Teoria della probabilità

SAPIENZA
UNIVERSITÀ DI ROMA

annarita.vestri@uniroma1.it

Il caso

L'esempio più semplice è quello della roulette.

Non è possibile sapere in anticipo, al lancio della pallina, se essa si fermerà sul colore rosso o sul colore nero (a meno che la roulette sia truccata: ma qui si assume che non lo sia). Il risultato di un singolo evento è puramente casuale. Tuttavia se si lancia la pallina diciamo un miliardo di volte, possiamo avere la "certezza" che la metà delle volte essa si fermerà sul rosso e l’altra metà delle volte essa si fermerà sul nero.

In altri termini, quando ci si trova davanti a evento singolo quello che prevale è il caso; quando gli eventi sono molto numerosi, sembra esservi una necessità alla quale gli eventi finiscono con l'ubbidire.

Questa è in estrema sintesi l'essenza della probabilità.

Intorno al XVII secolo lo studio della probabilità si staccò dai problemi di gioco per entrare in qualcosa di più “serio” e, grazie ad un commerciante di stoffe inglese, John Graunt, fu applicato alle Scienze sociali.

Graunt, insieme a Malthus, può essere considerato il fondatore della Demografia, cioè quella scienza che studia, da un punto di vista quantitativo, tutto ciò che riguarda i movimenti delle popolazioni.

Egli cominciò a consultare i cosiddetti Bills of mortality (Bollettini di mortalità) che fornivano la lista dei morti e delle nascite in alcuni quartieri di Londra, in cui erano indicate anche le cause di morte: questi bollettini erano spesso consultati dai ricchi londinesi per conoscere l'insorgere di eventuali epidemie di peste e, quindi, mettersi al sicuro lontano dalla città.

Per Graunt questi dati diventarono la base di un approfondito studio per compiere diverse analisi; per esempio, stimare la popolazione della capitale: egli scoprì che nel 1660 c'erano state circa 3 morti ogni 88 persone, per cui essendoci stati 19200 decessi stimò la popolazione in 563200 abitanti. Inoltre egli studiò le cause biologiche, sociali ed economiche della mortalità, lo studio del rapporto tra i sessi, la differenza tra nascite e morti in città e campagna, i flussi migratori.

ELEMENTI DI TEORIA DELLA PROBABILITA'

La TEORIA DELLA PROBABILITA' ci permette di studiare e descrivere i **fenomeni aleatori**.

DEFINIZIONE**:** un fenomeno è aleatorio quando di esso non si può <u>predire con certezza</u> il risultato.

Esempi:

- *Il risultato dell'estrazione al lotto*
- *Il risultato del lancio di una moneta*
- *Il risultato dell'esposiz. al bacillo di Cock*
- *Il risultato dell'esposiz. al fumo di sigaretta*

EVENTI E SPAZI CAMPIONARI

Un **esperimento** è un qualsiasi processo di osservazione o misurazione.

Esempi:

- *estrazione di un numero al lotto*
- *lancio di una moneta*
- *valutazione della presenza di un'infezione virale*
- *misura dell'altezza su un campione di bambini*

Ad ogni esperimento è associato uno **spazio campionario S** costituito dall'insieme dei possibili risultati. I singoli risultati dell'esperimento sono detti **elementi di S** o **eventi elementari**.

Prova, Evento

Concetti Primitivi: nozioni originarie ed intuitive.

Prova (o esperimento): è qualsiasi attività sviluppata in condizioni di incertezza. Gli esperimenti di cui si occupa il calcolo delle probabilità sono quelli nei quali i risultati non sono certi perché non univoci.

Evento: è uno dei possibili risultati della prova.

Gli eventi semplici di un esperimento sono mutuamente esclusivi

(o incompatibili)

Gli eventi composti (insieme di eventi semplici) non sono necessariamente mutuamente esclusivi

***Def.1**. Spazio dei Campioni.*

E' la totalità di tutti i possibili risultati di un esperimento concettuale. Verrà indicato con Ω.

***Def.2**. Evento Certo. Evento Impossibile.*

L'evento certo è quello che si verifica sempre, Ω.

L'evento impossibile è quello che non si verifica mai, ϕ.

***Def.3**. Spazio degli Eventi (o algebra di Boole).*

E' l'insieme di tutti i possibili sottoinsiemi di Ω.

Per la rappresentazione degli spazi campionari e dei loro elementi si utilizza la NOTAZIONE INSIEMISTICA.

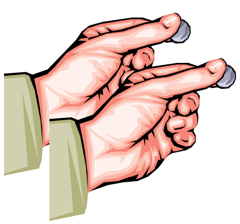
ESPERIMENTO	SPAZIO CAMPIONARIO
lancio di un dado	**{1, 2, 3, 4, 5, 6}**
misurazione della temp. corporea	**{x │ 34< x <42}**
lancio consecutivo di due monete	**{TT, TC, CT, CC}**
misurazione del sesso e del fumo in un passante	**{Mfum, M nn fum, F fum, F nn fum}**

E' possibile definire un qualsiasi evento A come combinazione di più eventi elementari

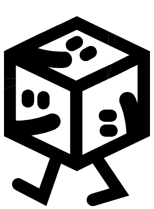
→ **EVENTO COMPOSTO**

SPAZIO CAMPIONARIO (S)

EVENTO (A)



'almeno una testa nel lancio di due monete' = {TT, TC, CT}

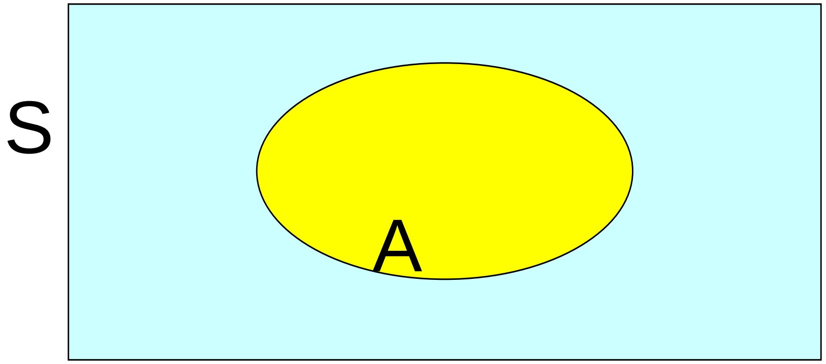


'numero dispari nel lancio di un dado' = {1, 3, 5}

Ogni evento A e’ un **SOTTOINSIEME** di S:

$$A \subset S$$

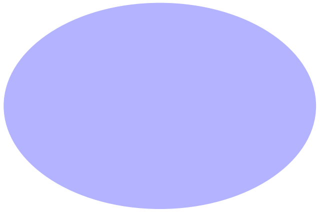
Gli eventi e gli spazi campionari possono essere rappresentati graficamente mediante i DIAGRAMMI DI VENN:



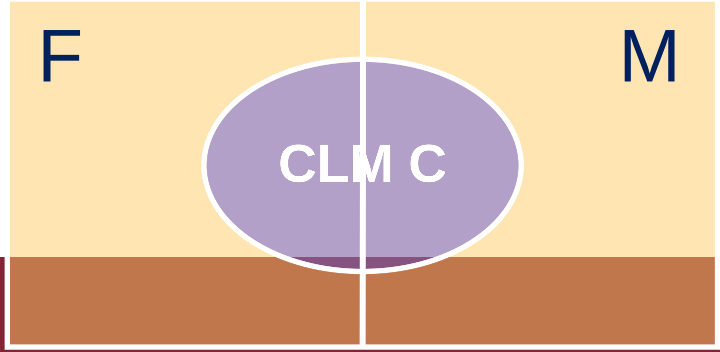
dove l’evento A è il sottoinsieme formato dagli eventi elementari in esso inclusi.

DIAGRAMMI DI VENN

Intera popolazione in studio
(studenti della facoltà di medicina)



sottogruppo (CLM C)



N.B.

Un evento è CERTO se comprende tutti gli elementi di S → $A = S$

Un evento è IMPOSSIBILE se non comprende alcun elemento di S → $A = \varnothing$

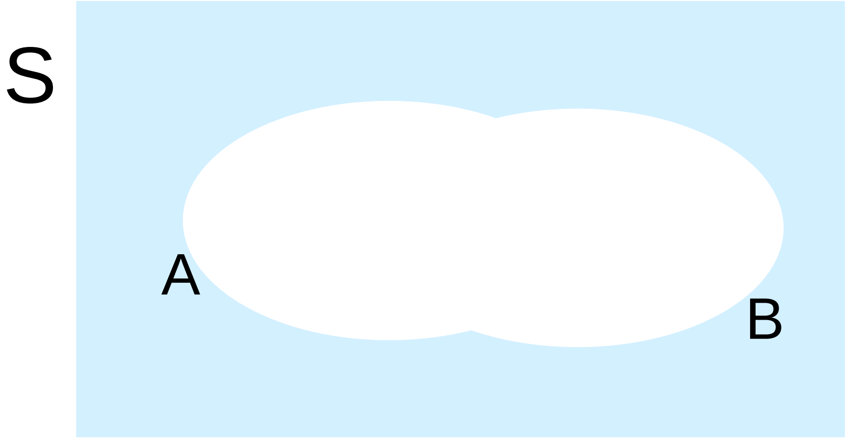
OPERAZIONI LOGICHE SUGLI EVENTI

Dato uno spazio campionario S e degli eventi A_i in esso inclusi, è possibile **definire** mediante operazioni logiche **nuovi eventi**.

Il valore di questi nuovi eventi rimane determinato noto quello degli eventi dati.

UNIONE DI EVENTI - SOMMA LOGICA: $A \cup B$

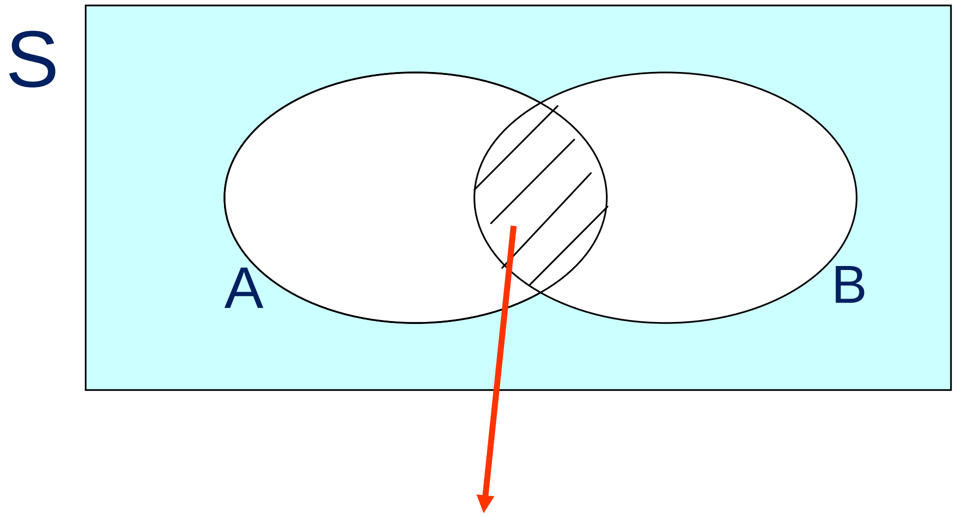
Siano A e B due eventi associati ad un esperimento: l'evento C è definito **unione di A e B** se comprende tutti gli elementi che appartengono ad A **oppure** a B.



$$C = A \cup B$$

INTERSEZIONE DI EVENTI - PRODOTTO LOGICO: $A \cap B$

Siano A e B due eventi associati ad un esperimento: l'evento C è definito **intersezione di A e B** se comprende tutti gli elementi che appartengono ad A **e contemporaneamente** a B.



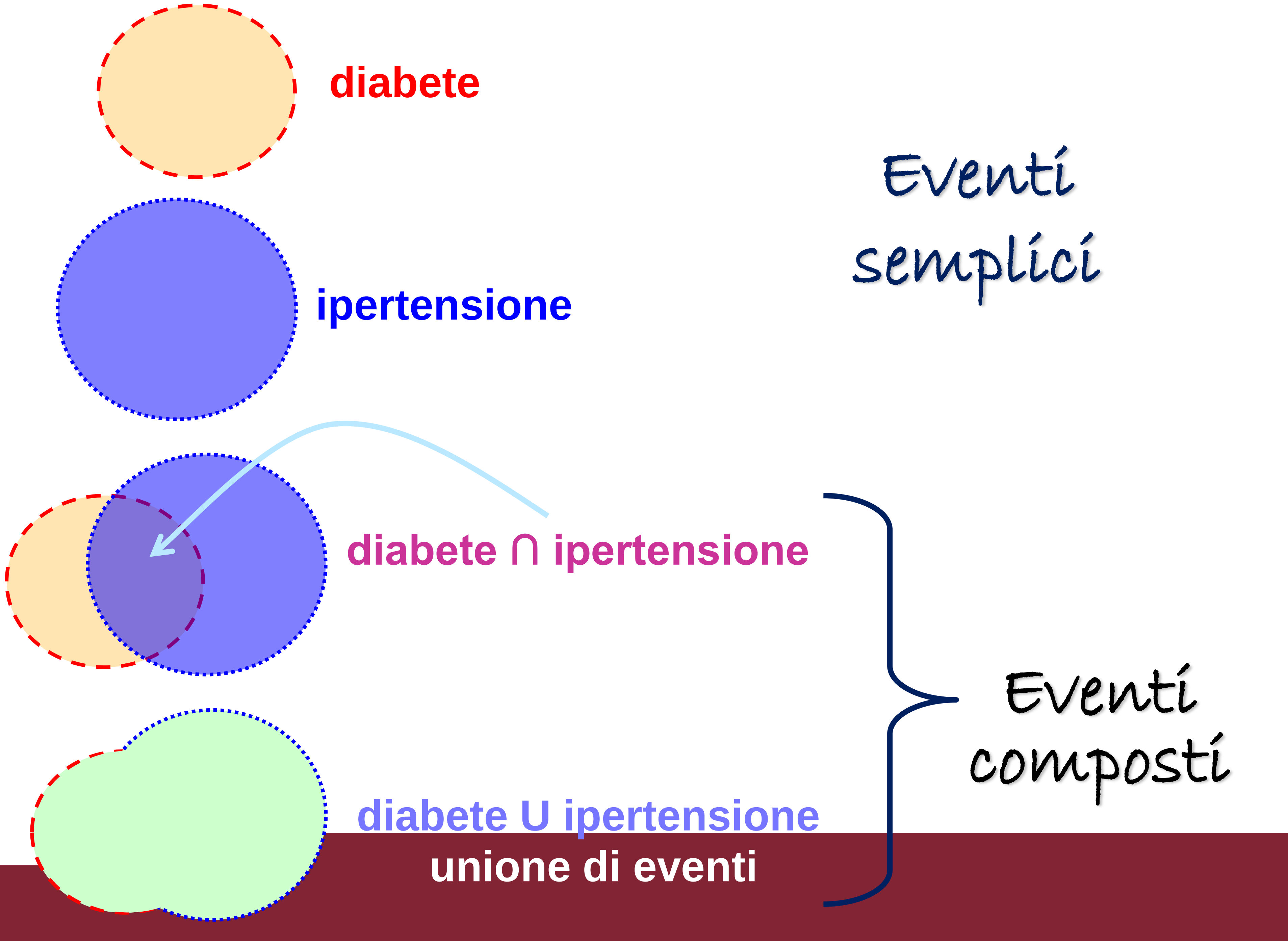
$$C = A \cap B$$

Esercizio: *nel lancio di un dado sia A = ‘numero pari’ e B = ‘numero $\geq$ 4’; si determini l’insieme unione e l’insieme intersezione.*

A = {2, 4, 6} **B = {4, 5, 6}**

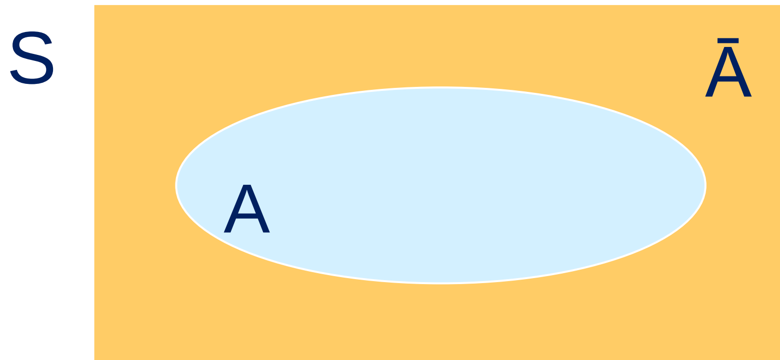
UNIONE → $A \cup B = \{2, 4, 5, 6\}$

INTERSEZIONE → $A \cap B = \{4, 6\}$


diabete
ipertensione
Eventi semplici
diabete ∩ ipertensione
diabete U ipertensione
unione di eventi
Eventi composti

NEGAZIONE DI UN EVENTO: $\bar{A}$

Dato un evento A, la sua negazione identifica un nuovo evento $\bar{A}$ costituito da tutti gli elementi di S non appartenenti ad A. $\bar{A}$ è detto **complemento di A in S**.



Segue che:

$$A \cup \bar{A} = S$$

Se due eventi A e B non hanno elementi in comune essi sono detti eventi **DISGIUNTI** o **MUTUAMENTE ESCLUSIVI** perché l'occorrenza dell'uno esclude l'altro.

Se A e B sono <u>mutuamente esclusivi</u> $\Rightarrow$ $A \cap B = \varnothing$

PROBABILITÀ

Lo spazio campionario S rappresenta l’insieme dei possibili risultati di un esperimento.

In genere, di fronte alla possibilità di un evento, esprimiamo la nostra **maggiore o minore fiducia** sul fatto che esso si verifichi

attribuiamo una maggiore o minore probabilità al verificarsi dell’evento

Esempi:

- *è <u>probabile</u> che oggi piova*
- *è <u>più probabile</u> che il tumore al polmone insorga in un fumatore che in un non fumatore*
- *è <u>più probabile</u> che il tumore al seno insorga in una donna che ha partorito per la prima volta dopo i 40 anni*

PROBABILITÀ

Il concetto di probabilità ci permette di graduare l'ambito delle possibilità o di precisare il grado di fiducia che abbiamo nel verificarsi di un evento.

La TEORIA DELLA PROBABILITA' ci permette di formulare delle **valutazioni numeriche di probabilità** e di ricondurle alle regole del calcolo matematico.

L'interpretazione di tali valori numerici dipende dal significato che viene attribuito al concetto di probabilità.

Definizioni di Probabilità

- Laplace: numero di eventi favorevoli sul
  numero di eventi possibili

- Von Mises: numero di successi sul
  numero di prove

- De Finetti: aspettativa soggettiva sul verificarsi di
  un evento

Definizione di probabilità.

Def. 4. *Classica*

La probabilità di un evento A è il rapporto tra il numero di casi favorevoli di A e il numero di casi possibili, ammesso che questi siano equiprobabili.

Def. 5. *Frequentista (o legge empirica del caso).*

In una serie di prove di un dato esperimento, ripetuto un gran numero di volte in circostanze più o meno simili, ciascuno degli eventi possibili si manifesta con una frequenza che è circa uguale alla sua probabilità. L'approssimazione si riduce al crescere del numero di prove.

Def. 6. *Soggettivista.*

La probabilità è la valutazione che il singolo individuo può coerentemente formulare, in base alle proprie conoscenze, del grado di avverabilità di un evento.

CONCEZIONE CLASSICA DELLA PROBABILITÀ

La probabilità di un evento A è il rapporto tra il numero di <u>casi favorevoli</u> al verificarsi di A (n) e il numero di <u>casi possibili</u> (N)

$$P(A) = \frac{n}{N}$$

Esempi:

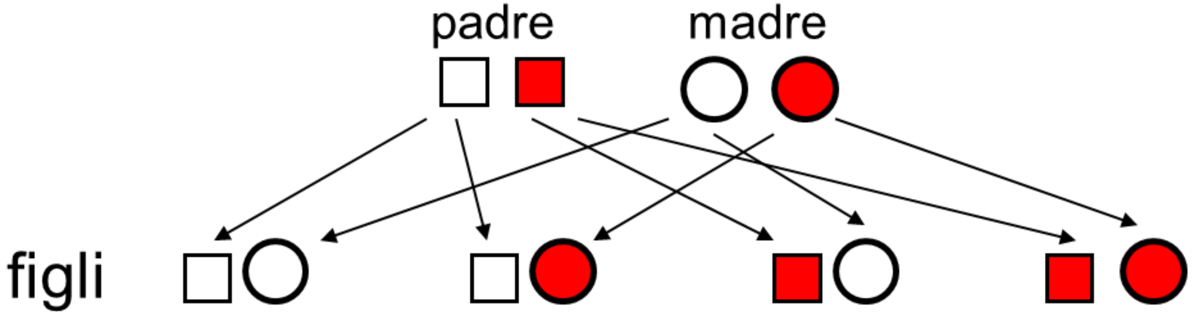
- **probabilità di estrarre un asso da un mazzo di 52 carte = 4/52 = 0.08**
- **probabilità di ottenere testa nel lancio di una moneta = 1/2 = 0.5**

- Tale definizione vale se i possibili risultati sono **equi-probabili** (gioco d'azzardo)

- Scarsamente applicabile a molte situazioni reali

Esempio di applicazione in medicina - Malattie genetiche

Se entrambi i genitori sono portatori sani del gene della talassemia o della fibrosi cistica, la probabilità di avere un figlio malato è una su quattro



CONCEZIONE FREQUENTISTA DELLA PROBABILITÀ

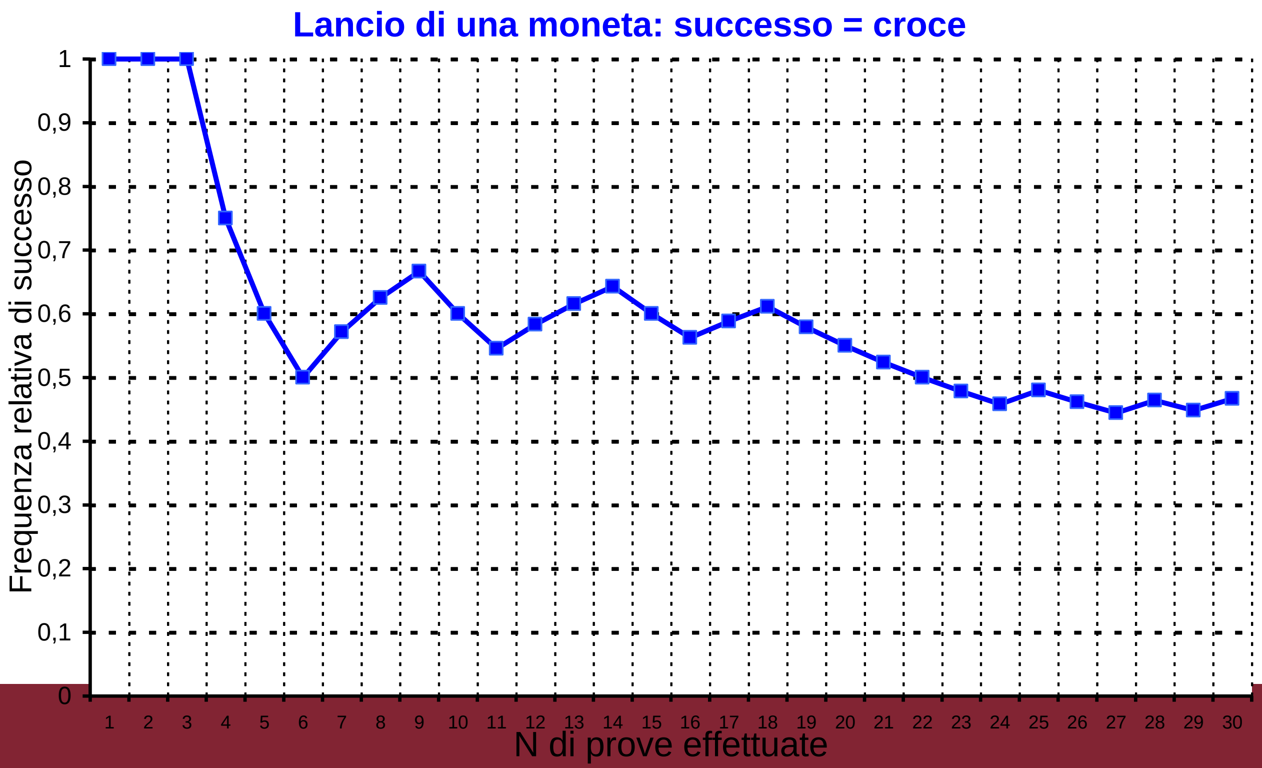
La probabilità di un evento A è la frequenza relativa di successo (occorrenza di A) in una serie **tendente all'infinito** di prove, ripetute **sotto identiche condizioni**:

$$P(A) = \lim_{N \to \infty} \frac{n}{N}$$

Nel caso della **concezione frequentista** la probabilità viene assegnata:

– **sulla base dei risultati di un esperimento ripetuto molte volte nelle stesse condizioni *(es. 1)***

Esempio 1: Frequenza dell'evento testa in una successione di lanci di una moneta.



Nel caso della **concezione frequentista** la probabilità viene assegnata:

- sulla base dei risultati di un esperimento ripetuto molte volte nelle stesse condizioni
- **sulla base di situazioni che possono essere ricondotte a tale contesto concettuale, ad esempio utilizzando statistiche correnti (*es. 2*).**

Esempio 2:

Prob. che un bambino italiano nasca morto nel 2000 =

$$= \frac{\textit{numero nati morti nel 2000}}{\textit{numero nati nel 2000}}$$

Non tutti gli eventi, pur valutabili in termini di probabilità, possiedono il requisito della **ripetitività sotto le stesse condizioni**

CONCEZIONE SOGGETTIVISTA DELLA PROBABILITÀ

La probabilità di un evento A è la **valutazione del grado di fiducia** che un individuo o un gruppo di individui può coerentemente formulare sull'occorrenza di A, **in base alle proprie opinioni e informazioni**

TEORIA BAYESIANA

- Riguarda quei fenomeni per i quali l'**attesa** o la **convinzione** rispetto all'esito influisce sull'evento stesso (*interventi chirurgici; eventi che dipendono dalla propria volontà, capacità, ...*)

- Riguarda per lo più **eventi unici o irripetibili**

ESEMPIO: QUAL È LA PROBABILITÀ CHE UN NEONATO SIA FEMMINA?

L'ecografista, alla decima settimana di gravidanza, dice ai genitori che *80 su 100* il neonato è femmina (l'ecografista, secondo le sue opinioni ed informazioni, esprime coerentemente il suo grado di fiducia nell'avverarsi dell'evento "nascita di una femmina").

(definizione SOGGETTIVISTA di probabilità)

ESEMPIO: QUAL È LA PROBABILITÀ CHE UN NEONATO SIA FEMMINA?

gravidanza

maschio femmina

1 su 2 = 50%

(definizione **CLASSICA** di probabilità)

Però nel mondo, in assenza di interventi dell'uomo (aborti o infanticidi selettivi, omessa denuncia) nascono 1057 maschi ogni 1000 femmine.

$$1000 / (1000+1057) = 48{,}6\%$$

(definizione **FREQUENTISTA** di probabilità)

Proprietà delle probabilità

1. $0 \leq P(E_i) \leq 1$ **La probabilità di un evento E_i è sempre un numero compreso tra 0 e 1**

2. $\Sigma_i P(E_i) = 1$ **La somma delle probabilità di tutti gli eventi $E_i \in$ spazio degli eventi è = 1**

3. **Proprietà Additiva della probabilità** **(eventi non disgiunti)**

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

MAZZO DI 52 CARTE

ESERCIZIO:

1. calcolare la probabilità di estrarre <u>l'asso di cuori</u>

   1 / 52 = 0.02

2. calcolare la probabilità di estrarre <u>una carta rossa</u>

   26 / 52 = 0.5

3. calcolare la probabilità di estrarre <u>una figura</u>

   12 / 52 = 0.23

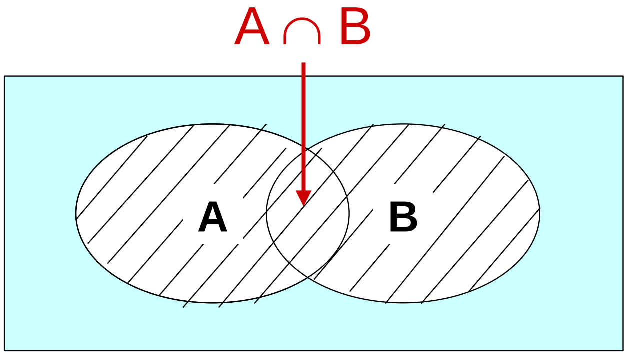
4. calcolare la probabilità di estrarre <u>un K</u>

   4 / 52 = 0.08

REGOLA DELL'ADDIZIONE

1. Se A e B sono due eventi in S tali che $A \cap B \neq \emptyset$ (**eventi non disgiunti**):

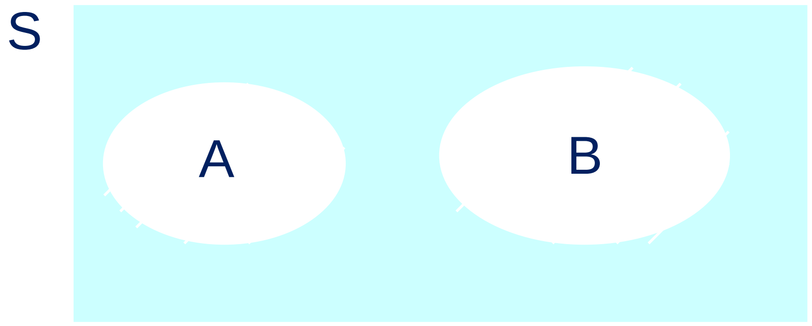
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



REGOLA DELL'ADDIZIONE

2. Se A e B sono due eventi in S tali che $A \cap B = \varnothing$ (**eventi disgiunti o incompatibili**):

$$\boxed{P(A \cup B) = P(A) + P(B)}$$



Esercizio: calcolare la probabilità di estrarre una carta rossa o una figura da un mazzo di 52 carte (eventi non disgiunti)

P(carta rossa) = 26 / 52 = ½ = 0.5

P(figura) = 12 / 52 = 0.23

P(carta rossa $\cap$ figura) = 6 / 52 = 0.11

P(carta rossa $\cup$ figura) =

= 0.5 + 0.23 -0.11 = 0.62

Esercizio: calcolare la probabilità di estrarre una figura o una carta compresa tra 3 e 6 da un mazzo di 52 carte (eventi disgiunti)

P(figura) = 12 / 52 = 0.23

P(carta 3-6) = 16 / 52 = 0.31

P(carta 3-6 $\cap$ figura) = 0 !!!

P(carta rossa $\cup$ carta 3-6) =

= 0.23 + 0.31 = 0.54

Proprietà Additiva della probabilità

TEOREMA DELLA SOMMA EVENTI disgiunti (incompatibili)

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

Esempio.

GRUPPI SANGUIGNI P(0) =0.46 P(AB)=0.04

$P(0 \cup AB) = P(0) + P(AB)= 0.46 + 0.04= 0.50$

Proprietà Additiva della probabilità TEOREMA DELLA SOMMA EVENTI NON MUTUAMENTE ESCLUSIVI (compatibili)

$P(E1 \cup E2) = P(E1) + P(E2) – P(A \cap B)$

GRUPPI SANGUIGNI $P(Rh+) = 0.85$

$P(0 \cap Rh+) = 0.39$

$P(Rh+ \cup 0) = P(Rh+) + P(0) - P(0 \cap Rh+) =$

$= 0.85 + 0.46 – 0.39 = 0.92$

Se Ā è il complemento di A in S:

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$$

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

Esempio: se la probabilità di morire nel I° anno dalla diagnosi per un paziente affetto da tumore al polmone è pari a 0.30, qual è la probabilità di sopravvivere al I° anno?

PROBABILITÀ CONDIZIONALE E REGOLA DELLA MOLTIPLICAZIONE

Talvolta è molto utile conoscere la probabilità di un evento A in S quando si è verificato un altro evento B in S $\rightarrow$ **PROBABILITA' CONDIZIONALE**

Esempio:

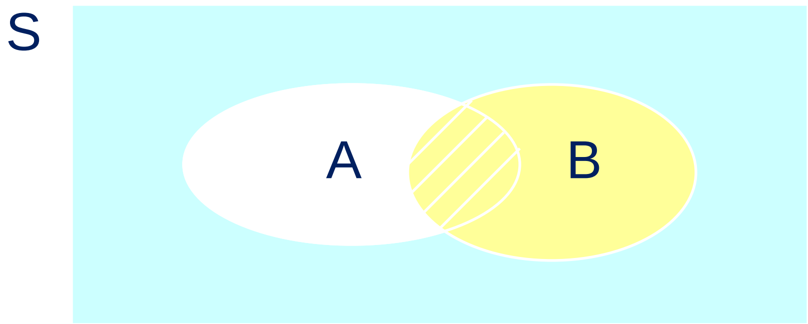
probabilità di uscita del 7 di quadri dato che è uscita una carta di quadri

probabilità di avere un tumore al polmone dato che si fuma

probabilità di avere il colera data la presenza di una gastroenterite acuta

Se A e B sono due eventi dello spazio campionario S, si definisce **probabilità condizionale di A dato B**:

$$P(A|B) = P(A \cap B) / P(B)$$



N.B.: *lo spazio campionario dell'evento B diviene il nuovo spazio campionario*.

Dalla definizione di probabilità condizionale segue la **REGOLA DELLA MOLTIPLICAZIONE**:

$$P(A \cap B) = P(B) \cdot P(A \mid B)$$

$$P(A \cap B) = P(A) \cdot P(B \mid A)$$

Se il verificarsi di B non condiziona la probabilità del verificarsi di A, segue che:

$$P(A / B) = P(A)$$

e i due **eventi** sono detti **indipendenti**, ovvero:

$$P(A \cap B) = P(A) \cdot P(B)$$

INDIPENDENZA

Due eventi **A** e **B** si dicono **indipendenti** se e solo se

$$P(A \cap B) = P(A) \times P(B)$$

ovvero: $P(A|B) = P(A)$ e $P(B|A) = P(B)$

Se due eventi sono indipendenti il verificarsi dell'uno non influenza la probabilità di verificarsi dell'altro

esempio: elevati *livelli di glicemia* sono indipendenti dalla presenza di *ulcera*

Esempio:

Qual è la probabilità di estrarre <u>senza reimbussolamento</u> ***due*** *palline gialle da un'urna che contiene tre palline gialle, due nere e una verde?*

A = estraz. Iª pallina gialla $\rightarrow$ $P(A) = 3/6$

B = estraz. IIª pallina gialla $\rightarrow$ $P(B \mid A) = 2/5$

$$P(A \cap B) = P(A) \bullet P(B \mid A) = (3/6)(2/5) = \mathbf{1/5}$$

Se l'estrazione fosse <u>con reimbussolamento</u>:

$$P(B \mid A) = P(B) = P(A) = 3/6$$

$$P(A \cap B) = (3/6)(3/6) = \mathbf{1/4}$$

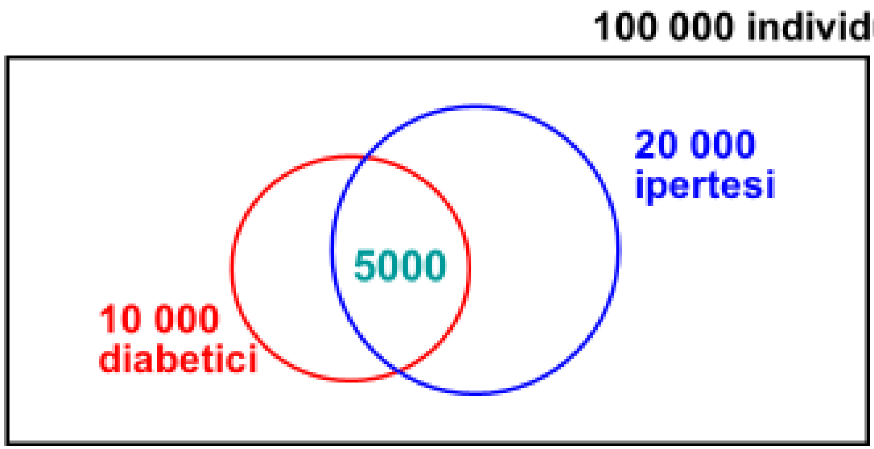
ESERCIZIO: CALCOLO DELLE PROBABILITA'

In una popolazione di **100 000 individui** vi sono:

10 000 diabetici (e 90 000 non-diabetici)

20 000 ipertesi (e 80 000 non-ipertesi).

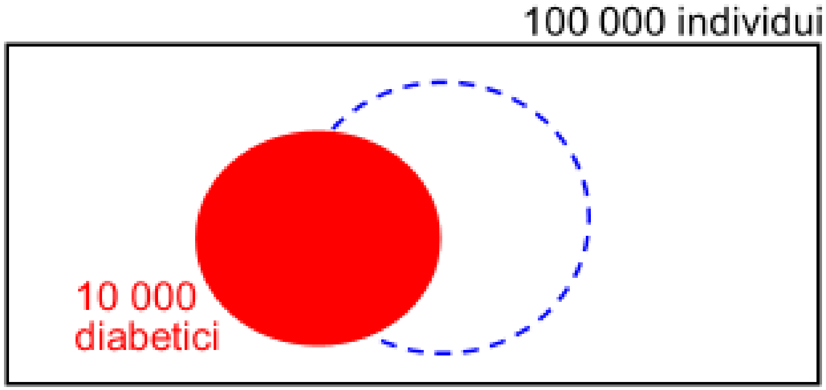
5000 persone che hanno sia il diabete che l'ipertensione



Qual è la probabilità di avere il diabete in quella popolazione?

Qual è la probabilità di avere l'ipertensione in quella popolazione?

Qual è la probabilità di avere il diabete in quella popolazione?

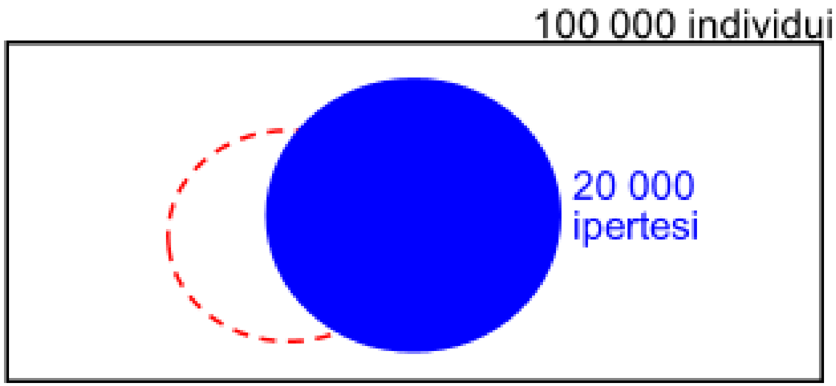


N.B.

E' stato usato l'approccio frequentista: la probabilità è stata stimata dalla frequenza relativa

p (diabete) = 10 000 / 100 000 = 0,1 = 10%

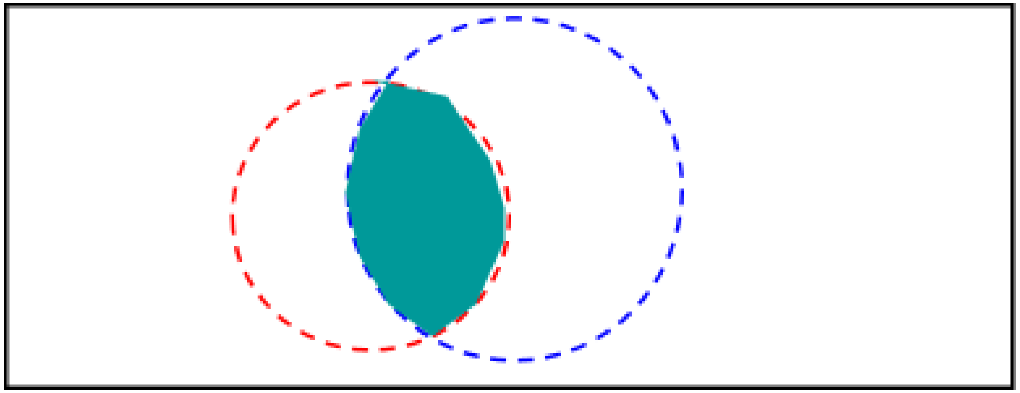
Qual è la probabilità di avere l'ipertensione in quella popolazione?



p (ipertensione) = 20 000 / 100 000 = 0,2 = 20%

Qual è la probabilità di avere il diabete **e** l'ipertensione
(**sia il diabete che l'ipertensione**)?

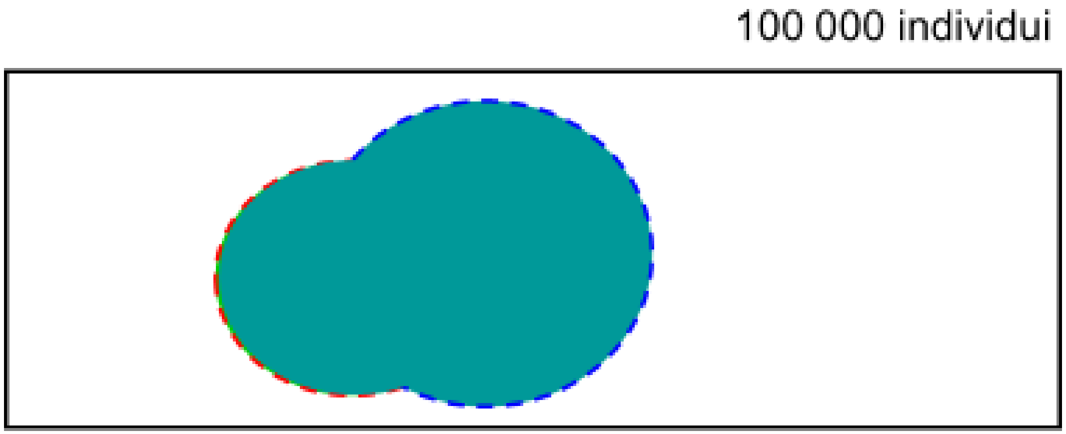
100 000 individui



p (diabete ∩ ipertensione) = 5 000 / 100000 = 0,05 = 5%

Qual è la probabilità di avere il diabete **O** l'ipertensione

(**solo il diabete o solo l'ipertensione o entrambi**)?



p(diabete U ipertensione) = (10000+20000-5000)/100000 = 25000/100000 = 0,25 = 25%

p(diabete U ipertensione) = p(diabete) + p(ipertensione) - p(diabete ∩ ipertensione) = 10% + 20% - 5% = 25%

Probabilità Condizionata

$$P(E_1|E_2)=\frac{P(E_1\cap E_2)}{P(E_2)}$$

		E1		
		si	no	
E2	si	e1∩e2		e2
	no			
		e1		1,0

Condizionare per "un evento" significa considerare quell'evento come il nuovo spazio degli eventi.

Per questo motivo si divide per la sua probabilità.
E' come dividere per il totale di riga o il totale di colonna quando nella tabella ci sono frequenze invece di probabilità.

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

		IMA		
		Si	No	
FUMO	Si	386	239	625
	No	231	144	375
		617	383	1000

52

Eventi condizionati

		IMA		
		Si	No	tot
FUMO	Si	386	239	625
	No	231	144	375
	tot	617	383	1000

Qual è la probabilità di avere IMA dato che fumo?

P(IMA)=617/1000=0.617

P(fumo)=625/1000=0.625

P(IMA|fumo)=0.617x0.625=0.386

0.386x1000=386

Eventi condizionati

		IMA		
		Si	No	tot
FUMO	Si	386	239	625
	No	231	144	375
	tot	617	383	1000

Qual è la probabilità di non avere IMA dato che fumo?

P(noIMA)=383/1000=0.383

P(fumo)=625/1000=0.625

P(noIMA|fumo)=0.383x0.625=0.239

0.239x1000=239

CALCOLO DEGLI ATTESI

Definita la probabilità di un evento o di una qualsiasi combinazione di eventi, è immediato definire il **numero di eventi attesi** in una serie di prove ripetute in modo casuale.

DEFINIZIONE: Se P(A) è la probabilità di comparsa dell'evento A, il **numero di eventi attesi E** in una serie di N prove sarà dato da:

$$\boxed{E = P(A) \cdot N}$$

NB: tanto più è grande N, tanto più il numero di successi osservati tenderà ad E

Esempio: *Se la probabilità di influenza per gli adulti tra i 20 e i 40 anni in un determinato mese dell'anno è* ***0.20****, quanti malati ci attendiamo in un campione casuale di* ***800*** *persone?*

$$E = 0.20 \cdot 800 = 160$$

(numero atteso di soggetti con l'influenza)

Esempio: *Si supponga che la probabilità di fumare sia* ***0.30*** *e quella di avere un tumore al polmone sia* ***0.01****. In un'indagine condotta su* ***1000*** *individui, sono stati osservati* ***15*** *soggetti fumatori con tumore al polmone. Secondo voi, i dati dimostrano che avere il tumore al polmone e fumare sono eventi indipendenti?*

$$E = (0.30 \cdot 0.01) \cdot 1000 = 3$$

(numero atteso di fumatori malati sotto l'ipotesi di indipendenza)

O = 15 (numero osservato di malati fumatori)

$\rightarrow$ **eccesso di 12 casi**

Prodotto di probabilità e sindrome plurimetabolica

Nello studio di Brunico *(Bonora et al, Diabetes 47: 1643-1649, 1998):*

N = 888

	Prevalenza
ridotta tolleranza glucidica	16,6%
dislipidemia	29,2%
iperuricemia	15,4%
ipertensione	37,3%

Se queste condizioni fossero indipendenti, la probabilità dell'intersezione (avere tutti e 4 i disturbi simultaneamente) sarebbe pari a:

0,166*0,292*0,154*0,373 = 0,0028 = **0,28%**

Prodotto di probabilità e sindrome plurimetabolica

p (avere tutte le patologie) = **0,28%**

Gli attesi (= soggetti con tutte e 4 le malattie sotto l'ipotesi di indipendenza) dovrebbero essere:

N * p = 888*0,0028 = **2,5**.

Invece se ne osservano 21!

Dal momento che gli osservati (21) sono molti di più degli attesi (2,5) si conclude che queste patologie non si riscontrano per caso negli stessi soggetti, ma rappresentano le diverse espressioni di una stessa patologia, la sindrome plurimetabolica.

TEST DIAGNOSTICO E TEOREMA DI BAYES

I sintomi più frequenti della malattia da coronavirus (Covid-19) sono febbre, stanchezza e tosse secca.
La maggior parte dei malati (circa l'80%) guarisce senza bisogno di cure particolari. Ma in alcuni casi la malattia può aggravarsi e portare anche al decesso.
Un paziente presenta questi sintomi: tosse, febbre e stanchezza.
Potrebbe essere malato di COVID?

Viene eseguito un tampone faringeo e si attende l'esito.
Il test diagnostico (tampone faringeo) e positivo.
Il medico sa che il tampone fornisce la risposta corretta nel 95% dei casi e fallisce nel restante 5%.

TEST DIAGNOSTICO E TEOREMA DI BAYES

Sulla base dell'esito, il medico sa con quale probabilità il paziente ha il Covid-19, dato che è risultato positivo al test.

I test di laboratorio sono utilizzati con finalità diagnostiche.
In ambito clinico si effettuano su persone che hanno, o si sospetta che abbiano, una determinata patologia.
Il test richiede la rilevazione del campione, l'indagine di laboratorio e la compilazione del referto.

Ma se l'esito del test è negativo, è possibile che il paziente in realtà sia positivo?

E se l'esito è positivo, può essere un errore del test?
Insomma quanto è affidabile un test?

TEST DIAGNOSTICO E TEOREMA DI BAYES

test	Malato (M+)	Sano (M-)	totale
Positivo (T+)	**VP**	**FP**	**VP+FP**
Negativo (T-)	**FN**	**VN**	**VN+FN**
totale	**VP+FN**	**FP+VN**	**TOTALE**

Sensibilità = VP/M+

Specificità = VN/M-

I soggetti sono classificati in due gruppi:

1. **gruppo dei pazienti sani (non malati): M+**
2. **gruppo dei pazienti malati: M+**

Il test e positivo (T+) se segnala la presenza della malattia e e negativo (T-) se non la segnala.

Interpretiamo la frequenza relativa come una probabilità
(o meglio, come una stima della probabilità).
VP/Totale rappresenta la probabilità che un individuo della popolazione in studio sia un Vero Positivo

TEST DIAGNOSTICO E TEOREMA DI BAYES

Il test diagnostico attualmente utilizzato per la diagnosi del Covid-19 è il test molecolare con metodo Real Time PCR per SARS-CoV-2 indicato dall'OMS (il cosiddetto tampone).
Il test ha Sensibilità e Specificità pari circa al 95%

	Malato (M^+)	**Sano** (M^-)	**Totale**
Test Positivo (T^+)	$P(M^+ \cap T^+)$	$P(M^- \cap T^+)$	$P(T^+)$
Test Negativo (T^-)	$P(M^+ \cap T^-)$	$P(M^- \cap T^-)$	$P(T^-)$
Totale	$P(M^+)$	$P(M^-)$	

$$P(T^+|M^+) = P(M^+ \cap T^+)/P(M^+) = 0.95$$
$$P(T^-|M^-) = P(M^- \cap T^-)/P(M^-) = 0.95$$

e, di conseguenza, $P(T^-|M^+) = P(T^+|M^-) = 0.05$

TEST DIAGNOSTICO E TEOREMA DI BAYES

La performance del test (sensibilità e specificità) è

$$P(T^+|M^+) = P(M^+ \cap T^+)/P(M^+) = 0.95$$
$$P(T^-|M^-) = P(M^- \cap T^-)/P(M^-) = 0.95$$

e, di conseguenza, $P(T^-|M^+) = P(T^+|M^-) = 0.05$

Ma non conosciamo la prevalenza P(M).

Alcune stime basate sulla sieroprevalenza oscillano tra 0.4% a 10%.

Ipotizziamo una prevalenza più alta 20%.

P(M+)=0.20

P(M-)=1-P(M)=0.80

Siamo interessati a P(M+|T **+**) (si chiama valore predittivo positivo)

TEST DIAGNOSTICO E TEOREMA DI BAYES

	Paziente Malato	Paziente Sano	Totale
Test Positivo	$P(M^+ \cap T^+)$	$P(M^- \cap T^+)$	$P(T^+)$
Test Negativo	$P(M^+ \cap T^-)$	$P(M^- \cap T^-)$	$P(T^-)$
Totale	$P(M^+) = 0.2$	$P(M^-) = 0.8$	

Sappiamo che $P(T^+|M^+) = P(M^+ \cap T^+)/P(M^+) = 0.95.$

Quindi $P(M^+ \cap T^+) = P(T^+|M^+)P(M^+) = 0.95 \times 0.2 = 0.19.$

Qual è la probabilità $P(M^+|T^+)$?

test	Malato (M+)	Sano (M-)	totale
Positivo (T+)	19	4	23
Negativo (T-)	1	76	77
totale	20	80	100

TEST DIAGNOSTICO E TEOREMA DI BAYES

Ricordando che

$$P(A|B) = P(A \cap B) / P(B)$$

$$\begin{aligned} P(M^+|T^+) &= \frac{P(M^+ \cap T^+)}{P(T^+)} \\ &= \frac{P(T^+|M^+)P(M^+)}{P(M^+ \cap T^+) + P(M^- \cap T^+)} \\ &= \frac{P(T^+|M^+)P(M^+)}{P(T^+|M^+)P(M^+) + P(T^+|M^-)P(M^-)} \\ &= \frac{0.95 \times 0.2}{0.95 \times 0.2 + 0.05 \times 0.8} = 0.826 \end{aligned}$$

Ho una probabilità del 82.6% di essere malato se il tampone è positivo (valore predittivo positivo)

TEST DIAGNOSTICO E TEOREMA DI BAYES

Con il tampone si ha:

$$P(T^+|M^+) = P(T^-|M^-) = 0.95$$

$$P(T^-|M^+) = P(T^+|M^-) = 0.05.$$

Vogliamo calcolare $P(M^-|T^-)$.

$$\begin{aligned} P(M^-|T^-) &= \frac{P(T^-|M^-)P(M^-)}{P(T^-|M^-)P(M^-) + P(T^-|M^+)P(M^+)} \\ &= \frac{0.95 \times 0.8}{0.95 \times 0.8 + 0.05 \times 0.2} = 0.987 \end{aligned}$$

Ho una probabilità del 98.7% di essere sano se il tampone è negativo (valore predittivo negativo)

TEST DIAGNOSTICO E TEOREMA DI BAYES

Un esame del sangue riconosce una certa malattia nel 99% dei casi quando essa è in atto.
Tuttavia, l'esame fornisce un falso positivo (esito positivo quando la malattia non è in atto) nel 2% dei pazienti.

Supponiamo che 0.5% della popolazione abbia la malattia.

Quale è la probabilità che una persona scelta a caso abbia effettivamente la malattia se il test è positivo?

Indichiamo rispettivamente con M+ ed T+ gli eventi
M+ = un soggetto estratto casualmente ha la malattia
T+= il test è positivo
Il testo ci dice che il test è affidabile al 99%, ossia fornisce un esito positivo quando il soggetto è effettivamente malato.
Ciò significa che **P(T+|M+) = 0.99**
Tuttavia, l'esame fornisce un falso positivo nel 2% dei casi, ossia **P(T+|M-) = 0.02**

Sapendo che P(M+)=0.005, per determinare P(M+|T+) possiamo utilizzare il teorema di Bayes come segue:

P(M+|T+) = P(T+|M+)P(M+) /(P(T+|M+)P(M+) + P(T+|M-)P(M-) = (0.99 * 0.005)/ ((0.99* 0.005) + (0.02 * 0.995)) = 0.199

Risulta quindi che una persona scelta a caso che ottiene risultato positivo al test ha una probabilità del 20% di avere effettivamente la malattia.

Come avere successo in amore

Massimizzerete la vostra probabilità di trovare la moglie (il marito) migliore se uscirete con circa il 37% delle candidate (candidati) che potete conoscere e poi decidete di stare con la prima (primo) che sarà migliore di tutte le (di tutti i) precedenti.

37% è una approssimazione del numero irrazionale 1/e, dove e è la base dei Logaritmi naturali (2,17828…)
Ovviamente la regola non garantisce il successo ma vi dà una probabilità del 37% di prendere la decisione migliore.

Come avere successo in amore

Supponiamo che prevediate di conoscere nel corso della vita, 100 candidate possibili:se sposate la prima la probabilità di avere effettivamente trovato la migliore di tutte è di 1/100, se invece aspettate di sposare la 100^ vi troverete ad avere respinto tutte le 99 venute prima e la probabilità che l'ultima che incontrerete sia anche la migliore sarà 1/100.

Le strategie migliori vi concedono di andare abbastanza avanti con la campionatura così da imparare un po' di cose sulle diverse candidate e fra queste strategie la più conveniente è quella che vi fa prendere un campione del 37% del totale, per poi scegliere la prima candidata che batterà tutte quelle venute prima.

Ovviamente c'è la possibilità di non trovarne mai una migliore di tutte quelle del 37% iniziale, ma in amore nulla è certo!

Se prevedete di conoscere 100 scapoli attraenti, dovreste lasciar perdere i primi 37 e sposare il primo, dopo di loro, che vi attragga più dei 37 con cui siete già uscite!