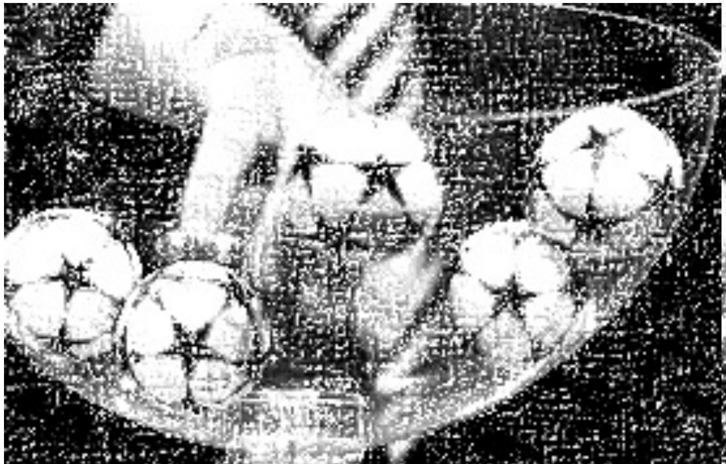
Probabilità condizionata, ovvero quanto ci si può fidare dell'intuito

"Professione Statistico"

Emanuela Sasso

19 Febbraio 2019

Liceo S.S. Enrico Fermi



Rispondiamo di “pancia”

1. Un’urna contiene 10 palline, 5 rosse e 5 gialle. Si eseguono 2 estrazioni, qual è la probabilità che la seconda pallina estratta sia gialla?
2. Il signor Rossi ha due figli: qual è la probabilità che entrambi i figli siano maschi? Se sappiamo che almeno uno di questi è maschio, qual è la probabilità che entrambi i figli siano maschi? Se sappiamo che il primogenito è maschio, qual è la probabilità che entrambi i figli siano maschi?
3. Problema di Monty Hall.

Definizioni di Probabilità

1. Classica: la probabilità di un evento è il rapporto tra il **numero dei casi** favorevoli e dei casi possibili, purchè questi ultimi siano tutti **equiprobabili** (Laplace).
2. Frequentista: la probabilità è il limite a cui tende la frequenza relativa dell'evento al crescere del numero degli esperimenti (von Mises)
3. Soggettivista: *la probabilità di un evento è il prezzo che un individuo ritiene equo pagare per ricevere 1 se l'evento si verifica, 0 se l'evento non si verifica* (De Finetti).
4. Assiomatica... (Kolmogorov)

Definizioni assiomatica

Esperimenti aleatori ed **eventi:** i primi indicano un esperimento di cui non è possibile conoscere con certezza la risposta prima di averlo compiuto (per esempio il lancio di un dado); gli eventi, invece, costituiscono un possibile esito dell'esperimento aleatorio (per esempio "la faccia in alto ha 5 pallini")

Lo **spazio campionario** Ω (o spazio delle osservazioni) come la totalità degli esiti possibili di un esperimento casuale.

I sottoinsiemi dello spazio campionario sono gli **eventi** che in genere sono indicati con le lettere maiuscole: $A, B, E, A_1, A_2, \ldots$ Gli eventi possono essere anche rappresentati con proposizioni. I sottoinsiemi formati da un solo elemento sono chiamati eventi elementari.

Definizioni assiomatica

La probabilità è una MISURA, ossia un’applicazione che va dall’insieme degli eventi possibili a un valore reale (normalmente compreso fra 0 e 1)

Due caratteristiche “ragionevoli” che richiediamo per questa funzione sono:

- la probabilità di tutti i possibili risultati deve essere uguale a 1
- la probabilità che si verifichi l’unione di eventi incompatibili (ciò significa che gli insiemi che rappresentano i due eventi sono disgiunti) deve coincidere con la somma delle probabilità dei singoli eventi.

Probabilità condizionata

Consideriamo due eventi A e B, con $P(B)>0$. Si dice **probabilità condizionata** di A sapendo B (e si indica con **P(A|B)**) il rapporto

$$P(A|B) = P(A \text{ e } B) / P(B)$$

La probabilità che si verifichi A e (anche) B, ossia P(A e B), è quindi data da

$P(A \text{ e } B) = P(B)\,P(A|B)$ oppure $P(A \text{ e } B) = P(A)\,P(B|A)$

Che cosa significa se si verifica: $P(A|B)= P(A)$?

Sono indipendenti, ossia

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

La definzione “torna”?

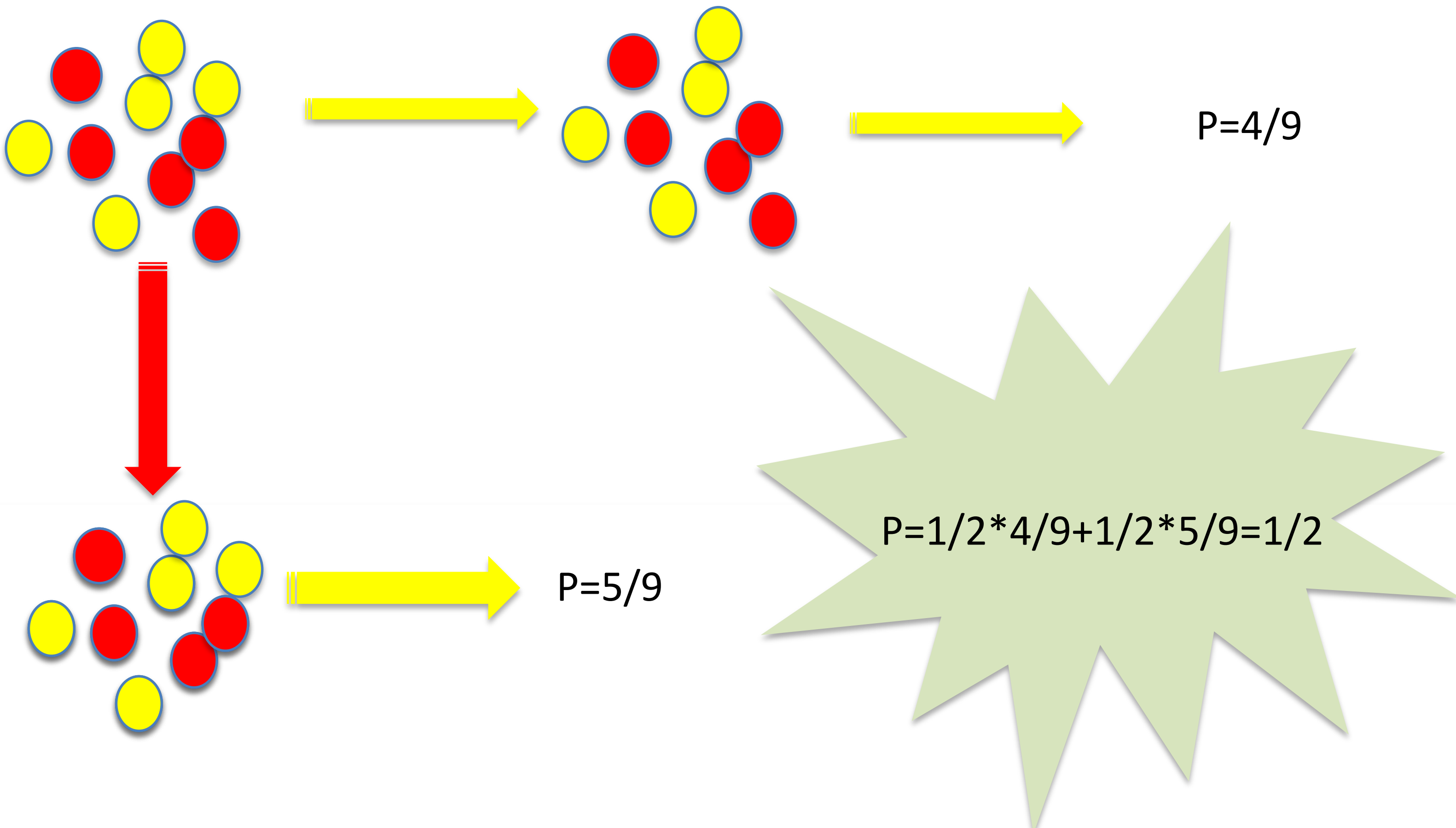
Prendiamo un mazzo di carte 40 carte e estraiamo 1 carta, qual è la probabilità che sia un asso (A)

P(A)=4/40=1/10

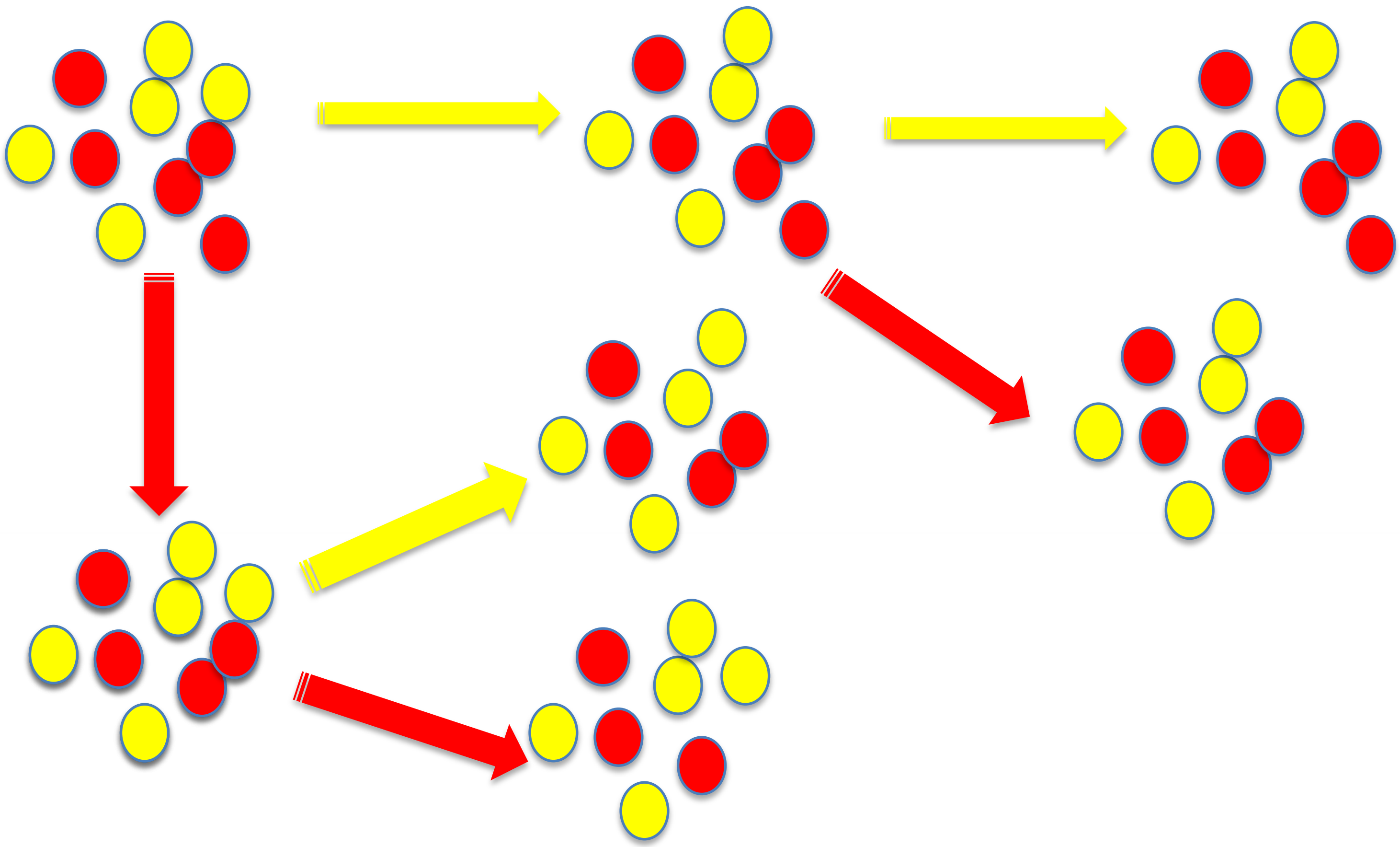
Ora se sappiamo che il giocatore prima di me ha estratto un asso di cuori (B)

P(A|B)=3/39=P(A e B)/P(B)=(1/40*3/39)/1/40

Domanda sulle palline: senza reinserimento



Domanda sulle palline: senza reinserimento

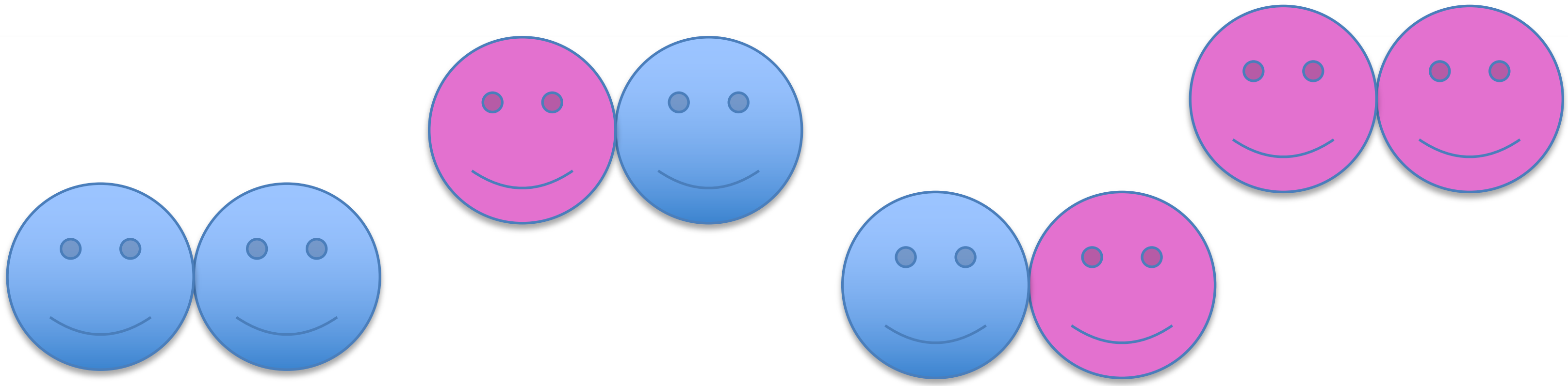


Domanda sulle palline: senza reinserimento

- ✓ Cosa cambia fra la situazione CON reinserimento e SENZA reinserimento?
- ✓ Cosa ci dice sui numeri ritardatari??

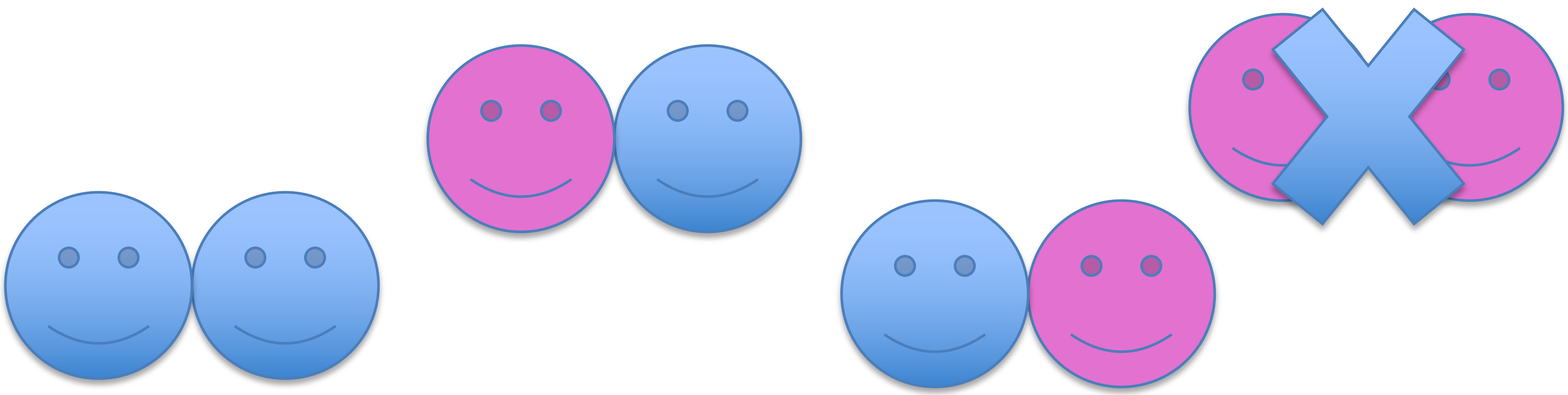
Domanda sui figli

- ✓ Il signor Rossi ha due figli: qual è la probabilità che entrambi i figli siano maschi? (1/4)
- ✓ Se sappiamo che almeno uno di questi è maschio, qual è la probabilità che entrambi i figli siano maschi?
- ✓ Se sappiamo che il primogenito è maschio, qual è la probabilità che entrambi i figli siano maschi?



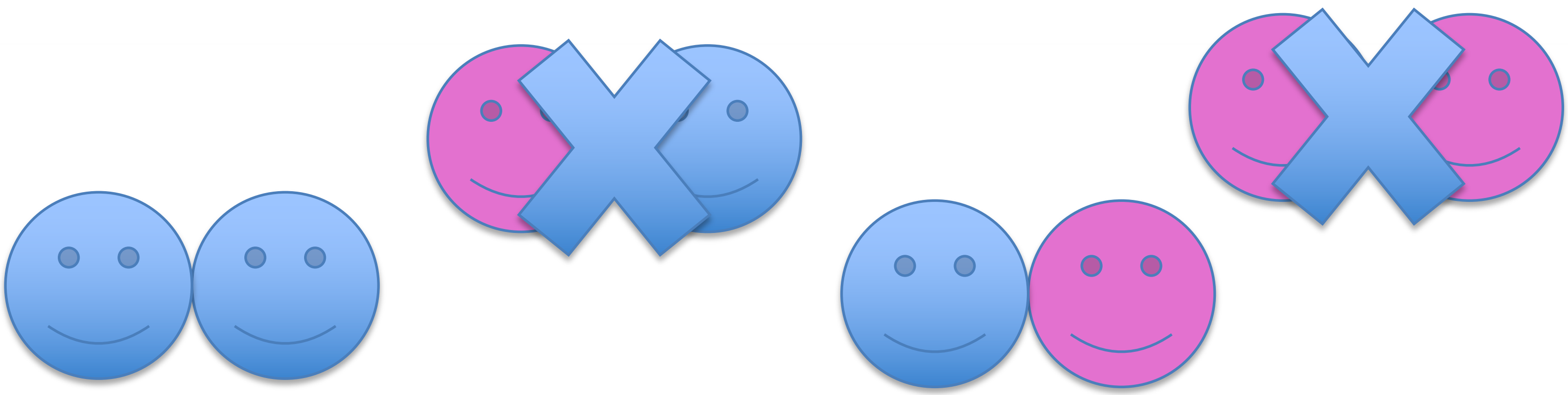
Domanda sui figli

- ✓ Il signor Rossi ha due figli: qual è la probabilità che entrambi i figli siano maschi? (1/4)
- ✓ Se sappiamo che almeno uno di questi è maschio, qual è la probabilità che entrambi i figli siano maschi? (1/3)
- ✓ Se sappiamo che il primogenito è maschio, qual è la probabilità che entrambi i figli siano maschi?

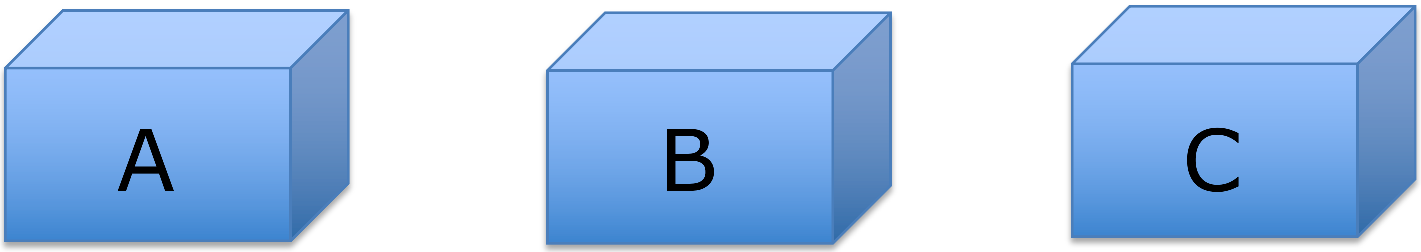


Domanda sui figli

- ✓ Il signor Rossi ha due figli: qual è la probabilità che entrambi i figli siano maschi? (1/4)
- ✓ Se sappiamo che almeno uno di questi è maschio, qual è la probabilità che entrambi i figli siano maschi? (1/3)
- ✓ Se sappiamo che il primogenito è maschio, qual è la probabilità che entrambi i figli siano maschi? (1/2)



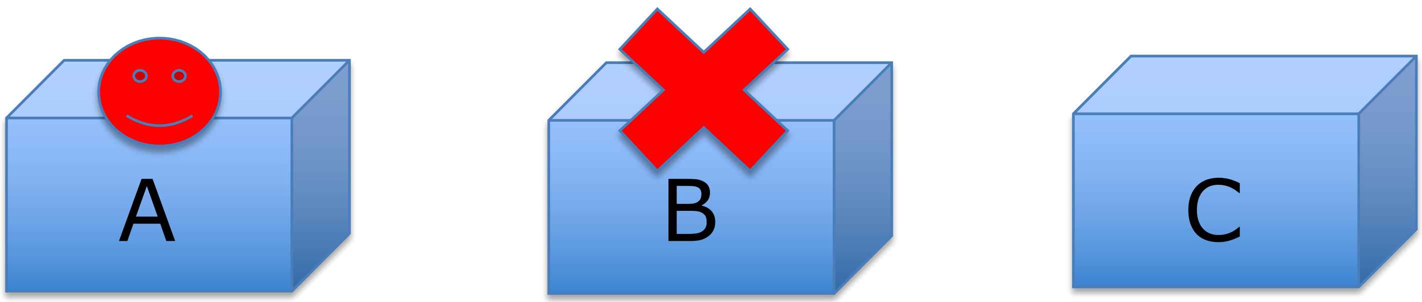
Monty Hall



- ✓ Solo in una scatola c'è il tesoro!
- ✓ Il giocatore è invitato a scegliere una scatola
- ✓ Il gestore apre l'altro contenitore vuoto
- ✓ Il gestore propone tre metodi per proseguire

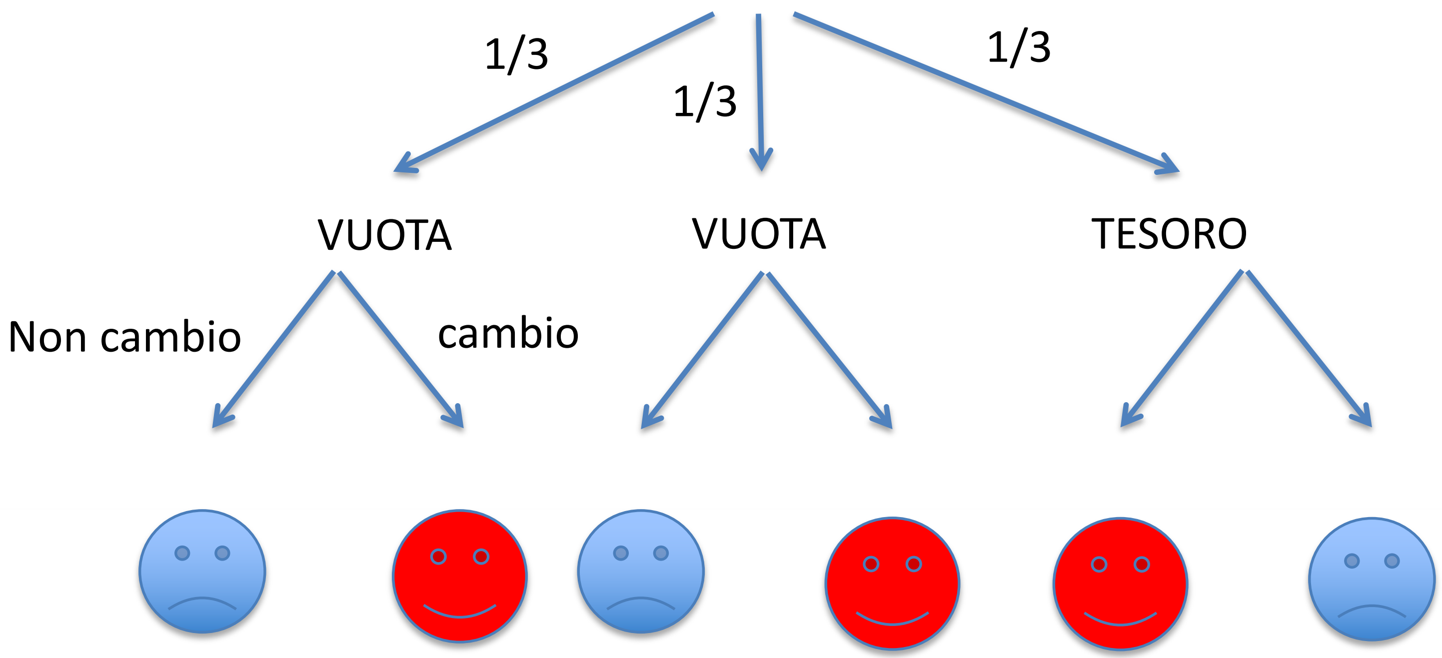
1. Il giocatore mantiene la sua scelta
2. Il giocatore cambia la scelta e indica l'altro contenitore chiuso
3. Il giocatore sceglie nuovamente a caso

Monty Hall



1. Il giocatore mantiene la sua scelta (1/3)
2. Il giocatore cambia la scelta e indica l’altro contenitore chiuso (2/3)
3. Il giocatore sceglie nuovamente a caso (1/2)

Monty Hall



La fallacia dell’accusa

Nella seconda metà degli anni ’90 il processo a Sally Clark in Gran Bretagna ha fatto discutere a lungo. Questa donna è stata accusata di aver ucciso il suo primo figlio di tre mesi e poi il suo secondo figlio di due mesi.

L’accusa si basava su un dato fornito da un’indagine ufficiale che aveva stimato la probabilità che un bambino scelto a caso morisse di morte naturale poteva considerarsi uguale a 1 su 1303. Se poi si consideravano le famiglie “simili” (per estrazione sociale, abitudini, ...) a quella di Sally Clark questa probabilità scendeva a 1 su 8500. Considerando due morti di cause naturali in una stessa famiglia come eventi indipendenti, la probabilità che i due bambini di Sally non fossero stati uccisi era di 1 su 73 milioni.

Sally Clark fu condannata nel 1999 e poi assolta quattro anni dopo.

La fallacia dell'accusa

Successivamente alla prima accusa, sempre sulla base di dati sperimentali, si è ipotizzato che la probabilità che due figli in una famiglia morissero di morte naturale è da 10 a 20 volte più grande rispetto al caso di morti indipendenti. Quindi indicando con N1 e N2 gli eventi "il primo figlio / il secondo figlio è morto di morte naturale" abbiamo che

$P(N1)=1/1300$ e

$P(N2|N1)=1/1300 *13=1/100$ (appross "da 10 a 20" con 13)

e quindi

$P(N2 \text{ e } N1)=1/130000$

Assumiamo – come ha fatto la difesa – che la probabilità di due figli morti per cause naturali sia 1 su 100 000 per famiglie "simili" a quelle di Sally Clark e indichiamo con N questo evento. Quindi $P(N)=1/100000$

La fallacia dell'accusa

Se indichiamo con M l'evento che due fratelli muoiano entro i primi mesi di vita, stiamo cercando

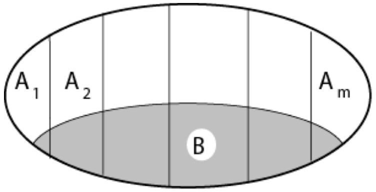
$$P(N|M)=P(N \text{ e } M)/P(M)=0{,}40$$

Si può assumere che in Gran Bretagna la mortalità infantile è circa del 5 per mille. Supponiamo gli eventi "morte di un figlio" e "morte di un secondo figlio" siano indipendenti, allora $P(M)=25/1000000$

$P(N|M)=P(N)/P(M)=0{,}40$….non così improbabile!

Formula di Bayes

Quando si può ricoprire lo spazio campionario con un numero finito di eventi *disgiunti* A_1, A_2,…,A_m.



$P(B) = P(B \text{ e } A_1) + P(B \text{ e } A_2) + \ldots + P(B \text{ e } A_m).$

Si può anche scrivere come:

$P(B) = P(B|A_1)\, P(A_1) + P(B|A_2)\, P(A_2) + \ldots + P(B|A_m)\, P(A_m)$

Tale **formula è detta della probabilità totale.**

Formula di Bayes

Come nell'esempio precedente, in alcuni contesti può essere più facile avere informazioni sulle probabilità condizionate $P(B|A_k)$ piuttosto che sulle probabilità congiunte $P(B \cap A_k)$.
Queste probabilità condizionate e la formula della probabilità totale ci permettono di calcolare "altre" probabilità condizionate.

Si ottiene così la **formula di Bayes.**

Falsi positivi e Falsi negativi

Sappiano che il 4% della popolazione in esame ha contratto la malattia M
Effettuando un test sperimentale per rilevare se un individuo è malato, si osserva che il test ha la seguente adabilita:

- Se l'individuo è malato, il test è + nel 99% dei casi
- Se l'individuo è sano il test è + positivo nel 2% dei casi

Qual è la probabilità che se il test è + , l'individuo sia effettivamente malato?

$P(M)=4\%$
$P(+|M)=0{,}99$
$P(+|NM)=0{,}02$

$$P(M|+)=P(M \text{ e } +)/P(+)=P(+|M)P(M)/P(+)$$

Falsi positivi e Falsi negativi

Sappiano che il 4% della popolazione in esame ha contratto la malattia M
Effettuando un test sperimentale per rilevare se un individuo è malato, si osserva che il test ha la seguente adabilita:

- Se l'individuo è malato, il test è + nel 99% dei casi
- Se l'individuo è sano il test è + positivo nel 2% dei casi

Qual è la probabilità che se il test è + , l'individuo sia effettivamente malato?

$P(M)=0{,}04$
$P(+|M)=0{,}99$
$P(+|NM)=0{,}02$

$P(M|+)=P(M \text{ e } +)/P(+)=P(+|M)P(M)/P(+)$

$P(+)=P(+|M)P(M)+P(+|NM)(1-P(M))=0{,}58$

$P(M|+)=0{,}\ 67$
$P(NM|+)=\ 0{,}33$ numero "alto" di falsi positivi