

6. Probabilità di eventi che dipendono da altri

La probabilità che dipende da altre¹, ovvero la probabilità condizionata, è già comparsa in varie situazioni esaminate nei capitoli precedenti del Percorso: dalle estrazioni senza reinserimento nell'urna fino ai test clinici.

Di comparsa appunto si è trattato, visto che il concetto è rimasto sullo sfondo e non è stato precisato tramite una definizione nonché una notazione specifica. Ciò è frutto di una scelta meditata e coerente con l'approccio seguito nelle altre fasi del Percorso: passare alla formalizzazione solo dopo che gli studenti hanno preso confidenza con i nuovi oggetti matematici, attraverso attività che permettono di scorgerne significato e portata. Anzi, proprio tali situazioni didattiche dovrebbero far nascere l'esigenza di comunicare in modo coinciso ed univoco, motivando così il ricorso alla formalizzazione.

In sintesi, riteniamo che l'attività didattica nel contesto² della probabilità condizionata andrebbe articolata in tre fasi: esplorativa, formale, applicativa. E ciò dovrebbe avvenire per varie ragioni, che in parte abbiamo già discusso: innanzitutto affinché i nuovi oggetti matematici si possano radicare con continuità sui saperi pregressi dello studente, in secondo luogo perché i ragazzi possano attribuire un senso a tali oggetti e non li percepiscano come inutili complicazioni imposte dall'esterno, ed infine affinché i contenuti e i metodi così sviluppati rimangano effettivamente disponibili anche a lungo termine.

In quest'ottica, si partirà dall'esame di alcune situazioni allo scopo di investigare gli aspetti didatticamente più significativi della probabilità che dipende da altre. Si proseguirà precisando formalmente il significato di probabilità condizionata, ad un livello che riteniamo adeguato per lo studente di scuola secondaria.

Infine, lo strumento matematico così introdotto verrà impiegato per descrivere e risolvere efficacemente alcuni interessanti problemi ambientati in contesti reali: test clinici, casi giudiziari...

¹ Più precisamente dovremmo parlare di "probabilità di eventi che dipendono da altri", ma la formula linguistica utilizzata ci sembra più espressiva.

² Come in diversi altri contesti didattici, in particolare in quelli esaminati nel nostro percorso.

6.1 Situazioni motivanti

Dunque iniziamo ad esaminare come il valore di probabilità che si assegna a un dato evento possa venir modificato quando si disponga di ulteriori informazioni su di esso. Precisamente ci proponiamo di schematizzare tali situazioni mediante opportune rappresentazioni grafiche e intendiamo motivare l'introduzione di notazioni e termini specifici che contribuiscano a sintetizzarle in modo espressivo.

a) Sopravvivenza e tavole

L'Istat pubblica periodicamente le tavole di mortalità della popolazione, elaborando in modo opportuno i dati anagrafici rilevati. Esse costituiscono un riferimento fondamentale anche per le assicurazioni sulla vita.

I dati in tabella sono estratti dalle tavole pubblicate per l'anno 2013³: si considera una popolazione iniziale fittizia di 100.000 nati; su ogni riga è riportato il numero di individui della popolazione che sono in vita all'età specificata nella colonna a sinistra, suddivisi per sesso.

età	numero viventi maschi	numero viventi femmine
0	100.000	100.000
40	98.120	98.976
70	82.269	89.970

Qual è la probabilità per **una quarantenne**, scelta a caso nella popolazione, di vivere almeno fino a 70 anni?

Per rispondere alla questione assumiamo innanzitutto che la popolazione sia stazionaria⁴ per quanto riguarda la durata della vita. Con questa ipotesi, possiamo interpretare la tabella dicendo che tra le 98.976 quarantenni della popolazione, 89.970 arriveranno a compiere 70 anni. Di conseguenza, la nostra valutazione della probabilità richiesta è

$$p(\text{"70 da 40"}) = \frac{89.970}{98.976} = 0,909 \dots$$

Osservazione

Con ciò abbiamo risposto, ma il nostro interesse per la questione si spinge oltre.

Consideriamo infatti l'evento "*una neonata vive almeno fino ai 70 anni*", e indichiamolo sinteticamente con "**70**"; coerentemente con quanto appena discusso, diciamo che la sua probabilità è

$$p(\text{"70"}) = \frac{89.970}{100.000} = 0,899 \dots$$

Abbiamo così ottenuto

$$p(\text{"70 da 40"}) > p(\text{"70"})$$

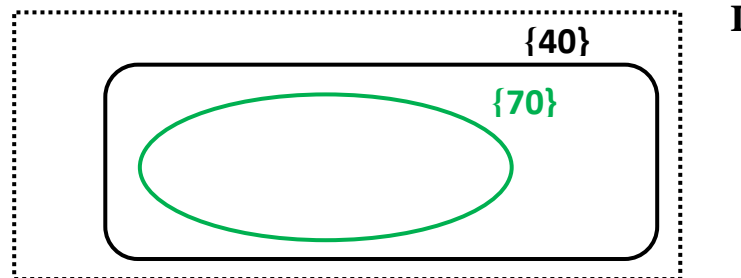
³ Tali dati non sono stati effettivamente rilevati su una popolazione seguita costantemente nell'arco di molti anni. Comunque per ottenerli ci si basa su rilevazioni anagrafiche (reali) relative ad un anno recente e le si elabora opportunamente.

⁴ Ossia il numero di nati è costante ogni anno e, per ogni classe di età, è costante il numero di individui che vi appartengono (questa seconda ipotesi si può anche esprimere in modo equivalente dicendo che, per ogni classe di età, il numero di ingressi è uguale al numero di uscite).

Perché i due valori di probabilità p ("70 da 40") e p ("70") sono diversi?

Le ragioni profonde di tale differenza risiedono nel fatto che nel valutare la probabilità dell'evento "70 da 40" si sfrutta un'informazione in più: l'individuo è arrivato a compiere i 40 anni. Come vedremo, in situazioni di questo tipo si parla di probabilità condizionata.

Più in dettaglio, possiamo schematizzare la situazione nel modo seguente



dove I è l'insieme della popolazione iniziale (femmine), $\{40\}$ l'insieme degli individui della popolazione I che arrivano a compiere 40 anni, e $\{70\}$ l'insieme degli individui della popolazione I che arrivano a compiere 70 anni.

- L'insieme dei casi **favorevoli** è $\{70\}$, sia per l'evento "70 da 40" che per l'evento "70"
- l'insieme dei casi **possibili** invece è diverso nelle due situazioni: per "70 da 40" "è" $\{40\}$, mentre per "70" è l'insieme I .
- perciò, anche le probabilità dei due eventi in esame dovranno essere diverse.

In particolare, visto che tra i due insiemi di casi possibili vale la relazione $\{40\} \subset I$, si conclude che p ("70 da 40") $>$ p ("70").

b) Fumatori

Sulla popolazione U dei 51,4 milioni italiani che hanno almeno 15 anni, ben 11,3 milioni sono fumatori, di questi 5,1 milioni sono donne⁵.

Assumendo che le donne costituiscano il 52,5% della popolazione U , qual è la probabilità che **la donna** (estratta a caso tra le donne di U) sia fumatrice?

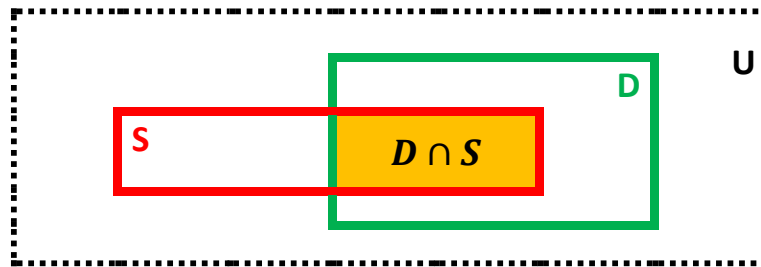
Per prima cosa osserviamo che abbiamo a che fare essenzialmente con tre insiemi di persone:

- l'insieme degli individui della popolazione, che denotiamo con U
- l'insieme dei fumatori della popolazione, che indichiamo con S
- l'insieme delle donne della popolazione, che chiamiamo D .

La prima frase del testo permette di affermare che l'insieme S e l'insieme D hanno intersezione non vuota. Inoltre, l'informazione contenuta nella seconda frase, consente di escludere che l'insieme S sia contenuto in D .

⁵ Dati da indagine Doxa del 2014 per conto dell'ISS, Istituto Superiore Sanità
<http://www.iss.it/binary/pres/cont/schedeFUMO.pdf>

Pertanto possiamo schematizzare la situazione mediante i diagrammi di Venn, nel modo seguente:



A questo punto non resta che valutare la probabilità richiesta, ossia la probabilità di “essere fumatore **per la donna** della popolazione”.

Dunque, qual è l’insieme dei “**casi possibili**” relativamente a tale richiesta? Non può essere l’intero insieme U, dato che all’insieme iniziale si è aggiunta l’ipotesi “per la donna della popolazione”, ma il più ristretto insieme **D**. Analogamente l’insieme dei “**casi favorevoli**” è **D ∩ S**.

Di conseguenza la probabilità richiesta è:

$$\frac{5,1 \cdot 10^6}{27,0 \cdot 10^6} \approx 0,189 \dots$$

Osservazione

Proviamo ora a togliere l’ipotesi aggiuntiva, ossia prendiamo in esame l’evento “essere fumatore per un individuo scelto sull’intera popolazione”. Con questa assunzione, l’insieme dei casi possibili diventa l’intero insieme iniziale U, l’insieme dei casi favorevoli l’insieme S e il valore di probabilità è

$$p(S) = \frac{11,3 \cdot 10^6}{51,4 \cdot 10^6} \approx 0,220$$

Come ci si aspetta, è un valore diverso da quello ottenuto considerando l’ipotesi aggiuntiva “per la donna”. In altre parole, l’informazione aggiuntiva modifica il valore della probabilità dell’evento “essere fumatore”, facendolo passare dal 22% al 18,9%.

Il divario risulta ancora più marcato se si confronta la probabilità di essere fumatore per una donna con quella di essere fumatore per un uomo: quest’ultima si attesta infatti sul 25,4%, come si può dedurre dai dati forniti nel testo del quesito. L’esame dei dati relativi al 2013 evidenzia una differenza ancora maggiore: 15,3% per le donne contro il 26,3% per gli uomini.

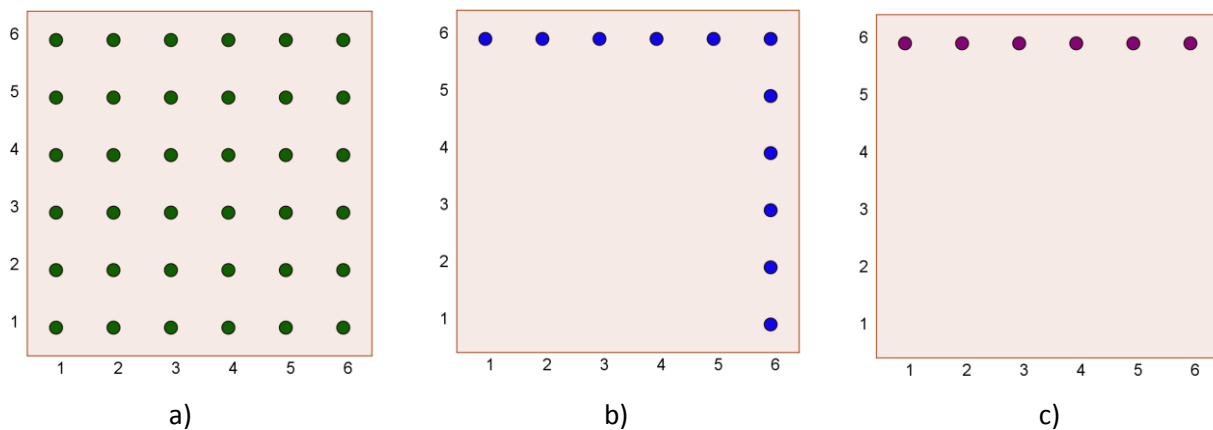
Possiamo riassumere quanto visto finora nella sezione, dicendo che la probabilità richiesta in questo quesito è una probabilità che dipende dal fatto che l’evento D si è realizzato e pertanto si indica spesso come “probabilità condizionata”. Il suo valore è diverso da quello della probabilità dell’evento “essere fumatore”, che usualmente si denota con $p(S)$. Dunque per la probabilità condizionata dovremmo utilizzare un simbolo diverso: come discuteremo in un contesto più generale in una delle prossime sezioni, utilizzeremo l’espressiva notazione $p_D(S)$ per indicare che la sua valutazione è condizionata al verificarsi dell’evento D.

c) Lancio di due dadi

Calcola la probabilità che in un lancio di due dadi, uno bianco e l'altro giallo, escano due "6"

- senza informazioni aggiuntive;
- sapendo che nel lancio è uscito **almeno** un "6";
- sapendo che l'esito del **dado giallo** è "6".

Le tre situazioni a), b), c) si possono modellizzare in modo espressivo mediante tabelle a doppia entrata, segnando con un puntino ogni cella che corrisponde ad uno dei casi possibili. Ecco un esempio di cosa si ottiene seguendo tale proposito:



Tali modelli si possono visualizzare in modo ancora più efficace mediante opportune rappresentazioni dinamiche, come quelle offerte dal software Geogebra⁶, per esempio usando il file [LancioDadi Condizionata.ggb](#).

Sulla base di tali rappresentazioni si dovrebbero determinare presto i valori di probabilità richiesti: essi sono

a) $\frac{1}{36}$

b) $\frac{1}{11}$

c) $\frac{1}{16}$

Osservazione

I tre schemi grafici mettono chiaramente in evidenza il fatto, a nostro avviso, didatticamente più rilevante: le nuove informazioni sugli esiti dei lanci modificano l'insieme dei "casi possibili".

⁶ Il file proposto è stato ideato dalla collega Francesca Arrigoni.