METODI STATISTICI E COMPUTAZIONALI

Stefania Spagnolo

Dipartimento di Matematica e Fisica, Univ. del Salento

LEZIONE 6

Richiami di Calcolo Combinatorio e Probabilità

Continua

CALCOLO COMBINATORIO E PROBABILITÀ

- ***Un insieme S contenente tutti i possibili risultati di un esperimento*** è detto ***spazio campione*** e ciascun risultato è detto elemento o punto di S.
- Un ***evento (A) in S*** è costituito da 1 o più elementi di S. ***L'insieme di tutti gli eventi in S è anche detto campo di Borel***

- Uno spazio campione può essere finito o infinito, continuo o discreto
- Esempi:
  - lancio di un dado == spazio discreto finito
  - serie storica a lotto == spazio discreto infinito
  - misura di una lunghezza == spazio continuo
- Esiste un evidente parallelismo tra l'**algebra degli eventi** e quella degli ***insiemi***

Unione	$A \cup B$	evento A oppure B
Intersezione	$A \cap B$	sia l'evento A che quello B
Complementare	$\overline{A}$	l'evento non A
Differenza	$A - B$	l'evento A ma non B

CALCOLO COMBINATORIO E PROBABILITÀ

- Esempio:
- Supponiamo di avere a che fare con una malattia rara che colpisce la popolazione con una frequenza di 2 malati ogni 100000 abitanti.
- P(malato)=2/100000=2 x 10^{-5}

NOTA: probabilità nota a priori

- Supponiamo di disporre di un test con le seguenti caratteristiche:
- ***Sensibilità***: probabilità, cioè, di un risultato positivo se la persona è effettivamente malata
  - P(pos | malato) = 0.993

Calibrazione. Conosco quale è l'effetto data la causa.

- ***Specificità***: probabilità, cioè, di un risultato negativo se la persona è sana
  - P(neg | sano) = 0.9999

Il test è molto buono. La Sensibilità mi sta dicendo che identifica un malato se questo lo è effettivamente con una probabilità del 99.3 % mentre la specificità mi dice che dichiara una persona sana se questa lo è effettivamente con una probabilità del 99.99%

CALCOLO COMBINATORIO E PROBABILITÀ

- Supponiamo di voler valutare se sia utile o meno sottoporre a screening la popolazione solo in base ai risultati di questo test.

- Caso pratico: sottopongo a test 10^6 soggetti (screening), cosa mi aspetto si ottenga dal test?

$$P(pos \mid malato) = 0.993 \quad \text{Sensibilità}$$

$$P(neg \mid sano) = 0.9999 \quad \text{Specificità}$$

- **Uso una simulazione per verificare le mie previsioni.**
- Prendo 10^6 soggetti e casualmente assumo che siano malati o sani in base alle conoscenze a priori. In base alla sensibilità e specificità del test simulo la risposta del test per ogni soggetto.
- ***Valuto quanti soggetti positivi al test sono effettivamente malati***
  - ***P (malato | pos)***

Teorema di Bayes

Sia A un evento di S con P(A)>0 e $\{B_1,..,B_n\}$ una famiglia di eventi dello spazio campione S mutuamente esclusivi ed esaustivi, allora

$$P(B_k \mid A) = \frac{P(A \mid B_k) P(B_k)}{P(A)}$$

$B_i \cap B_j = 0 \quad i \neq j$ Mutuamente esclusivi

$B_1 \cup B_2 \cup \ldots \cup B_n = S$ Esaustivi

Il teorema di Bayes è diretta conseguenza della definizione di probabilità condizionata

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

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

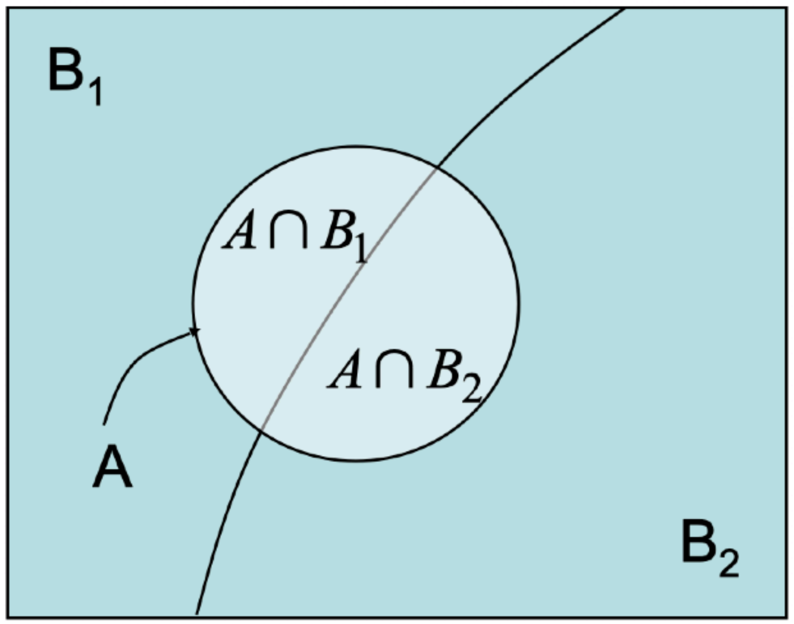
CALCOLO COMBINATORIO E PROBABILITÀ

Dividere lo spazio campione in n eventi mutuamente esclusivi e esaustivi equivale a dividerlo in n classi.

Es.: Classi== malati di cuore e non; Gruppi di massa in uno spettro di Raggi Cosmici, etc.

Il teorema di Bayes risponde alla domanda:
Supponiamo di aver ottenuto l'evento A qual è la probabilità che il mio elemento appartenga alla classe k?

CALCOLO COMBINATORIO E PROBABILITÀ



B_1 e B_2 esclusivi ed esaustivi

$$P(A) = P(A \cap B_1) + P(A \cap B_2)$$

Ma per la definizione di probabilità condizionata

$$P(A \cap B_1) = P(A \mid B_1)P(B_1)$$

Da cui si dimostra il

Teorema della probabilità totale
Sia A un evento di S e $\{B_1, \ldots. B_n\}$ una famiglia di eventi di S mutuamente esclusivi e esaustivi

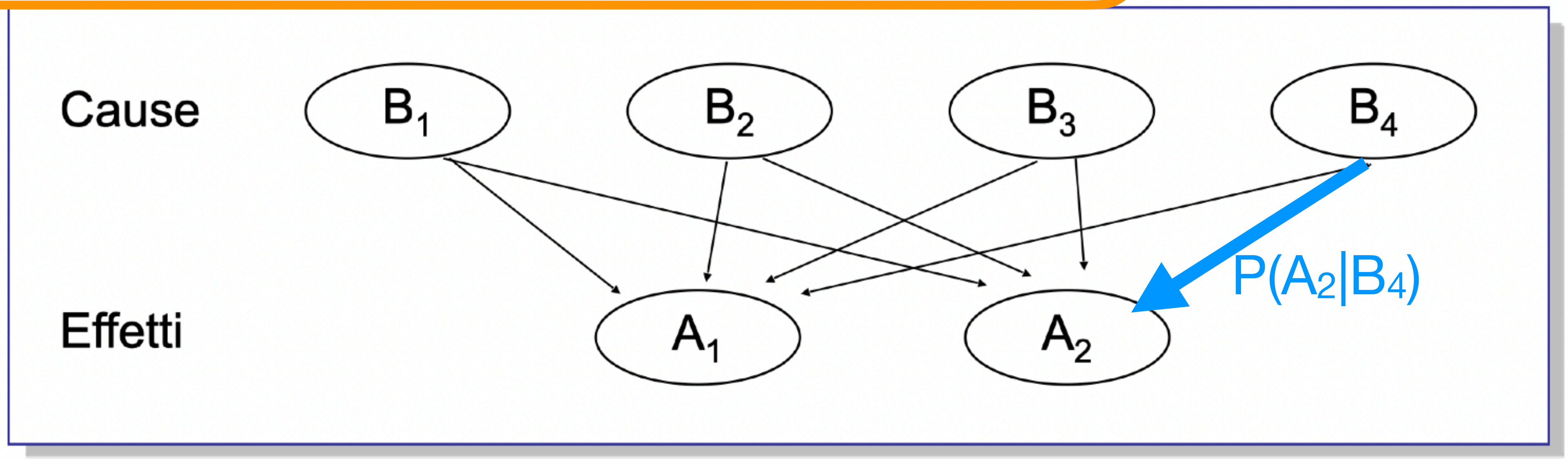
$$P(A) = P(A \mid B_1)P(B_1) + \ldots + P(A \mid B_n)P(B_n)$$

Usando il teorema della probabilità totale, il teorema di Bayes si può riscrivere come:

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{\sum_{k=1}^{n} P(A \mid B_k)P(B_k)}$$

CALCOLO COMBINATORIO E PROBABILITÀ

Analizziamo una misura in termini di causa ed effetto



Lo spazio degli eventi è separato in n classi mutuamente esclusive ed esaustive che rappresentano tutte le possibili cause (B_1, …, B_4 nella figura).

Ogni causa può dare origine con probabilità diversa ad alcuni effetti (A_1,A_2 nella figura).

La probabilità che l'effetto A_1 sia dovuto alla causa B_2 è normalmente noto e rappresenta la probabilità che si verifichi l'effetto A_1 quando si è verificata la causa B_2.

$P(A_1|B_2)$

In una misura tipicamente vedo (misuro) l'effetto (A_1) e mi chiedo a quale possibile causa sia dovuto.

$P(B_k|A_1)$.

Il teorema di Bayes mi aiuta in questo.

$$P(B_k \mid A_i) = \frac{P(A_i \mid B_k)\ P(B_k)}{\sum_j P(A_i \mid B_j)\ P(B_j)}$$

CALCOLO COMBINATORIO E PROBABILITÀ

$$P(B_k | A_i) = \frac{P(A_i | B_k)\ P(B_k)}{\sum_j P(A_i | B_j)\ P(B_j)}$$

Notiamo come sia la specificità che la sensibilità del test sono due esempi di probabilità condizionata.

Misura

Ora supponiamo di voler applicare il test ad un campione di persone a caso la domanda che ci poniamo è:
"*quanti di coloro che sono risultati positivi al test sono effettivamente malati e quanti invece sono sani?* "
Questo equivale ad identificare la probabilità (condizionata) che un soggetto positivo al test sia effettivamente malato

Questa è conoscenza a priori

$$P(malato | pos) = \frac{P(pos | malato) P(malato)}{P(pos | malato) P(malato) + P(pos | sano) P(sano)}$$

$$\boxed{P(malato \mid pos) = \frac{P(pos \mid malato)P(malato)}{P(pos \mid malato)P(malato) + P(pos \mid sano)P(sano)}}$$

$$P(pos \mid malato) = 0.993$$ Sensibilità del test

$$P(malato) = 2 \cdot 10^{-5}$$ Conoscenza a priori

$$P(pos \mid sano) = 1 - P(neg \mid sano) = 1 - 0.9999 = 1 \cdot 10^{-4}$$ Da specificità

$$P(sano) = 1 - P(malato) = 0.99998$$ Conoscenza a priori

Da cui

$$P(malato \mid pos) = 0.1656$$

Per cui di coloro che risultano positivi al test solo poco più del 16% sono effettivamente malati i rimanenti (84%) sono sani ma erroneamente indicati come malati (falsi positivi)

CALCOLO COMBINATORIO E PROBABILITÀ

Quant'è invece la frazione di coloro che risultano negativi al test e sono effettivamente sani
Questo equivale a calcolare la P(sano|neg)

$$P(sano \mid neg) = \frac{P(neg \mid sano)P(sano)}{P(neg \mid sano)P(sano) + P(neg \mid malato)P(malato)}$$

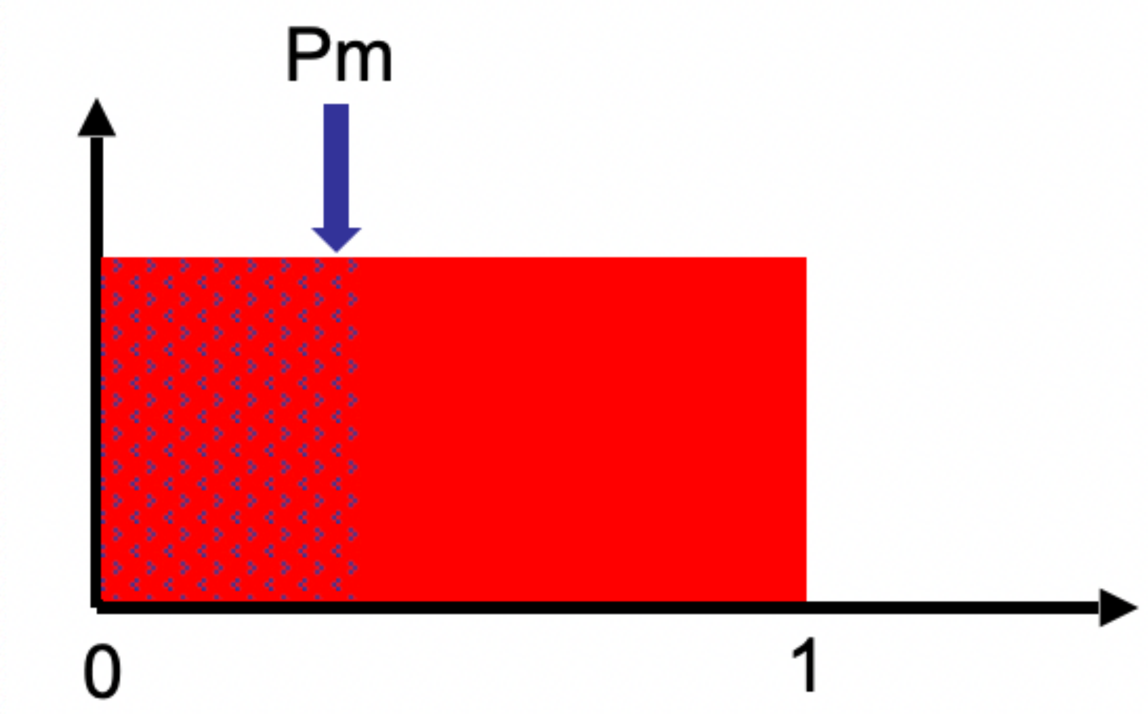
Con un calcolo analogo al precedente si trova che

$$P(sano \mid neg) = 0.9999999$$

Quindi se il risultato del test è negativo, la probabilità di essere sani è estremamente elevata o, inversamente, la probabilità che un malato risulti negativo al test (falso negativo) è estremamente bassa.

SIMULAZIONE

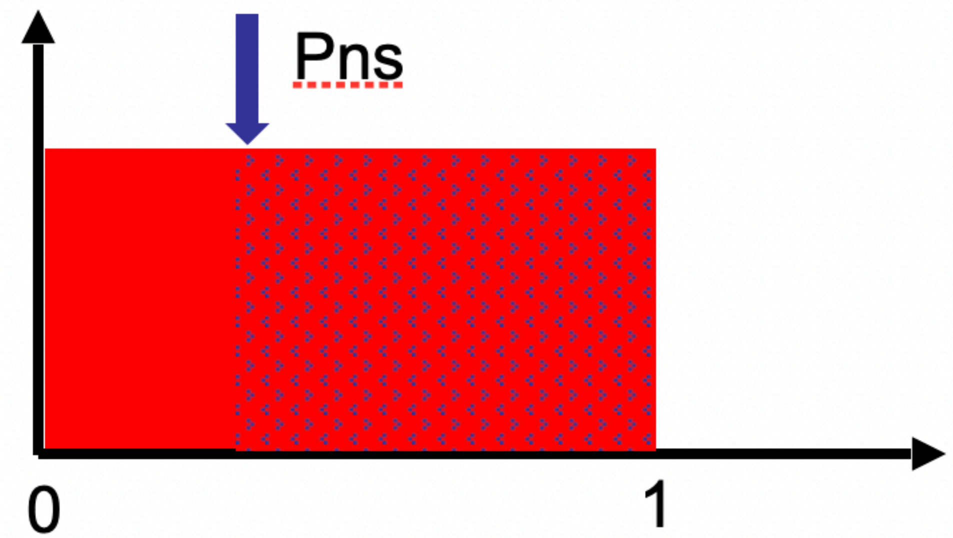
```
#include "TRandom.h"
void Bayes(long N=1000000){
   double Pm= 0.00002; // a priori
  //double Pm= 0.001; // a priori
   double Ppm=0.993;  // P(positivo|malato)
   double Pns=0.9999; // P(negativo|sano)
   TRandom3 R(12345);
   long Pos=0;        // contatore esiti positivi
   long Neg=0;        // contatore esiti positivi
   long MVero=0;      // contatore malati (veri)

   //long N=10000000000;
   //long N=100000000;
   long FalsiNegativi=0;
   long VeriNegativi=0;
   long FalsiPositivi=0;
   long VeriPositivi=0;

   for (int i=0; i<N ; i++) {
    int     Sog=0;  // =0 -> sano ; =1 -> malato
    double  x=R.Uniform();
    int R1=0;
    if (x<Pm) {
      Sog=1;   // malato
      MVero++;
    }
```

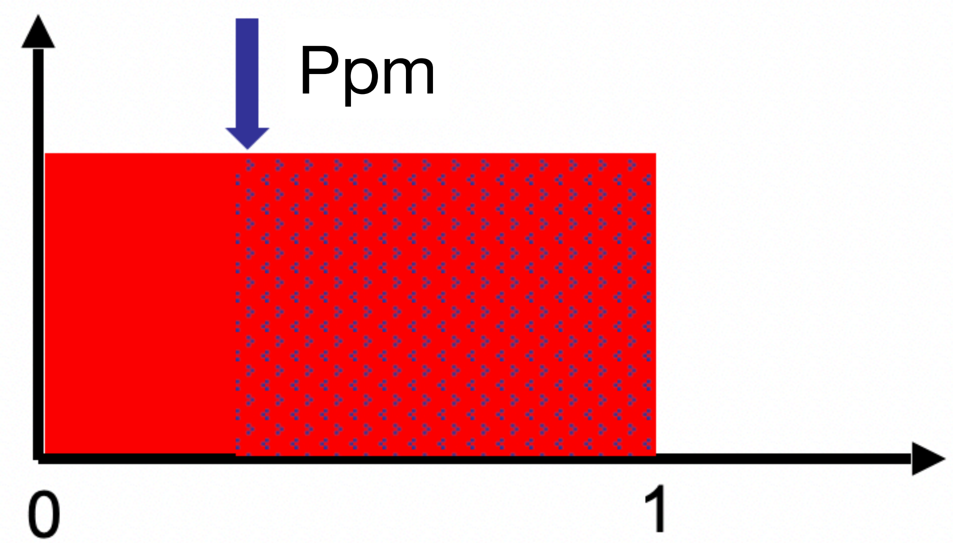


Sog=1 → Malato   Sog=0 → Sano

SIMULAZIONE

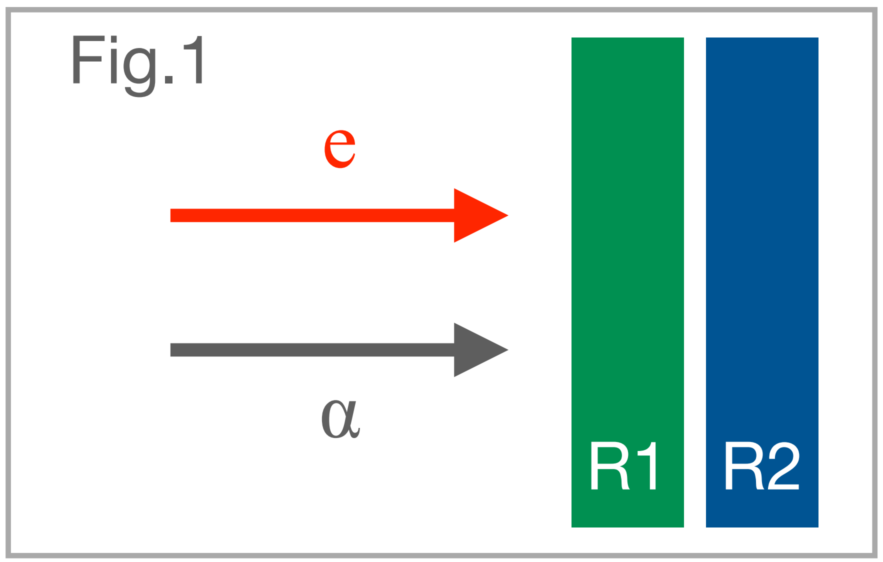
```
x=R.Uniform();
if (Sog==0)
  {
    // se soggetto sano
    if (x<Pns) {
      // esito negativo per soggetto sano
      R1=1;
      VeriNegativi++;
      Neg++; // esisto negativo
    }
    else
      {
        Pos++; // esisto positivo
        FalsiPositivi++;
      }
  }
else
  {
    // se soggetto malato
    if (x<Ppm) {
      // esito positivo per soggetto malato
      R1=1;
      Pos++;  // esisto positivo
      VeriPositivi++;
    }
    else
      {
        Neg++; // esisto negativo
        FalsiNegativi++;
      }
  }
```



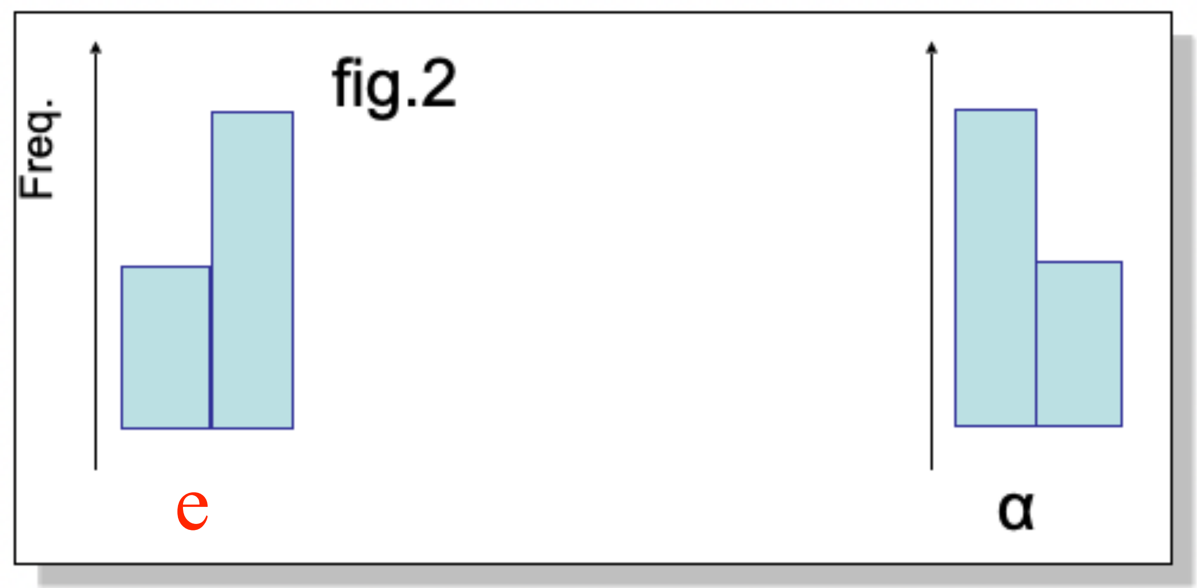


CALCOLO COMBINATORIO E PROBABILITÀ

- Un altro esempio: consideriamo un apparato sperimentale costituito da due rivelatori consecutivi che rivelano l'interazione di particelle (fig.1). Sull'apparato sperimentale arrivano 2 tipi di particelle: elettroni e particelle alpha (α). Se i due rivelatori hanno una diversa "risposta" (sensibilità) alle due tipologie di particelle posso determinare la composizione del fascio (quanti e, quante α) ?



Precedentemente avrò calibrato il mio rivelatore su fasci di particelle di soli elettroni e sole α per determinarne la risposta di R1 e R2 ai diversi tipi di particella (fig.2)



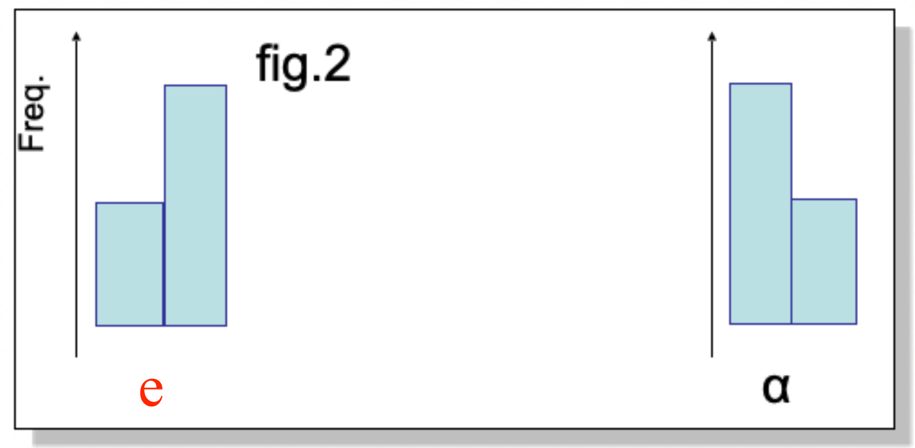
Cause, classi di eventi: e, α
Effetti: segnale in R1, segnale in R2

P(R1| e) = 0.1 P(R1| α) = 0.9
P(R2| e) = 0.8 P(R2| α) = 0.05

- Posso usare il rivelatore per studiare un processo di decadimento dal quale non so se emergono elettroni particelle alpha.

CALCOLO COMBINATORIO E PROBABILITÀ

- R1 ha efficienza 10 % per elettroni, 90% per particelle alpha
- R2 ha efficienza 80 % per elettroni, 5% per particelle alpha
- 4 esiti possibili:
  - R1 && !R2
  - R1 && R2
  - !R1 && !R2
  - !R1 && R2

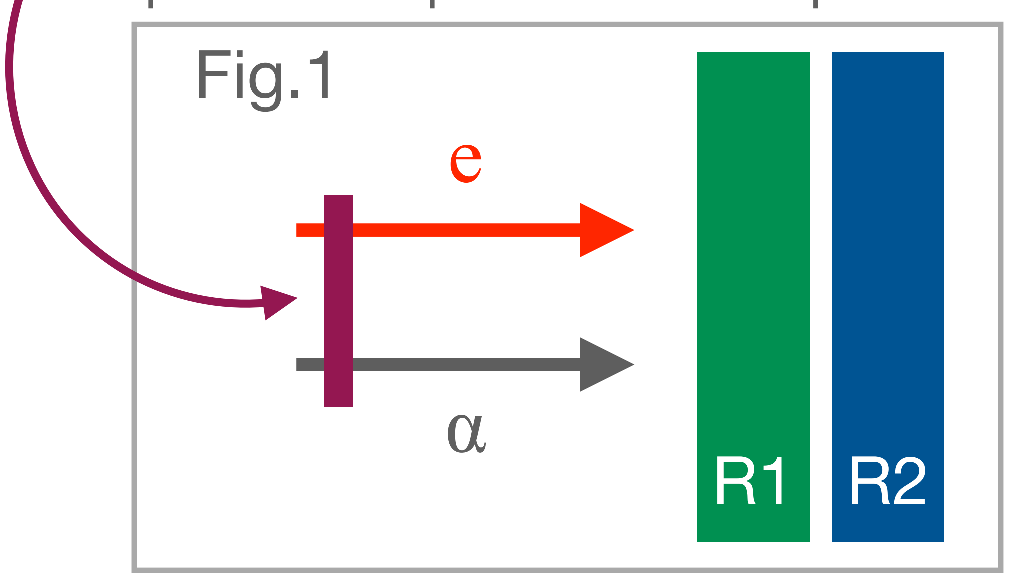


Cause, classi di eventi: e, α
Effetti: segnale in R1, segnale in R2

P(R1| e) = 0.1 P(R1| α) = 0.9
P(R2| e) = 0.8 P(R2| α) = 0.05
f(e) = 0.4, f(a) = 0.6 note

Tagger: indica il passaggio di una particella del fascio con la stessa probabilita' per elettroni e alpha

Fig.1



Per N=10000 particelle segnalate dal tagger osservo

N(R1 && !R2) = 8000
N(R1 && R2) = 200
N(!R1 && !R2) = 15
N(!R1 && R2) = 10000-8000-200-15 = 1785

Ignorare

CALCOLO COMBINATORIO E PROBABILITÀ

- R1 ha efficienza 10 % per elettroni, 90% per particelle alpha
- R2 ha efficienza 80 % per elettroni, 5% per particelle alpha
- 4 esiti possibili:
  - R1 && !R2
  - R1 && R2
  - !R1 && !R2
  - !R1 && R2

Cause, classi di eventi: e, α

Effetti: segnale in R1, segnale in R2

P(R1| e) = 0.1 P(R1| α) = 0.9

P(R2| e) = 0.8 P(R2| α) = 0.05

Per N=10000 particelle segnalate dal tagger osservo

N(R1 && !R2) = 8000

N(R1 && R2) = 200. ——>>>N(R1) = 8200

N(!R1 && !R2) = 15 ——>>>N(R2) = 1985

N(!R1 && R2) = 10000-8000-200-15 = 1785

- Se consideriamo solo N(R1) = 8200 e N(R2) = 1985
  - N(R1) = N x P(R1) = N x (fe x P(R1| e)) + N x (fa x P(R1| a)) Teo prob totale
  - N(R2) = N x P(R2) = N x (fe x P(R2| e)) + N x (fa x P(R2| a)) Teo prob totale
    - Due eq. in due incognite, fe e fa, se il processo fosse deterministico
- P(e | R2) = P(R2|e)P(e)/P(R2)
- P(a | R1) = P(R1|a)P(a)/P(R1)

CALCOLO COMBINATORIO E PROBABILITÀ

- R1 ha efficienza 10 % per elettroni, 90% per particelle alpha
- R2 ha efficienza 80 % per elettroni, 5% per particelle alpha
- 4 esiti possibili:
  - R1 && !R2
  - R1 && R2
  - !R1 && !R2
  - !R1 && R2

Cause, classi di eventi: e, α
Effetti: segnale in R1, segnale in R2

$P(R1|e) = 0.1 \quad P(R1|\alpha) = 0.9$

$P(R2|e) = 0.8 \quad P(R2|\alpha) = 0.05$

Per N=10000 particelle segnalate dal tagger osservo

N(R1 && !R2) = 8000
N(R1 && R2) = 200. ——>>>N(R1) = 8200
N(!R1 && !R2) = 15 ——>>>N(R2) = 1985
N(!R1 && R2) = 10000-8000-200-15 = 1785

- P(R1 && !R2 | e) = 0.1x(1-0.8) = 0.02
- P(R1 && R2 | e) = 0.1x 0.8 = 0.08
- P(!R1 && !R2| e) = (1-0.1)x(1-0.8) = 0.18
- P(!R1 && R2 | e) = (1-0.1)x0.8 = 0.72

Somma = 100% Teo prob totale

- P(R1 && !R2 | a) = 0.9x(1-0.05) = 0.855
- P(R1 && R2 | a) = 0.9x 0.05 = 0.045
- P(!R1 && !R2| a) = (1-0.9)x(1-0.05) = 0.095
- P(!R1 && R2 | a) = (1-0.9)x0.05 = 0.005

Somma = 100%

p(e | R1&&!R2) = 0.02xP(e)/p(R1&&!R2) = 0.02 P(e) / (0.02 P(e) + 0.855 P(a))
p(e | R1&& R2) = 0.08xP(e)/p(R1&&R2) = 0.08 P(e) / (0.08 P(e) + 0.045 P(a))
p(e | !R1&&!R2) = 0.18xP(e)/p(!R1&&!R2) = 0.18 P(e) / (0.18 P(e) + 0.095 P(a))
p(e | !R1&& R2) = 0.72xP(e)/p(!R1&& R2) = 0.72 P(e) / (0.72 P(e) + 0.005 P(a))

CALCOLO COMBINATORIO E PROBABILITÀ

OSSERRVAZIONE
p(R1&&!R2)   e’ stimato da 8000/10000 = 0.8
p(R1&&R2)   e’ stimato da 200/100000 = 0.002
p(!R1&&!R2)   e’ stimato da 15/100000 = 0.00015
p(!R1&& R2)   e’ stimato da 1785/100000 = 0.01785

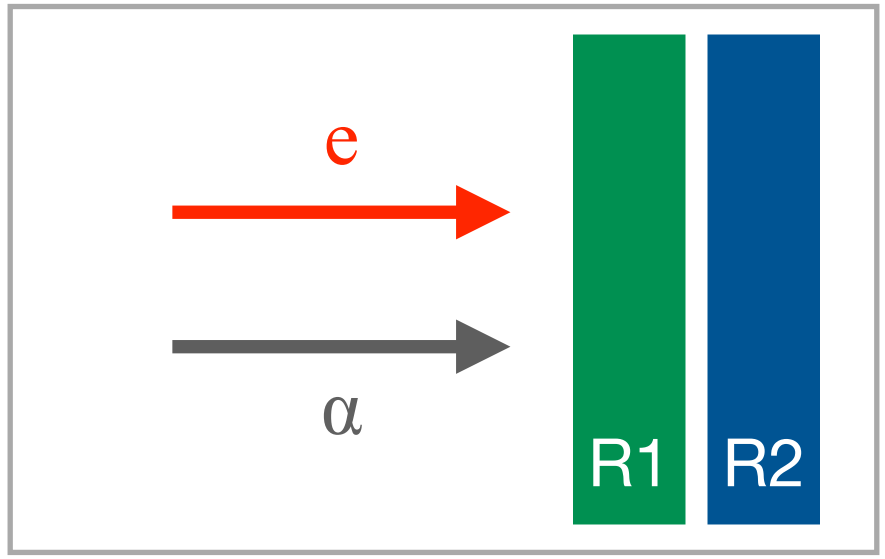
Cause, classi di eventi: e, α
Effetti: segnale in R1, segnale in R2
P(R1| e) = 0.1 P(R1| α) = 0.9
P(R2| e) = 0.8 P(R2| α) = 0.05

Per N=10000 particelle segnalate dal tagger osservo

N(R1 && !R2) = 8000
N(R1 && R2) = 200.   ——>>>N(R1) = 8200
N(!R1 && !R2) = 15   ——>>>N(R2) = 1985
N(!R1 && R2) = 10000-8000-200-15 = 1785

p(e | R1&&!R2) = 0.02xP(e)/p(R1&&!R2)   = 0.02 P(e) / (0.02 P(e) + 0.855 P(a))
p(e | R1&& R2) = 0.08xP(e)/p(R1&&R2)   = 0.08 P(e) / (0.08 P(e) + 0.045 P(a))
p(e | !R1&&!R2) = 0.18xP(e)/p(!R1&&!R2)   = 0.18 P(e) / (0.18 P(e) + 0.095 P(a))
p(e | !R1&& R2) = 0.72xP(e)/p(!R1&& R2)   = 0.72 P(e) / (0.72 P(e) + 0.005 P(a))

CALCOLO COMBINATORIO E PROBABILITÀ



Il mio esperimento: decadimento ignoto
Conto (in un modo indipendente da R1 e R2) i decadimenti totali
Conto le risposte di R1
Mi chiedo quanti decadimenti abbiano prodotto particelle α ?
Voglio insomma determinare P(α | R1)

- Le conclusioni del nostro esperimento dipendono dalla conoscenza a priori del processo in studio?
- Con che probabilità le particelle che interagiranno sul rivelatore 1 sono alpha?

CALCOLO COMBINATORIO E PROBABILITÀ

$$P(B_k \mid A_i) = \frac{P(A_i \mid B_k)\ P(B_k)}{\sum_j P(A_i \mid B_j)\ P(B_j)}$$

- NOTA: per poter stimare la probabilità che la misura (effetto) sia attribuibile ad una delle possibili cause devo introdurre una stima della probabilità che una delle cause sia presente (P(Bk)).
- Nell'approccio Bayesiano alla misura ho sempre una probabilità a priori (P(Bk)) e una probabilità a posteriori (P(Bk|A)).
- La misura altera la conoscenza del fenomeno in studio cambiando la convinzione su quali sono le cause che sono alla base dell'universo che osserviamo.
- La assoluta ignoranza di un processo oggetto di studio comporta la necessità di assegnare a tutte le possibili cause la stessa probabilità.

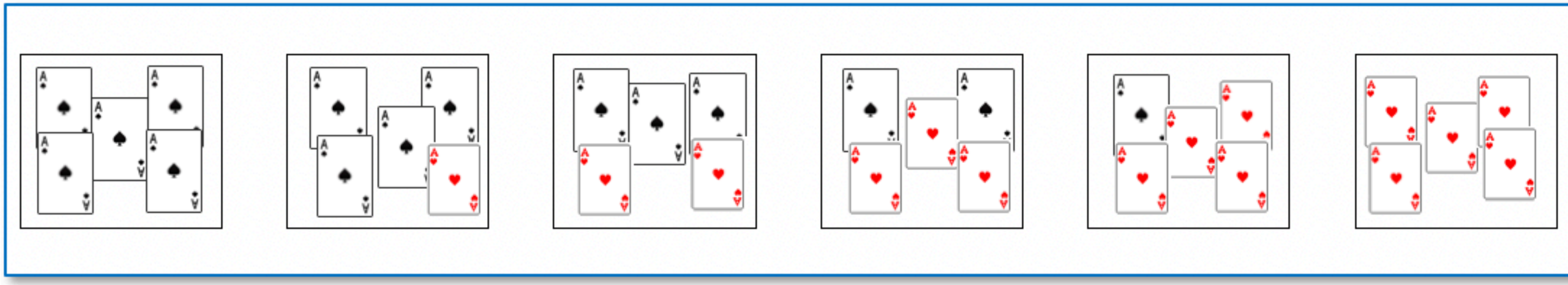
CALCOLO COMBINATORIO E PROBABILITÀ

$$P(B_k | A_i) = \frac{P(A_i | B_k)\ P(B_k)}{\sum_j P(A_i | B_j)\ P(B_j)}$$

- La assoluta ignoranza di un processo oggetto di studio comporta la necessità di assegnare a tutte le possibili cause la stessa probabilità.
- Consideriamo quindi tutte le possibili cause sono equiprobabili
  - P(Bk)=1/N ; k=1,…,N

- Tutte le nostre conoscenze sono frutto di probabilità a posteriori.
- Lo stesso concetto di probabilità assoluta non esiste, è solo un assioma matematico. In termini pratici possiamo parlare solo di probabilità a posteriori.

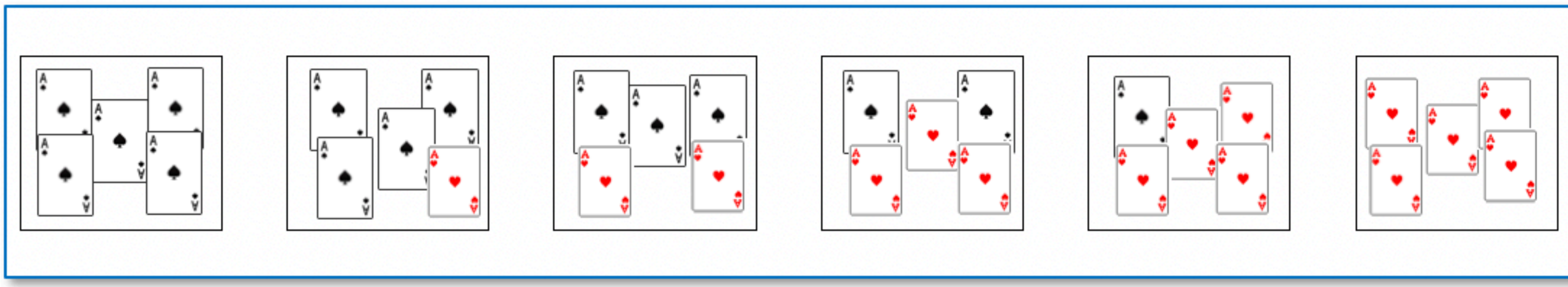
Il problema delle sei buste



- Quant'è la probabilità di estrarre una carta rossa da una delle buste prese a caso?
- Supponiamo di aver estratto una carta rossa da una busta e di averla rimessa dentro la stessa busta, quant'è, ora, la probabilità di estrarre una carta rossa dalla stessa busta?

CALCOLO COMBINATORIO E PROBABILITÀ

Il problema delle sei buste



- Quant'è la probabilità di estrarre una carta rossa da una delle buste prese a caso?

6 cause, ossia sei buste, equiprobabili;

p(rossa,B1) = 0 x 1/6

p(rossa,B2) = 1/5 x 1/6

p(rossa,B3) = 2/5 x 1/6

p(rossa,B4) = 3/5 x 1/6

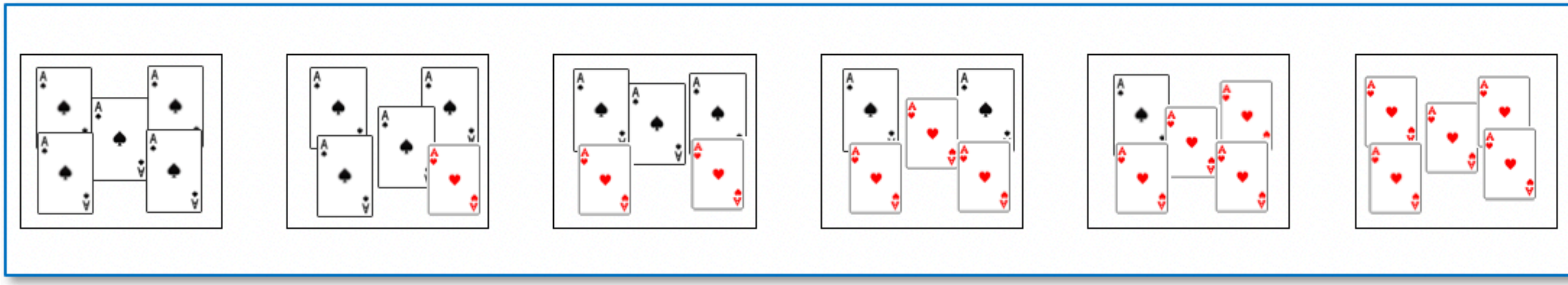
p(rossa,B5) = 4/5 x 1/6

p(rossa,B6) = 5/5 x 1/6

p(rossa, any B) = 15/5 x 1/6 = 1/2

CALCOLO COMBINATORIO E PROBABILITÀ

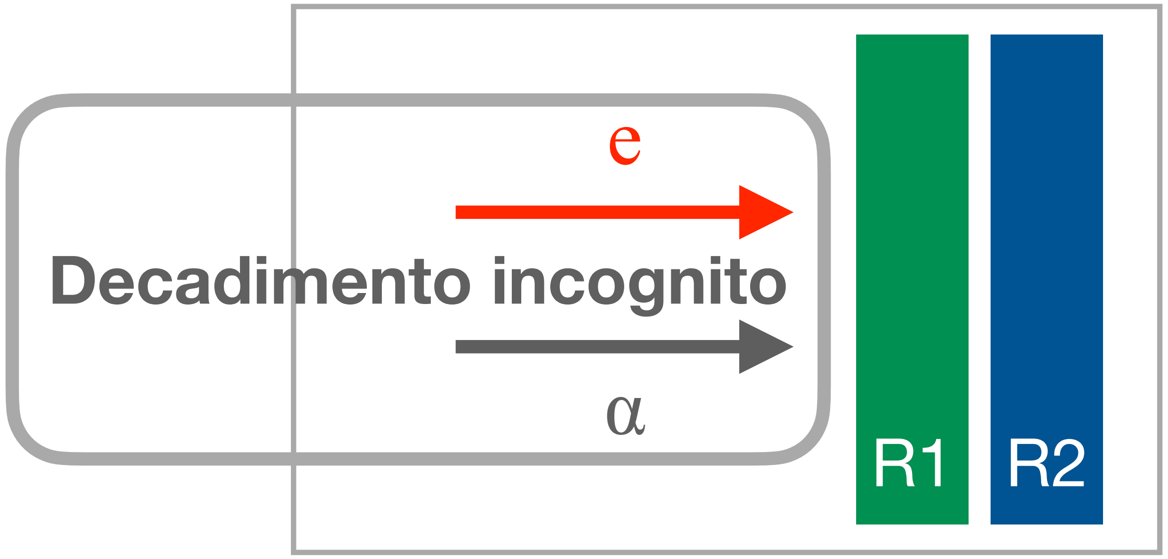
Il problema delle sei buste



- Supponiamo di aver estratto una carta rossa da una busta e di averla rimessa dentro la stessa busta, quant'è, ora, la probabilità di estrarre una carta rossa dalla stessa busta?

p(rossa,B2) = 1/5 x 1/5
p(rossa,B3) = 2/5 x 1/5
p(rossa,B4) = 3/5 x 1/5
p(rossa,B5) = 4/5 x 1/5
p(rossa,B6) = 5/5 x 1/5
p(rossa, any B but B1) = 15/5 x 1/5 = 3/5

CALCOLO COMBINATORIO E PROBABILITÀ



Le conclusioni del nostro esperimento dipendono dalla conoscenza a priori del processo in studio?
Con che probabilità le particelle che interagiranno sul rivelatore 1 sono alpha?

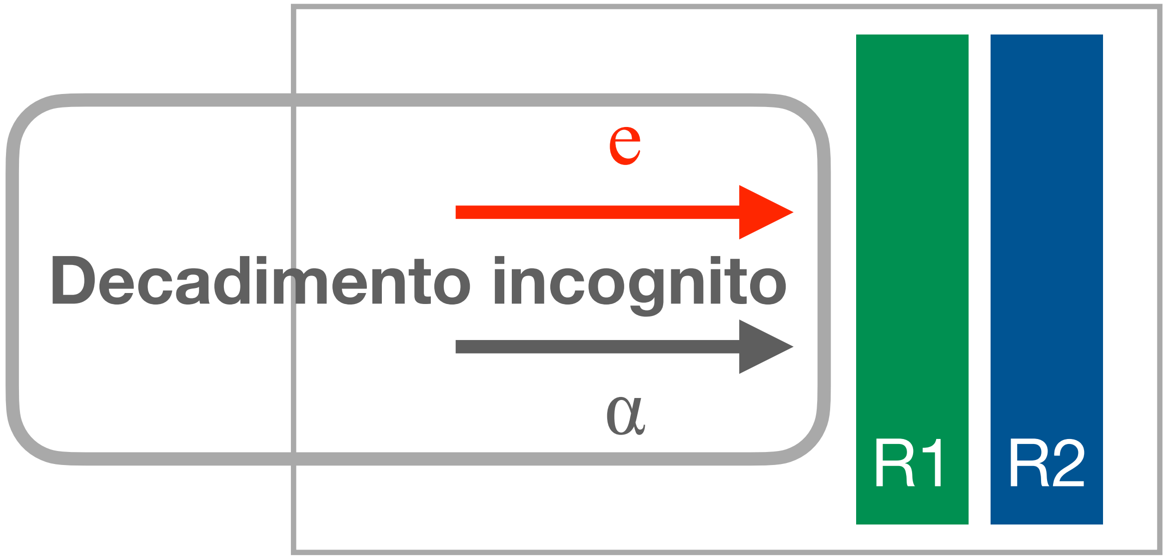
$$P(\alpha|R_1) = P(R_1|\alpha)P(\alpha)/P(R_1)$$

Probabilità a priori uniforme = 0.5 per ogni tipo di particella

$$P(R_1) = P(R_1|\alpha)P(\alpha)+P(R_1|e)P(e) = 0.9 \bullet 0.5 + 0.1 \bullet 0.5=0.5$$

$$P(\alpha|R_1) = 0.9 \bullet \frac{0.5}{0.5} = 0.9$$

CALCOLO COMBINATORIO E PROBABILITÀ



Le conclusioni del nostro esperimento dipendono dalla conoscenza a priori del processo in studio?
Con che probabilità le particelle che interagiranno sul rivelatore 1 sono alpha?

$$P(\alpha|R_1) = P(R_1|\alpha)P(\alpha)/P(R_1)$$

Probabilità a priori non uniforme 80-20

$$P(R_1) = P(R_1|\alpha)P(\alpha)+P(R_1|e)P(e) = 0.9 \bullet 0.8 + 0.1 \bullet 0.2 = 0.74$$

$$P(\alpha|R_1) = 0.9 \bullet \frac{0.8}{0.74} = 0.97$$

FINE