi, *cioè persone che, nonostante siano malate, risultano negative al test, e una certa percentuale di* falsi positivi, *cioè persone che, nonostante siano sane, risultano positive al test. Se risulto positivo al test, qual è la probabilità che io sia davvero malato?*

Quali sono le probabilità di essere malato *prima* e *dopo* il test?

Prima del test non ho alcuna informazione, se non l'incidenza della malattia sull'intera popolazione, che posso assumere come probabilità iniziale. Ad esempio, supponiamo che il 5% della popolazione sia affetto da questa malattia. Sia

$$M = \text{"sono malato"}$$
$$P = \text{"il test è positivo"}$$

La probabilità *iniziale* (senza altre informazioni) che io sia malato è $\Pr(M) = 5\%$. Eseguo il test, che risulta positivo. Voglio ora stimare la probabilità *finale* $\Pr(M|P)$.

Per il teorema di Bayes risulta

$$\Pr(M|P) \propto \Pr(P|M)\Pr(M)$$
$$\Pr(M^C|P) \propto \Pr(P|M^C)\Pr(M^C)$$

Ci servono i seguenti dati, che caratterizzano l'efficacia del test:

- $\Pr(P|M)$: la probabilità che un individuo malato risulti effettivamente positivo al test;
- $\Pr(P|M^C)$: la probabilità che un individuo sano risulti positivo al test (un *falso positivo*).

Queste informazioni sono note (si spera!) dall'analisi clinico-stastistica sull'affidabilità del test diagnostico; supponiamo che sia

$\Pr(P|M) = 0.95$: i *falsi negativi* sono solo il 5% dei malati che si sottopongono al test.

$\Pr(P|M^C) = 0.05$: i *falsi positivi* sono solo il 5% dei sani che si sottopongono al test.

Dunque

$$\Pr(M|P) \propto \Pr(P|M)\Pr(M) = 0.95\cdot 0.05$$
$$\Pr(M^C|P) \propto \Pr(P|M^C)\Pr(M^C) = 0.05\cdot 0.95$$

Senza ulteriori calcoli è evidente che $\Pr(M|P) = \Pr(M^C|P)$; il risultato potrebbe apparire sconcertante: se il test è positivo, la probabilità di essere davvero malato è *solo* del 50%!

In realtà dobbiamo ragionare diversamente: la probabilità iniziale $\Pr(M)$ era del 5% e ora, dopo il test, è diventata 10 volte tanto!

Piuttosto l'efficacia del test deve essere migliorata: le percentuali di falsi positivi e di falsi negativi devono essere significativamente minori dell'incidenza percentuale della malattia.

TEOREMA DI BAYES: TABELLE E GRAFI

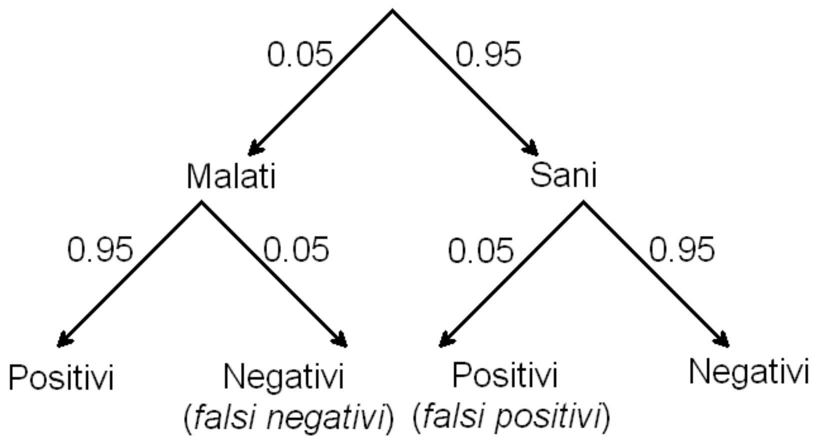
Utilizzando ancora i dati dell'esempio del test clinico, la tabella seguente mostra, su una popolazione di 2000 unità, il significato statistico delle diverse probabilità; in ogni cella è riportato il numero di individui che appartengono alla intersezione degli eventi.

	Positivi	Negativi	Totale
Malati	95	5	**100**
Sani	95	1805	**1900**
Totale	**190**	**1810**	**2000**

Su 2000 individui 100 sono malati e 1900 sani. Dei 100 malati il test ne individua 95 (i rimanenti 5 sono falsi negativi); dei 1900 sani il test ne individua 1805 (i rimanenti 95 sono falsi positivi). La probabilità di essere davvero malato se sono positivo al test è

$$\Pr(M|P) = \Pr(M \cap P)/\Pr(P) = 95/190 = 50\%.$$

Gli stessi risultati si ottengono mediante un *grafo*.



Ricordiamo che sulla seconda riga del grafo si leggono le probabilità condizionate

$$\Pr(P|M),\ \Pr(N|M),\ \Pr(P|S),\ \Pr(N|S)$$

Nel contesto del teorema di Bayes queste sono esattamente le verosimiglianze: per esempio $\Pr(P|M) = 0.95$ è la probabilità che il test risulti positivo su una persona effettivamente malata.

Poiché

$$\Pr(P\cap M) = \Pr(P|M)\Pr(M) = 0.95\cdot 0.05$$

allora

$$\Pr(M|P) = \frac{\Pr(M\cap P)}{\Pr(P)} = \frac{0.05\cdot 0.95}{0.05\cdot 0.95+0.95\cdot 0.05} = 50\%.$$

Generalizziamo. Supponiamo che il test preveda una stessa percentuale x di falsi positivi e di falsi negativi, e che sia m l'incidenza percentuale della malattia nella popolazione. Allora

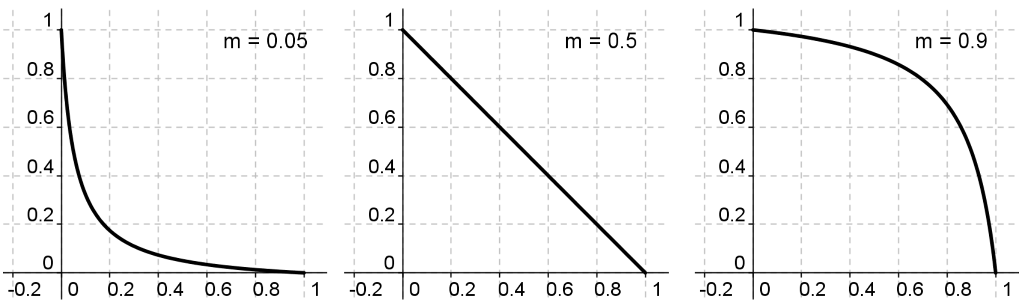
$$\Pr(M|P) \propto \Pr(P|M)\Pr(M) = (1-x)m$$

$$\Pr(S|P) \propto \Pr(P|S)\Pr(S) = x(1-m)$$

quindi

$$\Pr(M|P) = \frac{(1-x)m}{(1-x)m+x(1-m)}$$

Ecco i grafici di $\Pr(M|P)$ in funzione di x, con diversi valori di m: $m = 5\%$, 50%, 90%.



UN PROBLEMA CLASSICO: IL TEST DI AMMISSIONE

Ecco un problema del tutto equivalente al precedente, in un contesto differente.

Sapendo che gli studenti eccellenti sono una minoranza, circa il 5% del totale degli studenti, una nota università di Milano appronta un test di ammissione molto selettivo e molto mirato: solo il 6% degli studenti *mediocri lo passa e solo il 3% degli studenti eccellenti lo fallisce. Se ho passato il test, qual è la probabilità che io sia uno studente eccellente?*

Sia

E = "sono uno studente eccellente"
T = "ho passato il test"

Se passo il test non è detto che io sia eccellente; però è *verosimile* che lo sia: quanto verosimile?

I dati sono:

$\Pr(E) = 0.05$ *probabilità iniziale*
$\Pr(T^C|E) = 0.03$ *falsi negativi*
$\Pr(T|E^C) = 0.06$ *falsi positivi*

Allora la *probabilità finale* è

$$\Pr(E|T) = \frac{\Pr(T|E)}{\Pr(T)}\Pr(E) = \frac{\Pr(T|E)\Pr(E)}{\Pr(T|E)\Pr(E)+\Pr(T|E^C)\Pr(E^C)} =$$

$$= \frac{0.97\cdot 0.05}{0.97\cdot 0.05+0.06\cdot 0.95} \approx 46\%.$$

Se ho passato il test c'è una probabilità del 46% che io sia uno studente eccellente e dunque del 54% che non lo sia: suona strano, vero?

Una traduzione statistica del problema è la seguente: su una popolazione, per esempio, di 2000 studenti ci sono 100 eccellenti (il 5% di 2000); di questi, ben 97 passano il test (il 97% di 100); degli altri 1900, solo 114 passano il test (il 6% di 1900), con il risultato che tra gli studenti che hanno passato il test ce ne sono 97 eccellenti (il 46%) e 114 mediocri (il 56%).

	Test positivo	Test negativo	Totale
Eccellenti	97	3	**100**
Mediocri	114	1786	**1900**
Totale	**211**	**1789**	**2000**

Il risultato sembra paradossale, ma anche questa volta dobbiamo interpretarlo in un altro modo: nella popolazione iniziale che affronta il test c'è solo il 5% di studenti eccellenti; dopo il test la popolazione degli ammessi contiene il 46% di studenti eccellenti.

Allora una strategia vincente per l'Università è quella di sottoporre la popolazione già filtrata con il primo test ad un secondo test, che supponiamo avere le stesse caratteristiche selettive del primo; ora i dati sono i seguenti:

$\Pr(E) = 0.46$ *nuova probabilità iniziale*

$\Pr(T^C|E) = 0.03$ *falsi negativi*

$\Pr(T|E^C) = 0.06$ *falsi positivi*

Allora la *nuova probabilità finale* che io sia uno studente eccellente se ho superato anche il secondo test è

$$\Pr(E|T) = \frac{\Pr(T|E)\Pr(E)}{\Pr(T|E)\Pr(E)+\Pr(T|E^C)\Pr(E^C)} =$$

$$= \frac{0.97\cdot 0.46}{0.97\cdot 0.46+0.06\cdot 0.54} \approx 93\%.$$

Possiamo continuare in questo modo, utilizzando ogni volta come probabilità iniziale la precedente probabilità finale e otteniamo che gli studenti eccellenti che superano il terzo test sono il 99.5% dei partecipanti, e così via. La stessa strategia si utilizza per i test clinici: si ripete il test per innalzare la soglia di selettività.

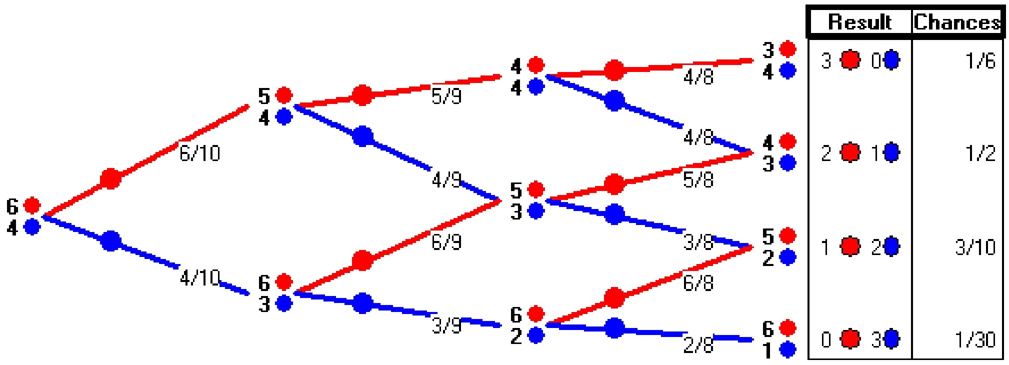
Il teorema di Bayes consente così, a ogni nuova informazione acquisita, di aggiornare la distribuzione di probabilità.

ESTRAZIONI CON E SENZA REIMMISSIONE

Sugli eserciziari di probabilità non manca solitamente un problema come il seguente. Spiegherò più avanti perché ritengo questo problema propedeutico per una lettura estremamente significativa del teorema di Bayes.

In un'urna ci sono $r = 6$ palline rosse e $b = 4$ palline blu. Si pescano, senza reimmissione, 3 palline. Quali sono le probabilità che di queste k siano rosse, con $k = 0, 1, 2, 3$?

Possiamo costruire il grafo del problema, aggiornando la composizione dell'urna ad ogni estrazione.



Tuttavia se il numero di palline estratte fosse alto, questa strada diventerebbe impraticabile. La soluzione generale, come è noto, coinvolge la cosiddetta *distribuzione ipergeometrica*: sia

R = "numero di palline rosse estratte"

da un'urna che contiene r palline rosse e b palline blu;

allora

$$\Pr(\mathrm{R}=k)=\frac{\binom{r}{k}\binom{b}{3-k}}{\binom{r+b}{3}} \qquad k = 0,\ 1,\ 2,\ 3$$

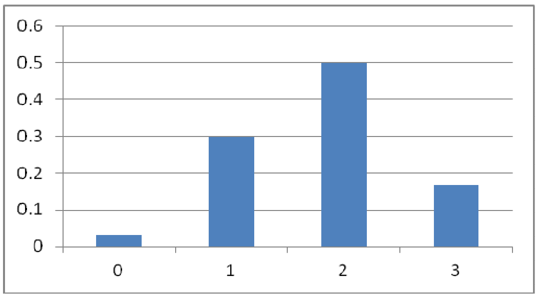
Risulta:

$$\Pr(\mathrm{R}=0)=\frac{\binom{6}{0}\binom{4}{3}}{\binom{10}{3}}=\frac{4}{120} \qquad \Pr(\mathrm{R}=1)=\frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}}=\frac{36}{120}$$

$$\Pr(\mathrm{R}=2)=\frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}}=\frac{60}{120} \qquad \Pr(\mathrm{R}=3)=\frac{\binom{6}{3}\binom{4}{0}}{\binom{10}{3}}=\frac{20}{120}$$

Possiamo sintetizzare la distribuzione di R con una tabella e un grafico.

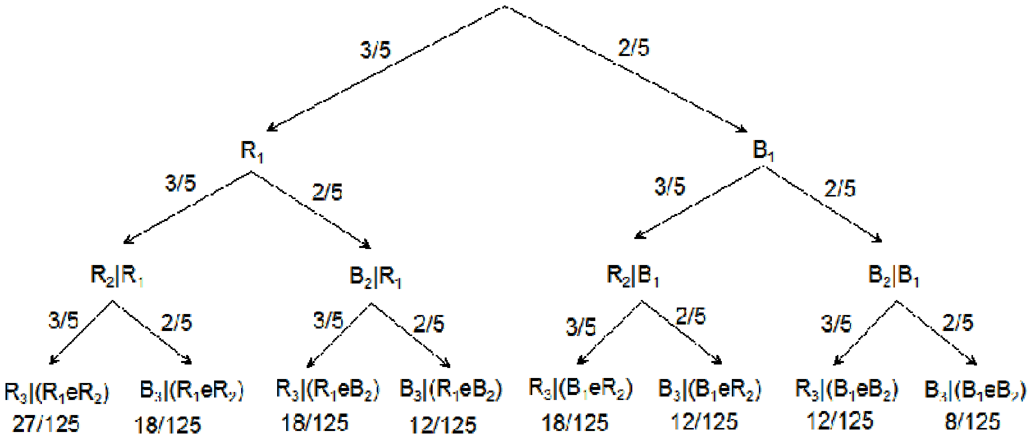
k	0	1	2	3
Pr(R = k)	1/30	3/10	1/2	1/6



Proviamo a enunciare lo stesso problema, però ora ogni pallina estratta è reimmessa nell'urna.

In un'urna ci sono $r = 6$ palline rosse e $b = 4$ palline blu. Si pescano, con reimmissione, 3 palline. Quali sono le probabilità che di queste k siano rosse, con $k = 0, 1, 2, 3$?

Questa volta il risultato di ogni estrazione è indipendente dalle estrazioni precedenti: la probabilità di pescare una pallina rossa è sempre 6/10=3/5. Il grafo ora è il seguente.



Anche in questo caso se il numero di palline estratte è alto, il grafo diventa impraticabile. La soluzione generale è data dalla *distribuzione binomiale*, che descrive il numero di "successi" in n prove in ciascuna delle quali si ha un "successo" con probabilità costante p. Nel nostro caso un "successo" è l'estrazione di una pallina rossa e p = 6/10; per il problema dell'urna in generale risulta $p = r/(r+b)$. Si ha dunque

$$\Pr(\mathrm{R}=k)=\binom{3}{k}\left(\frac{r}{r+b}\right)^{k}\left(\frac{b}{r+b}\right)^{3-k}=\binom{3}{k}p^{k}(1-p)^{3-k}\qquad k=0,\,1,\,2,\,3$$

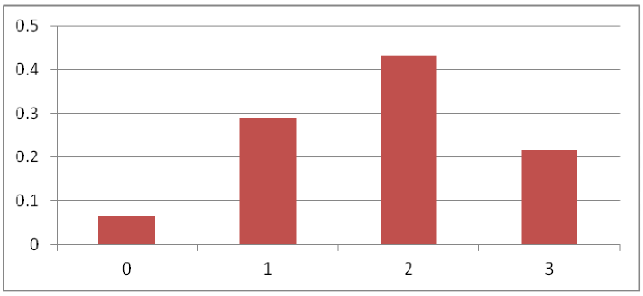
Nel nostro esempio risulta:

$$\Pr(\mathrm{R}=0)=\binom{3}{0}\left(\frac{3}{5}\right)^{0}\left(\frac{2}{5}\right)^{3}=\frac{8}{125}\qquad \Pr(\mathrm{R}=1)=\binom{3}{1}\left(\frac{3}{5}\right)^{1}\left(\frac{2}{5}\right)^{2}=\frac{36}{125}$$

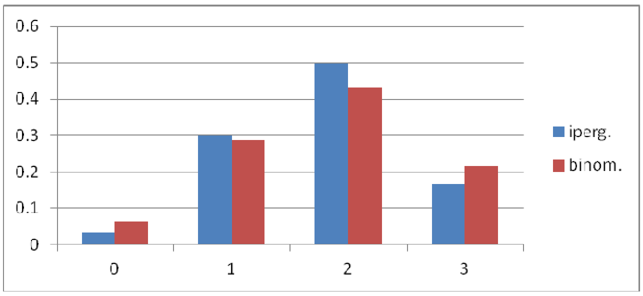
$$\Pr(\mathrm{R}=2)=\binom{3}{2}\left(\frac{3}{5}\right)^{2}\left(\frac{2}{5}\right)^{1}=\frac{54}{125}\qquad \Pr(\mathrm{R}=3)=\binom{3}{3}\left(\frac{3}{5}\right)^{3}\left(\frac{2}{5}\right)^{0}=\frac{27}{125}$$

Possiamo sintetizzare la distribuzione di R con una tabella e un grafico.

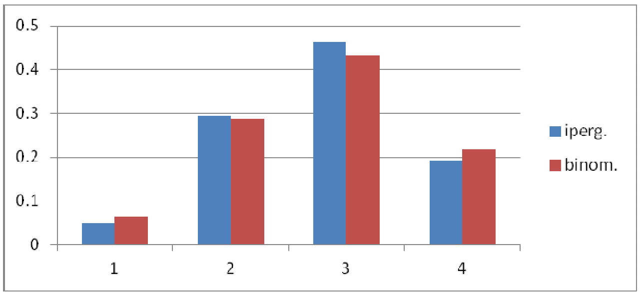
k	0	1	2	3
$\Pr(\mathrm{R}=k)$	8/125	36/125	54/125	27/125



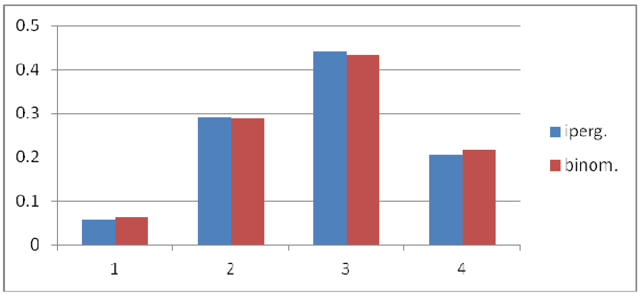
Ecco il confronto tra le due distribuzioni.



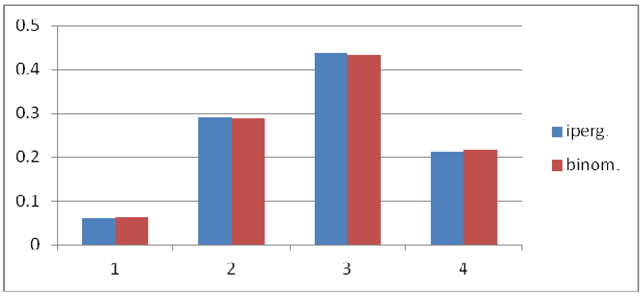
Le differenze tra le due distribuzioni ovviamente si attenuano se la popolazione è numerosa rispetto al numero di palline estratte: in questo caso le differenze tra i casi “senza” e “con” reimmissione sono trascurabili. Ecco per esempio i grafici per un’urna con 20 palline (di cui 12 rosse e 8 blu), 50 palline (30 rosse e 20 blu), 100 palline (60 rosse e 40 blu).



r = 12, b = 8



r = 30, b = 20



r = 60, b = 40

Può risultare utile, con Excel, utilizzare le funzioni DISTRIB.IPERGEOM e DISTRIB.BINOM.

Le corrispondenti sintassi, rispetto alla nomenclatura che abbiamo usato, sono le seguenti:

=DISTRIB.IPERGEOM(k;3;r;$r+b$)

per la distribuzione ipergeometrica, e

=DISTRIB.BINOM(k;3;p;FALSO)

per la distribuzione binomiale, dove $p = r/(r+b)$ e FALSO si riferisce alla possibilità di cumulare le probabilità (ottenendo così direttamente la funzione di ripartizione).

CAMPIONI E POPOLAZIONI: UN PROBLEMA DI VASTA PORTATA

Problemi come questi sono squisitamente "scolastici", nel senso che hanno lo scopo di esercitare all'uso di una formula. La loro struttura è semplice: si conosce la composizione dell’urna, chissà che cosa si pescherà se si estrae un campione. Nella realtà questo problema non ha alcun interesse perché i termini sono rovesciati: non si conosce la composizione dell’urna, mentre si può estrarre un campione e semplicemente osservare che cosa abbiamo pescato.

Il problema dell'estrazione da un'urna è il paradigma di un problema tipico, generale e di vasta portata: c’è una popolazione (l’urna) in cui un certo numero di individui (non si sa quanti) hanno una certa caratteristica: il numero di pazienti che hanno contratto una data malattia, il numero di studenti che passeranno un esame, il numero di elettori che voteranno SI al referendum, il numero di clienti che compreranno un certo prodotto, il numero di automobilisti che procureranno almeno un incidente quest'anno, il numero di persone che consumano un certo bene, il numero di treni che oggi arriveranno con più di 5' di ritardo, e così via. Il mondo è pieno di “popolazioni” delle quali vorremmo sapere in che misura è presente una certa proprietà. Non possiamo sapere quanti voteranno SI al prossimo referendum; però possiamo intervistare alcune persone e conoscere completamente la composizione del campione estratto.

Nei problemi scolastici con l’urna noi *facciamo finta di conoscere* la composizione dell’urna e *facciamo finta di non conoscere* i colori delle palline che abbiamo estratto. Nella realtà funziona esattamente a rovescio: noi sappiamo i colori delle palline estratte (basta guardarle!) e non conosciamo la composizione dell’urna. Traduciamo il problema: nella realtà noi possiamo conoscere la composizione di un campione (basta uscire dalle aule scolastiche!) e vorremo fare ipotesi sulla composizione della popolazione.

Nelle applicazioni reali l’urna è la popolazione: le palline rosse costituiscono gli individui della popolazione che hanno la proprietà individuata, le palline blu la frazione rimanente. Dunque nelle applicazioni reali r e b non sono noti, anzi rappresentano esattamente ciò che vorremmo conoscere.

Ecco allora la rivoluzione bayesiana: R (il numero di elementi della popolazione che hanno una certa proprietà) è un *numero aleatorio* di cui possiamo descrivere la distribuzione di probabilità. Per stimare la distribuzione di R ci affidiamo a un esperimento: estraiamo dall'urna un *campione* di palline, osserviamo il colore e cerchiamo di stimare quante siano le palline rosse nell'urna.

Traduciamo: per stimare la composizione della popolazione prendiamo un campione di persone a caso, intervistiamole. Dalla composizione del campione, mediante il teorema di Bayes, possiamo stimare la composizione della popolazione.

Per esempio: tu non sai quante persone della popolazione voteranno SI al prossimo referendum, ma puoi chiederlo a un po' di persone, questo sì che puoi saperlo.

Nei problemi scolastici abbiamo risolto il problema di stimare quale fosse la probabilità che su n palline estratte ce ne fossero k rosse sapendo che nell'urna ce n'erano r rosse e b blu. Ora vogliamo sapere quante palline rosse e quante blu ci sono nell'urna sapendo che estraendone n ne abbiamo trovate k rosse e $n–k$ blu.

Torniamo allora al problema precedente così riformulato, ipotizzando di conoscere il numero totale di palline, ma non la loro composizione.

In un'urna ci sono 10 palline, R palline rosse e B palline blu. Qual è il numero di palline rosse nell'urna, cioè qual è la distribuzione di R?

Inizialmente non abbiamo alcuna informazione sulla composizione dell'urna. Ragionevolmente adottiamo una stima iniziale *uniforme*; R può assumere i valori 0, 1, 2, …, 10 con la stessa probabilità comune, 1/11:

$$\Pr(R=k)=1/11 \qquad \text{con } k = 0, 1, 2, \ldots, 10$$

Ora procediamo a un esperimento: peschiamo 3 palline; sia

$$E_{3R} = \text{"le 3 palline estratte sono rosse"}$$

Vogliamo aggiornare la stima di probabilità su R=k sulla base di questo esperimento, vogliamo cioè stimare

$$\Pr(R=k|E_{3R})$$

Ma per il teorema di Bayes questa probabilità è proporzionale alla probabilità

$$\Pr(E_{3R}|R=k)$$

cioè la probabilità di pescare 3 palline rosse conoscendo la composizione dell'urna; questa è esattamente la probabilità che abbiamo calcolato con i problemi "scolastici" e che ora ci torna utile.

Vediamo i conti per bene:

$$\Pr(R = k|E_{3R}) = \frac{\Pr(E_{3R}|R = k)\Pr(R = k)}{\Pr(E_{3R})}$$

Qualunque sia *k*, il denominatore è sempre lo stesso, dunque passiamo alla proporzionalità:

$$\Pr(R = k|E_{3R}) \propto \Pr(E_{3R}|R = k)\Pr(R = k)$$

Qualunque sia *k*, il fattore Pr(R=*k*) è sempre lo stesso (abbiamo ipotizzato 1/11 per ciascun *k*), dunque passiamo alla proporzionalità:

$$\Pr(R = k|E_{3R}) \propto \Pr(E_{3R}|R = k)$$

Ecco di nuovo la magia del teorema di Bayes: la probabilità che le palline rosse siano *k*, dopo l'esito dell'esperimento, è proporzionale alla *verosimiglianza* di tale esperimento; se le palline rosse sono *k* (e sono 10–*k* le palline blu), quanto sarebbe probabile estrarne 3 rosse? Lo abbiamo già calcolato.

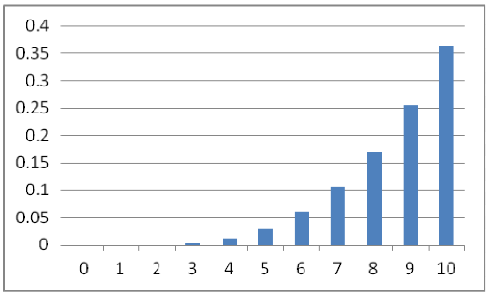
$$\Pr(E_{3R}|R = k) = \frac{\binom{k}{3}\binom{10-k}{0}}{\binom{10}{3}} = \frac{k(k-1)(k-2)}{720} \propto k(k-1)(k-2)$$

Le probabilità $\Pr(R = k|E_{3R})$ sono dunque proporzionali ai numeri $k(k-1)(k-2)$:

k	0	1	2	3	4	5	6	7	8	9	10
$k(k-1)(k-2)$	0	0	0	6	24	60	120	210	336	504	720

Dividiamo tutto per la loro somma, che è 1980, ed ecco la conclusione.

k	0	1	2	3	4	5	6	7	8	9	10
$\Pr(R = k\|E_{3R})$	0	0	0	1/330	2/165	1/33	2/33	7/66	28/165	14/55	4/11
$\Pr(R = k\|E_{3R})$	0	0	0	0.3%	1.2%	3%	6.1%	10.6%	17.0%	25.5%	36.4%



Si osservi che le probabilità sono crescenti con k: se abbiamo pescato solo palline rosse, l'evento più probabile è che siano *tutte* rosse.

Ora utilizziamo queste probabilità finali come nuove probabilità iniziali, e procediamo a un nuovo esperimento. Supponiamo di reimmettere le palline nell'urna e di estrarne altre 2: questa volta sono tutte e due blu. Chiamiamo E_{2B} questo esperimento. Risulta

$$\Pr(R = k|E_{2B}) \propto \Pr(E_{2B}|R = k)\Pr(R = k)$$

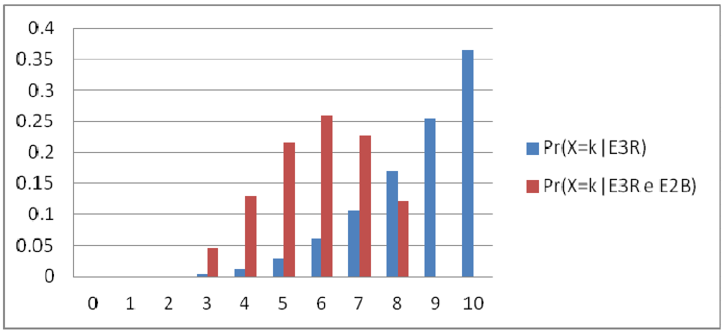
dove le $\Pr(R = k)$ sono le probabilità finali del precedente esperimento e ora diventano le nuove probabilità iniziali.

$$\Pr(E_{2B}|R = k)\Pr(R = k) = \frac{\binom{k-3}{0}\binom{10-k}{2}}{\binom{7}{2}}\,\frac{\binom{k}{3}\binom{10-k}{0}}{\binom{10}{3}} =$$

$$= \frac{k(k-1)(k-2)(k-9)(k-10)}{5400}$$

Le probabilità $\Pr(R = k|E_{2B})$ sono dunque proporzionali ai numeri $k(k–1)(k–2)(k–9)(k–10)$; dividendo ciascuno di essi per la loro somma si ottengono le nuove probabilità finali (tralasciamo le probabilità nulle).

k	3	4	5	6	7	8
$\Pr(R = k\|E_{1R2B})$	1/22	10/77	50/231	20/77	5/22	4/33
$\Pr(R = k\|E_{1R2B})$	4.5%	13%	21.6%	26%	22.7%	12.1%



Ora il numero più probabile di palline rosse è 6.

Possiamo chiederci quale sarebbe stata la conclusione se avessimo estratto in un solo esperimento 5 palline, di cui 3 rosse e 2 blu. Chiamiamo E_{3R2B} questo esperimento.

$$\Pr(R = k|E_{3R2B}) \propto \Pr(E_{3R2B}|R = k)\Pr(R = k) = \frac{\binom{k}{3}\binom{10-k}{2}}{\binom{10}{5}} \cdot \frac{1}{11}$$

È confortante (e diciamocelo, questa è la bellezza della matematica) osservare che saremmo arrivati alla stessa conclusione: a meno di costanti moltiplicative le espressioni

$$\frac{\binom{k-3}{0}\binom{10-k}{2}}{\binom{7}{2}} \frac{\binom{k}{3}\binom{10-k}{0}}{\binom{10}{3}} \quad \text{e} \quad \frac{\binom{k}{3}\binom{10-k}{2}}{\binom{10}{5}} \cdot \frac{1}{11}$$

sono uguali al variare di k: il teorema di Bayes garantisce la coerenza di più esperimenti con l'unico esperimento che li sintetizza.

CAMPIONI E POPOLAZIONI: IL CASO BINOMIALE

Supponiamo ora che la popolazione sia numerosa; come abbiamo visto, in questo caso la distribuzione ipergeometrica è ben approssimata dalla distribuzione binomiale. Affrontiamo dunque questo problema, che in un certo senso sintetizza l'intero articolo.

La popolazione svizzera verrà chiamata tra breve a pronunciarsi sulla rinuncia globale alle centrali nucleari. Come andrà il referendum? Vinceranno i SI che vogliono cancellare la scelta nucleare oppure tutto resterà come prima?

1. *Un sondaggio svolto su 10 persone rivela che 6 di queste sono per il SI. Qual è la probabilità che il SI vinca il referendum?*
2. *Un sondaggio svolto su 100 persone rivela che 60 di queste sono per il SI. Qual è la probabilità che il SI vinca il referendum?*
3. *Un sondaggio svolto su 1000 persone rivela che 600 di queste sono per il SI. Qual è la probabilità che il SI vinca il referendum?*

1. Gli elettori sono circa 6 milioni, ma si stima che solo i due terzi eserciteranno il diritto di voto, dunque 4 milioni di persone. Sia

N = n° di votanti = 4000000
S = n° di persone che voteranno SI al referendum
E_{10}, E_{100}, E_{1000} = risultati dei tre sondaggi

Ci interessa stimare la distribuzione di S subordinatamente a E_1; come al solito risulta

$$\Pr(S=k|E_{10}) \propto \Pr(E_{10}|S=k) = \binom{10}{6}\left(\frac{k}{N}\right)^6\left(1-\frac{k}{N}\right)^4 \propto k^6(N-k)^4$$

Normalizzando si ottiene

$$\Pr(S=k|E_{10}) = \frac{k^6(N-k)^4}{\sum_{k=0}^{N} k^6(N-k)^4}$$

e la probabilità che i SI vincano il referendum (indichiamo con V tale evento) è

$$\Pr(V|E_{10}) = \sum_{k=\frac{N}{2}+1}^{N} \Pr\left(S=k \mid E_{10}\right) = \sum_{k=\frac{N}{2}+1}^{N} \frac{k^6(N-k)^4}{\sum_{k=0}^{N} k^6(N-k)^4} =$$

$$= \frac{\sum_{k=\frac{N}{2}+1}^{N} k^6(N-k)^4}{\sum_{k=0}^{N} k^6(N-k)^4} \approx 73\%$$

Con Derive non è difficile calcolare il risultato, sia in forma simbolica sia in forma numerica:

$$\frac{\displaystyle\sum_{k=N/2+1}^{N} k^{6}\cdot(N-k)^{4}}{\displaystyle\sum_{k=0}^{N} k^{6}\cdot(N-k)^{4}}$$

$$\frac{9085913770054643618450514943388565457688610240334339644723093}{12522179429765636662188386326317360800287565308254919289446186}$$

$$\frac{\displaystyle\sum_{k=N/2+1}^{N} k^{6}\cdot(N-k)^{4}}{\displaystyle\sum_{k=0}^{N} k^{6}\cdot(N-k)^{4}}$$

$$0.7255856552$$

In modo analogo si ottiene:

$$\Pr(S=k|E_{100}) \propto \Pr(E_{100}|S=k) = \binom{100}{60}\left(\frac{k}{N}\right)^{60}\left(1-\frac{k}{N}\right)^{40} \propto k^{60}(N-k)^{40}$$

$$\Pr(S=k|E_{100}) = \frac{k^{60}(N-k)^{40}}{\sum_{k=0}^{N} k^{60}(N-k)^{40}}$$

$$\Pr(V|E_{100}) = \sum_{k=\frac{N}{2}+1}^{N} \Pr(S=k\,|\,E_{100}) = \sum_{k=\frac{N}{2}+1}^{N} \frac{k^{60}(N-k)^{40}}{\sum_{k=0}^{N} k^{60}(N-k)^{40}} =$$

$$= \frac{\sum_{k=\frac{N}{2}+1}^{N} k^{60}(N-k)^{40}}{\sum_{k=0}^{N} k^{60}(N-k)^{40}} \approx 98\%$$

Con Derive:

$$\frac{\sum_{k=N/2+1}^{N} k^{60}\cdot(N-k)^{40}}{\sum_{k=0}^{N} k^{60}\cdot(N-k)^{40}}$$

0.9769778296

E con E_{1000} si ottiene che la probabilità che i SI vincano il referendum è praticamente 1.

CAMPIONI E POPOLAZIONI: IL CASO CONTINUO

È chiaro che se la popolazione è molto numerosa allora conviene passare dal discreto al continuo. In questo caso il prodotto

$$\left(\frac{k}{N}\right)^6\left(1-\frac{k}{N}\right)^4$$

è sostituito dalla funzione

$$p^6(1-p)^4$$

dove

p = percentuale di persone che voterà SI al referendum

e $0 \le p \le 1$.

Si osservi che passando al continuo non siamo più costretti a utilizzare numeri "grandi" (per esempio N=4000000)

La funzione

$$f(p) = p^6(1-p)^4$$

è nulla in 0 e in 1: questo è in accordo con il fatto che, dato E_{10}, è molto improbabile che la percentuale di coloro che votano SI sia molto vicina a 0 o molto vicina a 1); f ha il punto di massimo proprio in 6/10 (la bellezza della matematica!), segnalando il fatto che la percentuale più probabile (la cosiddetta *moda* della distribuzione) sia proprio 0.6, cioè la stessa percentuale del campione.

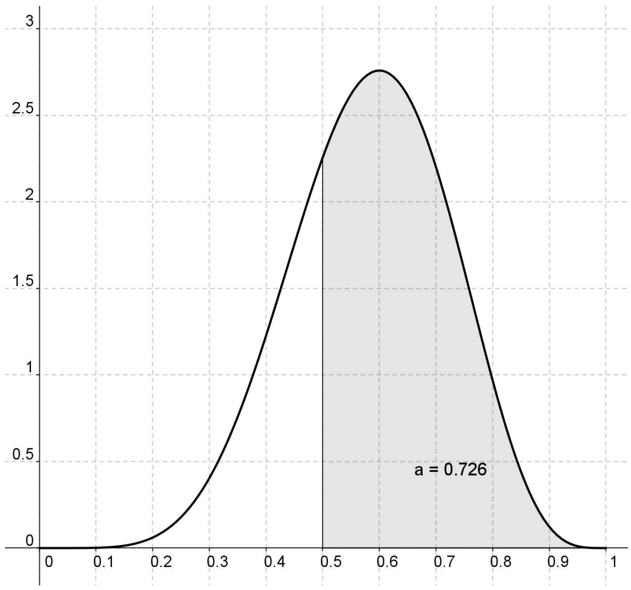


grafico di $p^6(1-p)^4$ normalizzato

Allora per ottenere la probabilità che i SI vincano il referendum sarà sufficiente sostituire alle somme i corrispondenti integrali:

$$\Pr(V|E_{10}) = \frac{\int_{0.5}^{1} p^6(1-p)^4\,\mathrm{d}p}{\int_{0}^{1} p^6(1-p)^4\,\mathrm{d}p} =$$

$$= \frac{743}{1024} \approx 72.6\%$$

Vediamo che cosa cambia se l'esperimento ipotizzato, anziché

E_{10} = “su 10 persone intervistate, 6 votano SI”

è ora

E_{100} = “su 100 persone intervistate, 60 votano SI”

Tutto rimane sostanzialmente uguale, tranne che la probabilità diventa proporzionale a

$$p^{60}(1-p)^{40};$$

ecco il relativo grafico.

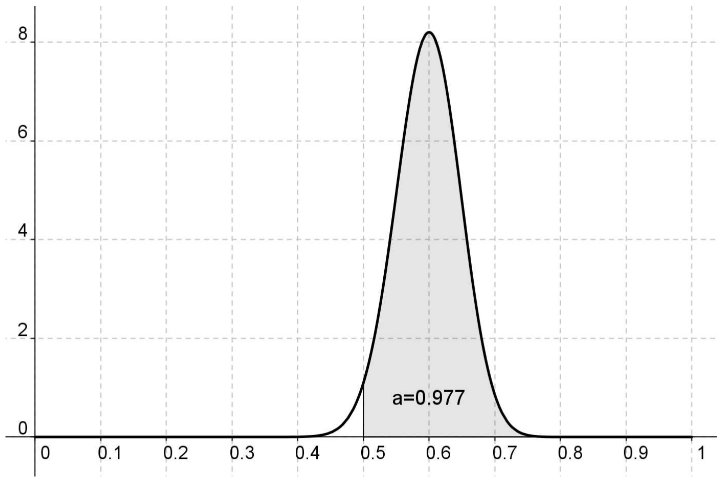


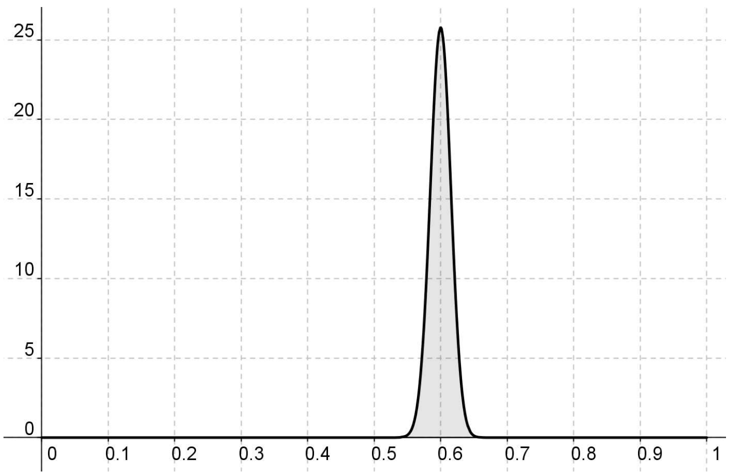
grafico di $p^{60}(1-p)^{40}$ normalizzato

Come si vede la distribuzione si "stringe" intorno a 0.6 e il suo integrale da 0.5 a 1 tende a occupare l'intera area. Risulta

$$\Pr(V|E_{100}) = \frac{\int_{0.5}^{1} p^{60}\left(1-p\right)^{40} \mathrm{d}p}{\int_{0}^{1} p^{60}\left(1-p\right)^{40} \mathrm{d}p} \approx 97.7\ \%$$

Infine se si sostituisce E_{100} con E_{1000} si ottiene la ragionevole certezza che i SI vinceranno:

$$P(V|E_{1000}) = \frac{\int_{0.5}^{1} p^{600}\left(1-p\right)^{400} \mathrm{d}p}{\int_{0}^{1} p^{600}\left(1-p\right)^{400} \mathrm{d}p} \approx 0.9999999998867.$$



CONCLUSIONI

Abbiamo visto come il teorema di Bayes permetta di aggiornare le stime di probabilità via via che si acquisiscono nuove informazioni: le stime di probabilità ottenute mediante un'informazione possono essere a loro volta utilizzate come probabilità iniziali per una nuova applicazione del teorema di Bayes, raffinando a ogni passo la soglia di incertezza.

Abbiamo anche mostrato che il risultato non dipende dall'ordine con cui acquisiamo le nuove informazioni.

Infine abbiamo applicato il teorema di Bayes a un problema di carattere generale, che in un certo senso lega probabilità e statistica: l'inferenza su una popolazione a partire da dati osservati su un campione. Spero che l'ultimo problema risolto, quello del referendum, metta in luce l'effettiva potenza del teorema di Bayes, applicato nella sua forma più semplice ed espressiva: la probabilità di un evento condizionato è proporzionale alla sua verosimiglianza, $\Pr(A|B) \propto \Pr(B|A)$.

A mio parere un percorso didattico che affronti in modo completo tale teorema si può collocare al primo anno di università; tuttavia il modo in cui le informazioni acquisite modificano la propria stima di probabilità è un tema al quale ci si può avvicinare per gradi già dalla scuola secondaria.

Per esempio l'analisi dei grafi nel gioco delle urne (senza reimmissione e con reimmissione) può essere proposto molto precocemente perché il prodotto delle probabilità lungo i rami del grafo può essere inizialmente giustificato in forma elementare e può risultare un efficace approccio a una comprensione via via più ricca del processo di aggiornamento. Si può poi tornare a livelli di complessità crescente sui grafi, per esempio mostrando in una seconda fase che il prodotto delle probabilità lungo i rami obbedisce alla relazione

$$\Pr(A|B)\Pr(B) = \Pr(A \cap B)$$

Nell'applicazione del teorema di Bayes si trova uno splendido condensato di temi e nozioni matematiche (non solo probabilità: operazioni tra insiemi, logica, coefficienti binomiali, somme e serie, funzioni, integrali), ma soprattutto si trova la presentazione di un *modello* matematico con il quale confrontare la propria intuizione nel classico percorso di andata e ritorno dalla realtà all'astrazione.

BIBLIOGRAFIA

Bayes T., 1763, *An essay towards solving a problem in the doctrine of chances*.

Barra M., 2003, Difficoltà nascoste nella didattica dei primi elementi di calcolo delle probabilità, *Progetto Alice*, N° 10, Ed. Pagine.

Barra M., 2003, Aspetti storici e pedagogici relativi al calcolo combinatorio, *Progetto Alice* N° 11, Ed. Pagine.

Barra M., 2005, Alcuni problemi dell'insegnamento del calcolo delle probabilità, *Progetto Alice*, N° 18, Ed. Pagine.

Barra M., 2006, La probabilità è nata in Italia, *Progetto Alice*, N° 19, Ed. Pagine.

Feyerabend P., 1979, *Contro il metodo*, Feltrinelli.

Kuhn T., 1962, *La struttura delle rivoluzioni scientifiche*, Einaudi.

Peccati L., 2002, *Metodi quantitativi per i giuristi*, EGEA.

Popper K., 1959, *La logica della scoperta scientifica*, Einaudi.

Prodi G., 1975, *La scoperta matematica*, D'Anna.

Scozzafava R., 2001, *Incertezza e probabilità*, Zanichelli.

Michele Impedovo