Corso di Statistica
Cenni di calcolo delle probabilità
Nozioni elementari

Domenico De Stefano

a.a. 2023/2024

Indice

Esempio di applicazione I

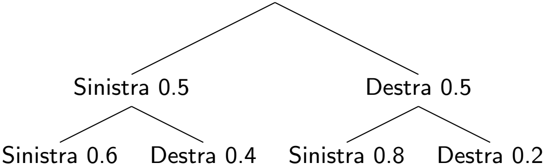
Il teorema delle probabilità totali si applica quando abbiamo a disposizione solo le probabilità condizionate di un certo evento e vogliamo conoscere la sua probabilità non condizionata.

Esempio: In un labirinto a T, ad un animale da laboratorio gli si dà la possibilità di andare a sinistra e ricevere cibo o andare a destra e ricevere una leggera scossa. Assumete che al tentativo numero 1, l'animale non abbia nessuna preferenza e vada a sinistra o a destra con la stessa probabilità. Tuttavia, se è andato a sinistra ed ha ricevuto il cibo al tentativo numero 2, l'animale andrà a sinistra e a destra con probabilità rispettivamente 0.6 e 0.4. Invece, dopo aver ricevuto la scossa andrà a sinistra e a destra con probabilità rispettivamente 0.8 e 0.2 (notare che le probabilità sommano sempre a 1, perché?). Qual è la probabilità che l'animale vada a sinistra al tentativo numero 2? E al tentativo numero 3?

Esempio di applicazione II

- Siano gli eventi: S_i = Sinistra all'i-mo tentativo e D_i = Destra all'i-mo tentativo
- Al primo tentativo sappiamo che $P(S_1) = 0.5$ e $P(D_1) = 0.5$
- Dopo aver ricevuto cibo (sinistra) avremo: $P(S_2|S_1) = 0.6$ e $P(D_2|S_1) = 0.4$
- Dopo aver ricevuto la scossa (destra), invece, avremo: $P(S_2|D_1) = 0.8$ e $P(D_2|D_1) = 0.2$
- Vogliamo sapere la probabilità $P(S_2)$ e poi analogamente la $P(S_3)$

Esempio di applicazione III



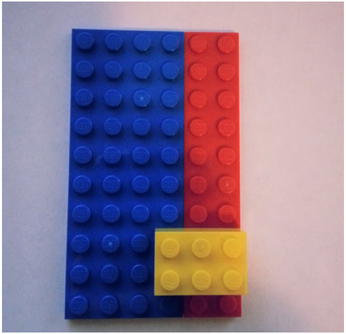
Esempio di applicazione IV

Il teorema della probabilità totali ci dice che, in una prova ripetuta possiamo calcolare la probabilità **non condizionata** di un evento a partire dalle probabilità condizionate in quanto (nel nostro esempio):

- $P(S_2) = P(S_2 \cap S_1) + P(S_2 \cap D_1)$
- e cioè: $P(S_2) = P(S_1)P(S_2|S_1) + P(D_1)P(S_2|D_1)$
- Sostanzialmente la probabilità non condizionata di un dato evento, ad esempio S_2, è dato dalla media ponderata delle sue probabilità condizionate, $P(S_2|S_1)$ e $P(S_2|D_1)$, con pesi $P(S_1)$ e $P(D_1)$. Se e solo se gli eventi condizionanti siano tutti i possibili eventi che si possono verificare e che siano tra loro incompatibili (cioè che tecnicamente rappresentino un partizione dello spazio di tutti i possibili eventi)
- Noi avremo: $P(S_2) = 0.5 \times 0.6 + 0.5 \times 0.8 = 0.7$

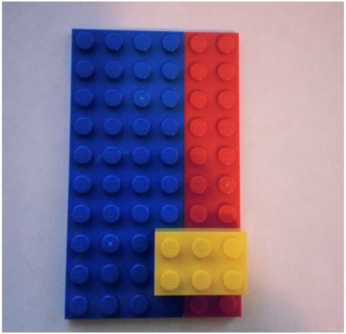
Indice

Teorema di Bayes e Lego /1



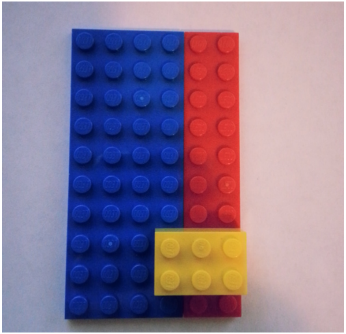
- Il teorema di Bayes è una di quelle idee che sono intutive ma allo stesso tempo "matematicamente" complesse (deriva dal teorema delle probabilità totali!)
- va molto al di là del sempice calcolo delle probabilità e ci dice con quanta probabilità possiamo stimare che una determinata causa abbia generato un certo evento!
- La formula è: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- Per capirlo, cerchiamo di arrivarci intuitivamente attraverso i mattoncini Lego!

Teorema di Bayes e Lego /2



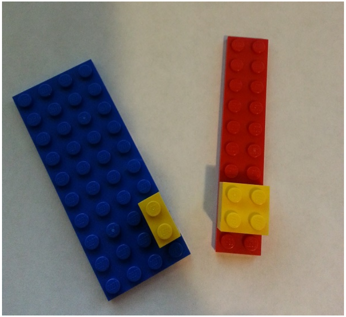
- Nella figura a fianco abbiamo un'area di 6×10 mattoncini (che rappresentano tutti i nostri possibili eventi... 60 per la precisione)
- Ci sono una serie di mattoncini blu, rossi e gialli (i gialli sono stati sovrapposti in un secondo momento)
- Passiamo alle probabilità... La probabilità di pescare a caso un mattoncino blu è (contando i mattoncini di quel colore sul totale, tralasciando per ora i gialli sovrapposti): $P(Blu) = 40/60 = 2/3$
- La probabilità di pescare a caso un mattoncino rosso è: $P(Rosso) = 20/60 = 1/3$
- Si noti che $P(Blu) + P(Rosso) = 2/3 + 1/3 = 1$
- Dato che le due probabilità sommano a 1, vuol dire che gli eventi blu e rosso esauriscono tutti i possibili eventi del nostro "esperimento"

Teorema di Bayes e Lego /3



- Ma cosa possiamo dire dei mattonicini gialli sovrapposti a parte dei blu e dei rossi?
- La probabilità di pescare a caso un mattocino giallo è: $P(Giallo) = 6/60 = 1/10$
- Tuttavia non possiamo sommare questa probabilità a quella degli eventi blu e rosso altrimenti avremo una probabilità maggiore di 1 (non ammissibile)!
- I mattoncini gialli saranno pescati sempre insieme ai blu o ai rossi (sono sovrapposti infatti!)
- Pertanto la probabilità di pescare un mattoncino giallo è condizionata a quella di pescare un mattoncino blu o rosso, ossia possiamo avere $P(Giallo|Blu)$ oppure $P(Giallo|Rosso)$

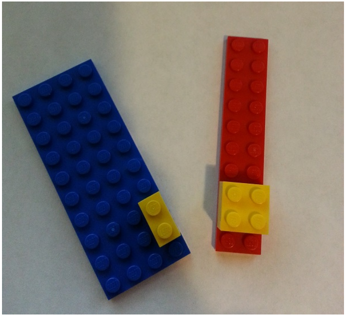
Teorema di Bayes e Lego /4



- Cerchiamo di calcolare allora la probabilità condizionata $P(Giallo|Rosso)$ attraverso i seguenti passaggi:

1. Separiamo tutti i mattonicini rossi dai blu
2. Contiamo solo i mattonicini rossi (è una probabilità condizionata quindi non ci serve il totale degli eventi ma solo il numero di eventi che verificano il condizionante!): sono $2 \times 10 = 20$
3. Contiamo quanti mattoncini gialli sono sovrapposti ai soli mattoncini rossi (4)
4. Dividiamo il numero di mattonicini gialli per l'area dei mattonicini rossi, e otteniamo: $P(Giallo|Rosso) = 4/20 = 1/5$

Teorema di Bayes e Lego /5



- Siamo arrivati alla probabilità condizionata $P(\mathit{Giallo}|\mathit{Rosso})$
- Ma che cosa succede se invertiamo il ragionamento?
- Ovvero cosa succede sei volessi sapere qual è la probabilità che dato il fatto di aver pescato un mattoncino giallo ci siano sotto i mattonicini rossi? cioè vorrei sapere l'altra probabilità condizionata: $P(\mathit{Rosso}|\mathit{Giallo})$
- Guardando alla figura è facile constatare che vi sono 6 mattonicini gialli in totale e che 4 di questi sono sovrapposti ai mattonicini rossi. Pertanto la probabilità di avere sotto dei mattonicini rossi se io ho pescato un mattoncino giallo è: $P(\mathit{Rosso}|\mathit{Giallo}) = 4/6 = 2/3$
- Se avete fatto questo ragionamento complimenti, avete scoperto in maniera autonoma il teorema di Bayes!

Il mistero del delitto dell'isola

- Causa una tempesta, l'isola di Bayes è isolata dalla terraferma
- Le 100 001 persone che si trovano sull'isola non possono perciò lasciarla, né qualcuno può arrivarvi.

Alberto Pullicino, XVIII secolo

Il mistero del delitto dell'isola

- Causa una tempesta, l'isola di Bayes è isolata dalla terraferma
- Le 100 001 persone che si trovano sull'isola non possono perciò lasciarla, né qualcuno può arrivarvi.

Alberto Pullicino, XVIII secolo

Il mistero del delitto dell'isola

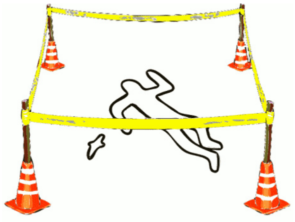
Chi ha ucciso?

Cadavere ritrovato nel bosco

T.B., un uomo di 50 anni che si trovava sull'isola di Bayes in gita di piacere è stato ucciso con un coltello da cucina nella notte di ieri. Il corpo è stato ritrovato da un pensionato che portava a passeggio il cane. Le circostanze fanno pensare a un delitto casuale, T.B. non conosceva nessuno sull'isola e quindi la ricerca di un movente pare vana. Chiunque potrebbe essere il colpevole.

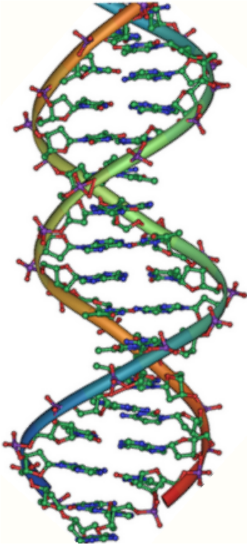
Le indagini

- A seguito dell'ispezione sul corpo della vittima del delitto, si è accertato che
  - il fatto è ad opera di un unico individuo,
  - vi è stata collutazione con la vittima,
  - sono state rinvenute tracce biologiche certamente dell'assassino sufficienti a estrarne il DNA.
- Si assume che solo un DNA sia compatibile



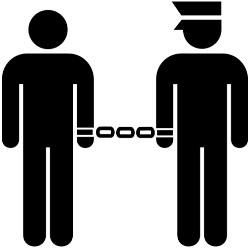
Si procederà perciò al confronto del DNA trovato con quello delle persone presenti sull'isola.

Il test del DNA



- Il DNA del colpevole è confrontato con quello di ciascuno dei 100 000 abitanti.
- Esito positivo al test significa che il DNA è compatibile secondo il test con quello del colpevole.
- Il test non è infallibile!
  - In un caso su 10000, un DNA non compatibile dà esito positivo. Ossia: $P(T|\text{non}\,C)$
  - In un caso su 100000, un DNA compatibile dà esito negativo. Ossia: $P(\text{non}\,T|C)$

Il test del DNA



- Il primo test dà esito "positivo", cioè il DNA del soggetto risulta compatibile.
- Il soggetto che ha dato esito positivo viene arrestato.
- Non vengono effettuati ulteriori test.

- Al processo, la prova del DNA è l'unico elemento portato dall'accusa, che però sostiene
- "Solo in un caso su 10000 la prova del DNA potrebbe essere sbagliata, l'imputato va condannato!"

Colpevole o innocente?

Qual è (circa) la probabilità che sia innocente?

A. Una su 100000
B. Una su 10000
C. Una su 10
D. Dieci su undici

Esiti attesi del test su tutti gli abitanti I

Supponiamo di fare il test a tutti, cosa succederebbe?

100000 abitanti

- 99999 innocenti
  - 99989 negativi
  - 10 positivi
- 1 colpevole
  - 1 positivo

I 99999 innocenti hanno tutti DNA incompatibile, ma per effetto dei falsi positivi, mediamente lo 0.1% è positivo, cioè 10.

Il colpevole risulterà positvo con elevata probabilità.

Esiti attesi del test su tutti gli abitanti II

Insomma, ci ritroveremmo con questa situazione

	COLPEVOLE	INNOCENTE
POSITIVO	1	10
NEGATIVO	0	99989

Cioè con 11 positivi, di cui uno solo colpevole, alla luce di questo, qual è la probabilità che il nostro positivo sia colpevole?

In termini di probabilità

Definiamo gli eventi (riferiti a un individuo)

- C : è colpevole (equivale a "il DNA è compatibile');
  - essendo tutti potenzialmente colpevoli è $P(C) = 1/100000 = 0.0001$
- T : risulta positivo al test
  - qui sappiamo che
    - $P(T|C) = 1 - 1/100000$
    - $P(T|\text{non } C) = 1/10000$
  - Possiamo mettere insieme le due cose e abbiamo

$$
\begin{aligned}
P(T) &= P(T|C)P(C) + P(T|\text{non } C)P(\text{non } C) = \\
&= \frac{99999}{100000}\frac{1}{100000} + \frac{10}{100000}\frac{99999}{100000} \approx \frac{11}{100000}
\end{aligned}
$$

Teorema di Bayes

Teorema di Bayes

Dati due eventi C e T si ha

$$P(C|T) = \frac{P(C)P(T|C)}{P(T)}$$

Applicandolo si ottiene

$$P(C|T) = \frac{\frac{1}{100000}\frac{99999}{100000}}{\frac{11}{100000}} \approx \frac{1}{11}$$

Esempio: elezioni, lega e Bayes I

Ricordiamo che

- $P(\text{Lega}) =$
- $P(\text{Lega}|\text{Nord}) =$, $P(\text{Lega}|\text{non Nord}) =$
- $P(\text{Nord}) = \frac{13292602}{35271541} =$

Sapendo che un elettore ha votato lega, qual è la probabilità che sia del Nord?

$$P(\text{Nord}|\text{Lega}) = ?$$

In base al teorema di Bayes

$$P(\text{Nord}|\text{Lega}) = \frac{P(\text{Lega}|\text{Nord})P(\text{Nord})}{P(\text{Lega})}$$

L'elettorato del Nord è il 35.74%, quindi $P(\text{Nord}) = 0.3574$. Si ha quindi, sostituendo

$$P(\text{Nord}|\text{Lega}) = \frac{0.092 \times 0.3574}{0.0394} = 0.8345$$

Esempio: elezioni, lega e Bayes II

Zona	Voti lega	Votanti	Elettori
Nord	1223131	13292602	16761578
Resto d'Italia	166883	21978939	30143576
Totale Italia	1390014	35271541	46905154

$$P(\text{Lega}) = \frac{1390014}{46905154} = 0.0394$$

$$P(\text{Lega}|\text{Nord}) = \frac{1223131}{16761578} = 0.092$$

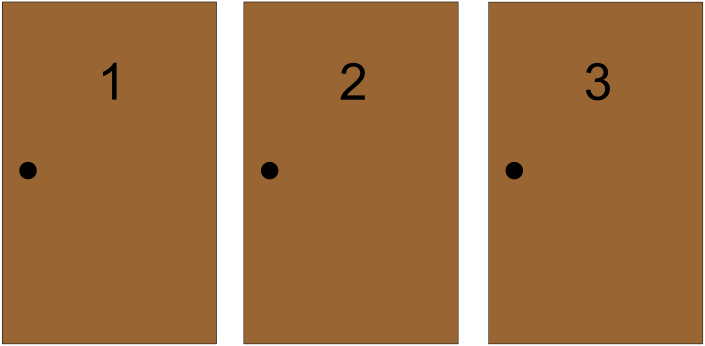
$$P(\text{Nord}) = \frac{16761578}{46905154} = 0.092$$

$$P(\text{Nord}|\text{Lega}) = \frac{1223131}{1390014} = 0.8345$$

Monty hall

- Il problema di Monty Hall fu proposto nel 1975 su *American Statistician* da *Steve Selvin*;
- ispirato al gioco a premi americano *Let's Make a Deal* (Monty Hall era lo pseudonimo del presentatore).

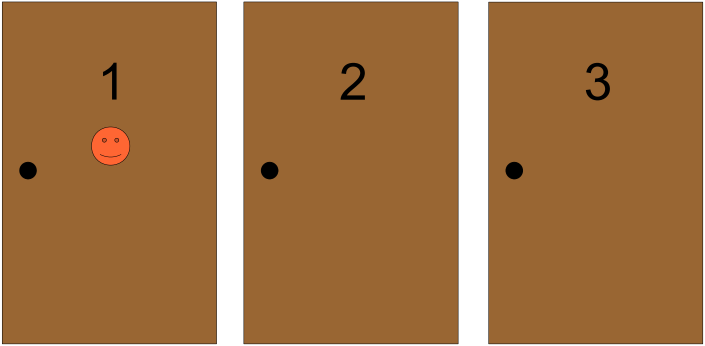
Monty hall



Scelgo 1 | Scelgo 2 | Scelgo 3

- Dietro una porta c'è un automobile, dietro le altre due capre.
- Il concorrente apre un aporta e vince quello che cè dietro.

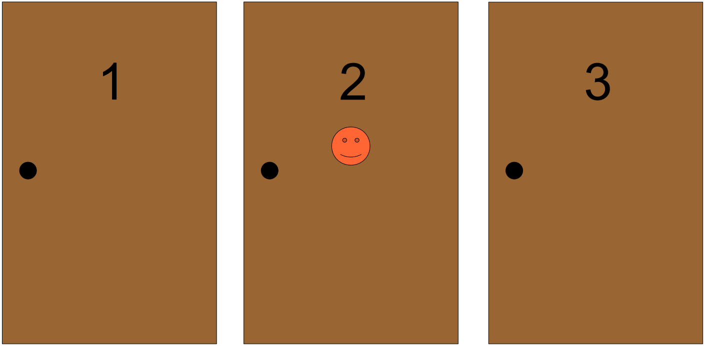
Monty hall



Apro 2 Apro 3

- È stata scelta la porta 1.
- Conduttore (che sa dove sono i premi): “Per aiutarla, ora le aprirò, delle altre due porte, una che cela una capra”

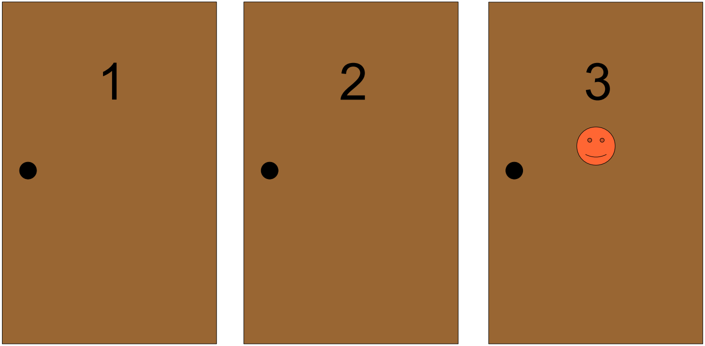
Monty hall



Apro 1 Apro 3

- È stata scelta la porta 2.
- Conduttore (che sa dove sono i premi): "Per aiutarla, ora le aprirò, delle altre due porte, una che cela una capra"

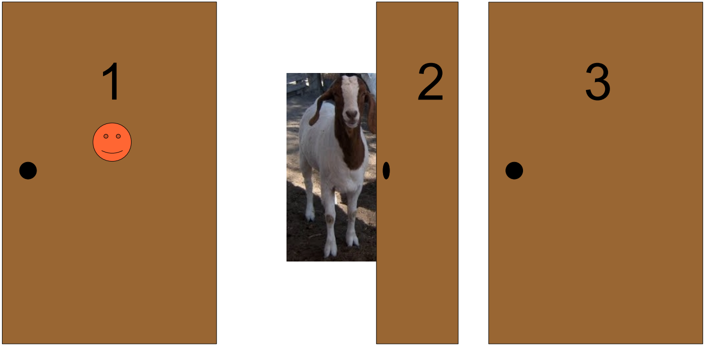
Monty hall



Apro 1 Apro 2

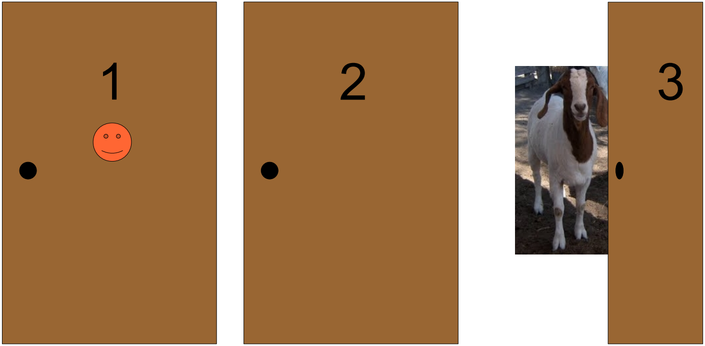
- È stata scelta la porta 3.
- Conduttore (che sa dove sono i premi): "Per aiutarla, ora le aprirò, delle altre due porte, una che cela una capra"

Monty hall



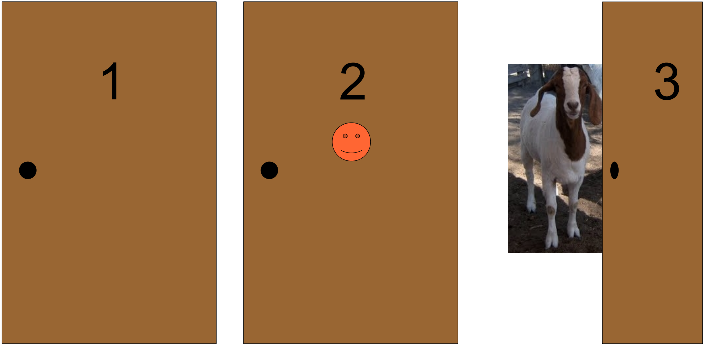
- Conduttore: “Ora, se vuole, può cambiare la sua scelta”
- Conviene cambiare? Soluzione

Monty hall



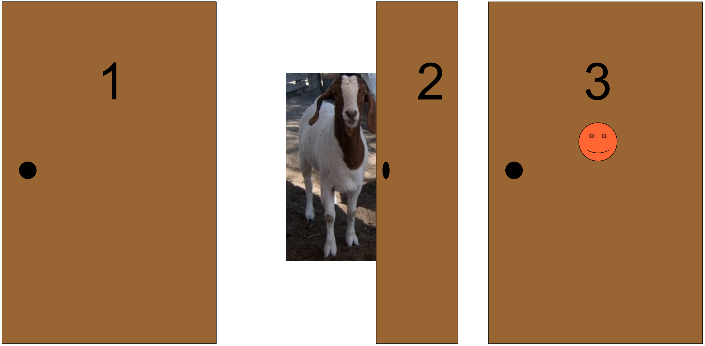
- Conduttore: "Ora, se vuole, può cambiare la sua scelta"
- Conviene cambiare? ▸ Soluzione

Monty hall



- Conduttore: “Ora, se vuole, può cambiare la sua scelta”
- Conviene cambiare? Soluzione

Monty hall



- Conduttore: "Ora, se vuole, può cambiare la sua scelta"
- Conviene cambiare? Soluzione

Monty hall: la risposta I

- $E_i =$"il premio è dietro la porta i";
- $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$;
- Il concorrente sceglie la porta 1, quindi vince se e solo se è vero E_1, con probabilità $1/3$.
- Consideriamo le due porte rimanenti
  - (a) se è vero E_1 il conduttore apre indifferentemente la porta 2 o 3;
  - (b) se è vero E_2 il conduttore apre la porta 3 e la porta non scelta che rimane chiusa è quella vincente;
  - (c) se è vero E_3 il conduttore apre la porta 2 e la porta non scelta che rimane chiusa è quella vincente;
- i tre casi sono equiprobabili, in due su tre conviene cambiare perché la porta non scelta e rimasta chiusa è quella vincente.
- Cambiando, si vince con probabilità $2/3$.

Monty hall: la risposta II

- La soluzione corretta è più intuitiva se consideriamo tante porte.
- Siano 100 le porte, una sola contiene il premio.
- Il concorrente sceglie una porta, (con probabilità 0.01 è vincente)
- il conduttore apre 98 delle restanti mostrando che hanno la capra.
- Alla fine rimangono due porte chiuse, quella scelta dal concorrente inizialmente e quella non aperta dal conduttore.

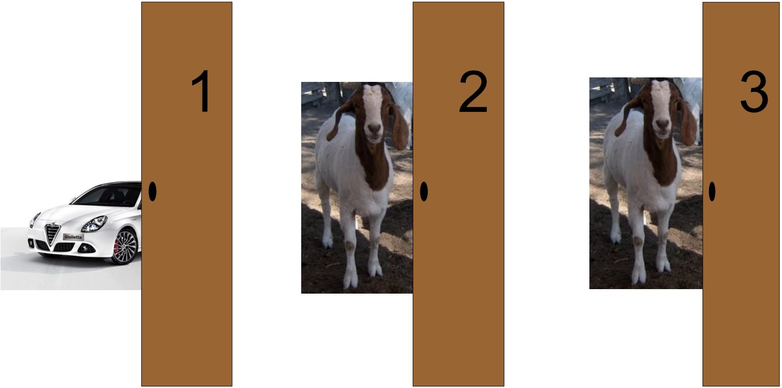
Monty hall: albero

È scelta la porta 1

$\mathbf{1}23$

- Auto in 1
  $\mathbf{1}23$
  $\left(\frac{1}{3}\right)$
  - Apre 2
    $\mathbf{1}\cancel{2}3$
    $\left(\frac{1}{3}\frac{1}{2}=\frac{1}{6}\right)$
  - Apre 3
    $\mathbf{1}2\cancel{3}$
    $\left(\frac{1}{3}\frac{1}{2}=\frac{1}{6}\right)$
- Auto in 2
  $\mathbf{1}23$
  $\left(\frac{1}{3}\right)$
  - Apre 3
    $\mathbf{1}2\cancel{3}$
    $\left(\frac{1}{3}1=\frac{1}{3}\right)$
- Auto in 3
  $\mathbf{1}23$
  $\left(\frac{1}{3}\right)$
  - Apre 2
    $\mathbf{1}\cancel{2}3$
    $\left(\frac{1}{3}1=\frac{1}{3}\right)$

Monty hall: porte aperte



Riepilogo: schema

non A	non si verifica A	$P(\text{non } A) = 1 - P(A)$
$A \cup B$	si verifica A oppure B	$P(A \cup B) = P(A) + P(B)$ se $A \cap B = \emptyset$ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$A \cap B$	si verificano sia A che B	$P(A \cap B) = P(A)P(B)$ se A, B indip $P(A \cap B) = P(A)P(B\|A)$
$A\|B$	si verifica A sapendo che si è verificato B	$P(A\|B) = P(A)$ se A, B indip $P(A\|B) = P(A \cap B)/P(B)$ $P(A\|B) = P(A)P(B\|A)/P(B)$