



UNIVERSITÀ  
DEGLI STUDI DI BARI  
ALDO MORO

# ELEMENTI DI PROBABILITÀ E STATISTICA: LEZ. 3

Prof. Roberto Capone  
A.A. 2023/24  
Corso di Laurea in Scienze Biologiche



# Probabilità: introduzione

## DEFINIZIONI

Un esperimento aleatorio è un fenomeno di cui non riusciamo a prevedere il risultato con certezza.

L'insieme  $U$  di tutti i possibili risultati di un esperimento aleatorio si chiama spazio campionario o universo.

Un evento è un qualunque sottoinsieme dello spazio campionario; un evento formato da un singolo risultato dell'esperimento è detto evento elementare.

Chiamiamo spazio degli eventi l'insieme di tutti gli eventi che si possono associare a un esperimento, cioè l'insieme delle parti di  $U$

# Definizione classica di probabilità

## DEFINIZIONE

La probabilità di un evento  $E$  è il rapporto fra il numero dei casi favorevoli  $f$  e quello dei casi possibili  $u$  quando sono tutti ugualmente possibili.

$$P(E) = \frac{f}{u}$$

Possiamo fare le seguenti osservazioni.

Poiché il numero  $f$  dei casi favorevoli è sempre minore o uguale al numero dei casi possibili,

$$0 \leq p(E) \leq 1$$

cioè la probabilità di un evento è sempre compresa tra 0 e 1.

Se  $f = u$ , poiché il numero dei casi favorevoli è uguale al numero dei casi possibili:

$P(E) = 1$  e l'evento è certo.

Se  $f = 0$ , poiché il numero dei casi favorevoli è nullo:  $p(E) = 0$  e l'evento è impossibile.

# Definizione classica di probabilità

## Evento contrario

Consideriamo un evento  $E$ . Il suo evento contrario  $\bar{E}$  è l'evento che si verifica se e solo se non si verifica  $E$ .

Per esempio, nel lancio di un dado l'evento contrario di  $E = \text{«esce un numero pari»}$  è  $\bar{E} = \text{«non esce un numero pari»}$ , ossia «esce un numero dispari». È vero che:

$$P(\bar{E}) = 1 - P(E)$$

## Esempio

Nel lancio di un dado l'evento  $\bar{E} = \text{«non esce il numero 6»}$  è l'evento contrario dell'evento  $E = \text{«esce il numero 6»}$ , quindi:

$$P(\bar{E}) = 1 - P(E) = 1 - \frac{1}{6} = \frac{5}{6}$$

# Definizione classica di probabilità

## Esempio

Nel lancio di un dado l'evento  $\bar{E}$  = «non esce il numero 6» è l'evento contrario dell'evento  $E$  = «esce il numero 6», quindi:

$$P(\bar{E}) = 1 - P(E) = 1 - \frac{1}{6} = \frac{5}{6}$$

Dal punto di vista degli insiemi, dato l'insieme corrispondente a un evento  $E$ , al suo evento contrario  $\bar{E}$  corrisponde l'insieme complementare di  $E$  rispetto a  $U$ .

Nell'esempio precedente,  $U = \{1, 2, 3, 4, 5, 6\}$ ,  $E = \{6\}$ ,  $\bar{E} = U - E = \{1, 2, 3, 4, 5\}$ .

Per quanto abbiamo detto, la somma della probabilità di un evento e di quella del suo evento contrario è 1

# Probabilità e calcolo combinatorio

## ESEMPIO

Da un'urna contenente 4 palline bianche e 6 nere estraiamo consecutivamente 5 palline, senza rimettere ogni volta la pallina estratta nell'urna.

Consideriamo l'evento:  $E_1 =$  «escono consecutivamente, nell'ordine, 2 palline bianche e 3 nere». Dobbiamo supporre di distinguere per l'ordine di uscita ogni possibile raggruppamento, anche se identico ad altri per composizione. I casi possibili si possono quindi calcolare con le disposizioni semplici:

$$D_{10,5} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5$$

I casi favorevoli sono tutti i gruppi formati da 2 palline bianche delle quattro contenute nell'urna e dai gruppi formati da 3 palline nere delle sei contenute nell'urna:

$$D_{4,2} \cdot D_{6,3}$$

$$P(E_1) = \frac{D_{4,2} \cdot D_{6,3}}{D_{10,5}}$$

# Definizione statistica di probabilità

Abbiamo un'urna che contiene palline colorate, ma non sappiamo né quali sono i colori, né quante sono le palline e, inoltre, non possiamo aprire l'urna per esaminarne il contenuto. L'unico procedimento che ci rimane per acquisire conoscenze è quello di estrarre a sorte una pallina un gran numero di volte, rimettendo ogni volta la pallina estratta nell'urna, in modo che ogni estrazione sia effettuata nelle stesse condizioni.

Effettuiamo consecutivamente 80 estrazioni e ogni volta prendiamo nota del colore uscito, e aggiorniamo il corrispondente numero di estrazioni nella seconda colonna della tabella a lato. Usando i valori ottenuti, calcoliamo il rapporto fra il numero delle volte in cui è uscito un determinato colore e il numero delle 40 prove effettuate tutte nelle stesse condizioni, cioè la sua frequenza relativa, e lo riportiamo nella terza colonna della tabella.

# Definizione statistica di probabilità

Effettuiamo consecutivamente 80 estrazioni e ogni volta prendiamo nota del colore uscito, e aggiorniamo il corrispondente numero di estrazioni nella seconda colonna della tabella a lato. Usando i valori ottenuti, calcoliamo il rapporto fra il numero delle volte in cui è uscito un determinato colore e il numero delle 40 prove effettuate tutte nelle stesse condizioni, cioè la sua frequenza relativa, e lo riportiamo nella terza colonna della tabella.

Colore	Numero estrazioni	Frequenza relativa
rosso	5	$\frac{1}{16}$
giallo	18	$\frac{9}{40}$
nero	22	$\frac{11}{40}$
verde	35	$\frac{7}{16}$
<b>Totale</b>	80	1

## DEFINIZIONE

La frequenza relativa  $f(E)$  di un evento  $E$  sottoposto a  $n$  esperimenti, effettuati tutti nelle stesse condizioni, è il rapporto fra il numero delle volte  $m$  in cui  $E$  si è verificato e il numero  $n$  delle prove effettuate.



# Definizione statistica di probabilità

I valori della frequenza relativa di un evento sono compresi tra 0 e 1:

$$0 \leq f(E) \leq 1.$$

Frequenza 0 non significa che l'evento è impossibile, ma soltanto che non si è mai verificato.

Per esempio, se nelle prove effettuate non è mai uscita una pallina blu, questo non significa che nell'urna non ve ne siano.

Frequenza 1 non significa che l'evento è certo, ma soltanto che in quella serie di esperimenti è stato sempre osservato.

Se ripetiamo l'esperimento, senz'altro otterremo valori diversi. Il valore  $m$  dipende infatti dal numero di prove  $n$  che effettuiamo. Ma se abbiamo la pazienza di aumentare il numero delle prove, rileviamo un fatto interessante: il valore della frequenza relativa

$$f(E) = \frac{m}{n}$$

tende a un valore costante che si può ritenere come la probabilità dell'evento

# Definizione statistica di probabilità

Per esempio, sappiamo che la probabilità di ottenere testa lanciando una moneta, secondo l'impostazione classica, è  $\frac{1}{2}$ .

Allo stesso valore tende la frequenza relativa se sperimentalmente lanciamo una moneta un numero elevatissimo di volte.

In generale, vale la seguente proprietà:

## Legge empirica del caso

Dato un evento  $E$ , sottoposto a  $n$  prove tutte nelle stesse condizioni, il valore della frequenza relativa

$$f(E) = \frac{m}{n}$$

tende al valore della probabilità  $p(E)$ , all'aumentare del numero  $n$  di prove effettuate.

La legge empirica del caso è alla base della definizione statistica o frequentistica della probabilità.

# Definizione statistica di probabilità

## DEFINIZIONE

La probabilità statistica di un evento  $E$  è la frequenza relativa del suo verificarsi quando il numero di prove effettuato è da ritenersi «sufficientemente alto»

Nell'impostazione classica il valore della probabilità è calcolato a priori, ossia prima che l'esperimento avvenga, mentre il valore della frequenza è un valore a posteriori.

Ci sono moltissimi eventi per i quali è difficile o impossibile calcolare la probabilità applicando l'impostazione classica, per esempio:

- produzione di un pezzo difettoso con un macchinario;
- trovare un posto di lavoro;
- contrarre una determinata malattia;
- incidente automobilistico;
- efficacia di un farmaco.

Per questi eventi occorre fondare il calcolo sui dati accumulati in passato e calcolare le relative frequenze, accettandole come probabilità degli eventi.

# Definizione statistica di probabilità

## ESEMPIO

Una ditta farmaceutica vuole sperimentare un nuovo vaccino antinfluenzale. Si sottopongono volontariamente al vaccino 10 000 persone e, nell'inverno successivo, 6750 di queste non contraggono l'influenza. Qual è la probabilità di non contrarre l'influenza con questo tipo di vaccino?

Poiché il numero di prove effettuate è 10 000 e il numero di esiti positivi è 6750:

$$f = \frac{6750}{10000} = 0,675$$

La probabilità statistica di non ammalarsi è del 67,5%

# Definizione assiomatica di probabilità

L'impostazione assiomatica nasce con lo scopo di superare questi aspetti particolari, per giungere a una formulazione rigorosa, utilizzando la teoria degli insiemi. Basandosi su di essa vengono fornite le definizioni fondamentali di spazio campionario, evento, evento elementare, spazio degli eventi, evento contrario, evento unione ed evento intersezione che abbiamo già esaminato, oltre a definire l'evento impossibile come quello corrispondente all'insieme vuoto e l'evento certo come quello relativo all'insieme  $U$  dello spazio campionario.

## DEFINIZIONE

Dato uno spazio campionario  $U$ , una funzione  $p$  che associa a ogni evento  $E$  dello spazio degli eventi un numero reale viene detta probabilità se soddisfa i seguenti assiomi:

1.  $P(E) \geq 0$
2.  $P(U) = 1$
3. Se  $E_1 \cap E_2 = \emptyset$ , allora  $p(E_1 \cup E_2) = p(E_1) + p(E_2)$

# Definizione assiomatica di probabilità

## ESEMPIO

L'insieme  $U$  è costituito da tre eventi elementari  $A$ ,  $B$ ,  $C$ , i quali hanno le seguenti probabilità:

$$p(A) = \frac{3}{8}, \quad p(B) = \frac{2}{5}, \quad p(C) = \frac{9}{40}$$

Tali valori soddisfano gli assiomi della probabilità. In particolare, essendo

$$A \cup B \cup C = U$$

e  $A \cap B = A \cap C = B \cap C = \emptyset$

$$p(A) + p(B) + p(C) = \frac{3}{8} + \frac{2}{5} + \frac{9}{40} = 1 = P(U)$$

# Definizione assiomatica di probabilità

## Proprietà

Dalla definizione assiomatica si deducono le seguenti proprietà:

$$p(\emptyset) = 0$$

$$0 \leq p(E) \leq 1$$

$$p(\bar{E}) = 1 - p(E)$$

Se gli eventi  $E_1, E_2, \dots, E_n$  sono una partizione di  $U$ , allora

$$p(E_1) + p(E_2) + \dots + p(E_n) = 1$$

$$p(E_2 - E_1) = p(E_2) - p(E_1 \cap E_2)$$

# Definizione assiomatica di probabilità

## ESEMPIO

Da un'indagine è risultato che, preso come campione un certo numero di italiani, la probabilità che essi trascorrano le prossime vacanze al mare è del 45%, in montagna è del 27%, sia al mare che in montagna è del 5%. Calcoliamo la probabilità che gli italiani:

- vadano solo al mare;
- non vadano in montagna;
- non vadano al mare né in montagna.

Consideriamo gli eventi:  $A$  = «andare al mare»;  $B$  = «andare in montagna».  
Conosciamo le probabilità:

$$P(A) = \frac{45}{100}; P(B) = \frac{27}{100}; P(A \cap B) = \frac{5}{100}$$

- L'evento considerato è  $A - B$ :

$$P(A - B) = p(A) - p(A \cap B) = 0,45 - 0,05 = 0,40 = 40\%$$



# Definizione assiomatica di probabilità

## ESEMPIO

Da un'indagine è risultato che, preso come campione un certo numero di italiani, la probabilità che essi trascorrano le prossime vacanze al mare è del 45%, in montagna è del 27%, sia al mare che in montagna è del 5%. Calcoliamo la probabilità che gli italiani:

- vadano solo al mare;
- non vadano in montagna;
- non vadano al mare né in montagna.

b. L'evento considerato è  $\bar{B}$ :

$$P(\bar{B}) = 1 - P(B) = 1 - 0,27 = 0,73 = 73\%$$

c. L'evento considerato è  $\overline{A \cup B}$ , per cui, dato che  $A \cup B = (A - B) \cup B$ , si ha:

$$P(\overline{A \cup B}) = 1 - [0,40 + 0,27] = 0,33 = 33\%$$

# Somma logica di eventi

## Eventi unione ed eventi intersezione

Ogni evento è un sottoinsieme dello spazio campionario, quindi possiamo utilizzare gli insiemi per definire eventi unione ed eventi intersezione.

### DEFINIZIONE

Dati due eventi  $E_1$  ed  $E_2$  di uno stesso spazio campionario:

- l'evento unione o somma logica è l'evento  $E_1 \cup E_2$  che si verifica quando è verificato almeno uno degli eventi  $E_1$  o  $E_2$ ;
- l'evento intersezione o prodotto logico è l'evento  $E_1 \cap E_2$  che si verifica quando sono verificati entrambi gli eventi  $E_1$  ed  $E_2$ .

La **o** e la **e** che abbiamo utilizzato nella definizione corrispondono alle operazioni logiche tra le proposizioni che descrivono gli eventi.

Infatti, date due proposizioni logiche  $p$  e  $q$ :

- la loro disgiunzione  $p \vee q$ , cioè  $p$  o  $q$ , è falsa quando entrambe  $p$  e  $q$  sono false, è vera negli altri casi;
- la loro congiunzione  $p \wedge q$ , cioè  $p$  e  $q$ , è vera quando entrambe  $p$  e  $q$  sono vere, è falsa negli altri casi.

L'evento unione viene anche detto **evento totale**, mentre l'evento intersezione è anche detto **evento composto**.

# Somma logica di eventi

## ESEMPIO

Estraiamo una pallina da un'urna che ne contiene 6, numerate da 1 a 6, e consideriamo:

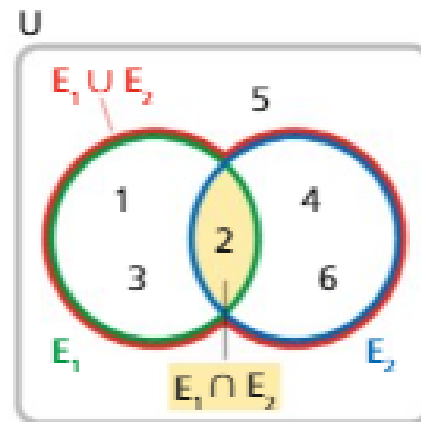
$E_1$  = «esce un numero minore di 4»,

$E_2$  = «esce un multiplo di 2»

Abbiamo che:

$E_1 \cup E_2$  = «esce un numero minore di 4 o multiplo di 2» = {1, 2, 3, 4, 6};

$E_1 \cap E_2$  = «esce un numero minore di 4 e multiplo di 2» = {2}



# Somma logica di eventi

## Eventi compatibili ed eventi incompatibili

### ESEMPIO

Estraiamo una pallina da un'urna contenente 12 palline numerate da 1 a 12 e consideriamo gli eventi

$E_1$  = «esce un numero pari»,

$E_2$  = «esce un numero maggiore di 7»

Osserviamo che gli eventi  $E_1$  ed  $E_2$  possono verificarsi contemporaneamente quando il risultato è un elemento di

$$E_1 \cap E_2 = \{8, 10, 12\}.$$

In casi come questo si dice che gli eventi sono compatibili

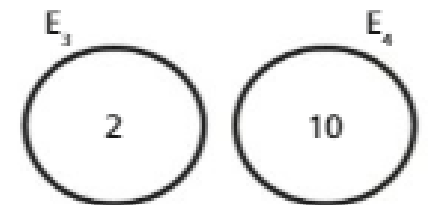
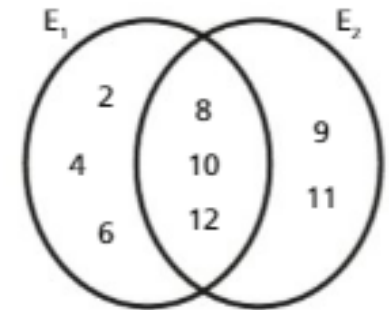
Consideriamo ora gli eventi:

$E_3$  = «esce il numero 2»;

$E_4$  = «esce il numero 10».

$$E_3 \cap E_4 = \emptyset$$

Questi due eventi, invece, non possono verificarsi contemporaneamente. Eventi di questo tipo sono incompatibili.

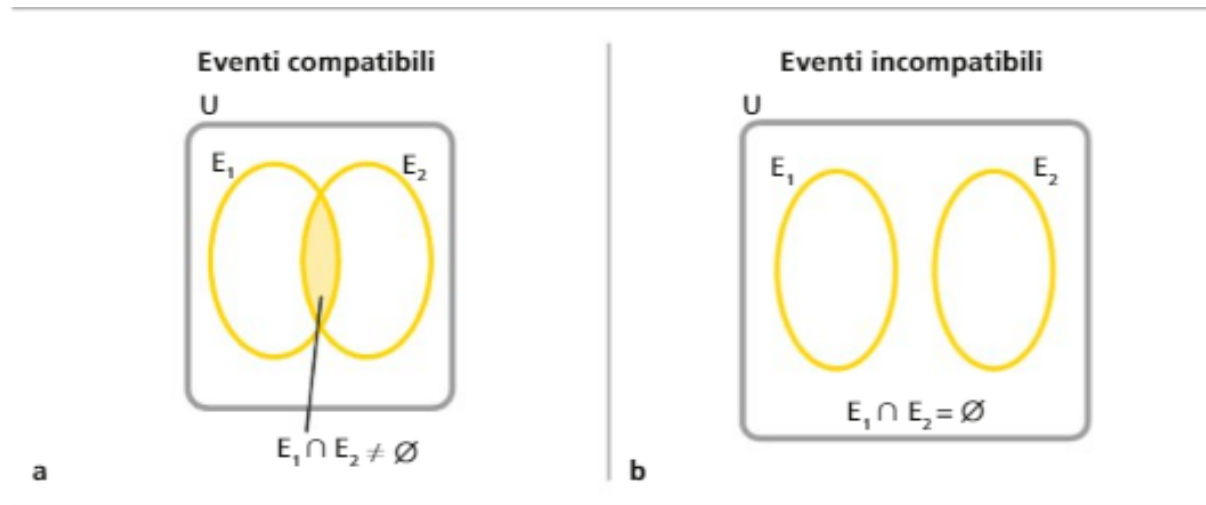


# Somma logica di eventi

## Eventi compatibili ed eventi incompatibili

### Definizione

Due eventi  $E_1$  ed  $E_2$  relativi allo stesso spazio campionario, sono incompatibili se il verificarsi di uno esclude il verificarsi contemporaneo dell'altro, cioè  $E_1 \cap E_2 = \emptyset$   
In caso contrario sono compatibili.



# Somma logica di eventi

## Probabilità della somma logica o unione di due eventi

### TEOREMA

La probabilità della somma logica di due eventi  $E_1$  ed  $E_2$  è uguale alla somma delle loro probabilità diminuita della probabilità del loro evento intersezione:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

In particolare, se gli eventi sono incompatibili:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

# Eventi indipendenti

## Precisazioni

La formula

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

esprime la probabilità dell'evento Unione in funzione della probabilità dei due eventi e dell'evento intersezione.

Cosa possiamo dire sulla probabilità dell'evento intersezione?

In generale poco! E la probabilità che avvengano entrambi gli eventi e dipenderà quindi da che relazione c'è tra i due eventi.

In generale, se 2 eventi non si influenzano a vicenda, allora la probabilità dell'evento intersezione è il prodotto delle probabilità dei due eventi trasformiamo questo suggerimento in una definizione:

diremo che due eventi  $E_1$  e  $E_2$  sono indipendenti se e solo se

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

La prima formula è una formula sempre vera valida per qualsiasi coppia di eventi la seconda invece è una condizione sui due eventi. Se è verificata diciamo che gli eventi sono indipendenti ma può benissimo non essere verificata

# Eventi indipendenti

## Precisazioni

Non confondere i concetti di eventi incompatibili e eventi indipendenti.

L'essere incompatibili è una relazione puramente insiemistica che non dipende dal calcolo delle probabilità:

due eventi  $E_1$  e  $E_2$  sono incompatibili se il solo si sono disgiunti come insiemi cioè  $E_1 \cap E_2 = \emptyset$ .

Invece l'essere indipendenti è una caratteristica che riguarda la probabilità degli eventi e non solo l'insiemistica:

Due eventi  $E_1$  e  $E_2$  Sono indipendenti se e solo se  $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$ .

Anzi due eventi incompatibili di probabilità non nulla non possono mai essere indipendenti



# Estrazioni con o senza rimbussolamento

## Esempio

Un cassetto contiene 12 calzini bianchi e quattro calzini neri. Qual è la probabilità di estrarre un paio spaiato?

Per rispondere a questa domanda chiariamo prima come vogliamo effettuare l'estrazione. Il modo naturale di procedere se vogliamo indossarli consiste nell'estrarre prima uno e poi l'altro; in questo caso si parla di estrazione senza rimbussolamento.

Se invece vogliamo limitarci ad annotare il colore dei calzini senza indossarli, possiamo estrarne uno, segnarcene il colore, rimetterlo dentro, ed estrarne il secondo. In questo caso si parla di estrazione con rimbussolamento.

Vediamo come cambia la probabilità.

Nel primo caso alla prima estrazione abbiamo una probabilità di  $12/16$  di estrarre un calzino bianco e  $4/16$  di estrarre un calzino nero.

Ma se abbiamo estratto già un calzino bianco la probabilità di estrarre poi un calzino nero è  $4/15$ .

Essendo le due estrazioni indipendenti la probabilità di estrarre prima un calzino bianco e poi uno nero è:

$$\frac{3}{4} \cdot \frac{4}{15} = \frac{1}{5}$$

# Estrazioni con o senza rimbussolamento

## Esempio

Un cassetto contiene 12 calzini bianchi e quattro calzini neri. Qual è la probabilità di estrarre un paio spaiato?

In modo analogo si vede che la probabilità di estrarre prima un calzino nero e poi un calzino bianco è ancora  $1/5$ .

Pertanto la probabilità di estrarre un paio spaiato è  $1/5$  più  $1/5$  ovvero  $2/5$ .

Nel caso con rimbussolamento, la probabilità di estrarre prima un calzino bianco e poi un calzino nero è:

$$\frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$$

Anche la probabilità di estrarre prima un calzino nero e poi uno bianco è  $3/16$  per cui la probabilità di estrarre con rimbussolamento un paio spaiato è  $3/16+3/16$  ovvero  $3/8$ , lievemente minore rispetto a quella senza rimbussolamento.

# Distribuzioni di probabilità

Adesso che sappiamo cosa sono gli eventi possiamo cercare di capire come attribuire delle probabilità agli eventi semplici.

Ci sono 2 metodi per attribuire una la probabilità agli eventi semplici:

- a priori, in funzione di ipotesi teoriche sulla natura degli eventi semplici (a volte si parla di probabilità teorica o stimata)
- a posteriori in funzione di misurazioni effettuate sul fenomeno che si sta studiando (a volte si parla di probabilità misurata)

In entrambi i casi l'attribuzione delle probabilità deve sottostare ad alcune semplici condizioni già viste:

1. L'evento certo ha probabilità 1, l'evento impossibile probabilità 0
2. tutti gli eventi hanno una probabilità compresa tra 0 e 1 inclusi
3. la probabilità che avvenga uno di due eventi incompatibili è la somma della probabilità dei due eventi

# Distribuzioni di probabilità

## Esempio 1

Nel lancio di un dado a sei facce non truccato lo spazio degli eventi è  $U = \{1,2,3,4,5,6\}$ . Dire che il dado non è truccato vuol dire che tutte le facce sono uguali; non c'è nulla che favorisca una faccia rispetto alle altre. Quindi è naturale attribuire a ciascun risultato la stessa probabilità. Siccome ci sono sei risultati possibili, a due a due incompatibili, ognuno di essi ha probabilità  $p$ . Dunque l'ipotesi dado non truccato si traduce nell'assegnare la stessa probabilità di  $1/6$  a tutti i possibili risultati del lancio di un dado

## Esempio 2

Nel lancio di due dadi distinti a sei facce non truccati lo spazio degli eventi è composto da 36 possibili risultati. Nel caso di due dadi la frase *non truccati* non significa solo che le facce di ciascun dado sono tutte uguali ma anche che i due dadi sono indipendenti, nel senso che il risultato di un dado non influenza in alcun modo il risultato dell'altro. Tutti i 36 risultati sono equi probabili. Pertanto ciascun caso possibile avrà una probabilità di  $1/36$

# Legge di disgiunzione di Mendel

Il corredo genetico di ogni individuo è composto da coppie di cromosomi. Un figlio riceve un cromosoma da ciascun genitore; quindi in caso di geni con più alleli il genotipo del figlio comprende un allele proveniente dal padre e un allele proveniente dalla madre.

Un'importante ipotesi teorica che sottende la trasmissione del genotipo da genitori a figli e la legge di disgiunzione di Mendel:

*Le possibili trasmissioni di alleli da genitori a figli sono tutte equiprobabili*

Vediamo come questa legge ci permette di stabilire la probabilità che un figlio abbia un certo genotipo conoscendo i genotipi dei genitori.

Prendiamo un gene con due possibili alleli che indicheremo con A e a.

Supponiamo che il padre abbia genotipo AA e la madre genotipo Aa. Allora il figlio può ricevere gli alleli dai genitori in quattro modi diversi:

Il primo allele A del padre e il secondo allele A della madre (genotipo del figlio AA)

Il primo allele A del padre e il secondo allele a della madre (genotipo del figlio Aa)

Il secondo allele A del padre e il primo allele A della madre (genotipo del figlio AA)

Il secondo allele A del padre e il secondo allele a della madre (genotipo del figlio Aa)

# Legge di disgiunzione di Mendel

La legge di disgiunzione di Mendel ci dice che queste quattro eventualità sono equiprobabili; quindi ciascuna di esse ha probabilità  $\frac{1}{4}$ . Siccome otteniamo come genotipo del figlio AA in due casi e Aa allora il figlio ha genotipo AA con probabilità  $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$  e genotipo Aa con probabilità  $\frac{1}{2}$ .

		Padre		
		AA	Aa	aa
Madre	AA	AA : 1 Aa : 0 aa : 0	AA : 1/2 Aa : 1/2 aa : 0	AA : 0 Aa : 1 aa : 0
	Aa	AA : 1/2 Aa : 1/2 aa : 0	AA : 1/4 Aa : 1/2 aa : 1/4	AA : 0 Aa : 1/2 aa : 1/2
	aa	AA : 0 Aa : 1 aa : 0	AA : 0 Aa : 1/2 aa : 1/2	AA : 0 Aa : 0 aa : 1

Distribuzione di  
probabilità dei genotipi

# Alleli dominanti e recessivi

Un allele è dominante quando la sua presenza nel genotipo causa la comparsa di una certa caratteristica (detta fenotipo) nell'individuo indipendentemente da quale sia l'altro allele; è recessivo negli altri casi.

Un tipico esempio di gene con due alleli di cui uno dominante e uno recessivo è quello che determina la presenza o meno del fattore  $Rh$  nel sangue. L'allele  $Rh^+$  è dominante sull'altro allele  $Rh^-$ : la sua presenza nel genotipo implica la presenza nel sangue del fattore  $Rh$  (fenotipo  $Rh^+$ ) mentre l'unico modo perché nel sangue non ci sia il fattore  $Rh$  (fenotipo  $Rh^-$ ) è che entrambi gli alleli siano  $Rh^-$ .

Se Leonardo e Silvia hanno fenotipo  $Rh^+$  possono avere genotipo omozigote  $Rh^+ Rh^+$  più oppure eterozigote  $Rh^+ Rh^-$ . Se anche uno solo dei due ha genotipo omozigote  $Rh^+ Rh^+$  la tabella della pagina precedente ci dice che la probabilità che abbiano un figlio con fenotipo  $Rh^-$  (e quindi genotipo  $Rh^- Rh^-$ ) è 0.

Invece se entrambi hanno genotipo eterozigote  $Rh^+ Rh^-$  la probabilità che abbiano un figlio con fenotipo  $Rh^-$  è  $\frac{1}{4}$ .

# Frequenza relativa - esempi

Se non abbiamo alcuna ipotesi teorica a cui aggrapparci per attribuire a priori delle probabilità ai possibili eventi del fenomeno che stiamo studiando dobbiamo metterci a misurare. Ripetiamo lo stesso esperimento un numero possibilmente grande di volte.

Diremo che la frequenza assoluta ho numero di successi di un evento e il numero di volte che l'evento si verifica ripetendo l'esperimento e la frequenza relativa è il rapporto fra la frequenza assoluta e il numero degli esperimenti o numero di tentativi effettuati.

Allora l'attribuzione a posteriori della probabilità consiste nell'assegnare come probabilità di ciascun evento la sua frequenza relativa

## Esempio 1

Vogliamo vedere qual è la probabilità che una cellula epiteliale di un ratto subisca una mutazione quando sottoposta ad una radioattività di 1 Roentgen (corrisponde a  $2.58 \cdot 10^{-4} C/kg$ ). Sottoponiamo 1000 e cellule epiteliali di ratto ha una radiazione di 1 R; si ne sono mutate 5 possiamo dire che la probabilità (a posteriori) di mutazione è  $5/1000$  cioè 0,005.



# Frequenza relativa - esempi

## Esempio 2

In un esperimento vengono pesate 15 cavie ottenendo i seguenti pesi in grammi:

28, 32, 37, 29, 31, 30, 32, 26, 32, 27, 29, 30, 28, 31, 31.

Qual è la probabilità che una cavia presa a caso fra queste pesi meno di 29 grammi?

Per rispondere a questa domanda contiamo: 4 cavie su 15 pesano meno di 29 grammi; quindi la probabilità (a posteriori) è  $4/15$ .

Analogamente 6 cavie su 15 pesano almeno 31 grammi per cui la probabilità che una cavia presa a caso fra queste pesi almeno 31 grammi è  $6/15$

# Frequenza relativa - esempi

## Esempio 3

All'ospedale San Camillo di Roma nel periodo 16 Aprile - 15 ottobre 2002 hanno registrato il peso dei bambini nati nel reparto maternità ottenendo i dati riassunti in tabella. Qual è la probabilità che un neonato preso a caso tra questi pesi più di 1500 grammi?

<i>Peso alla nascita</i>	<i>Frequenza assoluta</i>	<i>Frequenza relativa</i>
≤ 1000 g	9	$9/116 \approx 8\%$
1001-1500 g	21	$21/116 \approx 18\%$
1501-2500 g	50	$50/116 \approx 43\%$
> 2500 g	36	$36/116 \approx 31\%$
<b>Totale</b>	116	$116/116 = 100\%$

Siccome 86 neonati su 116 pesano più di 1500 grammi la risposta è  $86/116$  circa il 74%.

Nota che le probabilità essendo numeri compresi tra zero e uno sono spesso indicate come percentuali.

# La legge di Hardy - Weinberg

L'equilibrio di Hardy-Weinberg, o legge di Hardy-Weinberg, è un'applicazione alla genetica del concetto di eventi indipendenti.

È un modello della genetica delle popolazioni che postula che all'interno di una popolazione (panmictica ideale) vi è equilibrio delle frequenze alleliche e genotipiche da una generazione all'altra, ovvero queste non cambiano con il passare del tempo a meno che non intervengano fattori specifici atti a disturbare l'equilibrio stesso.

La legge di Hardy-Weinberg risponde alla domanda: "Come mai, se una mutazione è dominante, non siamo tutti soggetti a suddetta mutazione?". Punnet pose questa domanda a Hardy, uno dei più noti matematici inglesi, che dopo poco tempo trovò la soluzione. Semplificò il problema indicando che non ci fossero influenze esterne e concluse che la libertà che si manifestava nella mutazione aveva essenzialmente due conseguenze:

- Le frequenze degli alleli in una popolazione presa ad esame rimanevano costanti.
- Le relative frequenze dei diversi genotipi sarebbero rimaste le stesse dopo la prima generazione.

# La legge di Hardy - Weinberg

Hardy e il medico tedesco Weinberg formularono indipendentemente questa formula

$$p^2 + 2pq + q^2 = 1$$

o equivalentemente

$$(p + q)^2 = 1$$

Supponiamo di avere una popolazione composta da individui e consideriamo un gene con due possibili alleli A e a.

Vogliamo studiare come varia la distribuzione dei genotipi da una generazione all'altra.

I genomi che ci aspettiamo sono dunque:

- $f(AA) = p^2$ , probabilità che un uovo contenente l'allele "A" si incontri con uno spermatozoo contenente "A".
- $f(Aa) = 2pq$ , somma della probabilità che un uovo con l'allele "A" si incontri con un gamete maschile contenente "a" e della probabilità che un gamete femminile con "a" incontri uno spermatozoo avente "A".
- $f(aa) = q^2$ , probabilità che un gamete maschile e uno femminile, entrambi con l'allele "a", si incontrino.

# La legge di Hardy - Weinberg

La legge di H-W descrive la situazione più semplice di equilibrio genetico di una popolazione: nonostante le condizioni dell'equilibrio di Hardy-Weinberg sembrano difficili da ottenere, esse valgono per molti caratteri in parecchie situazioni

Le condizioni per cui un locus in una popolazione segue la legge di H-W sono le seguenti.

HW1 Popolazione praticamente infinita. Ciò è richiesto affinché si possa applicare la legge dei grandi numeri, e quindi le frequenze siano praticamente coincidenti con le probabilità. Sorprendentemente, basta una popolazione di poche centinaia di individui, pur essendo possibili (ma improbabili) fluttuazioni.

HW2 Assenza di immigrazione ed emigrazione. In questo modo il pool genetico è influenzato solo dalle sue dinamiche interne.

HW3 Panmissia (incrocio casuale). Significa che la probabilità che due individui si incrocino non è influenzata dal fenotipo del carattere in questione. In questo modo è come se i geni di tutti gli individui fossero mescolati nel pool genetico ed estratti a sorte per creare i genotipi dei nuovi individui. La panmissia manca, ad esempio, nel caso di forti preferenze matrimoniali all'interno di caste chiuse, specie se con diversa origine etnica.

# La legge di Hardy - Weinberg

HW4 Non selezione. Il successo riproduttivo medio degli individui (detto anche fitness) non deve essere influenzato dal genotipo per il carattere in questione. I due (o più) alleli devono quindi avere la stessa probabilità, una volta presenti, di essere trasmessi alle successive generazioni.

HW5 Non mutazione. Ovviamente le mutazioni alterano la composizione del pool genetico delle nuove generazioni. Sono comunque eventi rari.

La legge di Hardy-Weinberg stabilisce che nelle condizioni suindicate le frequenze geniche rimangono costanti e le frequenze genotipiche si stabilizzano in una generazione in modo che la frequenza degli omozigoti sia il quadrato di quella dell'allele, mentre quelle degli eterozigoti saranno il doppio prodotto delle frequenze degli alleli posseduti.

# La legge di Hardy - Weinberg

Immaginiamo una popolazione in cui sono presenti l'allele A con frequenza  $p$  e l'allele  $a$  con frequenza  $q = (1 - p)$ .

Avremo  $f(AA) = p^2$ ,  $f(aa) = q^2$ ,  $f(Aa) = 2pq$ . Il nuovo valore di  $p$  sarà

$$p' = p^2 + pq = p(p + q) = p$$

Vediamo la **dimostrazione** con due alleli: il caso generale si ricava facilmente. Vigendo le condizioni su esposte,  $p$  e  $q$  saranno anche le probabilità che un gamete contenga uno o l'altro allele. Perché nasca un individuo con genotipo AA servirà la "coincidenza" di un gamete maschile A e uno femminile pure A. Poiché i due fatti sono indipendenti, la probabilità congiunta è il prodotto delle due probabilità, per cui:

$$f(AA) = \Pr(AA) = \Pr(\text{gamete masch. A}) * \Pr(\text{gamete femm. A}) = p * p = p^2$$

e lo stesso varrà per tutti gli omozigoti.

Gli eterozigoti si possono ottenere in due modi: A dal padre e a dalla madre o viceversa. Ognuno dei due casi ha probabilità  $pq$  per cui:

$$f(Aa) = \Pr(Aa) = \Pr[(\text{masch. A e femm. a}) \text{ o } (\text{masch. a e femm. A})] = pq + qp = 2pq.$$

# La legge di Hardy - Weinberg

Un'applicazione del concetto di probabilità indipendenti e della legge di Hardy -Weinberg e il test del DNA.

La situazione è la seguente: la polizia è in possesso di un campione del DNA (ricavato da sangue sperma o altro materiale organico) del colpevole di un reato.

Viene prelevato un campione di DNA di un sospetto e si verifica che alcuni geni del sospetto sono uguali a quelli del colpevole.

Qual è la probabilità che si tratti di una coincidenza? In altri termini, qual è la probabilità che un individuo preso a caso nella popolazione abbia gli stessi marcatori genetici del colpevole?

Si parte dalla probabilità della presenza nella popolazione di ciascuno degli alleli dei marcatori genetici coinvolti. Supponendo che la popolazione sia di Hardy- Weinberg, si usano le due relazioni

$$P_{A_iA_i} = P_{A_i}^2 \quad (\text{probabilità di trovare un genotipo omozigote } A_iA_i)$$

$$P_{A_iA_j} = 2P_{A_i}P_{A_j} \quad (\text{probabilità di trovare un genotipo eterozigote } A_iA_j).$$

Supponendo che i vari marcatori genetici siano scelti in modo da essere indipendenti tra loro si moltiplicano le probabilità dei vari genotipi e si ottiene la probabilità che un individuo preso a caso abbia gli stessi genotipi del colpevole in quei marcatori genetici.

Da fare attenzione a come si usa il test del DNA nel decidere se un sospettato è colpevole o innocente. Il test del DNA ci permette di assolvere con certezza un sospettato: se il DNA è diverso da quello del colpevole non è stato lui. Tuttavia non ci permette di condannare con certezza un sospettato anche se la probabilità che un individuo presa a caso nella popolazione abbia lo stesso DNA del colpevole è minima